Signal Transduction proteins in *Streptococcus pneumoniae* and *Vibrio cholerae*

Luis Daniel Toribio Isaías

Doctoral programme in Biotechnology

Structural characterization of Signal Transduction proteins in *Streptococcus pneumoniae* and *Vibrio cholerae*

IRB Barcelona, Institute for Research in Biomedicine
IBMB-CSIC, Molecular Biology Institute of Barcelona

PhD student:

Luis Daniel Toribio Isaías

Thesis directors:

Prof. Miquel Coll Capella
Dr. Albert Canals Parera

Tutor:

Prof. Josefa Badia Palacin

Barcelona, 2019

This PhD thesis titled: “Structural characterization of Signal Transduction proteins in *Streptococcus pneumoniae* and *Vibrio cholerae*” was carried out in the Structural Biology of Protein & Nucleic Acid Complexes and Molecular Machines group at the Institute for Research in Biomedicine (IRB Barcelona) and the Molecular Biology Institute of Barcelona (IBMB-CSIC) and it was partially funded by the “Conacyt-Comecyt scholarship from 2014”.

牛逼

Ad astra per aspera

Index

List of Tables

List of Figures

List of abbreviations and symbols

%	Percentage
1D-NMR	One-dimensional Nuclear magnetic resonance
2D	Two-dimensional
3D	Three-dimensional
Å	Armstrong
AMP	Adenosine monophosphate
bp	Base pair
C-Terminal	Carboxyl-terminal
CTF	Contrast transfer function
CTX	Cholera toxin
CV	Column Volumes
Da	Dalton
ddH_2O	Double-distilled water
DLS	Dynamic light scattering
DM	Density modifications
DNA	Deoxyribonucleic acid
DR	Double Repeat
dsDNA	Double-stranded DNA
e^-	Electron
Ed.	Editor

E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Essay
ESRF	European Synchrotron Radiation Facility
eV	Electron Volt
f_0	Normal scattering term
f'	Anomalous scattering dispersion term
f''	Anomalous scattering absorption term
F_{obs}	Observed structure factor
F_{calc}	Calculated structure factor
FSC	Fourier shell correlation
g	Gram
g	Gravity acceleration constant
h	Hour
HGT	Horizontal Gene Transfer
His Tag	Histidine tag
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infra Red
Iter	Iteration
kb	kilo base
l	Liter

LB	Luria-Bertani
LC-MS/MS	Liquid chromatography tandem-mass spectrometry
LGT	Lateral Gene Transfer
LLG	Log-likelihood gain
µ	Micro (10^{-6})
M	Molar
m	Meter
MAD	Multi-wavelength anomalous dispersion
MALS	Multi-Angle Light Scattering
MES	2-(N-mopholino) ethanesulfonic acid
min	Minute
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacement with anomalous scattering
MS	Mass spectrometry
MR	Molecular replacement
MW	Molecular Weight
N	Newton
n	Nano (10^{-9})
N-terminal	Amino-terminal
NCS	Non-crystallographic symmetry

O.D.	Optical density
PAC	Automated Crystallography Platform
PAGE	Polyacrylamide gel
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethileneglycol
pRNA	Packaging RNA
Psi	Pounds per square inch
Pu	Purine
Px	Pixel
Py	Pyrimidine
RI	Refractive Index
RNA	Ribonucleic acid
rpm	Revolutions per minute
RMSD	Root mean square deviation
RR	Response Regulator
s	Second
S. pneumoniae	*Streptococcus pneumoniae*
SAD	Single-wavelength anomalous dispersion
SDS	Sodium dodecyl sulphate

SEC	Size Exclusion Chromatography
SEC-MALS	Size Exclusion Chromatography with Multi-Angle Light Scattering
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous scattering
SRF	Self-rotation function
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TEMED	N,N,N,N-tetramethylendiamine
Tfb I	Transformation buffer I
Tfb II	Transformation buffer II
TCP	Toxin co-regulated pilus
TFZ	Translation-function Z-score
TM	Transmembrane
ToxR-DBD	ToxR DNA binding domain
UV	Ultraviolet
V	Volt
V. cholerae	*Vibrio cholerae*
V_M	Matthews coefficient
[v/v]	Volume per volume
[w/v]	Weight per volume

w-HTH	Winged helix-turn-helix
WHO	World Health Organization
x*g*	Times of standard gravity
λ	Wavelength
^{o}	Degrees
^{o} C	Celsius degrees

Amino acid abbreviations

Amino acid	One letter code	Three letter
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Try
Tyrosine	Y	Tyr
Valine	V	Val

Acknowledgements

Foremost, I want to express my most sincere gratitude to my supervisor, Professor Miquel Coll, for his guidance, patience and the positive atmosphere in which he gave me the opportunity to work in. His support is a motivation that will inspire me to continue with my professional career.

I would also like to thank Dr. Albert Canals for mentoring and guiding my research and allowing me to take part in his project and for all the patience and support received from him. He was always willing to help and give his best opinion; I would possibly have cracked without him. I would like to thank Dr. Montserrat Fàbrega Ferrer for all her help and support and patience. I am thankful for all the effort and time they invested in me.

I would like to thank Prof. Eric Krukonis from the University of Detroit Mercy for providing us the purified ToxR-DBD protein as well as Dr. Simone Pieretti for setting all the groundwork to be able to compare his structure with the mutants used for this thesis.

I would like to thank CONACYT-COMECYT for providing the funding that allowed me to complete this PhD.

I am very grateful with all the members of the “Structural biology of protein & nucleic acid complexes and molecular machines” group as they always helped me with the issues I had when I approached them, trying to help out as they could without hesitating. I would like to thank Esther for all the support as a great lab technician, Rosa for all the advice; guidance and assistance, as well for all her work with the ToxR project; Cristina, for answering any doubts I had, Olga for helping me figure out one of my constructs was wrong, Mireia, Jorge and Sara for your friendship and camaraderie. I am really happy to have been able to work alongside with

these capable scientists. Thanks for everything; you helped me get through this.

I also want to thank Prof. Josefa Badia for all the help she provided while dealing with the university or visa issues. Also, Leonor for helping me fix the contract issues with the CSIC and extranjeria. I would like to also thank my “Tribunal Advisory Committee”: Prof. Ignasi Fita, Prof. Joan Guinovart and Prof. Manuel Palacin. I appreciate all the help provided by the IRB administration, especially Patricia Nadal and Leyre Caracuel. I want to thank Joan, Xandra and Roman from the PAC, for all the help throughout these years.

I would have not been able to reach this point without my mother's support; I am lucky to have you as my mum. I also would like to thank my family and close friends for all their love and support throughout this venture.

Abstract

The aim of this project is to structurally characterize proteins involved in bacterial signal transduction systems by applying X-ray crystallography.

Bacteria use signal transduction systems to react in response to any environmental changes detected. Bacterial signal transduction is divided into two categories, one-component systems and two-component systems. Two-component systems are composed by a Response Regulator (RR) and a Histidine Kinase (HK); the Histidine Kinase auto phosphorylates an inner domain, and soon after, it phosphorylates the receiver domain on the Response Regulator, activating the output domain; usually producing a physiological effect in the cell by activating a specific gene. While in one-component systems, one protein has both, a sensory and an output domain. An example of the one-component systems would be ToxR, and an example of the two-component systems would be ComD-ComE.

Bacterial transformation is a type of Horizontal Gene Transfer (HGT), which is a rapid evolutive mechanism in which entire genes can be transferred among bacterial cells. HGT is commonly deemed responsible for the appearance of antibiotic resistance, virulence factors and serotype switching.

Competence for genetic transformation in *Streptococcus pneumoniae* is a transient physiological state whose development is coordinated by a peptide pheromone (Competence Simulating Peptide or CSP) and its receptor, which activates transcription of two downstream genes, *comX* and *comW*, and 15 other “early” genes. ComD (HK) and ComE (RR) are involved in the quorum-signaling pathway that synthesizes the CSP. They help modulate the mechanism in which bacterial transformation occurs, by allowing the inclusion of naked DNA from the environment. We have successfully formed the binary (ComE+DNA) and ternary (ComD+ComE+DNA) complexes and

characterized them in the attempt of obtaining crystals to solve their 3-D structure.

On the other hand, to elaborate on one-component systems, like ToxR we can discuss cholera. Cholera is caused by the causative agent *Vibrio cholerae*. It is estimated that there are from 1.3 to 4.0 million cases out of which up to 143,000 result in cholera deaths annually. After ingesting the *V. cholerae*, it travels to the small intestine colonizing it and producing the cholera toxin. ToxR is a membrane-localized transcription factor that regulates the *toxT* promoter. The activation of the *toxT* promoter triggers the virulence cascade that leads to the secretion of toxin-coregulated pilus (TCP) and the expression of cholera toxin (CTX).

In recent years, our lab solved the crystal structure of the cytoplasmic domain of ToxR+20DNA, proposing molecular interactions between ToxR and the *toxT* promoter (Simone Pieretti's PhD Thesis). In this study, we want to determine the crystallographic structure of three mutants of the cytoplasmic domain of ToxR bound to the *toxT* promoter. According to biochemical data from our collaborator (Professor Eric Krukonis, from the university of Detroit), these mutations down regulate the activation of ToxR and we aim to analyze the structural changes that these mutants suppose. Using X-ray crystallography we solved the structure of three complexes; $ToxR^{Q78A}$+20DNA, $ToxR^{S81A}$+20DNA and $ToxR^{P101A}$+20DNA at 2.55, 2.95 and 2.95 Å resolution, respectively.

We have compared the final mutant structures with the wildtype, unveiling how the structural changes result in the decrease in activation of the *toxT* promoter.

Part I

ComD-ComE

Project I ComD-ComE

This PhD thesis is divided in two projects and organized in the different sections, the first project will focus on the ComD-ComE and their involvement in a mechanism for HGT.

. The **Introduction** summarizes previous knowledge about Bacterial Signal Transduction and establishes two case samples of the existing known types, while focusing on a two-component signaling system.

. The **Objectives** section lists the goals of this part of the project, which include the formation and characterization of complexes.

. The **Materials and methods** section describes techniques, materials and instruments used throughout the realization of the thesis. It is divided in 6 parts, the first half describe the sample preparation and the analysis before attempting to do structural studies, the next two deal with the creation of new mutants and the addition of DNA into the complexes and the last one is the attempt of structurally characterizing the proteins and complexes obtained.

. The **Results and discussion** chapter illustrates the generation of the experimental data and as well as an analysis of why we were not able to crystallize the proteins and protein complexes

. Lastly, the **Conclusions** section summarizes the main findings of this segment of the PhD thesis

1. Introduction ComD-ComE

Microbiota colonize the human body by digesting unprocessed dietary nutrients, producing vitamins, stimulating the immune system, inhibiting pathogen colonization amongst other functions in the body's physiology and health (Koropatkin *et al.*, 2012). There is a mutualistic relationship between mammals and their microbiota that allows both to survive. Host's physiological processes are shaped and modulated by the commensal bacteria from which the immune system and the intestinal barrier depend greatly on. The aggravation of this equilibrium could result in disease for either party of this mutualistic relationship (Natividad *et al.*, 2013). Sometimes it is necessary to fight bacterial infections such as cholera or *Streptococcus*. Both use bacterial signal transduction as means to either colonize the intestine or as a tool to obtain antibiotic resistance. It is of the most importance to be able to understand how these processes work and how to intervene.

1.1. Microwho?

Between 1665 and 1683, two fellows of the Royal Society, Antoni Van Leeuwenhoek and Robert Hooke discovered the existence of microscopic organisms (Gest, 2004). Van Leeuwenhoek was a Dutch merchant that in the late 1670's observed water under an early version of modern microscopes. His depictions were received as irrelevant yet interesting by the scientific community (Porter, 1976). It was not until after 200 years that the scientific community realized that microorganisms could be transmitted amongst people and cause disease.

1.1.1. Probiotics vs. pathogens

Microorganisms that benefit the host when they are introduced to internal ecosystem are called probiotics (Natividad *et al.,* 2013). Many probiotics have shown to provide an improvement in Inflammatory Bowel Disease, can reduce cholesterol, benefit the gut or even stimulate the immune system, as well as provide antimutagenic and anti-carcinogenic properties (Rivera-Espinoza *et al*., 2010). The human body contains approximately 10^{13} human cells and also about 10^{14} bacterial, fungal and protozoan cells; making it a

complex and thriving ecosystem, and without them we would cease to be as this symbiotic life-long relationship between our cells and our resident bacteria create a balance that starts at the moment of birth (Natividad *et al*., 2013; Brown *et al.,* 2004). Whenever this ecosystem is altered, it is called dysbiosis, and can lead to autoimmune disorders, increase in allergies and obesity amongst other (Koropatkin *et al*., 2012; Natividad *et al*., 2013; Ouwerkerk *et al.,* 2013; Zhang *et al*., 2009).

Pathogens differ from our normal microbial inhabitants because our normal flora only cause trouble if our immune systems are weakened or they access a place where they should not be. Opposite to pathogens that do not need the host to be immunocompromised, as they possess specialized mechanisms designed to cross biochemical and cellular barriers and to provoke specific reactions from the host that will aid with their survival and colonization (Johnson *et al.,* 2002).

1.1.2. Bacterium of interest

In this section, we will describe briefly the organism that naturally produces the proteins that are the focus of this part of the thesis.

1.1.2.1. Streptococcus pneumoniae (S. pneumoniae)

Albert Theodore Billroth used the word *Streptococcus* to refer to the organisms that had chain-like structures that formed in pus within human wounds. Etymologically speaking, *Streptococcus* comes from combining the words *streptós* (easily twisted) and the word *kókkos* (berry or grain) (Wilson, 1987). The genus Streptococcus includes pathogens such as *S. pyogenes, S. agalactie, S. pneumoniae, S. viridians,* and *S. mutans* as well as commensal species of human mouth, intestine, skin, and respiratory tract, which are commonly used in the production of milk derivatives, and *streptococcal* infections are commonly treated with penicillin derivatives (Patterson, 1996).

S. pneumoniae are lancet-shaped gram-positive anaerobic bacteria with over 90 known serotypes. They commonly inhabit the respiratory tract and could be isolated from 5-90% of the healthy population, depending on the population and setting. They are part of the flora in our upper respiratory

tract, yet if the host's immune system is suppressed, they can become pathogenic and cause disease (CDC, 2019). They are usually found in pairs, but sometimes can form short chains as shown in Figure 1.1 (Tettelin *et al.,* 2001).

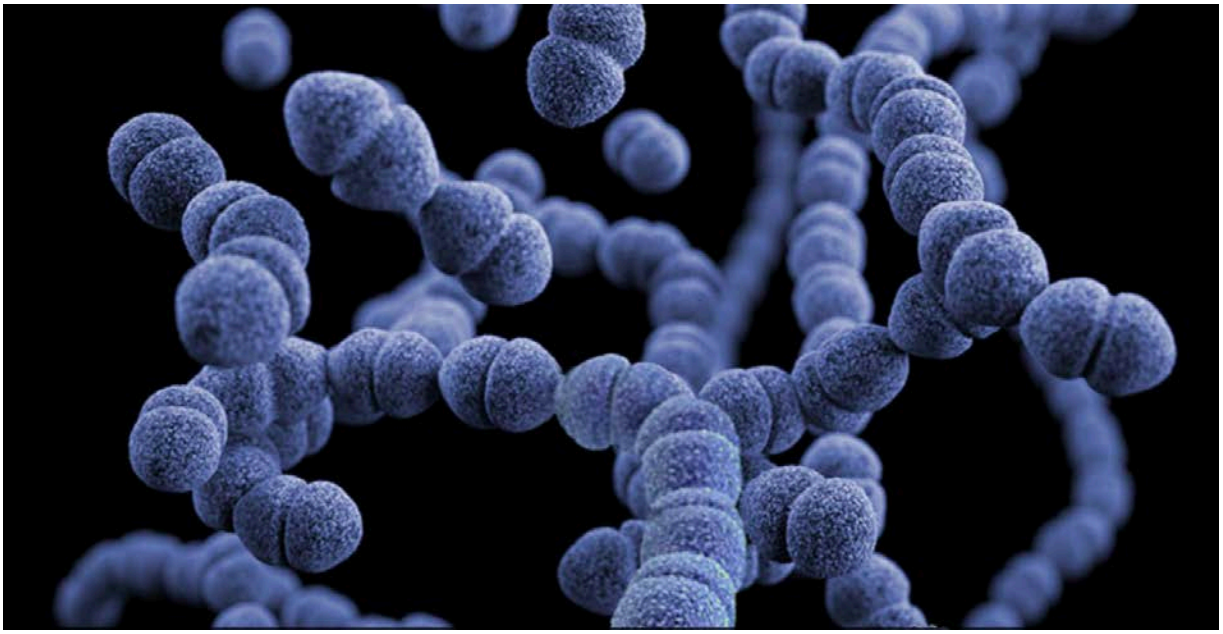


Figure 1.1 Illustration of *S. pneumoniae*
(Image retrieved from CDC, 2019).

1.2. Rise of antibiotics

Germ theory was structured and postulated by Koch in the second half of the nineteenth century, it was developed mainly for living agents (specially bacteria), as it has not been applicable to inanimate pathogens like infectious proteins or viruses (Walker *et al.,* 2006). This theory was the stepping of modern microbiology and the study of infectious diseases.

Initially, science had no mechanisms to fight off diseases, this was until Ehrlich, a German researcher that won the Nobel Prize of Medicine in 1908, discovered Arsenic compounds that had antimicrobial traits (Bosch *et al.*, 2008). In 1928 Alexander Fleming, a Scottish physician working with *Staphylococcus* saw the antibacterial properties of mold in his cultures, this was *Penicillum notatum*, he then isolated the compound responsible and called it penicillin (Bauman *et al*., 2011). Pennicillin was not utilized commonly until the 1940's until Florey and Chain were able to produce it on

a larger scale. The methods used for this, and the discovery of penicillin earned the three of them the 1945 Nobel Prize of Medicine (Ligon, 2004).

Waksman was the first to use the term antibiotic to refer to a compound that antagonizes the growth of microorganisms (Clardy *et al.,* 2009). The discovery of penicillin was the start of a new era of medicine as it motivated research in search of more compounds that had bacteriostatic and bactericidal qualities (Waksman & Woodruf, 1942).

Many antibiotics were discovered between 1945 and 1955, such as streptomycin or chloramphenicol, and the list grew exponentially (Aminov, 2010). An organism, like fungi, first produced them but techniques were developed to produce them synthetically (Borders, 2007). At this moment in time, the biochemical details of antibiotics were largely unknown, yet it was cleared that they had different chemical structures, mechanism of actions and that dosage (Waksman & Woodruf, 1942).

1.2.1. Antibiotic mechanisms

The direct effects and drug-targets interactions of antibiotics are well characterized, whilst bacterial responses to antibiotic treatments that provoke cell death have proven to be quite complex, as they involve different genetic and biochemical pathways and in some cases remain uncharacterized (Kohanski *et al.,* 2010).

Antibiotics are classified according to the system or cellular component they interact with, as well as if they have bactericidal properties (induce cell death) or bacteriostatic properties (inhibit cell growth). Nowadays antibiotics inhibit cell wall synthesis, protein synthesis, DNA synthesis, or RNA synthesis (Walsh, 2003). Table 1.1 lists some antibiotic classes, their chemical structure and other characteristics and examples of antibiotics (Borders, 2007; Peterson & Packer, 2013).

Table 1.1 Examples of antibiotic classes and their antibiotic mechanism (Borders, 2007; Peterson & Packer, 2013).

Antibiotic type	Mechanism	Examples
Aminoglycosides	Inhibit protein biosynthesis by binding to the ribosome	Streptomycin, neomycin, kanamycin, gentamicin, tobramycin, and amikacin
Ansamacrolides or ansamycins	Inhibit RNA polymerase	Rifampin
β-lactams	β-lactams inhibit bacterial cell wall synthesis, by binding to penicillin binding proteins and some induce the activity of autolysins	It includes penicillins, cephalosporins, carbapenems, mono-bactams, nocardicins, and clavulanic acid
Chloramphenicol	Inhibits protein synthesis by binding to the 50s ribosomal subunit and blocks the peptidyltransferase reaction	
Glycoconjugated molecules	Can completely inhibit the adhesion of pathogenic bacteria to membrane proteins	Fucoidan, KappaZinc
Glycopeptides	Inhibit cellular wall biosynthesis by binding to alanyl-D-alanine units of the lipid-boud precursor found in the cell walls	Vancomycin, avoparcin, and teicoplanin
Licosamides	They interfere with protein biosynthesis through binding to the 50s ribosomal unit	Lincomycin and celesticetins
Lipopeptides	Disrupts bacterial cell wall membrane	Daptomycin
Macrolides	Inhibit protein synthesis by binding to the 50s ribosomal unit	Azithromycin, erythromycin
Polyethers	Interact with bacterial cell membranes and cause cell death by passing cations through	
Tetracyclines	Inhibit protein synthesis by binding to the 30s ribosomal unit	
Synthetic antibacterial agents	Inhibit DNA gyrase and oxazolidinones and interfere with mRNA translation	Amprolium ethopabate , morantel citrate.

1.2.2. Life finds a way

At first, society thought that antibiotics would be the answer to control infections, but many bacterial strains found a way to become resistant to antibiotics (Appelbaum, 1992). Many bacterial strains became resistant to

antibiotics, this, partially due to the misuse of antibiotics, and in 2010, it translated into a cost of approximately 1,500 million euros a year in sanitary costs (Aminov, 2010). For the World Health Organization, antibiotic resistance is one of the biggest three health threats of the 21st century (WHO, 2014); nowadays, this represents a challenge for society, as it is estimated that in 2015, within the EU and the European Economic Area 671,689 infections with antibiotic-resistant bacteria occurred, accounting for an estimated 33,110 deaths. What this represented in 2015 to the EU and the European Economic Area is illustrated in Figure 1.2 (Cassini *et al.,* 2011).

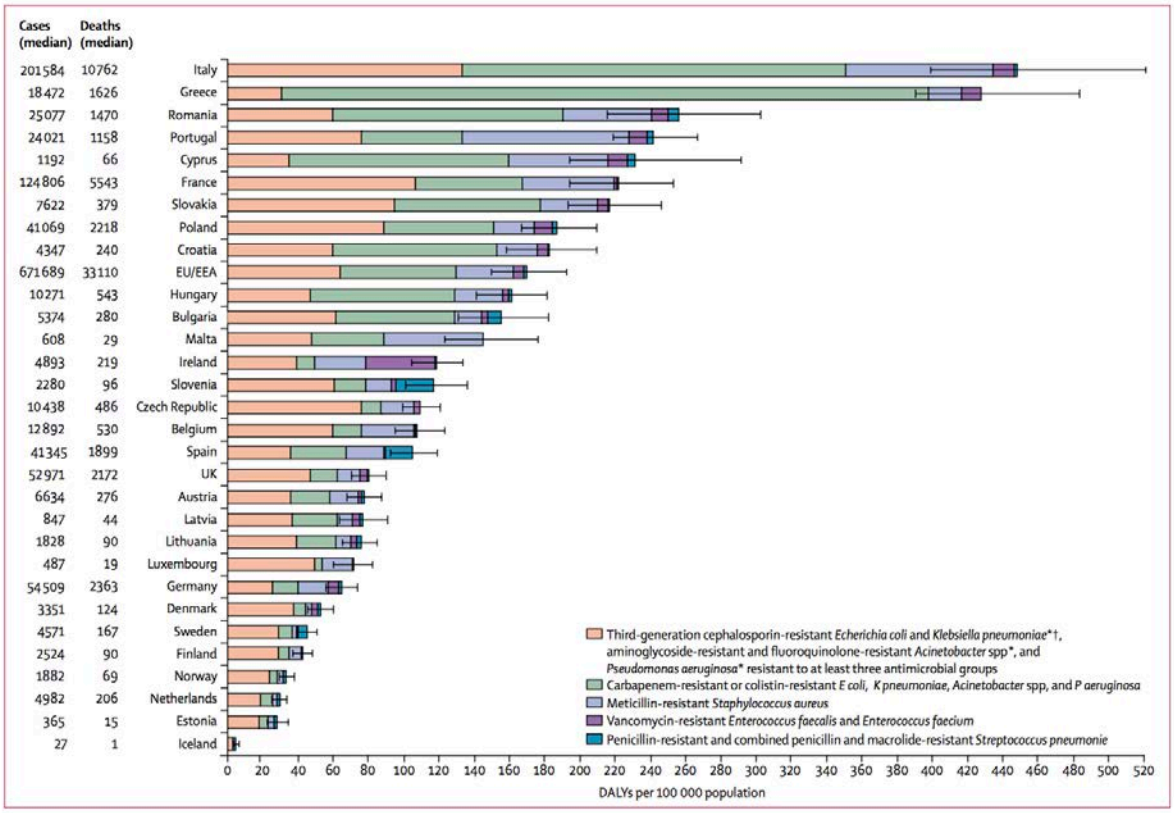


Figure 1.2 Burden of infections with antibiotic-resistant bacteria in DALYs, EU and European Economic Area, 2015
DALYs stands for Dissability Adjusted Life Years (Image retrieved from Cassini *et al.,* 2011).

It has been proven that the repercussion of antibiotics remains in the environment for extended periods of time, even if they have a short-term impact on the human micro biome. Antibiotic resistant strains can endure in the human host environment in dearth of selective pressure (Jernberg *et al.,* 2010).

This is the consequence of the enormous genetic flexibility of bacterial pathogens that produce changes in gene expression, the procurement of foreign genetic material or adaptational mutations, survival of the fittest at its best. This has generated antibiotic resistance, and this is why it is important to understand the genetic and biochemical mechanisms to be able to create novel therapeutic solutions for drug resistant pathogens (Munita & Arias, 2016). Today, the threat that arises from antibiotic resistant pathogens joint with the reduced number of new antibiotics discovered recently needs to be solved. This can be done by utilizing the knowledge that comes from studying the mechanisms in which bacteria respond to drug-stress, as well as the newfound information acquired from new clinical treatments and approaches (Kohanski *et al*., 2010).

Antibiotic resistance can be acquired through any of the following mechanisms: i) by mutating the gene or genes associated with the mechanism of action of the compound or ii) through new mutations in their genome or through Horizontal Gene Transfer (HGT) (Madigan *et al.,* 2019; Munita & Arias, 2016).

1.2.3. Mutational resistance

With this mechanism, susceptible bacterial cells develop mutations in genes where that interact with the compound, which translates into the survival in presence of said compound. Once a resistant mutant appears, the antibiotic kills all sensitive bacteria and the resistant mutant prevails (Munita & Arias, 2016). Mutations that arise from these mechanism usually affect the antibiotic compound in the following manners: a) Producing new enzymes that deactivate or destroy the compound; b) Diminishing or eliminating the entrance of said compound; c) By modifying the compound's target and decreasing or eliminating the binding with the compound; d) Altering the metabolic chemistry in which it is sensitive to the compound; and finally e) By expulsing the compound from the cell before it is active (Bauman *et al*., 2011).

1.2.4. Horizontal Gene Transfer (HGT)

HGT is a rapid evolutive mechanism by which entire genes are transferred among bacterial cells, thus enabling an almost immediate adaptation to new environmental conditions (Ochman *et al*., 2000). It happens so often that it is considered the potential precursor in pathogenic and non-pathogenic bacteria as a potential threat (Wright, 2007).

Acquiring foreign DNA material through HGT is one of the paramount drivers of bacterial evolution and often terminates in the development of antimicrobial resistance. Antimicrobial agents commonly derive from products found naturally in the environment. There is evidence the bacteria that coexist with these molecules nurse inherent genetic determinants of resistance, giving birth to an "environmental resistome" (Munita & Arias, 2016).

1.2.4.1. Models of Horizontal Gene Transfer (HGT)

There are three ways in which prokaryotes can obtain DNA from unfamiliar organisms: transduction, conjugation and transformation.

1.2.4.1.1. Conjugation

First discovered in 1946, by Lederberg and Tatum while studying two strains of *E. coli* with distinct nutritional requirements (Lederberg & Tatum, 1946). Conjugation is a process where DNA is transferred from one organism to another through direct contact, pores or through the creation of a conjugation bridge (Griffiths *et al.,* 2015). This process is illustrated in Figure 1.3.

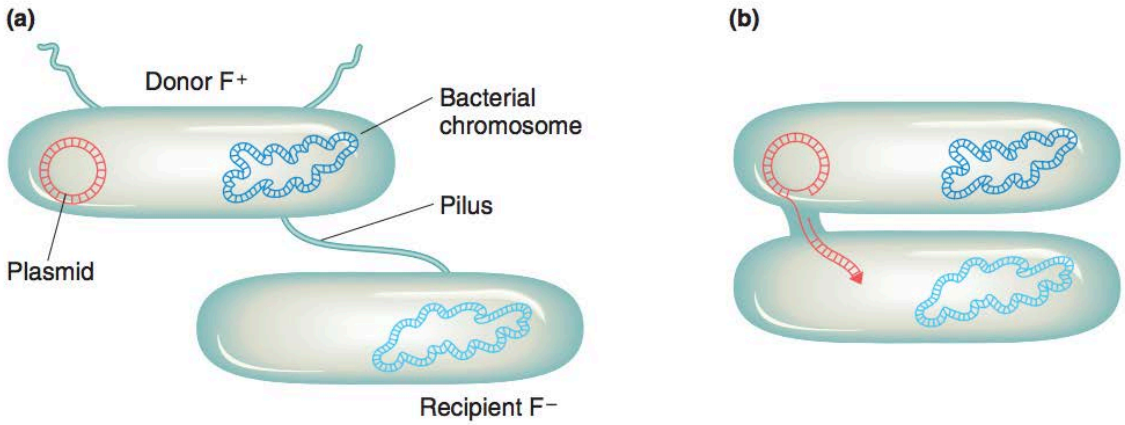


Figure 1.3 Mechanism of conjugation
1) The pilus pulls two bacteria together. 2) A bridge forms between the two cells, then one strand of plasmid DNA enters the recipient cell and each strand becomes double stranded again (Image retrieved from Griffiths *et al.,* 2015).

1.2.4.1.2. Transduction

First discovered by Norton Zinder and Joshua Ledererg in 1952 in Salmonella (Zinder & Ledereg, 1952). It is a process in which a virus or a viral vector introduces foreign DNA into an organism. This mechanism is consequence of an error; while the bacteriophage was trying to integrate its own DNA into the bacterial chromosome and create new viral particles, it included part of the bacterial DNA; which also gets transported into another bacteria and integrated into the new viral envelope. (Kokjohn *et al.*, 1989). This only happens in a small minority of the phage progeny, only 1 in 10,000 carry donor genes (Griffiths *et al.,* 2015). Bacterial transduction is illustrated in Figure 1.4.

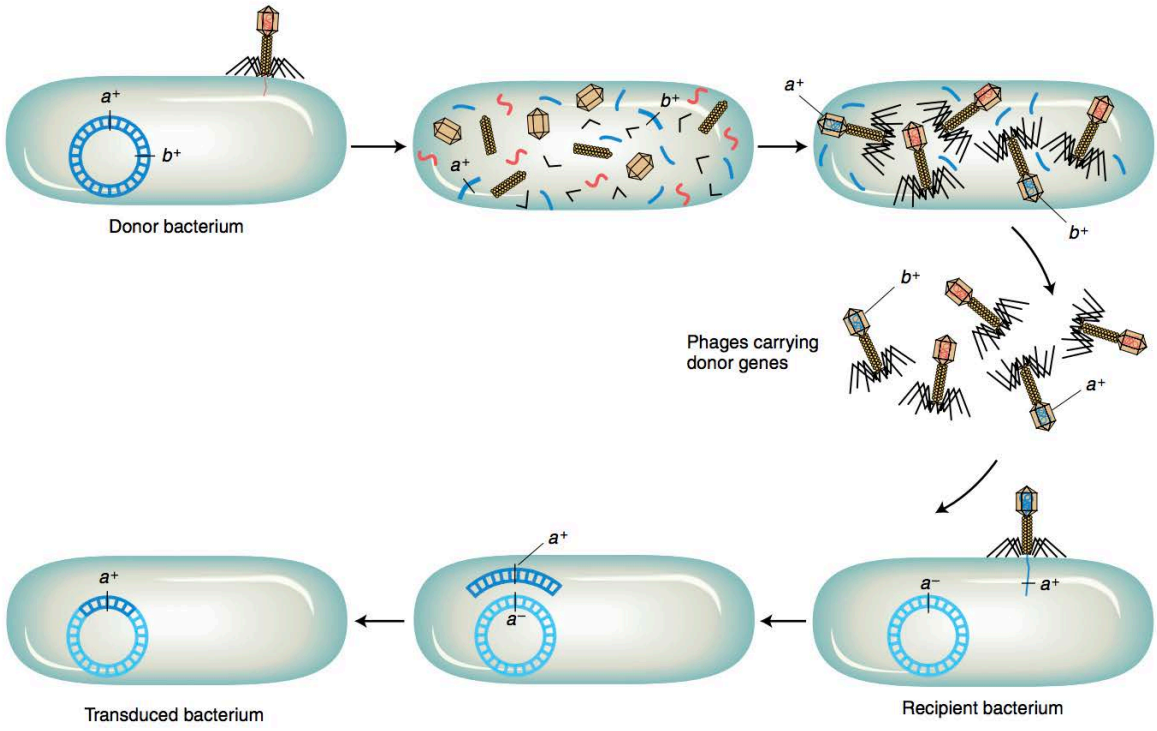


Figure 1.4 Mechanism of generalized transduction
1)Virus infects donor cell. 2) Donor cell integrates viral DNA into its own genome and creates new viral envelopes and viral DNA. 3) Viral DNA is packaged into the viral envelopes but some contain bacterial DNA as well. 4) Recipient cell is infected by virus carrying bacterial DNA. 5) Recipient cell integrates viral DNA and the bacterial DNA attached into its own genome (extracted from Griffiths *et al.*, 2015).

1.2.4.1.3. Transformation

Frederick Griffith first described this mechanism in 1928, using *S. pneumoniae* as a model organism (Lorenz & Wackernagel, 1994). Bacterial transformation is the process by which bacteria take up naked DNA from their environment and integrate it to their chromosome (Chen & Dubnau, 2004; Krüger & Stingl, 2011). Transformation does not require proximity between cells or between cells and viruses (Chen *et al.*, 2005). Figure 1.5 describes this mechanism.

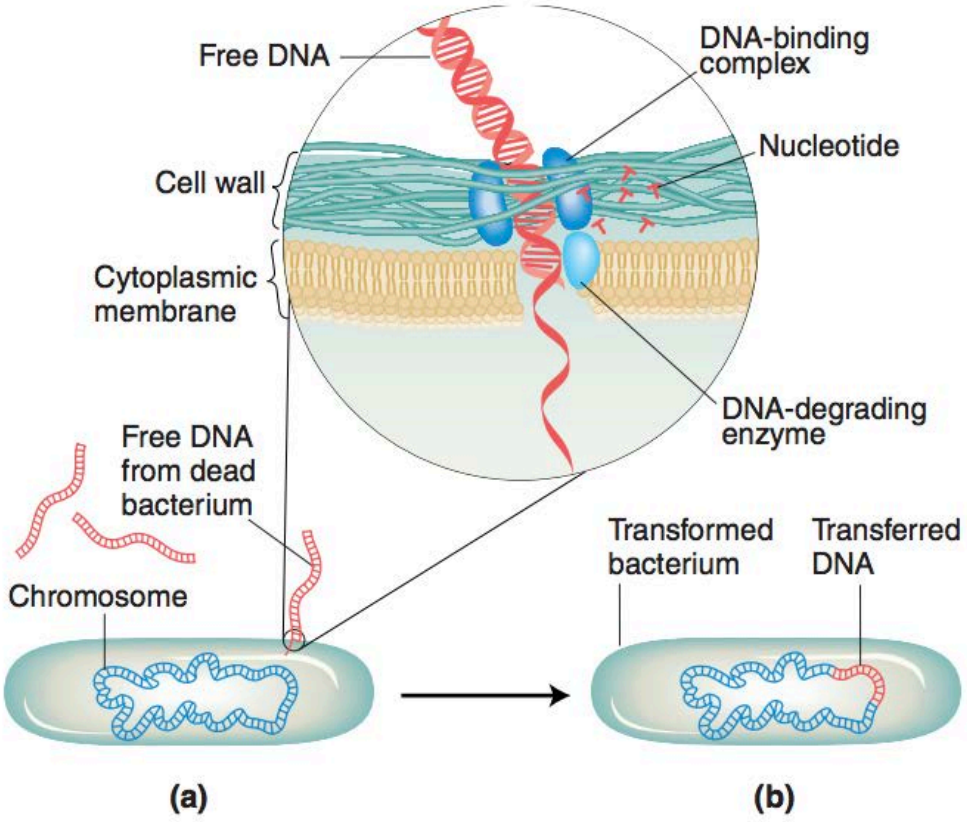


Figure 1.5 Mechanism of Transformation
1) Bacterium undergoing transformation picks up the foreign DNA from the environment. 2) While DNA binding complexes on the bacterial surface take up the DNA, one of the strands of it is broken down into nucleotides, and then a derivative of the other strand integrates into the now transformed bacterium (extracted from Griffiths *et al.,* 2015).

1.2.4.1.3.1. Natural Transformation

Natural transformation occurs in many different species of bacteria as well as in other organisms, to our knowledge 67 prokaryote species are capable of undergoing natural transformation (Johnsborg *et al.*, 2007), but they need to become competent for transformation to happen. Competence is a special physiological state that grants the ability of taking up DNA from the environment, and it is tightly regulated by specific physiological and environmental conditions (Chen & Dubnau, 2004); an example of this is *B. subtilis*, as the development of competence in this bacterium requires the activation of over 40 genes (Solomon & Grossman, 1996).

Natural transformation comprises five stages: (i) induction to competence; (ii) DNA binding; (iii) DNA fragmentation; (iv) DNA entrance; and (v) integration of the new DNA into the genome or into a plasmid (Averhoff, 2004).

The structure of most proteins involved in these five stages is still unknown, and hence the precise molecular mechanisms in which they are involved remain elusive. DNA uptake proteins (DUPs) are the translocating machineries used by bacteria to incorporate naked DNA from the environment to the cytoplasm during bacterial transformation as shown in Figure 1.6.

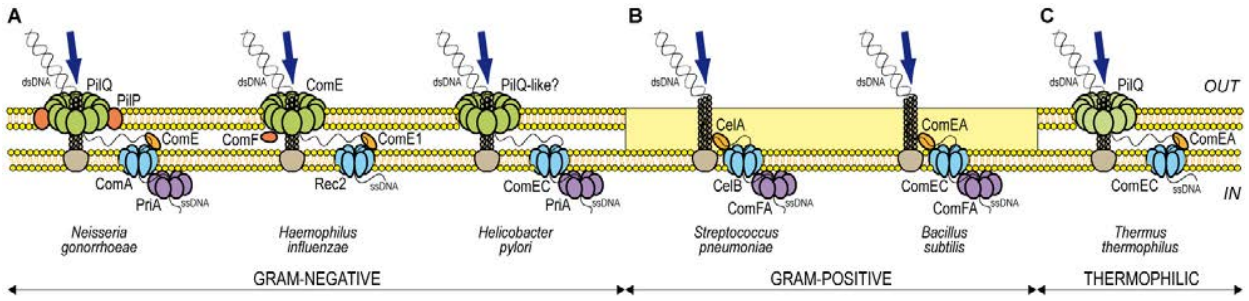


Figure 1.6 Diagram of the DUPs from six model organisms
In gram-negative bacteria (Figure 1.5 A) DNA is transported across the external membrane through a Type IV pilus (Tfp) channel, being the dodecameric, membrane-embedded PilQ (900 kDa) the portal protein, which is in some species assisted by ancillary proteins (PilP or ComF). The DNA is subsequently transferred to an inner membrane associated complex comprising a DNA-binding protein, a porine and an ATPase: the DNA-binding protein (ComE or ComE1) leads the DNA across the periplasm to the inner membrane channel (ComA, ComEC orRec2) through which one single strand of the DNA enters the cytoplasm by being pulled by the ATPase (PriA). In gram-positive bacteria (Figure 1.5 B) external DNA binds to a pseudopilus that spans the cell wall. Disassembly/retraction of the pseudopilus opens a hole in the cell wall, thereby enabling DNA to diffuse from the surface to an inner-membrane translocating complex that resembles that of the gram-negative bacteria. The DNA receptor (ComEA, or CelA) is anchored to the membrane and the possibly-dimeric channel protein (ComEC or CelB) is associated to a hexameric ATPase (ComFA). Thermophiles (Figure. 1.5 C) use a system similar to that of gram-negative bacteria (Averhoff, 2004).

1.3. Bacterial Signal Transduction

Bacterial use signal transduction systems to receive signals from environmental changes and create a response. These signal transduction systems can be divided into two groups: Two-component systems and one-component systems (Laub *et al.,* 2007).

One-Component Systems are composed by one protein that contains a sensory domain and an output domain (Ulrich *et al.,* 2005).

Two-Component Systems (TCS) are comprised by a sensor Histidine Kinase (HK) and a Response Regulator (RR) (Stock *et al.,* 2000). The HK auto-phosphorylates its inner domain, after this, the receiver domain of the RR is phosphorylated and this in turn, activates the outer domain, triggering the expression of a certain gene or a cellular physiological process through protein-protein or protein-DNA interactions. This is why this is such a paramount system for bacteria to adapt to changes in the environment. Figure 1.7 illustrates a two-component system (Hoch *et al.,* 2005).

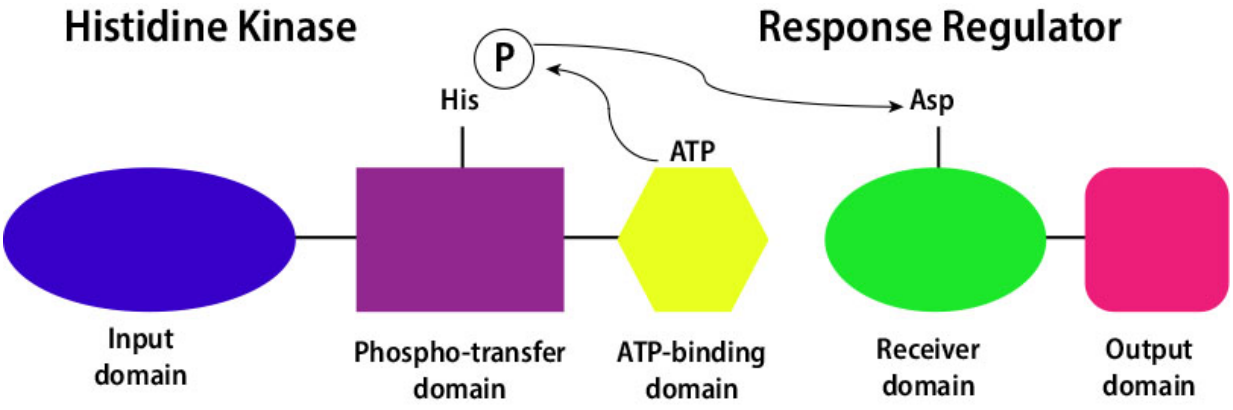


Figure 1.7 Two-component System schematic
After sensing the environmental signal, the input domain of the HK activates the autophosphorylation of the phospho-transfer subdomain of the autokinase; then the high-energy phosphoryl group is transferred to the aspartate residue in the receiver domain of the RR resulting in the activation of a specific gene.

There are three known architectures that extend beyond linear pathways for TCS systems that elaborate on the complexity of these systems in the mechanisms of survival in bacteria as illustrated in Figure 1.8 (Goulian, 2010).

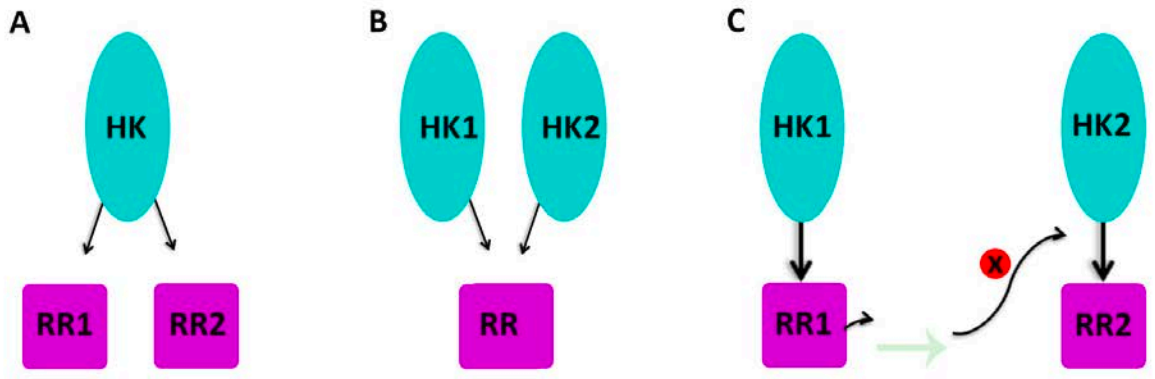


Figure 1.8 Two-component System non-linear architecture
A) One-to-many pathway. One HK can phosphorylate two RR. B) Many-to-one pathway. Two HK phosphorylate the same RR. C) Connector-mediated pathway. The HK1/RR1 TCS activates a connector protein X, which modulates the activity of the HK2/RR2 system. X can act either on the HK or RR by decreasing or increasing phosphorylation.

1.3.1. Elements of a Two-Component System (TCS)

As mentioned before, a TCS is conformed by a Histidine Kinase and a Response Regulator. In this section we will describe them.

1.3.1.1. Histidine Kinases (HK)

HK are a large family of signal transduction enzymes that have the capacity of autophosphorylating a conserved histidine residue that is used to phosphorylate the conserved aspartate domain of the RR they form a TCS with. They are consistently formed by a C-terminal cytoplasmic signaling domain and a N-terminal extracellular sensing domain with a transmembrane receptor (Wolanin *et al.,* 2002).

Generally speaking, there are two classes of HKs, orthodox and hybrid kinases. Orthodox HKs usually function as periplasmic membrane receptors, a signal peptide and possess transmembrane segments that separate the protein into a highly conserved C-kinase core and a periplasmic N-terminal sensing domain. While Hibrid kinases have multiple phosphoacceptor and phosphodonor sites and instead of using a single phosphoryl transfer, they use a multi-step phosphorelay system. A sensor domain, a kinase core, a

receiver domain and a His-containing phosphotransfer domain compose them (Parkinson & Kofoid, 1992; Wolanin *et al.,* 2002).

1.3.1.2. Response Regulator (RR)

Response regulators usually have a receiver domain and one or more output domains. After being phosphorylated by the HK, they suffer a conformational change that alters the function of the output domains, which can be translated into an increase in the activation of target genes (Stock *et al*., 2000; West *et al*., 2001). There are over 80,000 RR proteins annotated in the databases, they display a great diversity in type and structure of their output domains. The RR superfamily can be classified according to the nature of how their output domain exerts its response in: 1) DNA-binding, 2) RNA-binding, 3) enzymatically active, 4) single domain RR proteins and finally 5) protein-binding. The percentage distribution of the subclasses within the RR superfamily is illustrated in Figure 1.9 (Zschiedrich *et al., 2016).*

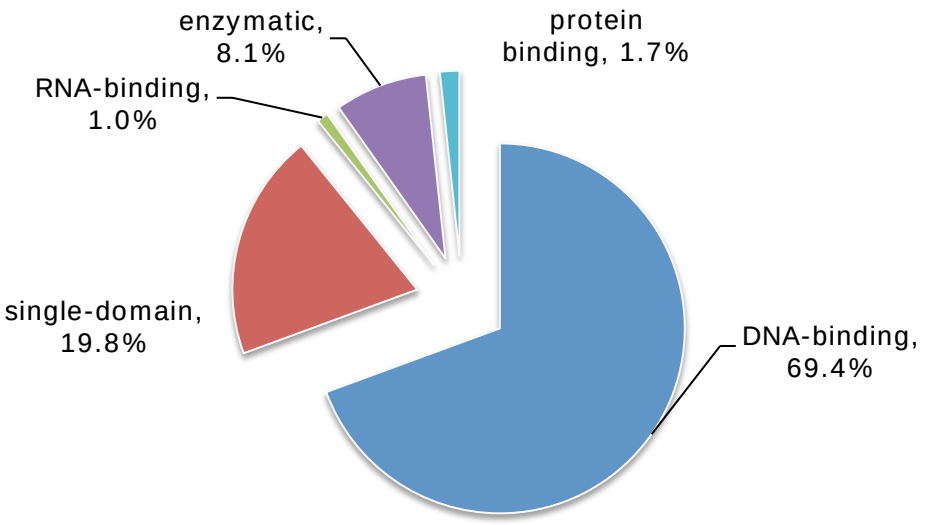


Figure 1.9 Percentage distribution of subclasses within the RR superfamily
(Adapted from Zschiedrich *et al.,* 2016).

1.4. Transformation in *S. pneumoniae*

Competence is developed in *S. pneumoniae* under the control of a quorum-sensing signaling pathways that involves the ComD-ComE two component signal transduction system formed by a ComD. which is a Histidine Kynase (HK) and ComE, that in turn is a Response Regulator (RR) and the auto synthesized competence-stimulating peptide (CSP) pheromone (Boudes *et al.*, 2014). ComE belongs to the LytTR subfamily of the RRs (Nikolskaya and Galperin, 2002). Competence is also the moment in which the transformasome is created, this dynamic engine implicates cytosolic and membrane proteins, while incorporating, protect and process transforming DNA (Claverys *et al*., 2009). This thesis will extend on the implications of this molecular engine and its mechanism for bacterial transformation.

According to current understanding, ComD senses the CSP leading to ComD being activated, and autophosphorylating residue H248, and then ComE is phosphorylated in residue D58; after, ComE ~P then binds to two imperfect direct repeats DR1 and DR2 (9 bp each, separated by a 12-mer link) in the promoter regions and induces the expression of the *comAB, comCDE* and c*omX* operons (Claverys & Håvarstein, 2002). These operons and their sequence are shown in Figure 1.10.

Figure 1.10 Sequence alignment of the three double site DR1-DR2 promoters of ComE
Conserved bases are shown in red (adapted from Claverys & Håvarstein, 2002).

These promoter regions also share a great homology amonst different species of streptococci as shown in Figure 1.11 (Ween *et al.,* 1999).

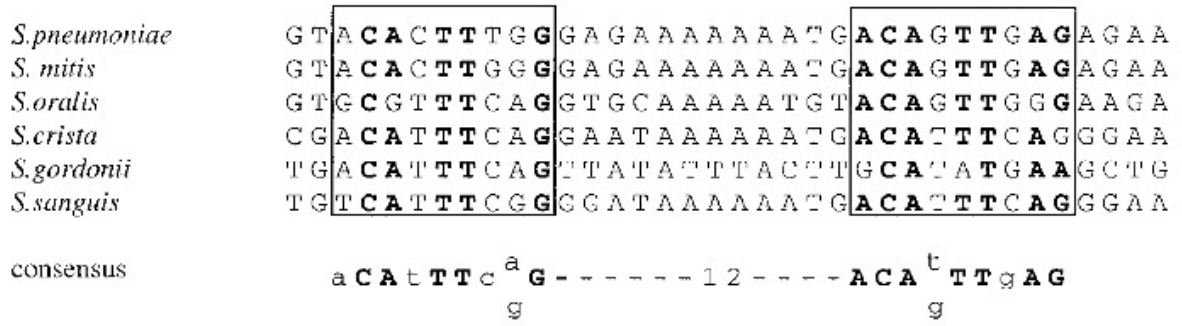


Figure 1.11 Sequence alignment of six different species of streptococci DR1-DR2 promoters of ComE

Alignment of imperfect direct repeat motifs present in the *comCDE* promoter regions of six different species of streptococci (extracted from Ween *et al.,* 1999).

Two mutants have been designed to mimic phosphorylated and nonphosphorylatable versions of ComE (ComE^{D58E} and ComE^{D58A}, respectively), this has shown that the D58A mutant abolishes basal *comCDE* expression, whilst the D58E mutant display a full competence state (Sanchez *et al.,* 2015). In addition to this, the supplement of CSP does not stimulate *comCDE* in these mutants, indicating that the transcription of the *comCDE* operon is regulated by ComE (Ween *et al.,* 1999).

1.5. Bacterial Signal Transduction: A target for antimicrobial therapy

The world is facing a post-antibiotic era and the number of chemotherapeutic agents and vaccines available to battle infectious disease is limited (Worthington *et al.,* 2013). TCSs are exceptional targets for the creation of therapeutics against bacterial infections; this is because bacterial TCS are very important signaling mechanisms for the diffusion, survival and colonization of pathogenic bacteria. Some compounds that target TCSs that have been predicted and experimentally validated are listed in Table 1.2 (Tiwari *et al.,* 2017).

Table 1.2 Two-Componen Systems targeted y molecular compounds

Bacteria	Inhibitor	TCS	Component of TCS	Mechanism of action	Reference
Pseudomonas aeruginosa	Thiazole derivatives	Algr1/ Algr2	Sensor protein	Inhibition of phosphorylation/dephosphorylation of Algr2	Roychoundhury *et al.,* 1993
Enterococcus faecium	Thiazole derivatives	VanR/ VanS	Sensor protein	Inhibition of autophosphorylation	Ulijasz & Weisblum, 1999
Bacilus subtilis	Walkmycin B and Waldiomycin	WalK/ WalR	Sensor protein	Binds to the HK cytoplasmic domain or the inhibition of the autophosphorylation	Okada *et al.,* 2010
Staphylococcus aureus	Walkmycin B and Waldiomycin	WalK/ WalR	Sensor protein	Binds to the HK cytoplasmic domain or the inhibition of the autophosphorylation	Okada *et al.,* 2010
Staphylococcus aureus	Salicyanilide	KinA/ SPO0F	Sensor protein	Affects membrane fluidity, disturbing signal transduction	Hilliard *et al*., 1999
Bacilus subtilis	Unsaturated fatty acids	KinA	Sensor protein	Causes non-competitive inhibition of ATP-dependent autophosphorylation	Barrett & Hoch, 1998
Pseudomonas aeruginosa	Thiazole derivatives	Algr1/ Algr2	RR	Inhibition of DNA-binding activity of Algr1	Roychoundhury *et al.,* 1993
Coryneacterium pseudotuberculosis	Rhein	PhoP/PhoR	RR	Inhibition of conserved receiver domain of PhoP	Tiwari *et al.,* 2014
Methicillin-resistant Staphylococcus aureus	Bis-phenol	VanR/ VanS	RR	-	Barrett & Hoch, 1998
Salmonella enterica	NSC9608 (8 compounds NCI library)	PhoP/ PhoR	RR	Inhibition of formation of the PhoP-DNA complex	Tang *et al.,* 2012

Cross-talk is when cross-phosphorylation occurs between two distinct signal transduction pathways, and it can be another approach on modulating the regulation of TCSs (Goulian, 2010). Therefore, it is also possible to speculate that another TCS can be used as a “control system” to regulate a target TCS.

Although it might be easy to predict inhibitors, because of the conservation of domains present in HK and sensor proteins, there is significant similarity between eukaryotic related proteins and bacterial TCSs. This generates

concerns regarding the effectiveness of TCS-targeting drugs and their effects over the host's well-being. This is why, understanding the nature of the interactions between a TCS, promoters, and its targeting compounds would benefit from the improved molecular structure awareness by creating a ligand with an increased specificity (Tiwari *et al.,* 2017).

2. Objectives ComD-ComE project

The objectives of this section were:

2. The production and preliminary biochemical characterization of the proteins and protein complexes composed by ComD and ComE with and without DNA.

3. To lay the ground work for the structural characterization of these proteins and complexes for bacterial transformation in the gram-positive bacterium *Streptococcus pneumoniae,* as a model species with proteins likely to be amenable for their structural characterization.

3. General Methods for ComD-ComE

Methods that are commonly used in molecular biology were used throughout this thesis. They will be explained in this section.

3.1. Culture media, bacterial strains and vectors

In this section we will present the culture media, bacterial strains and vectors used to perform the experiments for this part of the thesis.

3.1.1. Culture media

Composition of the culture media and additives used as shown:

- Luria-Bertani Broth Culture media (LB): Triptone 1%, NaCl 1%, Yeast extract 0,5%. (Autoclave sterilization)
- Double Luria-Bertani Broth Culture media (2× LB): Triptone 2%, NaCl 2%, Yeast extract 1%. (Autoclave sterilization)
- LB plates: Triptone 1%, NaCl 1%, Yeast extract 0,5%, bacterial-agar 1,5%. (Sterile conditions)
- IPTG Stock Solution 1M. (Sterilizing filtration)
- Antibiotics (Sterilizing filtration):
  - Kanamycin (Stock solution 50 mg/ml in ddH_2O, final concentration: 50 µg/ml)
  - Ampicillin (Stock solution 100 mg/ml in ddH_2O, final concentration: 100 µg/ml)
  - Chloramphenicol (Stock solution 35 mg/ml in ddH_2O, final concentration: 35 µg/ml)

- Autoinduction broth (Studier, F.W.; 2005)
  - 1M $MgSO_4$ (Filtered)
  - 1000× Trace Metals (1 l) (Filtered)

5M HCl	8 ml
$FeCl_2 \cdot 4H_2O$	5 g
$CaCl_2 \cdot 2H_2O$	184 mg
H_3BO_3	64 mg
$CoCl_2 \cdot 6H_2O$	18 mg
$CuCl_2 \cdot 2H_2O$	4 mg
$ZnCl_2$	340 mg
$Na_2MoO_4 \cdot 2H_2O$	605 mg
$MnCl_2 \cdot 4H_2O$	40 mg

  - 50× 5052 (200 ml) (Heat sterilized)

25% (v/v) glycerol	54.47 ml (stock 87%)
2.5% (w/v) glucose	5 g
10% (w/v) lactose	20 g

  - 50× M (250 ml) (Filtered)

1.25 M Na_2HPO_4	44.36 g
1.25 M KH_2PO_4	42.52 g
2.5 M NH_4Cl	33.43 g
0.25 M Na_2SO_4	8.87 g

3.1.2. Bacterial strains

Different *E. coli* strains were used for cloning and expression (Table 3.1).

Table 3.1 Bacterial cell strains
Bacterial cell strains used during cloning and protein expression.

Cell strain	Genotype	Remarks	Reference
BL21(DE3)	*B F⁻dcm ompT hsdS(r_B-m_B) gal λ(DE3)*	Used for protein expression. BL21-derived with a chromosomal copy of the gene for lacIq and for the T7 RNA polymerase	Invitrogen
BL21(DE3) T1^R	*B F⁻dcm ompT hsdS(r_B-m_B) gal λ(DE3) tonA*	Strain for protein expression. BL21-derived with a chromosomal copy of the gene for lacIq and for the T7 RNA polymerase. Resistant to bacteriophages T1 and T5	Invitrogen
BL21(DE3)pLysS	*B F⁻dcm ompT hsdS(r_B-m_B) gal λ(DE3) [pLysS Cam^r]*	Strain for protein expression. It expresses low levels of T7 lysozyme and inhibits basal leves of T7 RNA polymerase.	Invitrogen
DH5α	F⁻ϕ80d*lac*ZΔM15 Δ(lacZYA-argF)U169 *end*A1 *rec*A1 *hsd*R17(r_K-m_K+) *deo*R*thi*-1 *sup*E44λ^- *gyr*A96 *rel*A1	Used for general cloning	Invitrogen
XL1-Blue	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M1 5]* hsdR17(r_K- m_K+)	Supercompetent cells. Used for cloning	Agilent

3.1.2.1. Preparing competent cells

Buffers were prepared and sterilized as follows:

- **Transformation buffer 1 (Tfb1)**: 30 mM KOAc, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 3 mM $[Co(NH_3)_6]Cl_3$, 15% glycerol.
- **Transformation buffer 2 (Tfb2)**: 10 mM 3-(*N*-morpholino) propanesulfonic acid (MOPS), 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol.

E.coli cells were grown in LB at 37 ºC at 220 rpm overnight (o/n). Resulting culture was diluted by a factor of 10^2 and re-grown until optical density (O.D.) reached between 0.25 and 0.3. Cultures were then left on ice for 5 minutes, then the sample was centrifuged for 5 mins at 4,000 × g and 4 ºC. Pellet was resuspended in 1/5 of the initial volume of Tfb1 and supernatant was discarded. Once again, cells were placed on ice for 5 minutes and centrifuged again for 5 mins at 4,000 × g and 4 ºC. Pellet was gently resuspended in Tfb2 (1/10 of the volume of Tfb1) and incubated on ice for 15 minutes. Cells are now ready to be aliquoted and flash frozen in liquid nitrogen, before being stored at -80 ºC.

3.1.2.2. Bacterial transformation

Around 50 ng of vector DNA were added to 50 µl of competent cells in a small 1.5 ml eppendorf tube, gently mixed with a pippette and then placed on ice for 30 minutes. The tube was then heat shocked for 45 seconds at 42 ºC by using a thermoblock. Then sample was placed on ice for 5 minutes, then 1 ml of LB was added and the tube was incubated at 37 ºC and 220 rpm for 60 minutes. Sample was centrifuged and resuspended in 100 µl of LB and plated on LB-agar + antibiotic (depending on the resistance of the vector) for positive clone selection and kept at 37 ºC o/n.

3.1.3. Vectors

Each gene was inserted into a pOPIN vector using the In-fusion technique with a N-terminal-tag (Table 3.2). The protein was expressed fused to the

sequence contained in the vector. Vectors were obtained from the Protein Expression Core Facility at IRB.

Table 3.2 Expression vectors
Vectors used for protein overexpression (Berrow NS *et al.*, 2007).

	pOPINF	pOPINJ	pOPINM	pOPINS
Used for	ComE, ComX	ComE	ComE	ComD, ComW
Length (bp)	5457	5457	5457	5906
Promoter	T7 promoter	T7 promoter	T7 promoter	T7 promoter
Terminator	T7 terminator	T7 terminator	T7 terminator	T7 terminator
Protein tags	N-terminal His-3C-lacZ	N-terminal His-GST-3C-lacZ	N-terminal His-MBP-3C-lacZ	N-terminal His-SUMO-POI
Fused sequence	N.terminal : MAHHHHHH SSGLEVL	N.terminal :MAHH HHHH_GST_LEVLF QGT	N.terminal :MAHHHHHH SSG_MBP_LEVLFQGP	N.terminal : :MAHHHHHH_SUMO_ LEVLFQG
5´ cloning site	*Nco*I restriction site	*Nco*I restriction site	*Nco*I restriction site	*Nco*I restriction site
3´ cloning site	*Msc*I restriction site	*Kpn*I restriction site	*HindIII* restriction site	*Kpn*I restriction site
Antibiotic resistance	Ampicillin	Ampicillin	Ampicillin	Kanamycin
Reference	Addgene	Addgene	Addgene	Addgene

Expected protein parameters were computed with ProtParam (Gasteiger *et al.*, 2005). Protein secondary structure prediction based on position-specific scoring matrices was carried out using the Psipred server (Buchan *et al.*, 2019).

3.2. Protein expression

Initially, protein expression was performed using BL21(DE3) *E. coli* cells, this was changed for BL21(DE3) $T1^R$ to avoid bacteriophage contamination for all the target proteins except for ComW for which BL21(DE3) pLysS was used because it increased yield. Pre cultures were grown with 100 ml of LB in 500 ml Erlenmeyer flasks. One bacterial colony was added to the flask with 100 µl of the corresponding antibiotic (depending on the vector's resistance). This flask was incubated o/n at 220 rpm and 37 ºC. The next day, this was added to 1.5 l of LB in 5 l Erlenmeyer flasks and 1.5 ml of the corresponding antibiotic and placed in an Innova 44 (New Brunswick Scientific) incubator at

220 rpm and 37 ºC. When an O.D. of 0.6 was reached, 0.4 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. Then temperature was lowered to 16 ºC and kept incubating for 20 h.

If using auto-induction the media was prepared as follows:

For 500 ml of media:

o	Tryptone	5 g
o	Yeast extract	2.5 g
o	MiliQ Water	475 ml
o	1M $MgSO_4$	1 ml
o	1000× Trace metals	1 ml
o	50× 5052	10 ml
o	50× M	10 ml
o	Antibiotics	0.5 ml
o	Pre-culture	20 ml

Pre culture and antibiotics would be added to the previously described media, and grown for 3 hours at 220 rpm and 37 ºC. Then temperature was slowly lowered from 37 ºC to 16 ºC and the cells were incubated for 36 hours.

In both cases, to retrieve the bacterial pellet with the expressed proteins, the cells were centrifuged at 5,000 × g for 20 mins at 4 ºC using a Beckman Coulter Avanti J-26 XPI centrifuge. Supernatant was discarded and pellet kept at -20 ºC.

3.3. Protein purification

The process in which we obtained a pure sample to use for protein characterization and crystallization is described in this section.

3.3.1. Cellular lysate extraction

Pellets frozen at -20 ºC, obtained from the protein expression cultures, were resuspended in buffer A (500 mM NaCl, 30 mM imidazole, 5% v/v Glycerol, 5 mM β-mercaptoethanol, 20 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA). Then, 50 µl of DNase I and One Complete protease inhibitor cocktail tablet (Roche) were added to the resuspended sample. Cells were lysed by operating a cell disruptor (Constant Systems Ltd.) at 25 kpsi. Sample was run through the cell disruptor twice, and then centrifuged at 25,000 × g for 30 minutes. Subsequently, supernatant obtained was filtered with 0,22 µm filters (Millex) and pellet was discarded. Sample was kept on ice to avoid denaturation.

3.3.2. Chromatography steps

Protein purification was achieved through a multi-step chromatographic protocol using ÄKTA Purifier systems (GE Healthcare) or NGC Chromatography Systems (Bio-Rad) using the columns below, following manufacturer instructions (GE Healthcare) at 4 ºC:

- HisTrap HP (5 ml) (Immobilized metal ion affinity chromatography (IMAC)):

Column was washed with 5 Column Volumes (CV) of H_2O, then washed with 5 CVs of buffer A, another 5 of buffer B (500 mM NaCl, 500 mM imidazole, 5% v/v glycerol, 5 mM β-mercaptoethanol, 20 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA), and finally 10 CVs of buffer A. Protein sample was loaded into the system and then column was washed with 10 CVs of buffer A, after this, protein was eluted using a linear gradient of 15 CV of buffer B. A 5 CV wash of buffer B was performed to ensure that all protein was released from the column.

- Protease treatment

Protease treatment for ComD and ComW was performed by measuring the total yield of protein after the first step of purification (Histrap) and then adding 1:100 SUMO protease (IRB Protein Expression facility). Temperature, time and shaking conditions were optimized for each protein. Sample was left o/n at 20 ºC divided in 10 ml falcon tubes and then centrifuged to remove any precipitate. Then the sample was loaded again into the column

to do a reverse Histrap and the flowthrough concentrated and injected into the Size Exclusion Chromatography column.

- Histrap Heparin (5 ml)

Column was washed with 5 Column Volumes (CV) of H_2O, then washed with 5 CVs of buffer AH (200 mM NaCl, 5% v/v glycerol, 5 mM β-mercaptoethanol, 20 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA), another 5 CVs of buffer H (1000 mM NaCl, 5% v/v Glycerol, 5 mM β-mercaptoethanol, 20 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA), and finally 10 CVs of buffer AH. Protein sample was loaded into the system and then column was washed with 10 CVs of buffer AH. After this, protein was eluted using a linear gradient of 15 CV of buffer H. A 5 CV wash of buffer H was performed to ensure that all protein was released from the column.

- Superdex 200 gl (10/300) increase (Gel filtration)

Sample was filtered or centrifuged at 16,000 × g for 15 minutes at 4 ºC. Sample was introduced and eluted using 1.2 CV of buffer AH. Molecular weight (MW) standards were used to analyze the eluted peak according to manufacturer instructions (GE Healthcare).

Protein purity was checked by the SDS-PAGE gel of each chromatography and by analyzing the gel filtration chromatogram.

3.3.3. Protein visualization and verification

The steps followed verify the presence of the target proteins are described in this section.

3.3.4. Protein electrophoresis

A denaturating protein acrylamide gel electrophoresis was performed in the presence of Sodium Dodecyl Sulphate Polyacrylamide Gel (SDS-PAGE) to visually analyze protein fractions obtained from the chromatography steps. Gels were prepared as follows:

Stacking gel: 5% [w/v] acrylamide, 0.13% [w/v] bis-acrylamide, 0.1% [w/v] SDS in 0.13 M Tris-HCl (pH 6.8). Polymerization in presence of 0.75%

[w/v] initiator ammonium persulfate (APS) and 0.125% [v/v] crosslinking reagent N, N, N', N'-tetramethylethylene-diamine (TEMED).

Separative gel: 10% [w/v] acrylamide, 0.27% [w/v] bis-acrylamide, 0.1% [w/v] SDS in 0.38 M Tris-HCl (pH 8.8). Polymerization in presence of 0.5% [w/v] APS and 0.05% [v/v] TEMED.

Protein samples were diluted with loading buffer (bromophenol blue in 25 mM Tris-HCl (pH 6.8), 5% [w/v] SDS, 10% [v/v] glycerol and 5% [v/v] β-mercaptoethanol) and heated at 95 ºC for 5 minutes.

SDS-PAGE was performed using electrophoresis tanks and a power source (Bio-Rad) with running buffer (25 mM Tris-HCl, 0.2M glycine and 0.1% [w/v] SDS) at 200 V. PageRuler Prestained Protein Ladder (Thermofisher) was used as a protein MW marker.

Gels were removed from the tanks and stained with "one step blue" (Biotium) for 60 mins, and then transferred to water and kept until scanning.

3.3.5. Protein concentration and quantification

Protein samples were concentrated using Vivaspin concentrators (GE Healthcare) of 0.5 ml, 4 ml and 15 ml with 10,000 Da MW cut-off. Protein concentration was measured using the Bradford quantification method (Bradford, 1976), using the Bradford reagent (Bio-Rad); or using a NanoDrop 1000 spectrometer and following manufacturer's instructions (ThermoFisher Scientifics).

3.3.6. Western blots

Western blots were performed using the Trans-Blot SDSemi-Dry Electrophoretic Transfer Cell (Bio-Rad) following manufacturer's recommendations. Transfer was performed using the following materials:

- Western blots with the Semi-Dry Electrophoretic Transfer Cell.
- Whatmann 3MM paper (GE Healthcare).
- Nitrocellulose or PVDF membranes (GEHealthcare).
- iBlot Dry Blotting System (Invitrogen): Nitrocellulose transfer stacks and blotting pads.
- Buffers
    - Transfer buffer (200 ml Methanol, 3 g TRIS, 3 g glycine, 3 ml 10% SDS, up to 1l H_2O).
    - TBS-Tween buffer (50 mM Tris, 150 mM NaCl and 0.05% Tween 20).
- Protein markers
    - Benchmark protein ladder (Invitrogen)
    - MW markers: PageRuler prestained protein ladder (Thermofisher)
- Anti-His primary antibody (1:1000 dilution in Tween/PBS)
- Anti-rabbit secondary antibody (1:10000 dilution in Tween/PBS)

After protein transfer (o/n at 0.4 V at 4 ºC), membrane was blocked by incubating with TBS-Tween buffer with 10% skim milk for an hour at room temperature, and washed thrice with TBS-Tween buffer for 10 minutes. Then it was incubated with TBS-Tween buffer with 5% skim milk, supplemented with 1 µl (1:1000 dilution) of the primary antibody for another hour. Once again membrane was washed thrice with TBS-Tween buffer, and then incubated the TBS-Tween buffer with 5% skim milk, supplemented with 0.5 µl of the secondary antibody (1:1000 dilution) for an extra hour and washed again thrice with the TBS-Tween buffer. Membrane was then imaged through quimioluminiscence using the ECL plus substrate (GE Healthcare) and developed using Kodak photographic paper using a Hyper processor revealing machine (Amersham Pharma Biotech). Membrane could also be imaged by scanning the membrane using the Odissey Infrared Imager (Li-Cor) by following the manufacturer's recommendations.

3.4. Protein co-expression

We tested to see if by co-expressing both proteins yield would increase and purification would be easier, the process is described in this section.

3.4.1. Construct design

To increase yield and facilitate purification, the co-expression of both coupled proteins (ComDE and ComWX) was attempted by using the pETDuet-1 (Table 3.3) as it is designed to co-express two target genes, and has Ampicillin resistance as well as the pBR322-derived ColE1 replicon and the *lacI* gene.

Table 3.3 Expression vector used for protein co-expression

Vector name	**pETDuet-1**
Promoter	T7 promoter × 2
Terminator	T7 terminator × 2
Protein tags	N-terminal His•Tag, C-terminal S•Tag
Antibiotic resistance	Ampicillin
Reference	Novagen

3.4.2. PCR and cloning

ComD, *comE*, *comW* and *comX* genes were amplified from *Streptococcus pneumoniae R6* contained in the pOPIN vectors. Primers used to clone full-length proteins (ComE, ComW and ComX) and the cytoplasmic domain of ComD are shown in Table 3.4. Restriction sites were introduced according to the vector cloning sites.

Table 3.4 Primers used for cloning

Primers	Sequence
SSumoDuet	AGGAGATATACCATGGGTAGCAGCCATCACCATCAT
SHistagDuet	AGGAGATATCATATGGCACACCATCACCACCATCA
SComW080	ATGGTCTAGAAAGCTTTAACAAGAAATAAACCCCCGATT
SComD441NotI	ATGGTCTAGAGCGGCCGCCTTTATTCAAATTCCCTCTTAAATCTAATG ATTTGTCTAAATGTACTGCCT
SComE250	ATGCTCGAGAAAGCTTTACTTTTGAGATTTTTTCTCTAAAATATCTTTT AATTTTCCTAATTTTTAATCTTG
SComX159	ATGGTCTCGAGAGCTTTAATGGGTACGGATAGTAAACTCCTTAAA
SComE250NoEnd	ATGCTCGAGCTTTTGAGATTTTTTCTCTAAAATATCTTTTAATTTTCCT AATTTTTAATCTTGAAATA
SComX159NoEnd	ATGGTCTAGCTCGAGATGGGTACGGATAGTAAACTCCTTA

3.4.3. Polymerase Chain Reaction and cloning

PCRs were performed in an Eppendorf Mastercycler Gradient Thermal Cycler using the PfuUltra II Fusion HS DNA polymerase (Agilent). Reaction mix was composed of the reagents shown in Table 3.5.

Table 3.5 Reaction mix for PCR

Component	Amount
10× PfuUltra II reaction buffer	5.0 µl
dNTP mix (25 mM each dNTP)	0.5 µl
DNA template (100 ng/µl)	1 µl
Primer #1 (10 µM)	1 µl
Primer #2 (10 µM)	1 µl
PfuUltra II fusion HS DNA polymerase	1 µl
Nuclease-free water	Up to 50 µl
	————————
Total reaction volume	50 µl

PCR cycle was defined according to manufacturer's recommendations (Table 3.6).

Table 3.6 PCR cycle conditions

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	95 °C	2 minutes	1
Denaturation	95 °C	20 s	30
Annealing	45-72 °C	20 s	
Extension	72 °C	60 s per kb	
Final Extension	72 °C	3 minutes	1
Hold	4 °C	∞	

3.4.4. DNA restriction digests

Restriction digests were done with *New England Biolabs* or *Fermentas* restriction enzymes. Digestion was performed according to manufacturer's instructions (37 ºC for 1h in the digest buffers indicated by the manufacturer). The restriction enzymes that were used were:

- NcoI
- HindIII
- NdeI
- NotI
- XhoI
- RsrII

Sample mix components were as follows (Table 3.7):

Table 3.7 Reaction mix for restriction enzyme digestion

COMPONENT	50 µl REACTION
50 µM plasmid	40 µl
10× digest buffer	5 µl
Enzyme 1	1 µl
Enzyme 2	1 µl
ddH_2O	Up to 50 µl

3.4.5. Ligation

The following reaction (Table 3.8) was set up on ice in a microcentrifuge tube. This was left at 16 °C o/n and heat inactivated (65 °C) for 10 minutes. Then 2 µl of the sample were used to transform XL-Blue competent cells.

Table 3.8 Reaction mix for ligation

COMPONENT	20 µl REACTION
T4 DNA ligase buffer (10×)*	2 µl
Vector DNA (4 kb)	50 ng (0.020 pmol)
Insert DNA (1 kb)	37.5 ng (0.060 pmol)
Nuclease-free water	to 20 µl
T4 DNA ligase	1 µl

3.4.6. Cloning verification

By using Taq polymerase (*Stratagene*), colony PCR was performed to assess successful cloning. Colonies were picked with a sterile pipette tip and gently shaked inside a PCR eppendorf tube containing the sample mix detailed in the following table (Table 3.9).

Table 3.9 Reaction mix for PCR with Taq polimerase

Component	Amount
10× Taq polymerase buffer	2.0 µl
25 mM $MgCl_2$	0.4 µl
10 mM dNTPs	0.4 µl
10 µM 3'primer	0.4 µl
10 µM 5'primer	0.4 µl
Taq polymerase	0.25 µl
Nuclease-free water	Up to 20 µl
	_____
Total reaction volume	20 µl

PCR steps are listed on Table 3.10. Primers used depended on the PCR product that was being analyzed, these are listed on Table 1.3.4.

Table 3.10 Taq PCR cycle conditions

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	95 °C	30 s	1
Denaturation	95 °C	15-30 s	30
Annealing	45-68 °C	15-60 s	
Extension	68 °C	60 s per kb	
Final Extension	72 °C	5 minutes	1
Hold	4 °C	∞	

3.4.7. DNA electrophoresis

Agarose gel electrophoresis was used to analyze the size of the plasmids and fragments resulting from restriction enzyme digestion as well as for purification for DNA PCR fragments.

By using the Bio-Rad electrophoresis tanks and power sources (Power Pac Basic). DNA electrophoresis was performed at 100 V. Gels were stained for 30 minutes and then imaged with an UV light/blue light source. Reagents used are the following:

3.4.7.1. Gel preparation and running

- Agarose gels: 0.7-1% [w/v] agarose (Sigma) in 1 × TAE
- Tris-acetate- EDTA (TAE) buffer (stock solution 50×): 2 M Tris, 1 M acetic acid, 50 mM EDTA, pH: 8.4
- 6× DNA Loading dye (Fermentas)
- Molecular weight markers: GeneRuler 1 Kb DNA ladder (Fermentas)

3.4.7.2. Gel staining

Staining solutions: Syber green and Syber safe (Invitrogen) (stock solution 10000× in DMSO, working concentration 1× diluted in TAE).

3.4.8. Plasmid purification and linear DNA purification

Agarose gel electrophoresis was used to analyze the size of the plasmids and fragments resulting from restriction enzyme digestion as well as for purification of DNA PCR fragments.

Following the manufacturer's guidelines (*Qiagen*), we purified plasmids from individual samples using the *Qiagen Miniprep kit* by using an eppendorf table-top centrifuge.

Pure linear DNA fragments from PCR products and cleaved vectors were obtained with the PCR and Gel Purification Kit (*Qiagen*) or the Illustra GFX PCR DNA and Gel Band Purification Kit (*GE Healthcare*) by following manufacturer's guidelines and using an eppendorf table-top centrifuge.

3.4.9. Determination of DNA concentration, purity and sequence

DNA concentration and purity was analyzed by using a NanoDrop 1000 spectrometer and following manufacturer's instructions (ThermoFisher Scientifics). DNA fragments were checked by DNA sequencing (Macrogen).

3.4.10. Site-directed mutagenesis and creation of truncated variants

We created mutants that mimic the phosphorylated and unphosphorylatable states of ComE, this was done by changing the amino acid 58, which according to literature is the one that phosphorylates (Stock *et al.,* 2016). To mimic the phosphorylated state it was necessary to change D58E and for the unphonsphorylatable state the change was D58A. To accomplish this, we used the Quickchange Multi Site-Directed mutagenesis kit (Agilent) and we followed manufacturer's guidelines. XL-blue cells were transformed with the resulting vectors, which were sent to Macrogen for sequencing.

Trying to explore more of the phosphorylation qualities of ComE, we also decided to create a vector that included only the LyTr domain of the protein, which is the one on charge of DNA interaction. To do so, we amplified this domain using primers (see Table 3.11) starting at the N-terminus of this domain. To create the final vector we used the In-Fusion® Cloning system (Nick Berrow, Protein Expression Facility IRB) following manufacturer's guidelines. Resulting vectors were sent to Macrogen for sequencing.

Table 3.11 Primers used for mutants

NAME	SEQUENCE
ComED58A_Forward	GAAGTAAATCAGCTTTATTTCCTAGCTATCGATATTCATGGAATTGAGAAA
ComED58A_Reverse	TTTCTCAATTCCATGAATATCGATAGCTAGGAAATAAAGCTGATTTACTTC
ComED58E_Forward	GAAGTAAATCAGCTTTATTTCCTAGAGATCGATATTCATGGAATTGAGAAAAA
ComED58E_Reverse	TTTTTCTCAATTCCATGAATATCGATCTCTAGGAAATAAAGCTGATTTACTT
ComELyTr_Forward	AGAAAGCTTTTAGATTTTTTCTCTAAAATATC
ComELyTr_Reverse	ATACCATGGACTACAATTACAAGGGAAATG
Infussion_ComELyTr_Forward	AGGAGATATACCATGGACTACAATTACAAGGG
Infussion_ComELyTr_Reverse	GTGATGGTGATGTTTAGATTTTTTCTCTAAAATATC
Forward_NEB_GAGGLU_E	ATTTCCTAGAGATCGATATTCATG
Reverse_NEB_GAGGLU_E	AAAGCTGATTTACTTCATCATTTTC
Forward_NEB_GCTALA_A	TATTTCCTAGcTATCGATATTCATGG
Reverse_NEB_GCTALA_A	AAGCTGATTTACTTCATCATTTTC

3.5. Protein and complex characterization

In this section, we will describe the steps used to characterize the proteins used throughout this thesis.

3.5.1. MW estimation using protein standards

A Size-Exclusion Chromatography (SEC) was performed using the MW standards (GE Lifesciences) mixed according to manufacturer's instructions to be able to estimate for ComD and ComE to determine if the protein elutes

as an oligomer. MW standards are shown in table 3.12. The Void volume was calculated by using the Blue dextran as a marker using this equation **Kav = (Ve-Vo)/(Vt-Vo).** Then, these values were graphed against the log (MW) and the resulting line created an equation in which we substituted the elution volume for our proteins to obtain the calculated MW.

Table 3.12 Molecular Weight approximation standards

Standard name	MW (kDa)	log MW	Elution volume (ml)	Kav = (Ve-Vo) /(Vt-Vo)
Blue dextran	2000	**VOID VOLUME**	7,726	**VOID VOLUME**
Conalbumin	75	1,88	9,824	0,13
Carbonic anhydrase	29	1,46	11,998	0,26
Ribonuclease A	13,7	1,14	13,609	0,36
Aprotinin	6,5	0,81	15,707	0,49

3.5.2. Dynamic Light Scattering (DLS)

By calculating the size and size distribution of solution samples in the submicron region we were able to verify the aggregation state of the sample for ComD and ComE, as well as their homogeneity (Stetefeld *et al.,* 2016) by performing DLS with a Zetasizer Nano ZS from the Automated Crystallography Platform of Barcelona Science Park. Experiments were performed according to manufacturer's instructions (Malvern).

3.5.3. Mass Spectrometry (MS) analysis

Peptide mass fingerprinting (PMF) was performed to identify if the final pure sample was the protein of interest. This experiment was performed through the Proteomics and Genomics services for the Centro de Investigaciones Biológicas (CIB) CSIC in Madrid.

3.5.4. dsDNA stock preparation

Oligonucleotides designed and used for crystallization (Table 3.13) were purchased from biomers.net.

Table 3.13 Primers used crystallization oligonucleotides

NAME	SEQUENCE
DR1_16A	CAGTACACTTTGGGAG
DR1_16B	CTCCCAAAGTGTACTG
DR2_16A	CTGACAGTTGAGAGAG
DR2_16B	CTCTCTCAACTGTCAG
DR12_40A	CAAGTACACTTTGGGAGAAAAAAATGACAGTTGAGAGAAG
DR12_40B	CTTCTCTCAACTGTCATTTTTTTCTCCCAAAGTGTACTTG
DR12_40C	CAAGTACACTTTGGGAGAAAAAAATGACAGTTGAGAGAGG
DR12_40D	CTCTCTCAACTGTCATTTTTTTCTCCCAAAGTGTACTTGC
DR12_38A	CAGTACACTTTGGGAGAAAAAAATGACAGTTGAGAGAG
DR12_38B	CTCTCTCAACTGTCATTTTTTTCTCCCAAAGTGTACTG

DNA oligonucleotides were diluted to a concentration of 2 mM in ddH_2O. After mixing complementary primers, the sample was denatured at 80 °C for 30 min in a water bath, and annealed overnight by slowly cooling down after shutting off the water bath. Resulting sample was then kept at 4 °C.

To calculate the final number of nmoles, the next equation was used:

$$\frac{Y}{100\ \mu l + x\ \mu l} = Total\ nmoles\ of\ dsDNA$$

3.5.5. Electrophoretic Mobility Shift Assay (EMSA)

To detect protein–nucleic acid interactions we performed an EMSA with ComE, its mutants and the DNA that interacts with ComE. Gels used for this experiment were done with the following reagents, shown in Table 3.14.

Table 3.14 EMSA reaction mix for DNA

COMPONENT	20 ml REACTION
H_2O	15 ml
Acrylamide 30%	3,7 ml
TAE Buffer 10×	1 ml
PSA 10%	200 µl
Temed	80 µl

Protein:DNA ratios were tested to observe in which one the interaction started and which one to use for the following experiments. Samples were incubated for 3 hours previous to performing the EMSA. Samples were mixed with 6× DNA loading dye (Fermentas) and the oligonucleotide without protein was loaded as a negative control. 0.05 nM of DNA were loaded onto the gels. EMSA was performed using a power source and electrophoresis tanks (Bio-rad) with a running buffer (TAE 0.5%) at 200 V and 4 °C for 20 or 25 minutes, for the 20 bp oligos used and the 40 bp oligos used respectively. Gel was then dyed with Syber safe (Invitrogen) stock solution 10000× in DMSO, working concentration 1× diluted in TAE) for 20 minutes before being imaged with an UV light/blue light source.

3.5.6. Size Exclusion Chromatography of complexes

SEC were performed to visualize the interaction between ComD, ComE, ComE mutants and the oligonucleotides that contain the Direct Repeats (DR) that interact with ComE. EMSA Results were used as guidelines for the amount of protein and DNA to include. Runs for the individual components of the complexes were performed to visualize the differences in the elution volume between them and the final complexes.

3.5.7. Size Exclusion Chromatography – Multi Angle Light Scattering (SEC-MALS)

This technique allows calculating the MW and radius more accurately, also indicating the homogeneity of the sample as well as the amount of aggregation. This experiment was carried out with the help of Roman Bonet

from the Automated Crystallography Platform (PAC).

3.5.8. One-dimensional Nuclear Magnetic Resonance (1D-NMR)

NMR was performed with the help of the NMR Unit of the Cientific and Technological centers of the University of Barcelona (CCiT). This experiment would help us visualize if the protein or protein complexes are folded correctly.

3.6. Crystallization and X-ray diffraction analysis

X-ray crystallography is the technique that has provided the highest number of resolved protein structures up till now as shown in Figure 3.1 as well as contributing with the highest-resolution data available (Egli, 2016).

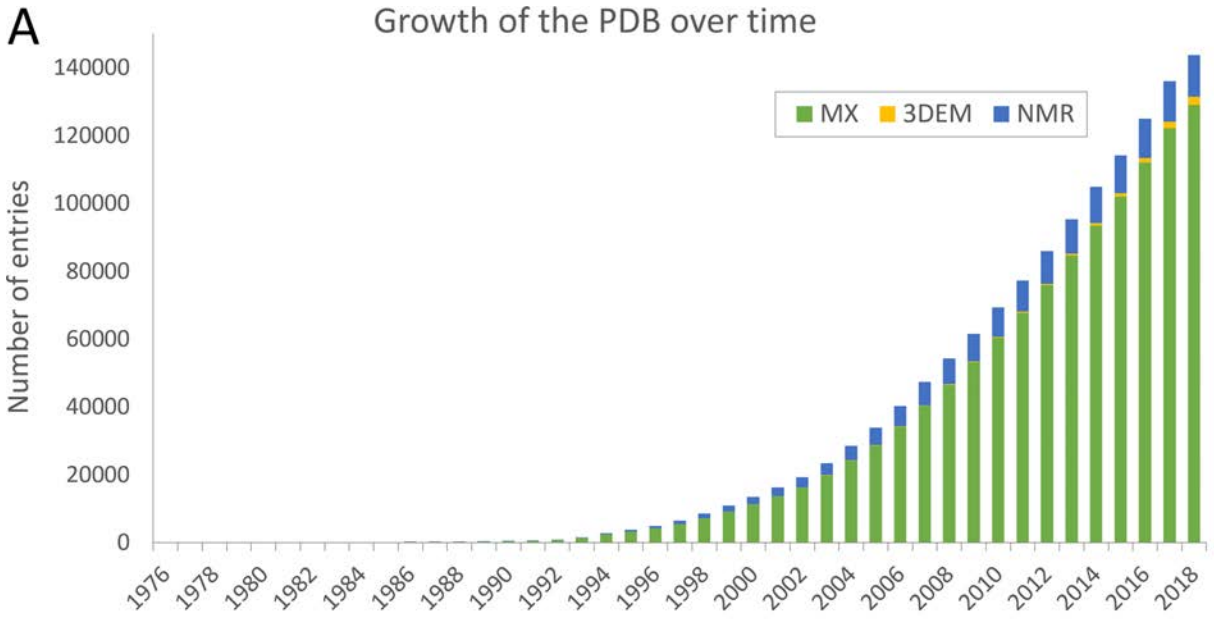


Figure 3.1 Growth of the pdb over time
(extracted from Egli, 2016).

This approach is outlined ahead:

3.6.1. Sample preparation.

A considerable amount of pure protein is required, and molecules should be well folded.

3.6.2. Crystallization

In this thesis, crystallization via vapor diffusion was performed; it is one of the most common crystallization protocols used. Crystallization relies on having a closed environment in which there is a reservoir well containing a precipitant and a drop with that same chemical at a lower concentration mixed with the pure protein. Water diffuses into the reservoir from the drop increasing the concentration of the precipitant and lowering protein solubility causing protein to change into a supersaturated phase from an unsaturated state, allowing nucleation to occur. By staying in the metastable zone, crystals will grow (Figure 3.2). It demands a broad screening of conditions in which crystals that diffract grow.

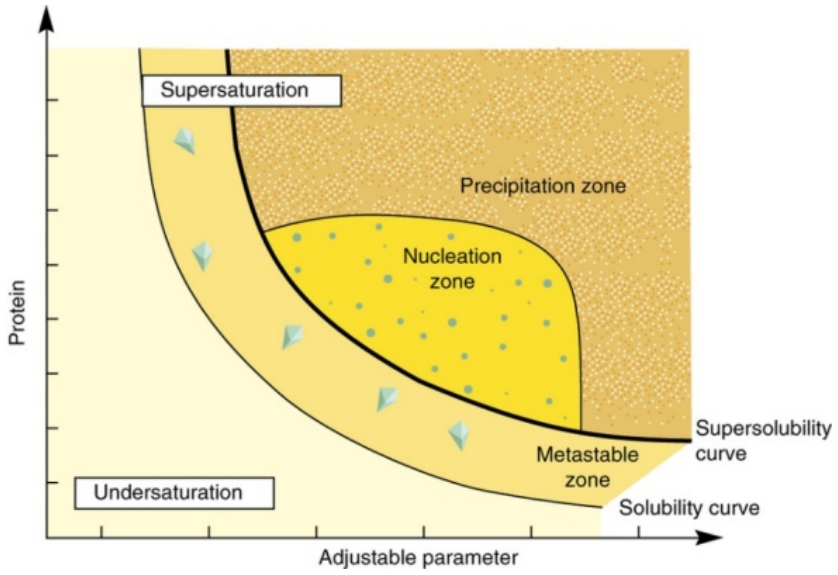


Figure 3.2 Protein crystallization phase diagram
Protein concentration is represented on the y-axis, while the adjustable parameter on the x-axis can be for instance the precipitant concentration. (Image extracted from Khurshid *et al.*, 2014.)

3.6.3. Protein crystallization screening

Crystal screening experiments were carried out at the Automated Crystallography Platform (PAC) of the IBMB/IRB at the Barcelona Science Park. 96-well sitting drop MRC plates (Molecular Dimensions) were used for screenings. Drops were prepared by mixing 100 nl of protein mix and 100 nl of the reservoir solution. Reservoirs contained 100 µl of Reservoir solution. Reagents used as reservoirs were prepared using a Freedom EVO (TECAN) robot. To facilitate the set up of the 96 well crystal plates a Crystal Phoenix (Art Robbins Instruments) robot was used. The crystal screenings used are shown in Table 3.15. Resulting plates were incubated at 4 ºC and 20 ºC in the crystal farms (Bruker AXS) provided by the PAC.

Table 3.15 List of crystal screenings used
Names of the screens and commercial screens in which they are based.

Name	Screen	Conditions	Reference
PAC1	Crystal Screen I Crystal Screen II	48 48	Hampton Research
PAC2	Wizard I Wizard II	48 48	Emerald Bio
PAC3	Index	96	Hampton Research
PAC4	Salt RX	96	Hampton Research
PAC9	Natrix Complex screen	48 40	Hampton Research PAC Platform IBMB
PAC10	Protein-DNA Screen	96	PAC Platform IBMB
PAC21	PACT premier HT-96	96	Molecular Dimensions
PAC22	Pi-PEG Screen	96	Jena Bioscience
PAC23	Pi – minimal screen	96	Jena Bioscience
PAC Plus	JCSG-Plus	96	Jena Bioscience
PACTOP 96	TOP 96	96	Anatrace Microlytic

When crystals or promising conditions appeared, the condition in which this happened was then optimized, trying to obtain crystals that could diffract.

3.6.4. Protein crystallization optimization

A condition became promising when small crystals started to appear. This could be checked through UV using the Crystalmation (Rigaku) imager or by visualizing it thanks to the Crystal Farm Navigator. When a promising condition was identified, optimization was required for the appearance of better crystals. Reagents, such as precipitant concentrations or buffer pH could be varied to improve the shape and size of the crystal. The technique could also be modified to try and improve the crystals, so 24-well hanging drop plates or sitting drop plates were also used to optimize conditions.

The most promising conditions were then optimized in 24-well hanging drop plates (Jena Biosciences). Reservoirs in these plates now contained 1 ml of the crystallization condition, and the crystal drops initially were created by manually mixing 1 μl of protein and 1 μl of reservoir mix but could be varied depending on the desired size or concentration of the drop. These drops were then observed using a SMZ 800 Nikon microscope and further optimization plates were done as needed, until crystals that could be collected appeared.

3.6.5. Crystal mounting and freezing

By using nylon cryo-loops (Molecular Dimensions) it was possible to fish protein crystals that appeared in the crystal drops. Loop size varied according to the size of the crystal. Fished crystals were cryo-protected by soaking them in reservoir solutions that contained a cryo-protectant. The amount of the cryo-protectant had to be optimized to adequately protect the crystals. The crystals were flash frozen using liquid nitrogen and were now ready to be taken to the synchrotron.

3.6.6. ComD-ComE data collection

Crystals are irradiated with a beam of X-rays, producing a diffraction pattern (reflections) that corresponds to the sample electrons that have been scattered by the X-rays, which are then collected. This is done at

synchrotrons that emit X-rays in a tangential sense relative to a closed circle where the accelerated electron beams are travelling. Crystals are cryo-cooled with Nitrogen to protect from radiation damage and avoid crystal desiccation. Cryo-protectants are required to avoid the crystallization of interference materials (solvents) with the protein diffraction.

No data-sets were obtained with the ComD-ComE project.

4. Results and Discussion ComD-ComE

In this section of the thesis, we will analyze the results of the experiments performed throughout the PhD project.

4.1. Construct design

We designed constructs that would express the target proteins inside *E. coli* and performed the secondary structure prediction of the proteins to have an integral analysis of ComD and ComE.

4.1.1. ComD

4.1.1.1. ComD sequence

Streptococcus pneumoniae strain R6 ComD full-length (Uniprot reference: Q8DMW4) is composed by a transmembrane helix domain (1-206) and a cytoplasmic domain (207-441). Figure 4.1

```
        10         20         30         40         50         60
MDLFGFGTVI VHFLIISHSY HFICKGQINR KELFVFGAYT LLTEIVFDFP LYILYLDGLG

        70         80         90        100        110        120
IERFLFPLGL YSYFRWMKQY ERDRGLFLSL LLSLLYESTH NFLSVTFSSI TGDNFVLQYH

       130        140        150        160        170        180
FPFFFVVTVL TYFVTLKIIY YFHLELAYFD EDYLYPFLKK VFFALLLLHI VSFVSDMVST

       190        200        210        220        230        240
IKHLNSFGSI LSSIVFISLL LTFFAMNSHK VQMEKEIALK QKKFEQKHLQ NYTDEIVGLY

       250        260        270        280        290        300
NEIRGFRHDY AGMLVSMQMA IDSGNLQEID RIYNEVLVKA NHKLRSDKYT YFDLNNIEDS

       310        320        330        340        350        360
ALRSLVAQSI VYARNNGVEF TLEVKDTITK LPIELLDLVR IMSVLLNNAV EGSADSYKKQ

       370        380        390        400        410        420
MEVAVIKMET ETVIVIQNSC KMTMTPSGDL FALGFSTKGR NRGVGLNNVK ELLDKYNNII

       430        440
LETEMEGSTF RQIIRFKREF E
```

Figure 4.1 Sequence of full-length ComD

4.1.1.2. ComD secondary structure prediction

S. pneumoniae strain R6 ComD full-length secondary structure was predicted using the Psipred server. Prediction showed the existence of an abundant number of helixes in the transmembrane helix domain as shown in Figure 4.2.

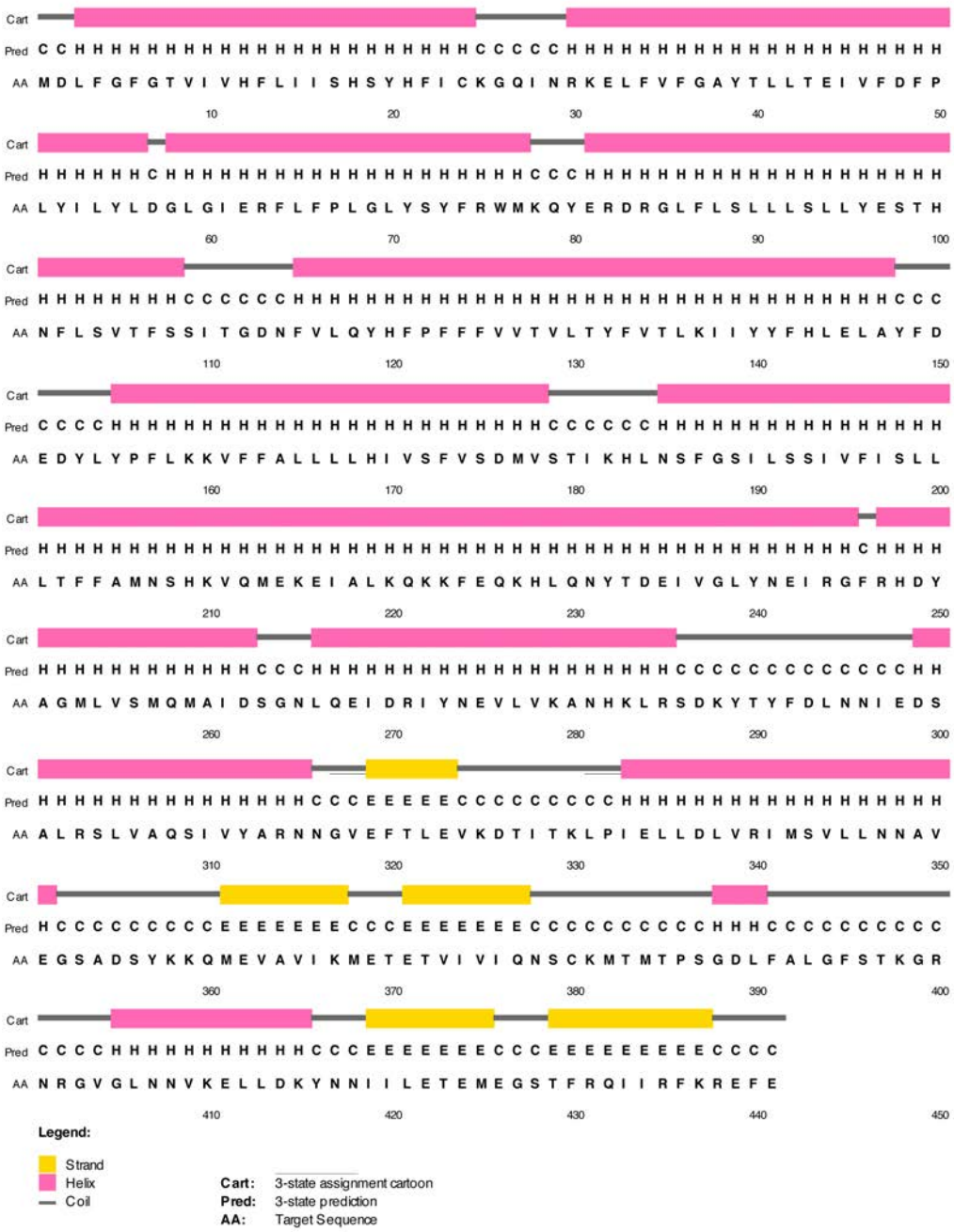


Figure 4.2 *S. pneumoniae* strain R6 ComD full-length secondary structure prediction

4.1.1.3. ComD secondary extracellular-cytoplasmic schematic

Furthermore, the Psipred server was used to predict the extracellular-cytoplasmic schematic of ComD full-length. Figure 4.3 shows the presence of an extracellular segment of the protein as well as the distribution of the helixes in the transmembrane helix domain.

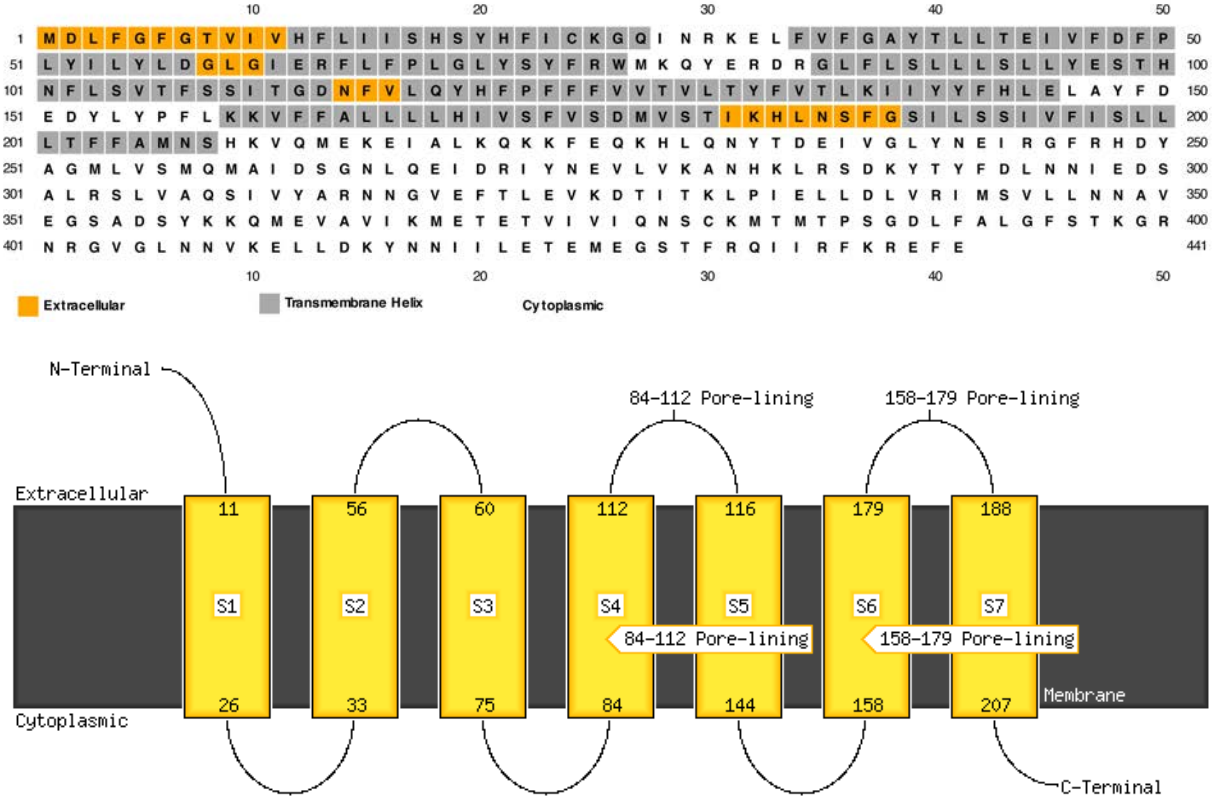


Figure 4.3 *S. pneumoniae* strain R6 ComD transmembrane domain schematic

4.1.1.4. Sumo-ComD construct

A construct that includes the segment of the cytoplasmic domain of ComD joint with the SUMO-Histag on the N-terminal to increase solubility and facilitate expression was designed and the sequence is shown in Figure 4.4.

```
        10         20         30         40         50         60
MGSSHHHHHH GSDSEVNQEA KPEVKPEVKP ETHINLKVSD GSSEIFFKIK KTTPLRRLME

        70         80         90        100        110        120
AFAKRQGKEM DSLRFLYDGI RIQADQTPED LDMEDNDIIE AHREQIGGNS HKVQMEKEIA

       130        140        150        160        170        180
LKQKKFEQKH LQNYTDEIVG LYNEIRGFRH DYAGMLVSMQ MAIDSGNLQE IDRIYNEVLV

       190        200        210        220        230        240
KANHKLRSDK YTYFDLNNIE DSALRSLVAQ SIVYARNNGV EFTLEVKDTI TKLPIELLDL

       250        260        270        280        290        300
VRIMSVLLNN AVEGSADSYK KQMEVAVIKM ETETVIVIQN SCKMTMTPSG DLFALGFSTK

       310        320        330        340
GRNRGVGLNN VKELLDKYNN IILETEMEGS TFRQIIRFKR EFE
```

Figure 4.4 Sequence of SUMO-His-ComD cytoplasmic domain construct (207-441)

4.1.2. ComE

4.1.2.1. ComE sequence

Streptococcus pneumoniae strain R6 ComE full-length (Uniprot reference: Q8DMW5) is composed by a Response Regulator domain (RR) (1-125) and a DNA binding domain (LyTr) (142-246). Figure 4.5

```
        10         20         30         40         50         60
MKVLILEDVI EHQVRLERIL DEISKESNIP ISYKTTGKVR EFEEYIENDE VNQLYFLDID

        70         80         90        100        110        120
IHGIEKKGFE VAQLIRHYNP YAIIVFITSR SEFATLTYKY QVSALDFVDK DINDEMFKKR

       130        140        150        160        170        180
IEQNIFYTKS MLLENEDVVD YFDYNYKGND LKIPYHDILY IETTGVSHKL RIIGKNFAKE

       190        200        210        220        230        240
FYGTMTDIQE KDKHTQRFYS PHKSFLVNIG NIREIDRKNL EIVFYEDHRC PISRLKIRKL

       250
KDILEKKSQK
```

Figure 4.5 Sequence of full-length ComE

4.1.2.2. His-ComE secondary structure prediction

S. pneumoniae strain R6 ComE full-length secondary structure was predicted using Psipred, shown in Figure 4.6.

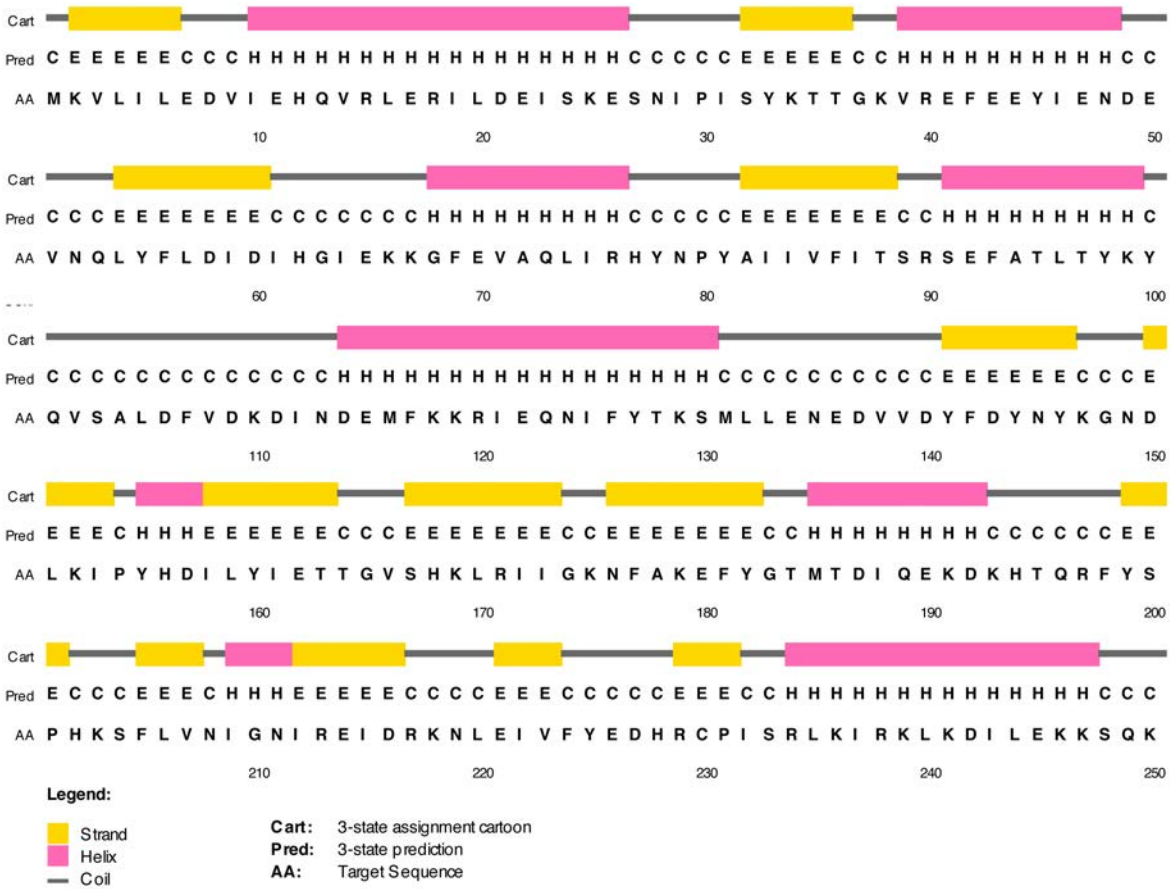


Figure 4.6 *S. pneumoniae* strain R6 ComE full-length secondary structure

4.1.2.3. His-ComE construct

A construct that includes ComE full-length joint with a Histag on the N-terminal to increase solubility and facilitate expression was designed and the sequence is shown in Figure 4.7.

```
        10         20         30         40         50         60
MAHHHHHHSS GLEVLFQGTV DPTGKRAMKV LILEDVIEHQ VRLERILDEI SKESNIPISY

        70         80         90        100        110        120
KTTGKVREFE EYIENDEVNQ LYFLDIDIHG IEKKGFEVAQ LIRHYNPYAI IVFITSRSEF

       130        140        150        160        170        180
ATLTYKYQVS ALDFVDKDIN DEMFKKRIEQ NIFYTKSMLL ENEDVVDYFD YNYKGNDLKI

       190        200        210        220        230        240
PYHDILYIET TGVSHKLRII GKNFAKEFYG TMTDIQEKDK HTQRFYSPHK SFLVNIGNIR

       250        260        270
EIDRKNLEIV FYEDHRCPIS RLKIRKLKDI LEKKSQK
```

Figure 4.7 Sequence of His-ComE construct

4.2. Protein purification

Purification protocols used to obtain a pure sample to be used for the analysis and crystallization trials performed for this thesis are shortly described ahead.

4.3. ComD purification

ComD was purified through a nickel-affinity chromatography followed by an inverse nickel-affinity chromatography after removing the His-SUMO tag, and finally a Size Exclusion Chromatography (SEC) was performed. ComD purification scheme is shown in Figure 4.8.

Figure 4.8 ComD purification scheme

ComD SUMO tag removal had to be optimized to obtain the highest yield after purification. Removal conditions effects are shown as an SDS-PAGE of the samples in Figure 4.9.

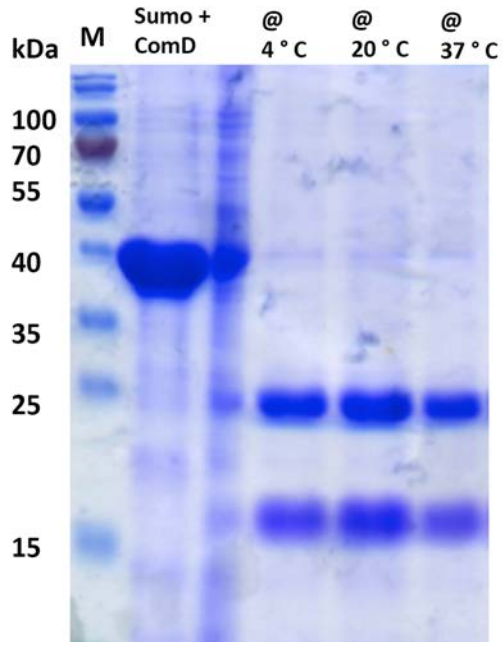


Figure 4.9 ComD SUMO-tag removal optimization SDS-PAGE

The temperature that was chosen was 20 ° C, because of the low quantity of precipitation and the reproducibility (20 ° C room availability), even though the tag was cut at different temperatures; 4° C, 20° C and 37° C.

Verification of the expression and purification of our target protein ComD was performed first via a western blot shown in Figure 4.10. As expected, there is signal throughout the gel, except in the last lane because the tag has been removed.

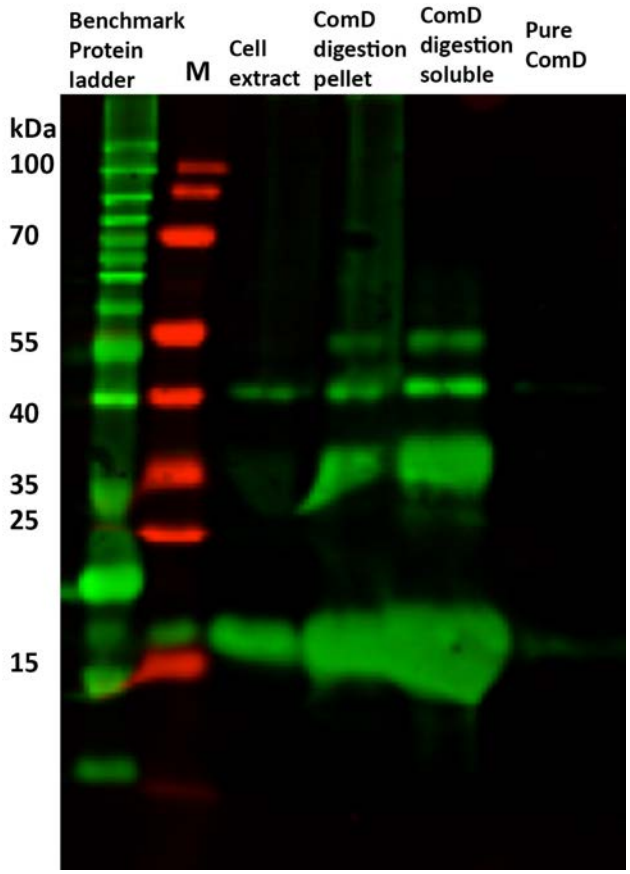


Figure 4.10 ComD western blot

4.4. ComE purification

ComE was purified through a nickel-affinity chromatography followed by a heparin-affinity chromatography and finally a Size Exclusion Chromatography (SEC) was performed. ComE purification scheme is shown in Figure 4.11.

Figure 4.11 ComE purification scheme

Verification of ComE was done via western blot as shown in Figure 4.12.

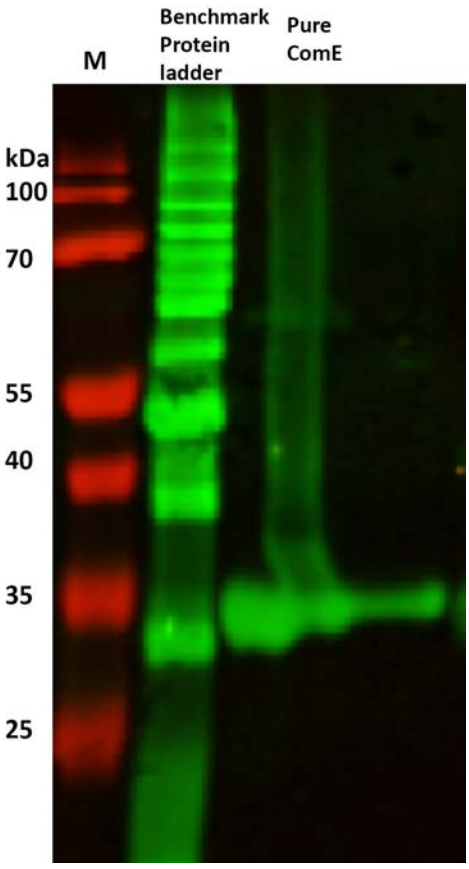


Figure 4.12 ComE western blot

Purification was performed individually to be able to remove the SUMO tag from ComD by performing a reverse Histrap after protease treatment. Final chromatographs and SDS-PAGE (Figure 4.13) show that the sample was pure. To be certain that we were working with our target proteins, peptide mass fingerprinting was performed by the Proteomics and Genomics services for the CIB CSIC in Madrid. Results confirmed that samples were ComD and ComE respectively.

4.5. ComD and ComE characterization

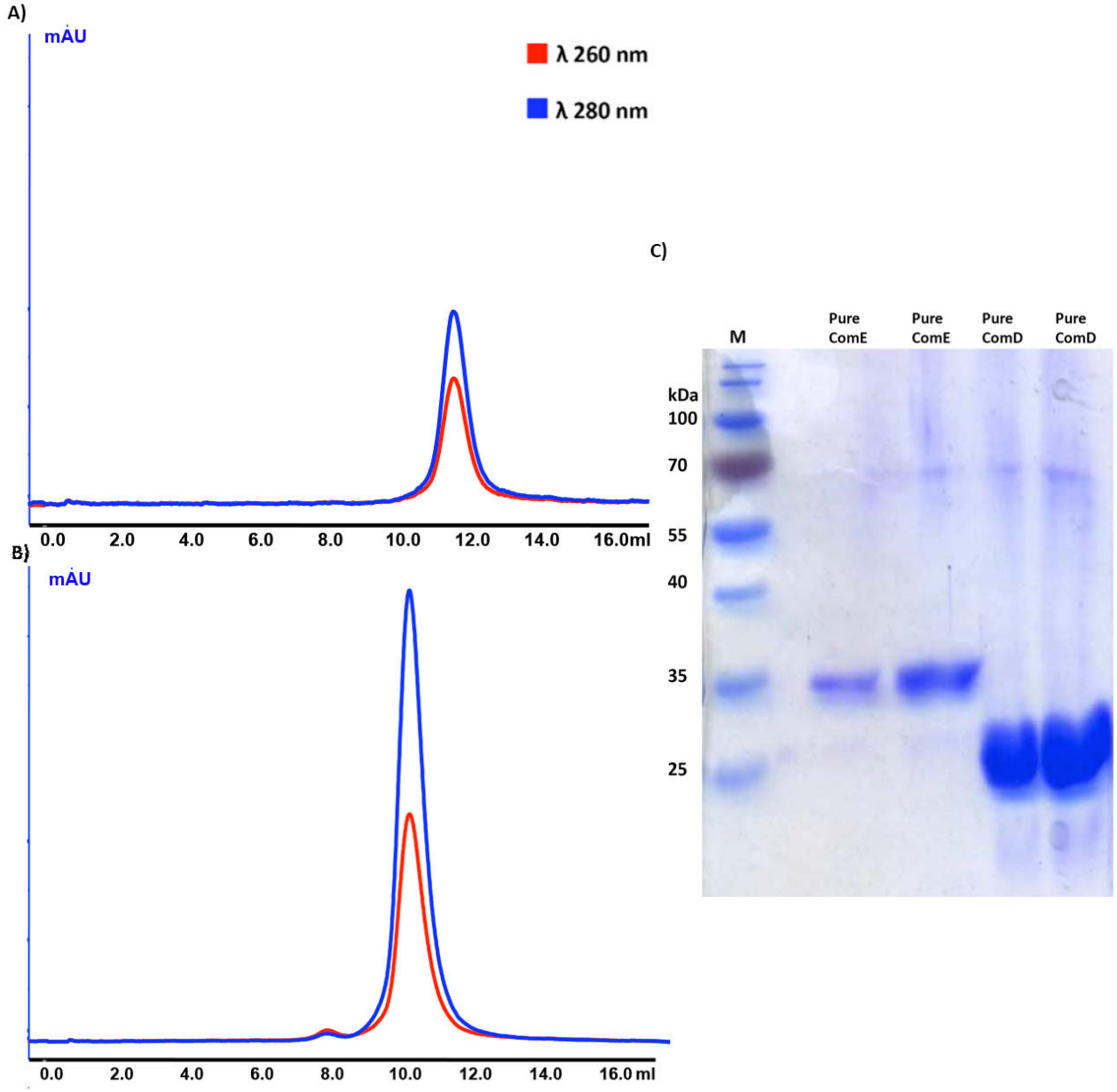


Figure 4.13 Pure ComE (Top) and pure ComD (Bottom) chromatogram and SDS-PAGE

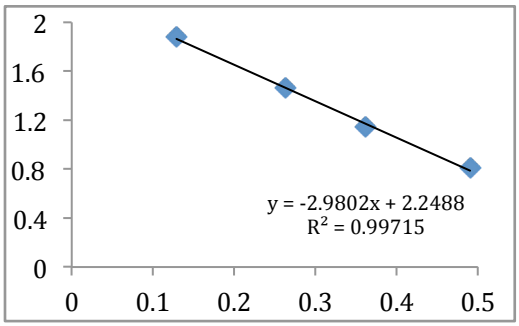
A) Chromatograph of final purification step of ComD, B) Chromatograph of final purification step of ComE, C) SDS-PAGE of the peak fractions obtained from A and B.

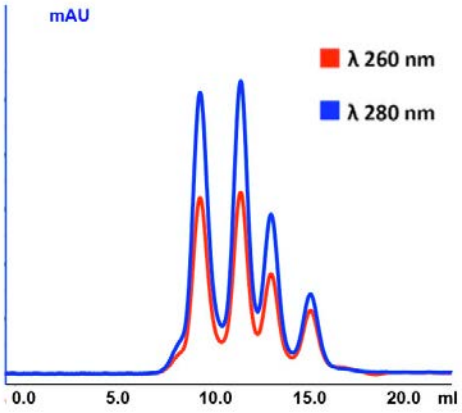
After optimizing production, it was possible to obtain a protein yield of 1.7 mg/l of culture of ComD and 2 mg/l of culture of ComE.

Pure ComD was comprised of 235 residues, a MW of 27,029.04 Da and a theoretical pI of 6.17, while pure ComE was comprised of 277 residues, a

MW of 32,927.60 Da and a theoretical pI of 6.4. Using MW standards, an approximation of the MW of both was performed, reaching an experimental value of 55.3 kDa for ComD and 31.96 kDa for ComE. With this information, (Figure 4.14) it is possible to hypothesize that ComD behaves as a dimer and ComE as a monomer. Monodispersity of both protein samples was verified using DLS.

Blue dextran	2000	VOID VOLUME	7.726	VOID VOLUME	n/a
Conalbumin	75	1.88	9.824	0.13	1.88
Carbonicanhydrase	29	1.46	11.998	0.26	1.46
Ribonuclease A	13.7	1.14	13.609	0.36	1.14
Aprotinin	6.5	0.81	15.707	0.49	0.81





SAMPLE	Elution volume (ml)	Kav = (Ve-Vo) / (Vt-Vo)	log MW	logMW= -2.980x + 2.248	Theoretical MW (kDa)	
His+ComE	11.78	0.25	1.50	31.96	32.92760	**Monomer**
ComD	10.48	0.17	1.74	55.3	27.02904	**Dimer**

Figure 4.14 MW approximation calculations using a Size Exclusion Chromatography

4.6. Protein-protein complexes

4.7. Protein-protein complex formation and analysis

Complex formation was achieved by mixing ComD and ComE to obtain a 1:1 molarity ratio and incubating for 3 hours at 4°C.

The binary protein-protein complex was analyzed and studied in the attempt of obtaining a sample that could be used for crystallization trials.

4.8. ComD-ComE complex SEC

Sample was injected into a SEC 200 10/300 and the chromatogram shows the formation of a protein-protein complex as shown in Figure 4.15, yet there is part of the protein that elutes very close to the exclusion limit, indicating a possible bigger complex or the formation of aggregates. It is possible to see that both ComD and ComE are present throughout the eluting peaks as shown in Figure 4.16.

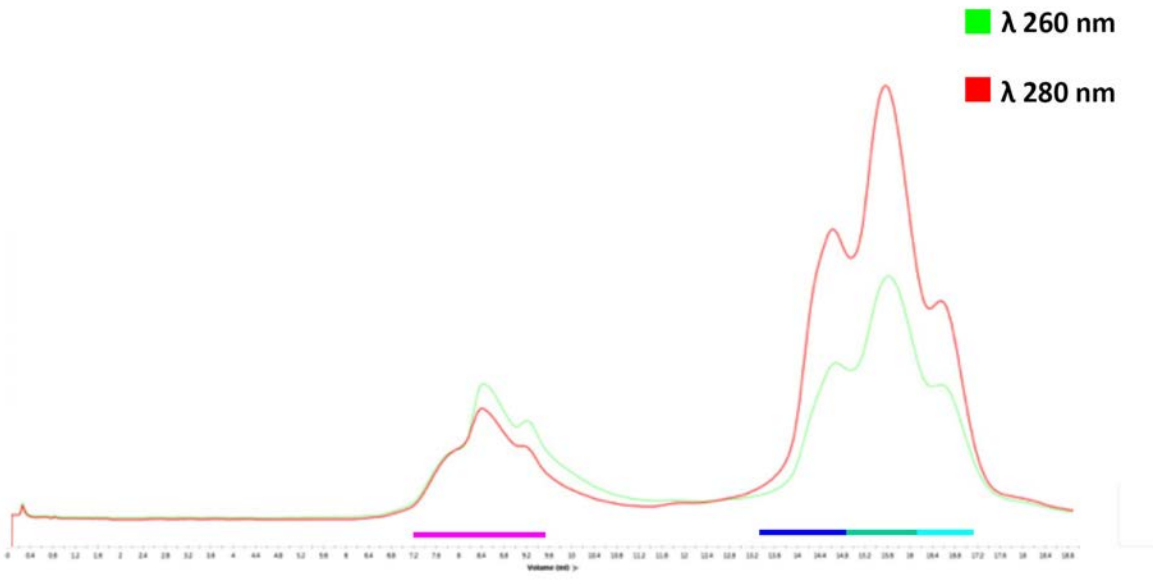


Figure 4.15 ComD-ComE complex SEC chromatogram
Fractions were loaded in the SDS PAGE and are marked with the colored bars.

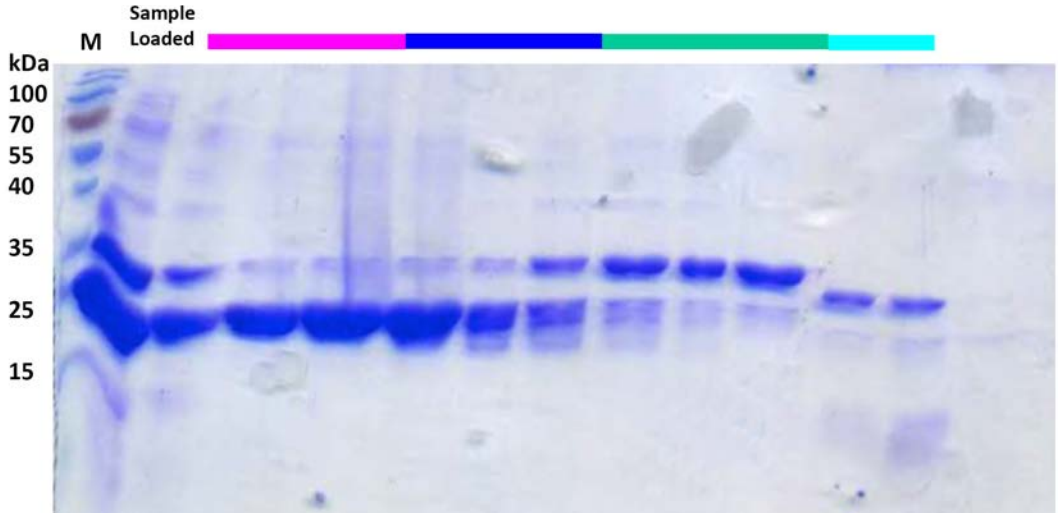


Figure 4.16 ComD-ComE complex SEC SDS-PAGE
Lanes are marked using the color bars as in Figure 4.15

4.9. ComD-ComE complex double SEC

We decided to perform two tandem SEC 200 10/300 to attempt and separate the complex better, but when we did this, maybe because it took a long time due to the pressure limits, it seems that the complex is so weak that the proteins eluted separately as shown in Figure 4.17. Color bars are used to identify fractions both in the Chromatogram and the SDS-PAGE.

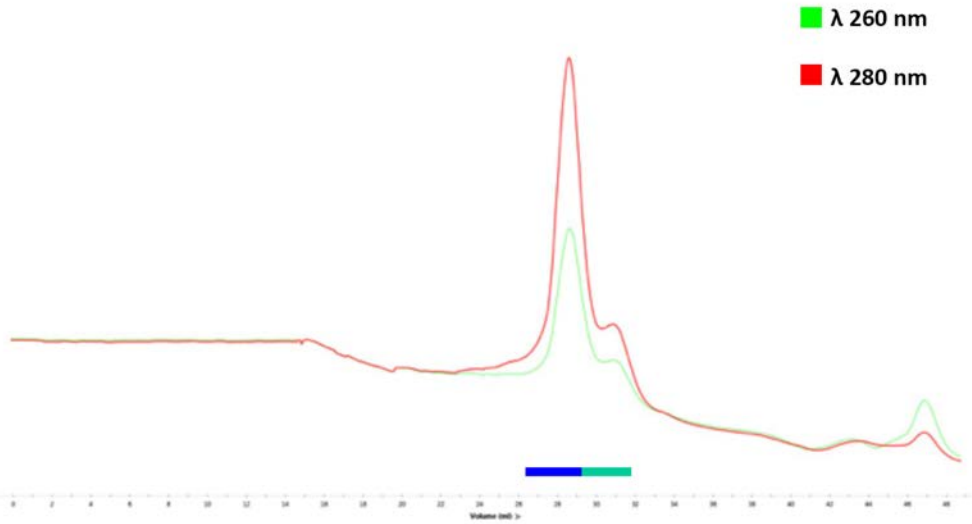


Figure 4.17 ComD-ComE complex double SEC chromatogram

The first peak is symmetrical and contains ComD predominantly, while the second peak has a higher presence of ComE (Figure 4.18). This could suggest that the complex does not have the strength to amalgamate through the double SEC. Therefore, we opted to pursue crystallization trials by adding pure ComD and pure ComE.

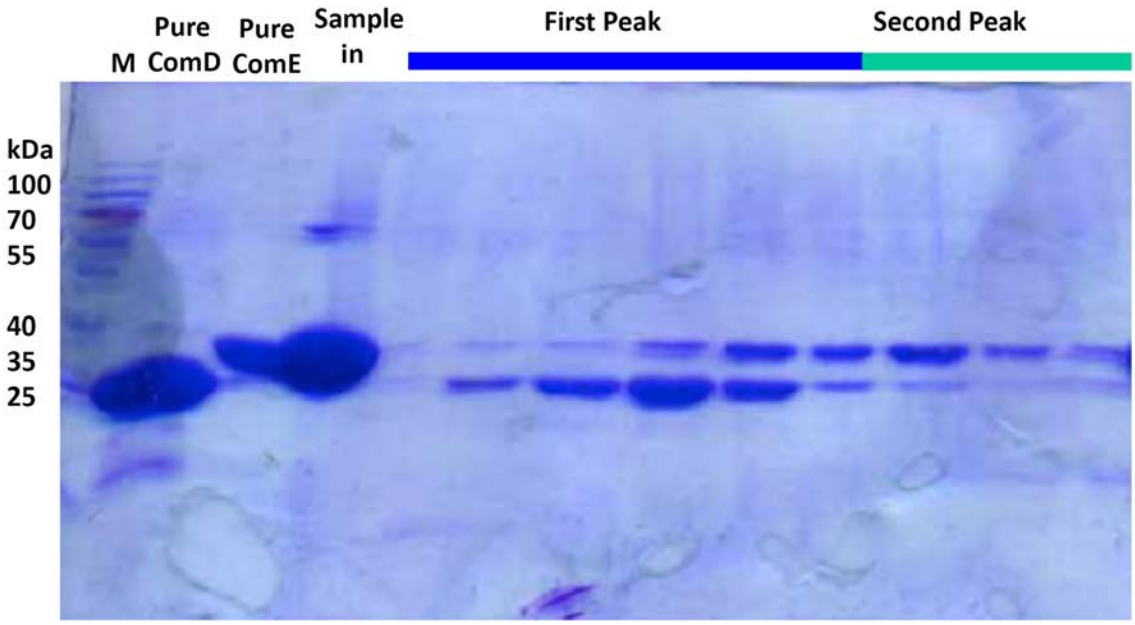


Figure 4.18 ComD-ComE complex double SEC SDS-PAGE

4.10. Protein Crystallization screenings

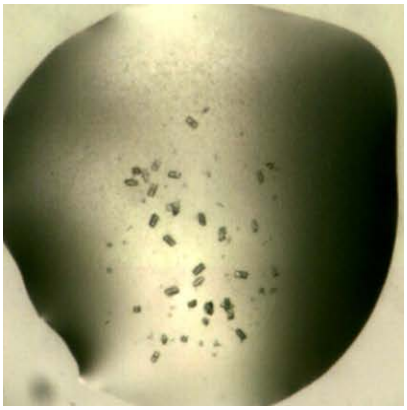
General crystallization screenings were set up with the intention of obtaining promising hits that would result in the optimization and appearance of diffraction-quality crystals to be used to solve the structure of the protein or complex.

4.11. ComD Crystallization screenings

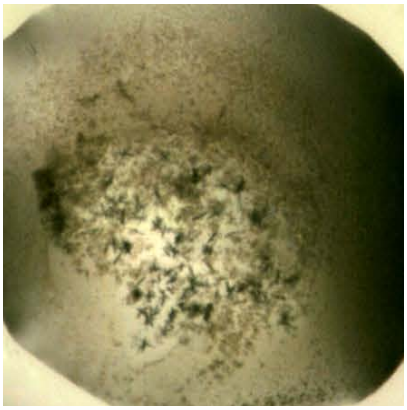
There is no available solved structure of ComD, this is why we attempted to crystallize it by itself. Several general crystallization screening plates were set up at different conditions, we had some possible hits but even after many rounds of phase diagrams and modification of variables, no crystals appeared.

4.12. ComD-ComE crystallization screenings

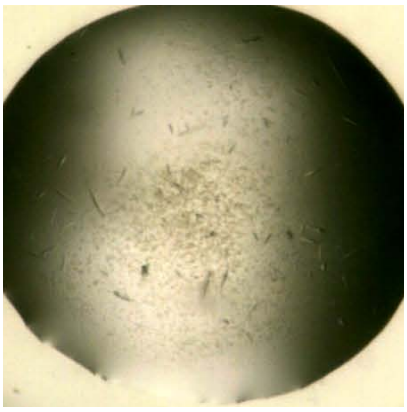
Many general crystallization screening plates were set up at different conditions; we observed promising conditions in PAC 96 H10 well (0.2 M Ammonium sulfide, 0.1 M TRIS at pH 8.5 and 10 %w/v PEG 3350), and PAC 2 H7 (0.2 M Magnesium chloride, 0.1 M TRIS at pH 8.5 and 25 %w/v PEG 8000). Phase diagrams, optimizations and additive screens were set attempting to improve these crystals as shown in Figure 4.19. In some conditions, some needles or crystal nucleus did appear with the additives (B02 is Praseodymium(III) acetate hydrate and D07 is Guanidine hydrochloride) but they could not be optimized nor scaled up.



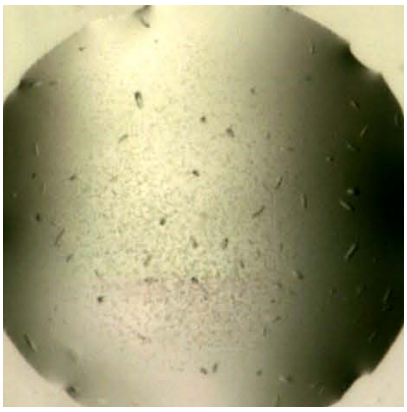
PAC 2 H7 B02



PAC 96 H10 B02



PAC 96 H10 B02



PAC 96 H10 D07

Figure 4.19 ComD-ComE + additive crystal plates with small crystals

4.13. ComD-ComE crystal diffraction

Many crystals were tested at ALBA synchrotron in the BL13-XALOC beamline and a couple did diffract but they were not usable because they were clusters of crystals and the resolution was very low. One of these clusters diffraction pattern suggests it was a protein crystal, but bigger crystals did not appear even after many rounds of optimization (Figure 4.20).

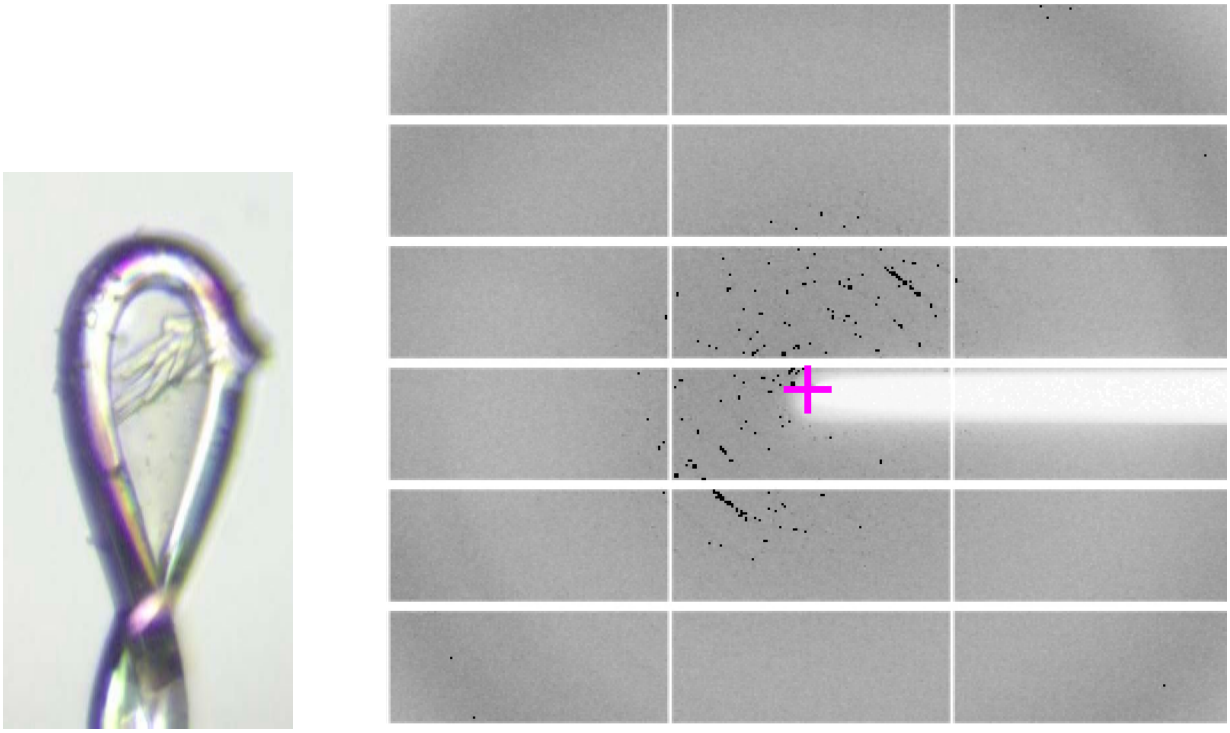


Figure 4.20 ComD-ComE crystal cluster and its diffraction pattern

We obtained crystals with the additive screens (Figure 4.21), but when tested they diffracted like a small molecule, suggesting that what crystallized was the additive itself.

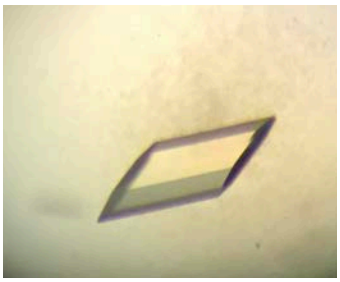


Figure 4.21 Small molecule crystal

4.14. Protein-DNA complex formation

We utilized different oligonucleotides based on the promoter region of *S. pneumoniae*, which is conserved amongst six different species of *streptococci*. These oligonucleotides shown in Figure 4.22, were designed to test the affinity of ComE and the *comCDE* promoter and to attempt complex crystallization. Oligonucleotides conserve the Direct Repeats found in the *comCDE* promoter in the *S. pneumoniae* genome but in some cases to increase stickiness and for design purposes the last base was changed to a cysteine. Oligonucleotides were also extended on the one end to allow a better binding to the protein.

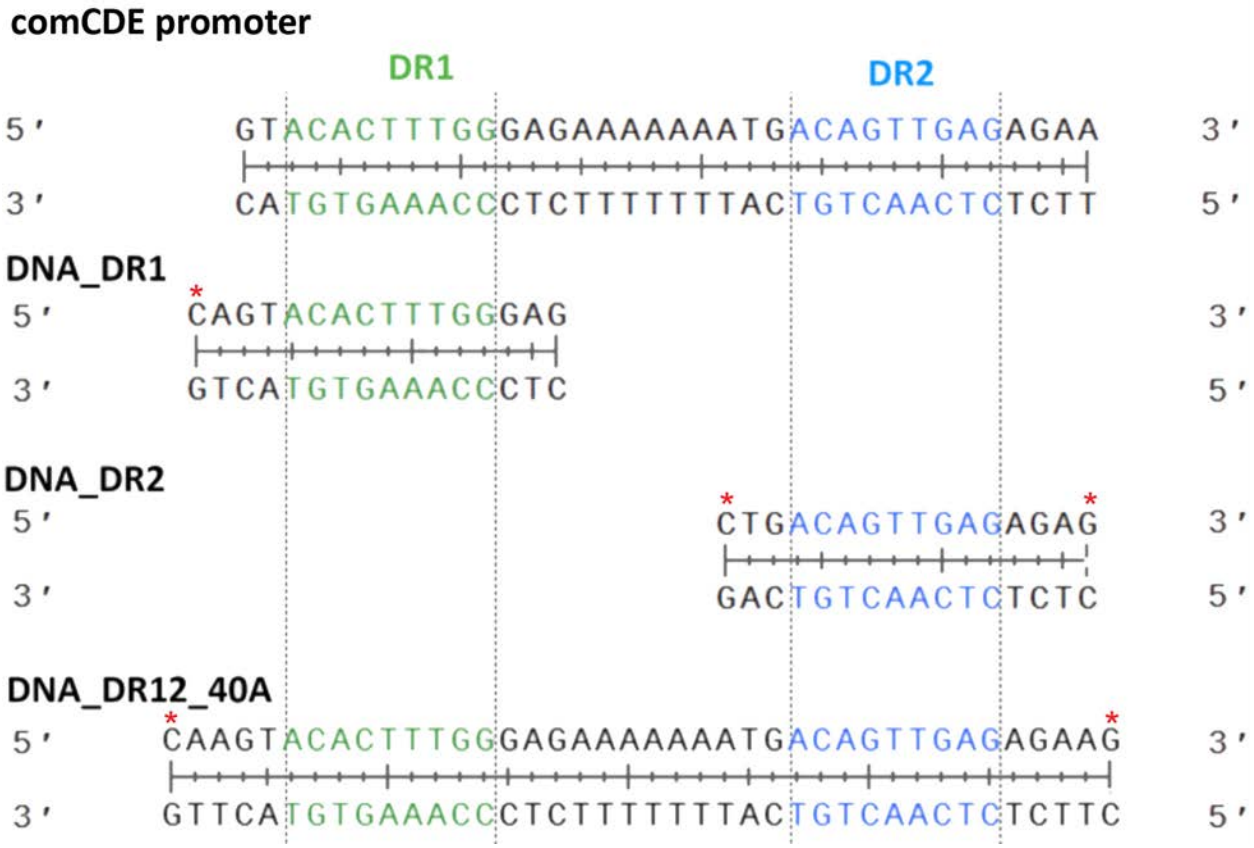


Figure 4.22 DNA sequence of the *comCDE* promoter and the oligonucleotides that were designed for this thesis

DR1 binding site is shown in green, DR2 binding site is shown in blue, modified ends are marked with a red asterisk.

4.15. Mimicking phosphorylated and unphosphorylatable ComE states

We decided to create mutants that mimic the phosphorylated and unphosphorylatable states of ComE. This was achieved by changing the aspartic acid in position 58 into a glutamic acid to mimic the phosphorylated state, and by changing the same amino acid into an alanine to mimic the unphosphorylatable state of the protein. Additionally, we decided to create another construct that included only the DNA binding domain (the LyTr domain) of ComE, to pursue a deeper exploration of how this mechanism works.

After receiving confirmation that the mutations were successful, we expressed and purified both mutants and attempted to do so with the new construct as well. Purification pathway was the same as used for the wildtype. ComE^{D58E} behaved quite similar to the wildtype, while ComE^{D58A} appeared to run faster in the SDS page as shown in Figure 4.23. They were both tested via Peptide mass fingerprinting by the Proteomics and Genomics services for the CIB CSIC in Madrid.

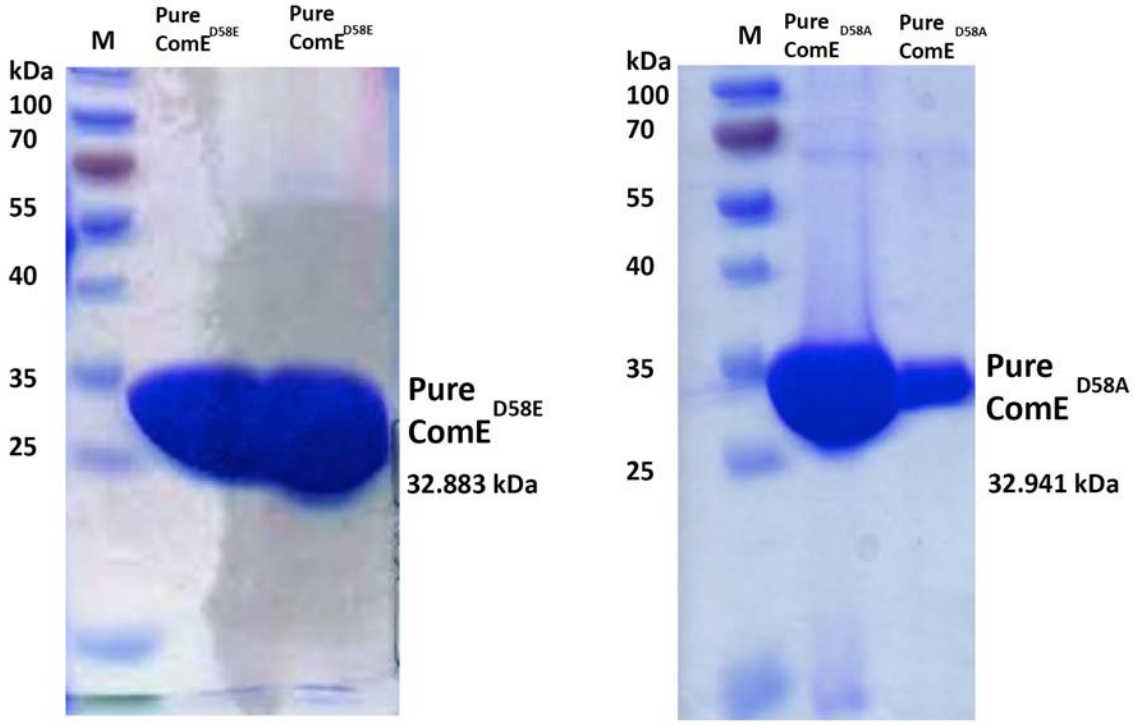


Figure 4.23 Pure ComE^{D58E} and ComE^{D58A} SDS-PAGE

The LyTr domain protein, which functions as the DNA-binding domain (residues 142 – 246) with a MW of 12,933 kDa shown in Figure 4.24 was also tested via peptide mass fingerprinting by the Proteomics and Genomics services at the CIB CSIC in Madrid, yet it could not be purified as most of it degraded and the rest precipitated very quickly

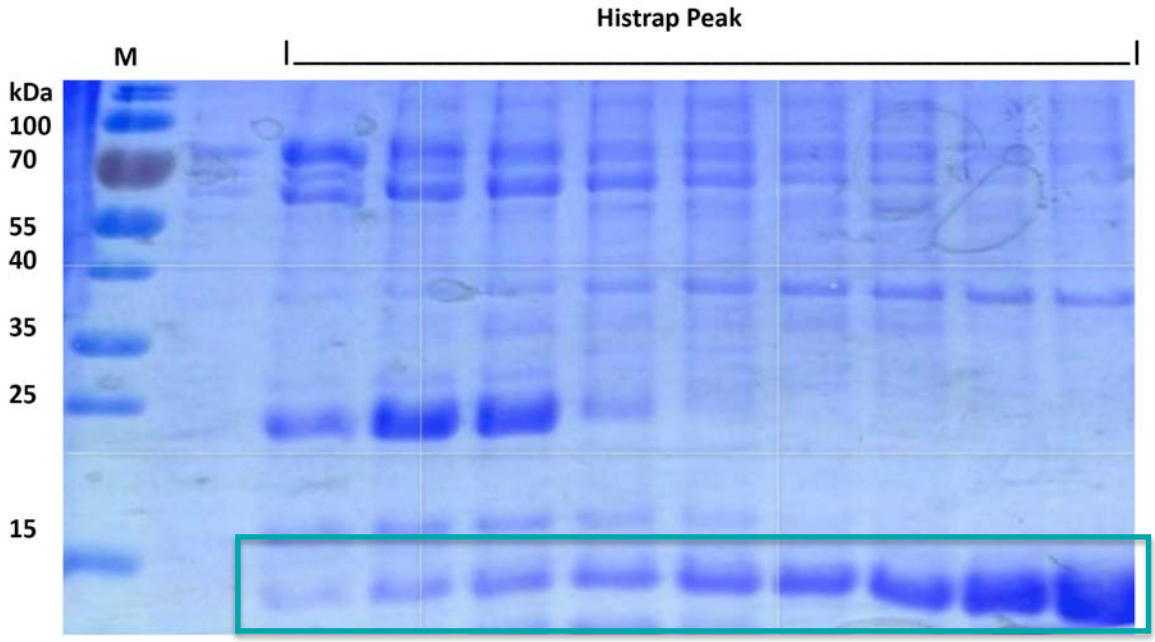


Figure 4.24 ComE LyTr Histrap SDS-PAGE
ComE LyTr histidine affinity chromatography peak is shown inside the turquoise rectangle.

4.16. Electrophoretic Mobility Shift Essay (EMSA)

Purified protein was incubated with the different oligonucleotides for three hours at different protein:DNA ratios to visualize at what concentration did the DNA shift. It was also done to visualize the differences between $ComE^{Wildtype}$, $ComE^{D58E}$ and $ComE^{D58A}$ and their affinity to the DNA_DR1 (Figure 4.25), DNA_DR2 (Figure 4.26) and the DNA_DR12_40A (Figure 4.27) oligonucleotides.

4.17. EMSA for DNA_DR1

The EMSA experiment performed using the short oligonucleotide that contained the first Direct Repeat (DNA_DR1) and different ratios of $ComE^{Wildtype}$, $ComE^{D58E}$ and $ComE^{D58A}$ is shown in Figure 4.25.

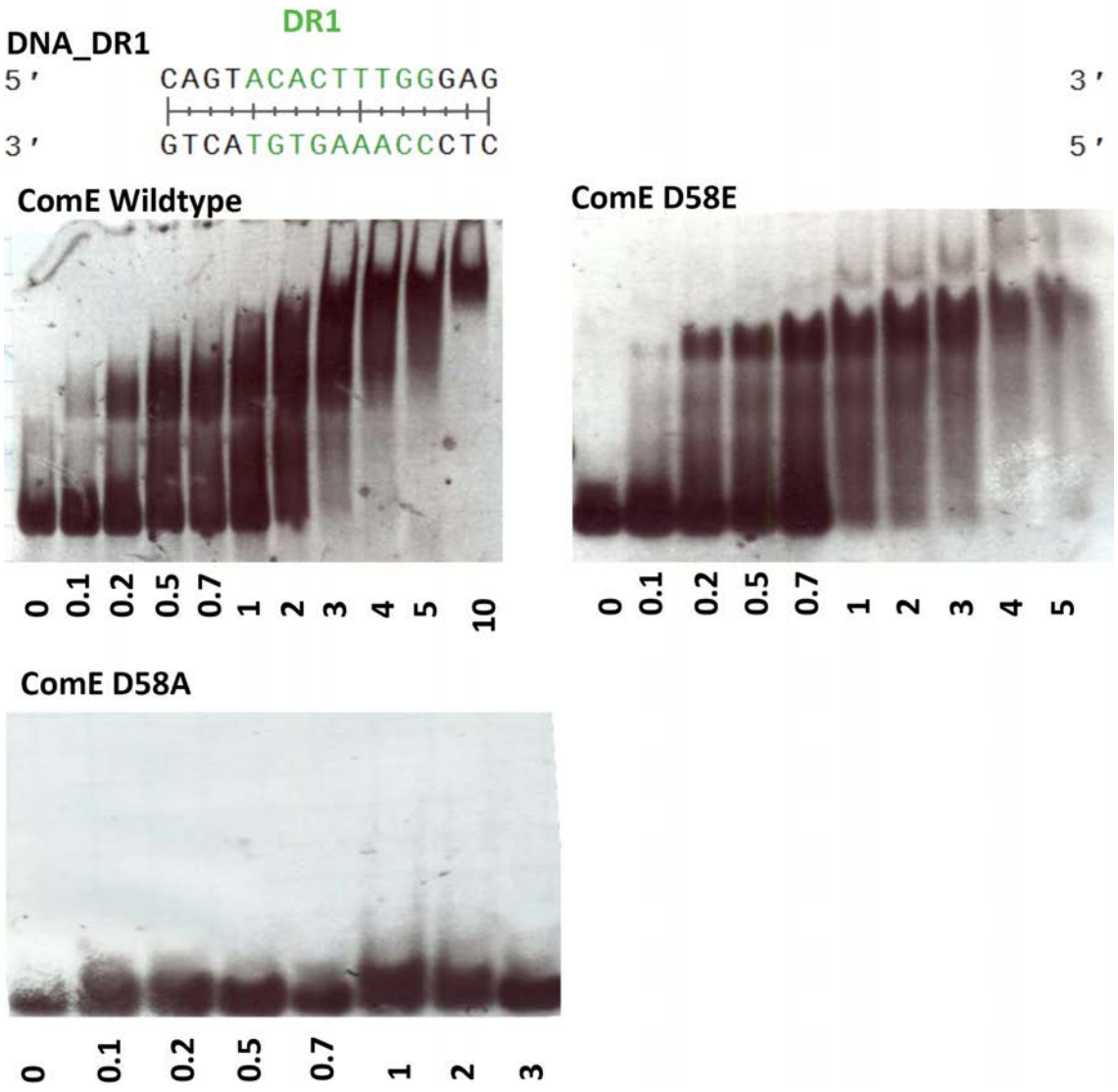


Figure 4.25 EMSA of the "DNA_DR1" oligo with ComE^{Wildtype}, ComE^{D58E} and ComE^{D58A}

There was no binding with the unphosphorylatable mutant (ComE^{D58A}). Unexpectedly, binding was slightly greater with the wildtype in comparison to the phosphorylated mimicking mutant. This suggests that the binding of the LyTr domain with DR1 is independent by the phosphorylation mimicking done by the mutation of this residue in the RR domain. By analyzing the DNA-shift in the gels can propose that the protein:DNA ratio for successful binding would be 3:1 for the ComE^{Wildtype} and 4:1 for ComE^{D58E}.

4.18. EMSA for DNA_DR2

The EMSA experiments performed with ComE^{Wildtype}, ComE^{D58E} and ComE^{D58A} and the short oligo that contained the second Direct Repeat (DNA_DR2) is shown in Figure 4.26.

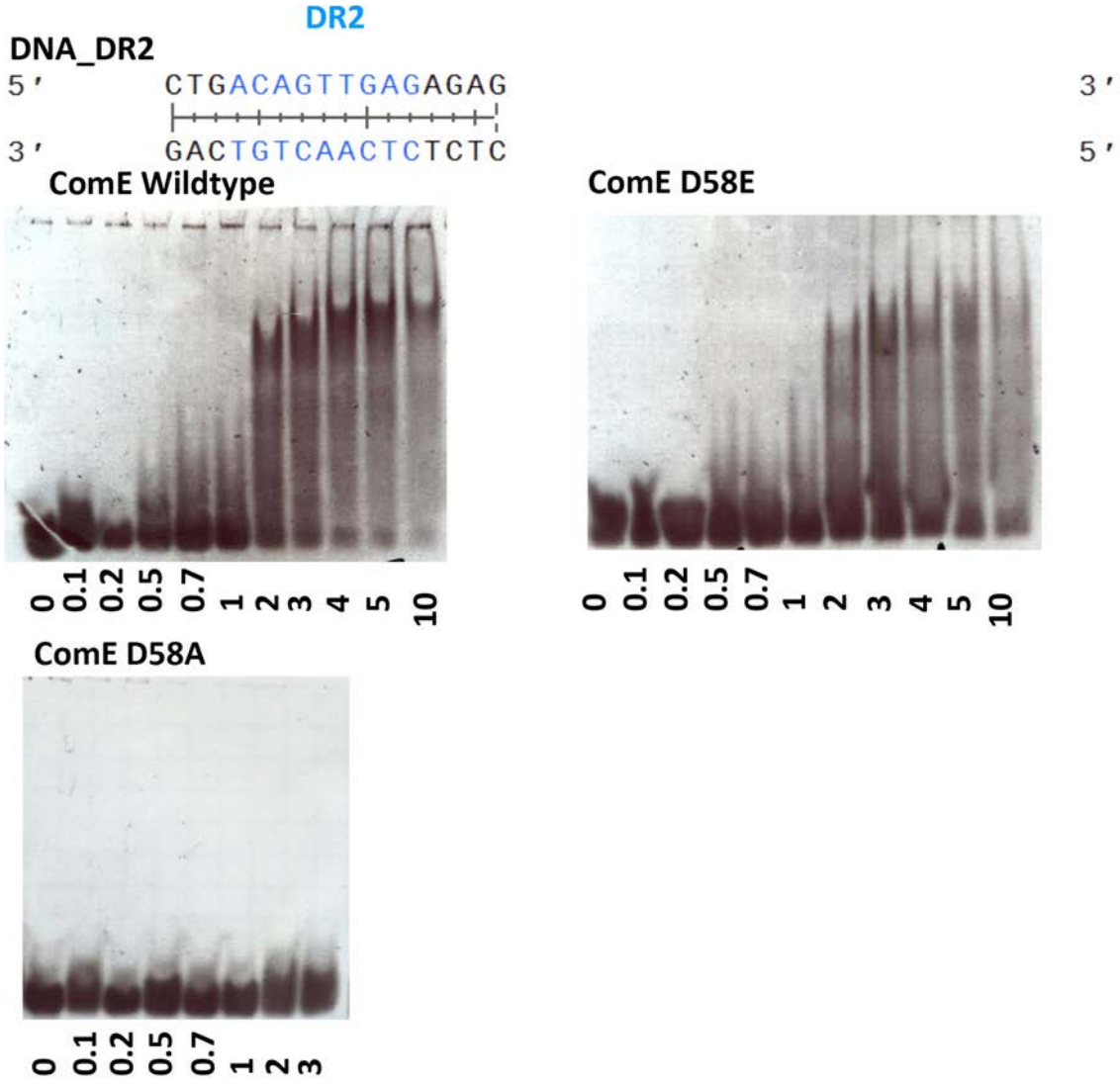


Figure 4.26 EMSA of "DNA_DR2" oligo mixed with ComE^{Wildtype}, ComE^{D58E} and ComE^{D58A}

Once again, there was no binding with the ComE^{D58A} mutant and it was greater with the wildtype in comparison to the ComE^{D58E} mutant.

Binding with DNA_DR2 was weaker than the one with DNA_DR1; therefore, we can theorize that DNA_DR1 has more affinity with ComE in both cases, ComE^{Wildtype} and ComE^{D58E}.

4.19. EMSA for DNA_DR12_40A

An EMSA performed with a longer oligonucleotide (40 bp) that contained both direct repeats (DNA_DR12_40A) and ComE^{Wildtype}, ComE^{D58E} and ComE^{D58A} is shown in Figure 4.27.

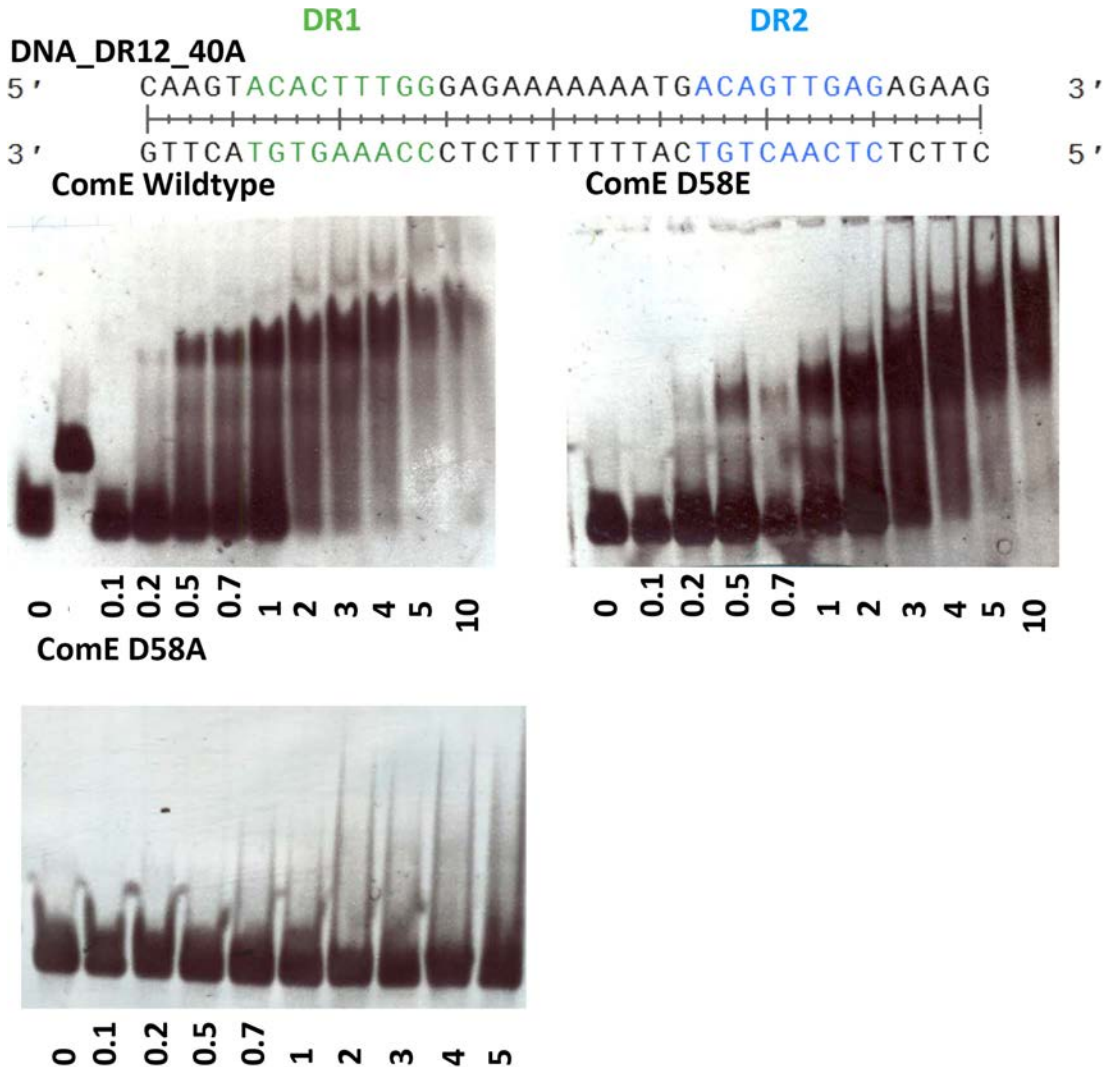


Figure 4.27 EMSA of "DNA_DR12_40A" with ComE^{Wildtype}, ComE^{D58E} and ComE^{D58A}

The DNA_DR12_40A oligonucleotide was able to bind with all three versions of ComE. It is clear that its binding was best with ComE^{Wildtype} followed by ComE^{D58E}, but in contrast with the shorter oligos we can see some displacement as the protein increases with the ComE^{D58A} mutant. This suggests that this oligonucleotide is the best candidate for crystallization trials. In comparison with DNA_DR1, the amount of protein needed for a good binding was lower for the longer nucleotide indicating a higher affinity of the protein.

4.20. Protein-DNA complex preparation for structural studies

Protein-DNA complex was prepared and characterized to produce a sample that could be used to set up crystallization screening plates.

4.21. Protein-DNA complex with ComE^{Wildtype}

After analyzing the EMSA results, we decided to focus on ComE^{Wildtype} and ComE^{D58E} as well as DR1 and DR12_40A. We incubated the protein-DNA complex using a 3:1 ratio (protein:DNA) for DR1 and a 2:1 ratio (protein:DNA) for DR12 for three hours. To verify the formation of the complex, we ran a SEC analysis as shown in Figure 4.28 for ComE^{Wildtype}.

We can observe the elution volume for all the individual components and compare them with both complex peaks, which as predicted elute before. We can observe that as predicted, the complex forms with both oligonucleotides, yet it seems more stable and the resulting peak more homogenous for the DR12_40A oligonucleotide. The ComE^{D58E}-DNA_DR12_40A complex shows that although it is formed, there is still a peak of unbound protein which is not homogenous suggesting that there is also unbound DNA; this is confirmed because of the shift in the elution profile at 260 nm, in the beginning of the second peak. The homogeneity of the ComE^{Wildtype}-DNA_DR12_40A complex peak was promising, so crystal screenings were set.

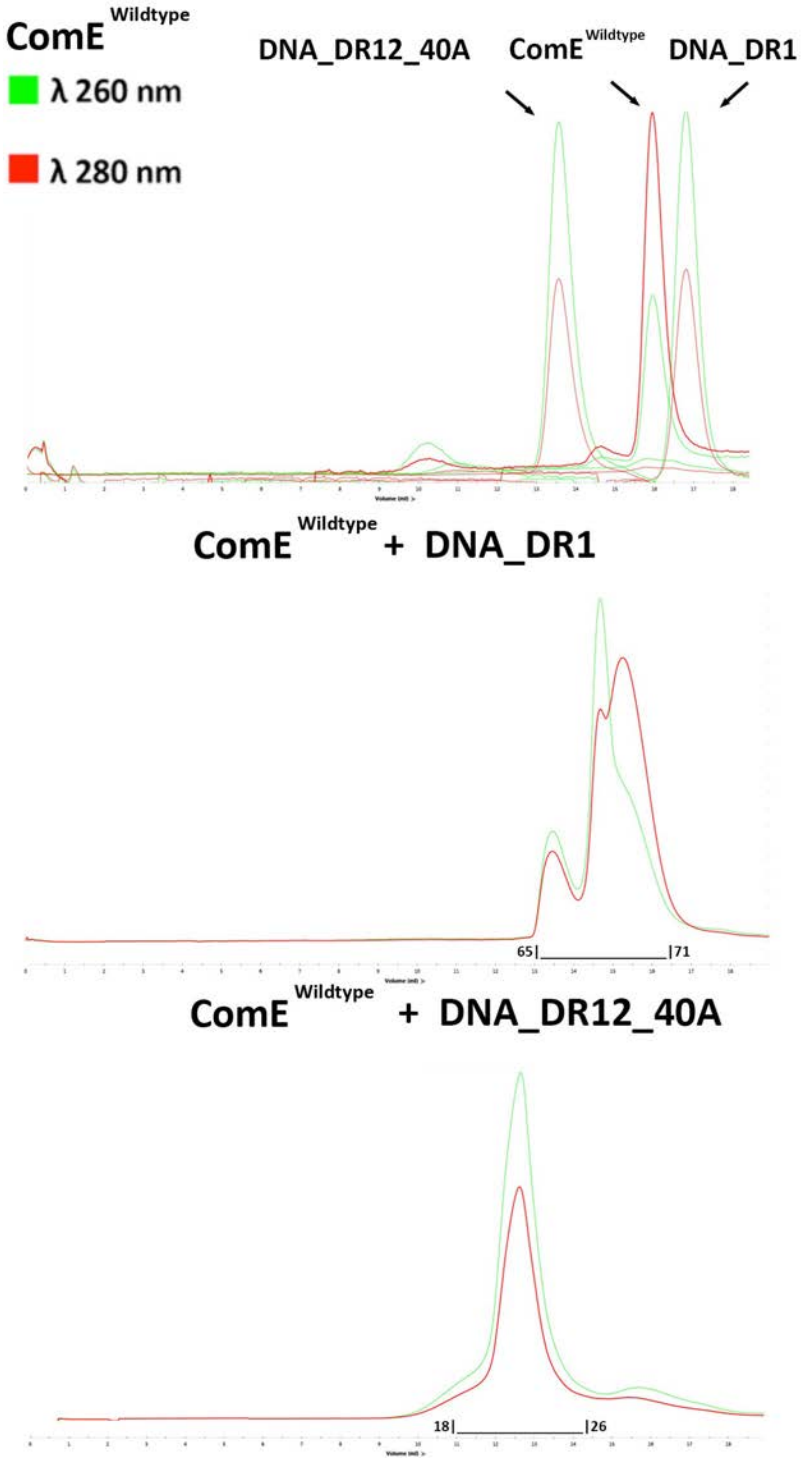


Figure 4.28 ComE^{Wildtype} protein-DNA complex trials with DNA_DR1 and DNA_DR12_40A chromatogram
Signal was normalized to the highest peak of the samples.

To confirm the presence of the protein and the oligonucleotide in the sample, a SDS-PAGE was performed, stained and imaged as shown in Figure 4.29.

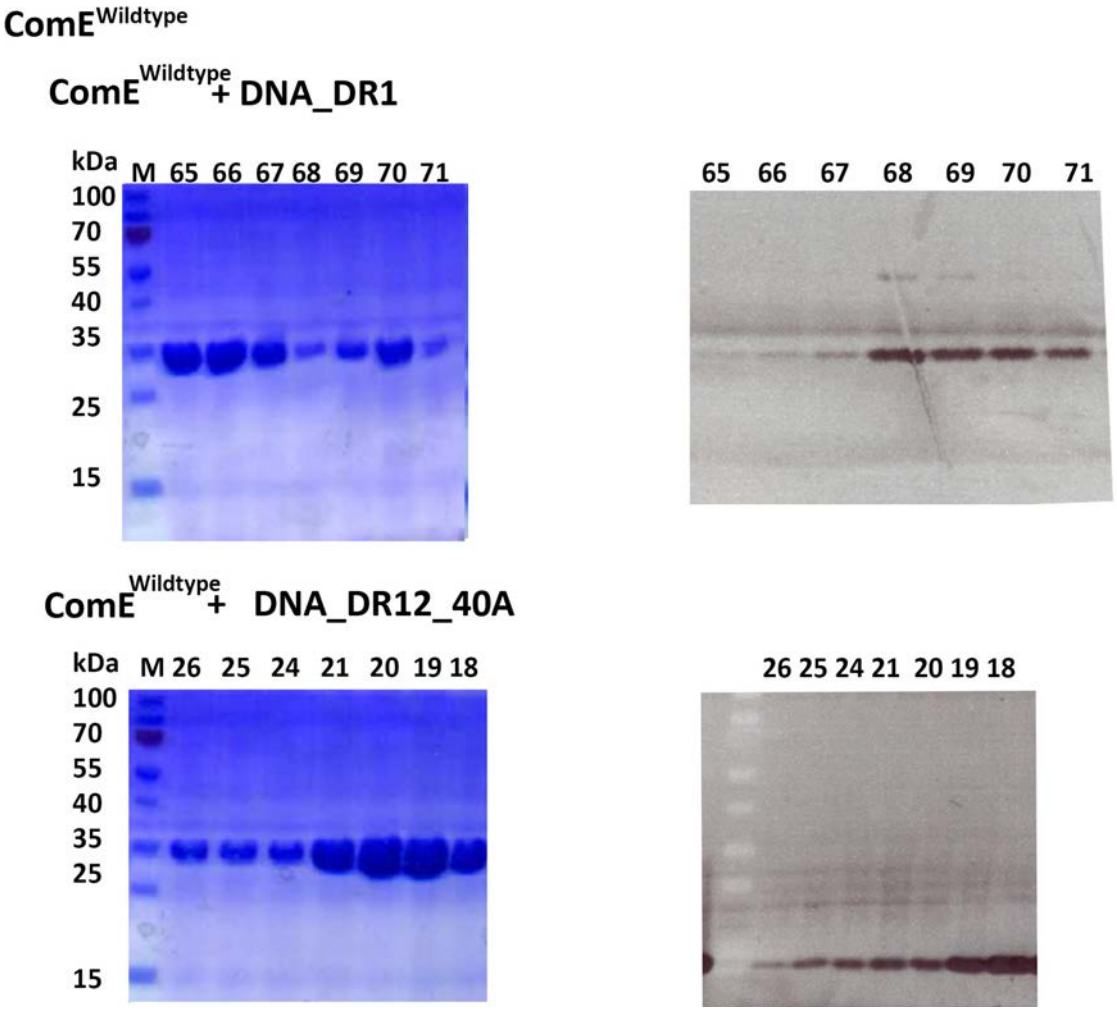


Figure 4.29 ComE^{Wildtype} protein-DNA complex trials with DNA_DR1 and DNA_DR12_40A SDS-PAGE

SDS PAGE on the left dyed with "one step blue" protein dye and on the right dyed with Syber safe dye to visualize DNA.

4.22. Protein-DNA complex with ComE^{D58E}

After forming the ComE^{D58E}-DNA complexes, these were loaded into a SEC performed to analyze the samples as shown in Figure 4.30.

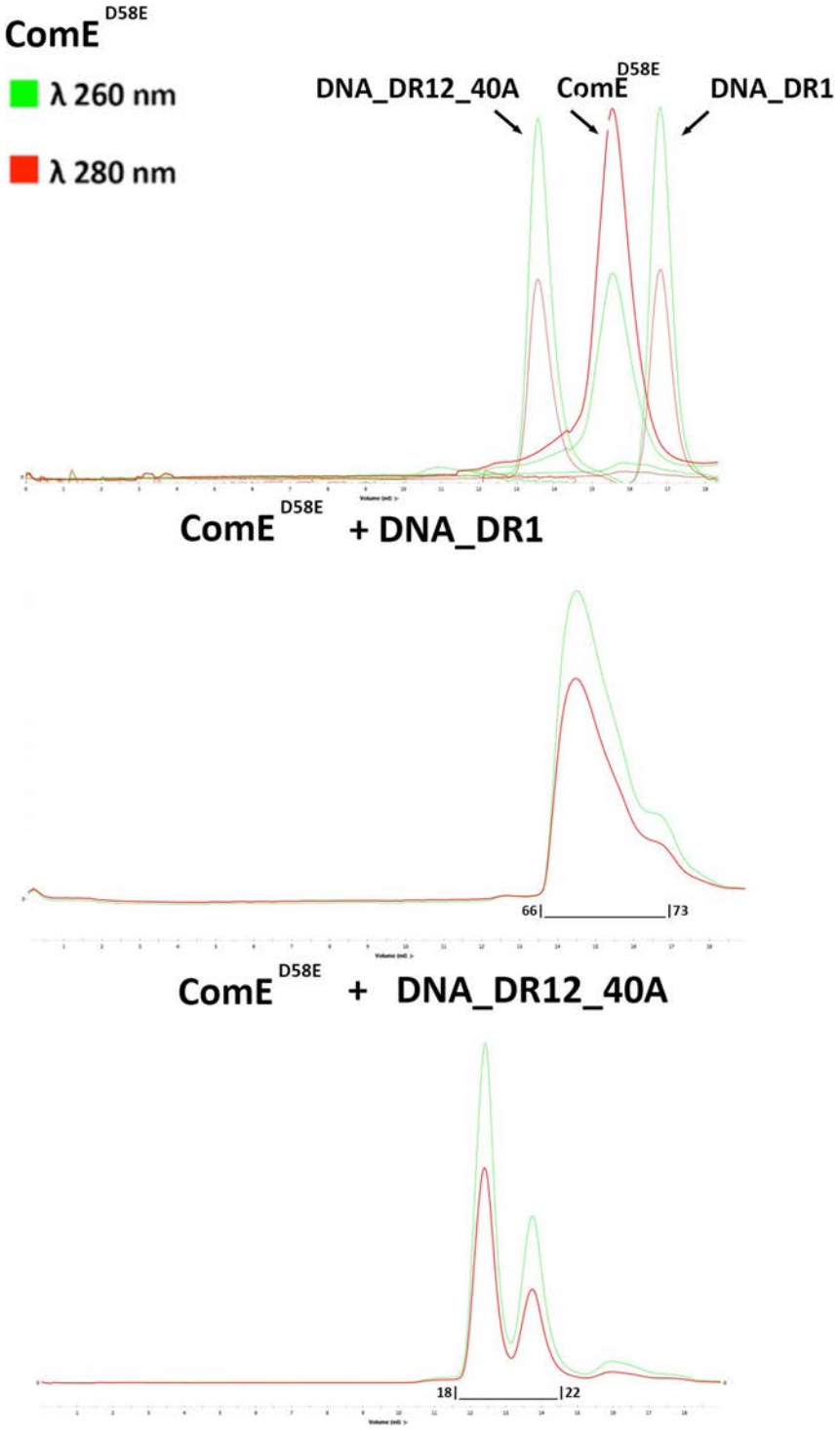


Figure 4.30 ComE^{D58E} protein-DNA complex trials with DNA_DR1 and DNA_DR12_40A

First chromatogram corresponds to the individual samples analysis. Signal was normalized to the highest peak of the samples.

Once again, presence of the protein and the oligonucleotide in the SDS-PAGE was verified by staining and imaging as shown in Figure 4.31.

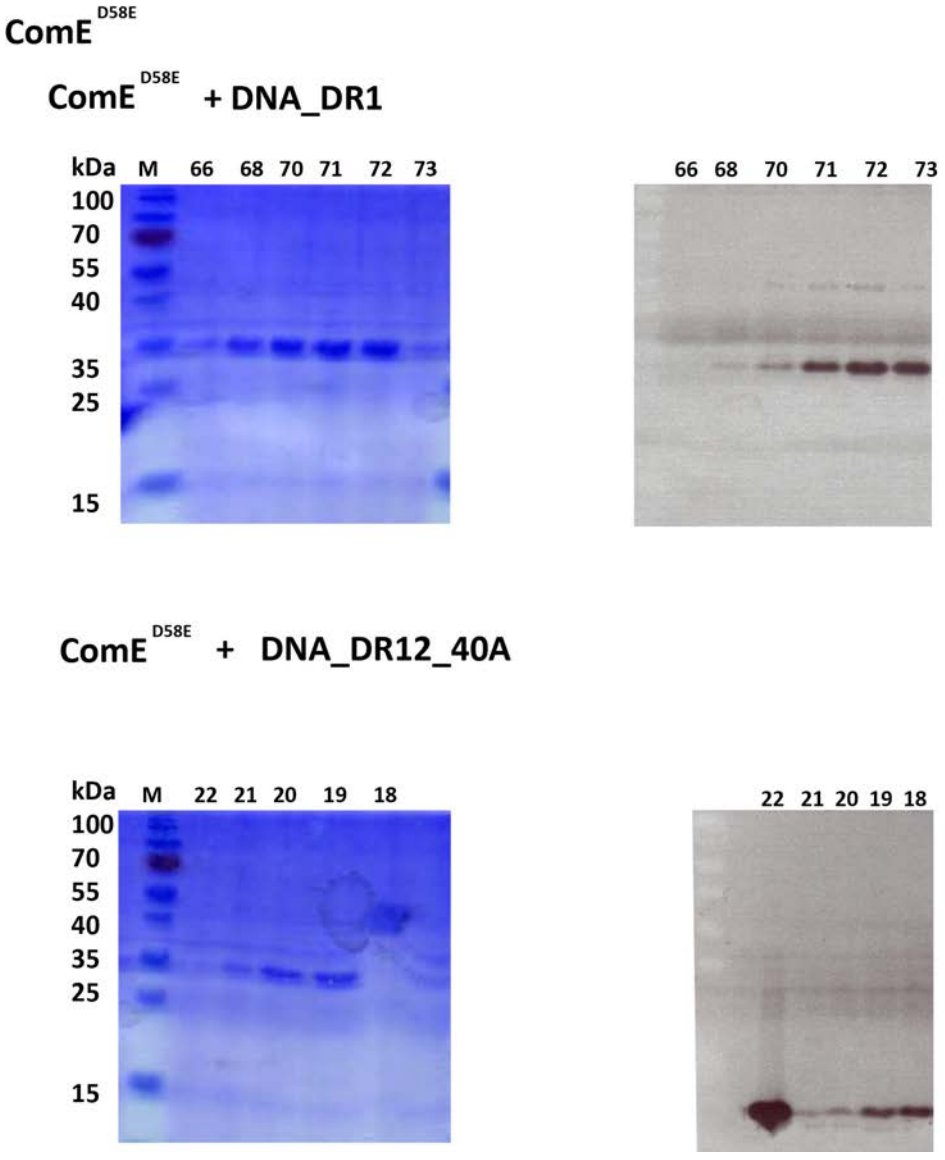


Figure 4.31 ComE^{D58E} protein-DNA complex trials with DNA_DR1 and DNA_DR12_40A

SDS PAGE on the left dyed with "one step blue" protein dye and on the right dyed with Syber safe DNA dye.

4.23. Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS)

First, the samples were run and analyzed separately as shown in Figure 4.32. Signal was normalized to the highest peak of the samples.

Intriguingly, MALS and RI signals do not overlap perfectly, indicating that MW is not stable, this also shows that there is more than one species present as shown in Figure 4.32; but the fact that the MW range between values obtained from this experiment and the theoretical values for both proteins (shown in Table 4.1) is under 10 kDa suggests that the equilibrium is shifted towards monomers (ComE 39 kDa vs. 33 kDa and ComD 31.5 kDa vs. 27 kDa).

Table 4.1 Individual MALS-RI SEC calculations for ComD, ComE and DNA_DR12_40A

	Mp (kDa)	MW (kDa)	Theoretical MW (kDa)	Calc mass (µg)	Mass fraction (%)
ComE	41.55 ± 0.02	39.25 ± 0.03	32.93	1620.50	95.73
ComD	33.12 ± 0.02	31.45 ± 0.04	27.03	63.39	92.48
DNA_ DR12_40A	26.20 ± 0.10	27.07 ± 0.16	24.6	9.02	100.00

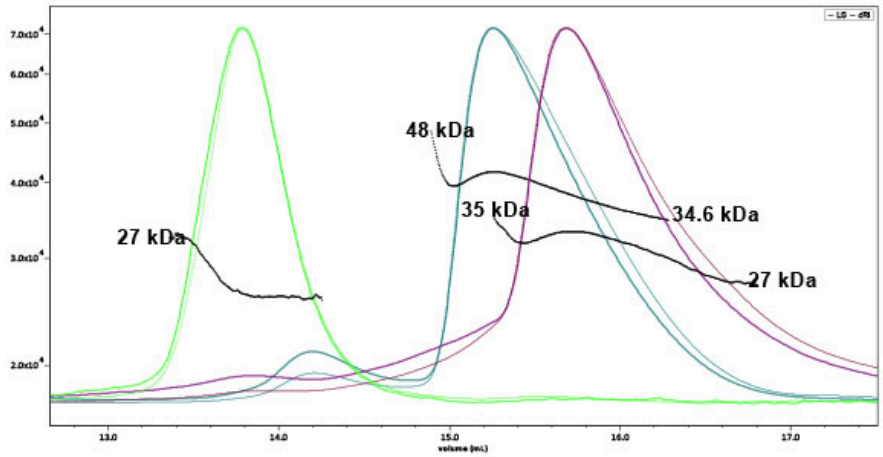


Figure 4.32 Individual MALS-RI SEC profiles for ComD, ComE and DNA_DR12_40A

ComD is shown in magenta, ComE in turquoise and DNA_DR12_40A in lime green. Signal was normalized to the highest peak of the samples.

To elaborate more on the monomer equilibrium, we performed a test to see if the amount of protein injected had an influence on the equilibrium. After performing the ComE injection, an initial fraction was injected, followed by a peak fraction and a final fraction to observe peak profiles for each part of the initial ComE peak. Strikingly the peak profile (Figure 4.33) was extremely similar as well as the calculated values for MW. In the initial ComE injection there is a second peak whose MW suggest that might be the dimer state of ComE. This suggests that when a threshold is surpassed the equilibrium shifts towards the appearance of a dimer state of the protein.

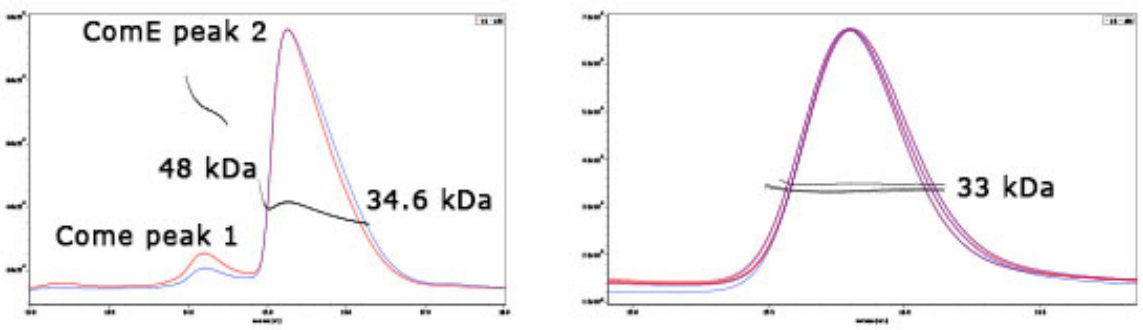


Figure 4.33 Individual MALS-RI SEC profiles for ComE

a) MALS-RI SEC Initial profile for ComE. b) Individual Initial peak ComE profile, Central peak ComE profile and Final peak ComE profile.

In Figure 4.34 we can observe that the protein-DNA complex is formed, and that the second peak is composed of ComE as it matches its theoretical weight (33 kDa). The calculated MW's suggest the dimerization of ComE joint with DNA_DR12_40A. This is in agreement with literature (Boudes *et al.,* 2014) as they also theorize the dimerization of ComE.

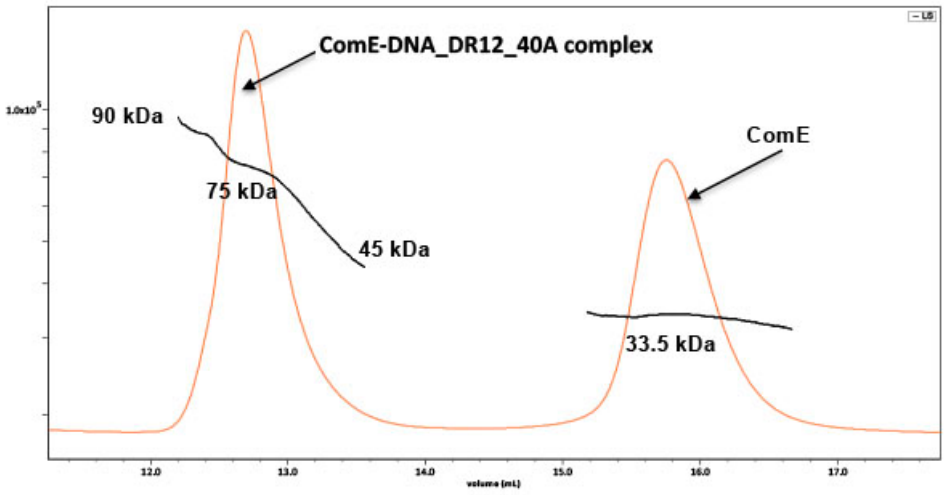


Figure 4.34 MALS-RI SEC profile for ComE-DNA_DR12_40A complex

The UV profiles obtained from the MALS experiments confirmed the presence of DNA in the sample, and suggest the formation of the binary (ComE + DNA) and ternary (ComD + ComE + DNA) complexes as it had a greater signal 260 nm (where DNA absorbs the most).

In Figure 4.35 we can observe appearance of the ternary protein-DNA complex, and that the second peaks MW suggests that it is composed by ComD + ComE as mentioned before. The MW of the ternary complex would be 90 kDa approximately but the average calculated value is 76.5 kDa. As its behavior is similar to the single protein profiles, it suggests equilibrium of species.

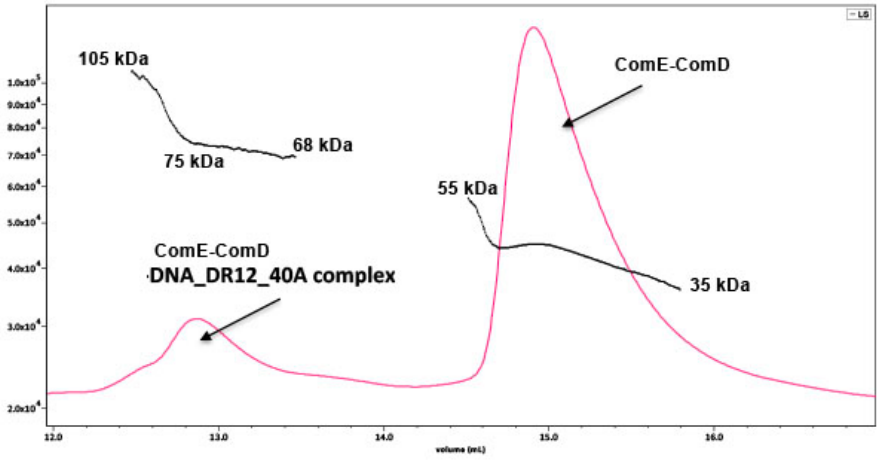


Figure 4.35 MALS-RI SEC profile for ComD-ComE-DNA_DR12_40A complex

4.24. Protein-DNA complex crystallization trials

It is obvious that the protein-DNA complex is binding better with the longer oligonucleotide. Nonetheless we tried many general crystallization screening plates for all three oligonucleotides. We observed promising conditions only with $ComE^{Wildtype}$. We tried crystalizing the binary and the ternary complexes and the best condition for both was PAC 1 B02 (0.1 M HEPES pH 7.5, 0.28 % PEG 400 and 0.2 M Calcium Chloride). The micro crystals that appear are quite different for both as shown in Figure 4.36.

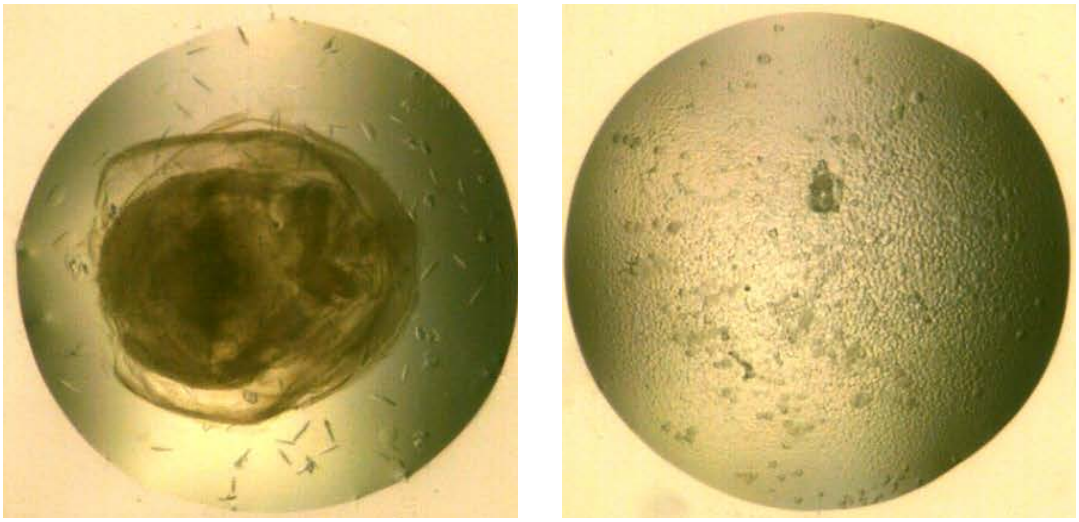


Figure 4.36 Binary (ComE+DNA_DR12_40A) and tertiary (ComD+ComE+DNA_DR12_40A) complex crystals
Binary complex is shown on the left and Tertiary is shown on the right.

4.25. Protein-DNA crystal diffraction

A couple of micro crystals were retrieved from a PAC 1 B02 Optimization (0.1 M MES pH 6, 0.23 % PEG 400 and 0.2 M Calcium Chloride) and tested at the Alba Synchrotron BL13-XALOC beamline. They diffracted at very low resolution (15 Å) (Figure 4.37). Even though several optimization plates were set up, it was not possible to obtain bigger crystals.

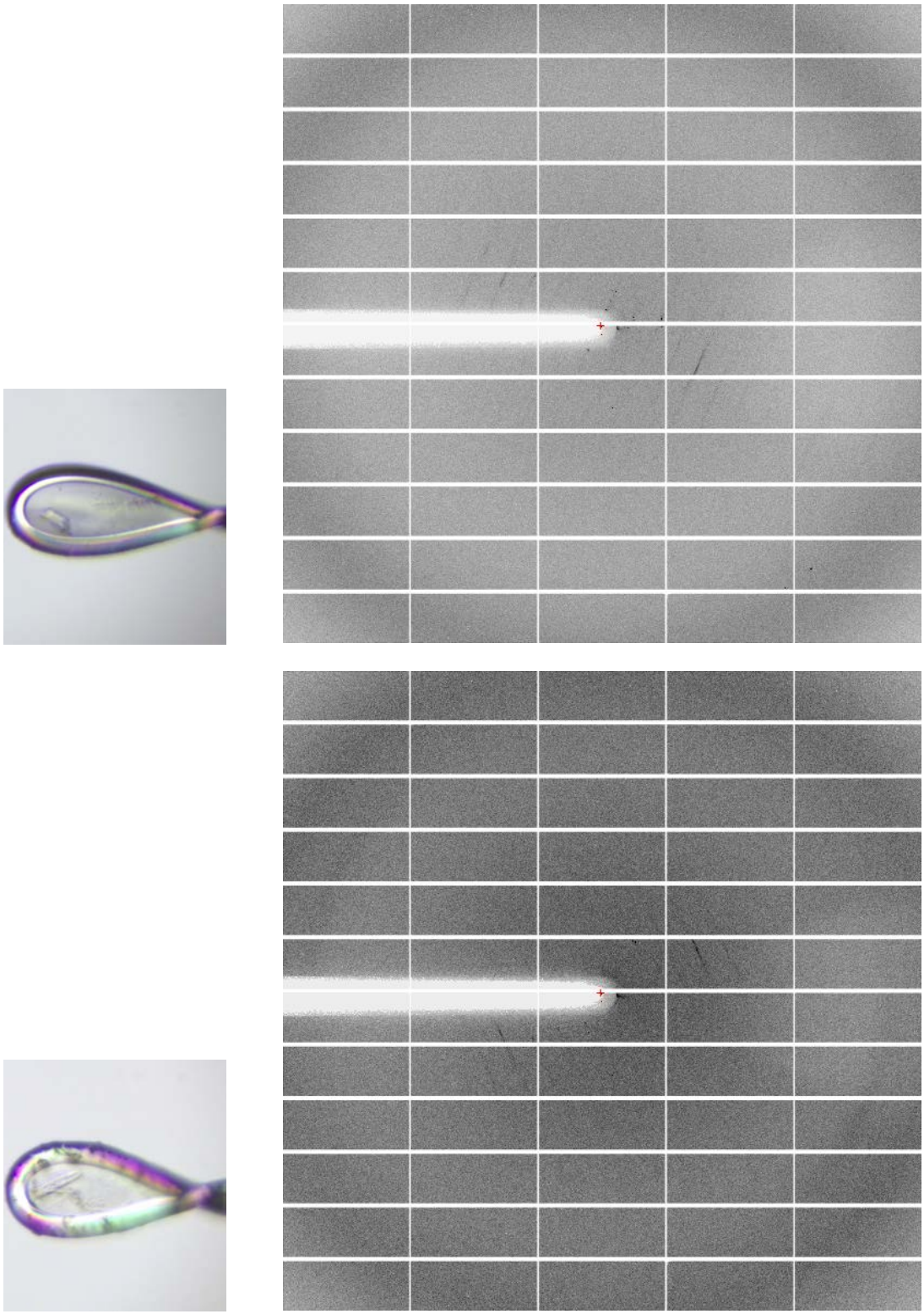


Figure 4.37 ComE+DNA_DR12_ 40A micro crystals and their diffraction pattern

4.26. Oligonucleotide redesign and crystallization trials

The oligonucleotide was redesigned in the attempt to see if changing the DNA structure would allow for the crystal packing to occur in an ordered manner.

4.27. Oligonucleotide redesign

While trying to crystalize $ComE^{Wildtype}$ and DNA_DR12_40A complex we observed the formation of round granules (Figure 4.38). These crystalline round granules appeared within hours of mixing the drop; many adjustment trials were performed like changing protein concentration, changing the protein to crystal buffer proportion or adding additives in order to allow crystals to appear.

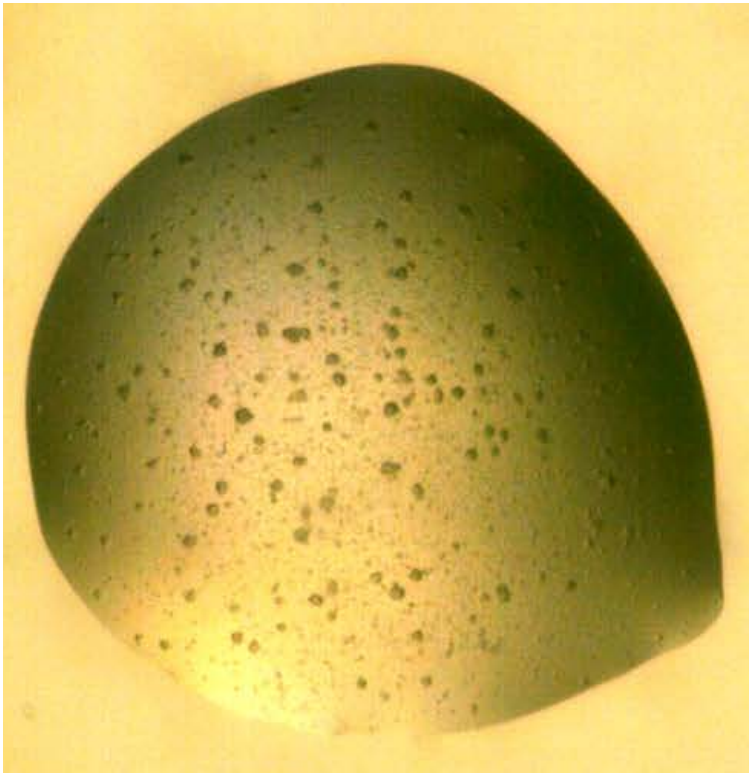


Figure 4.38 ComE+DNA_DR12_40A complex micro round nodules crystals

The fast appearance of these crystals led us to hypothesize that slightly modifying the oligonucleotides would allow the complex to crystalize properly, so we decided to design new oligonucleotides that

could possibly crystalize more effectively with ComE^{Wildtype}. The changes done to the design could improve crystal packing, which would translate into the appearance of protein-DNA complex crystals. These are shown in Figure 4.39.

Figure 4.39 DNA sequence of the *comCDE* promoter and the second round of oligonucleotides

DR1 binding site is shown in green, DR2 binding site is shown in blue, modified ends are marked with a red asterisk.

4.28. Redesigned oligonucleotide EMSA

We decided to test the binding of ComE^{Wildtype} and these newly designed oligonucleotides through EMSA (Figure 4.40). It is possible to conclude that the affinity is better for the original long DNA oligonucleotide that includes both direct repeats.

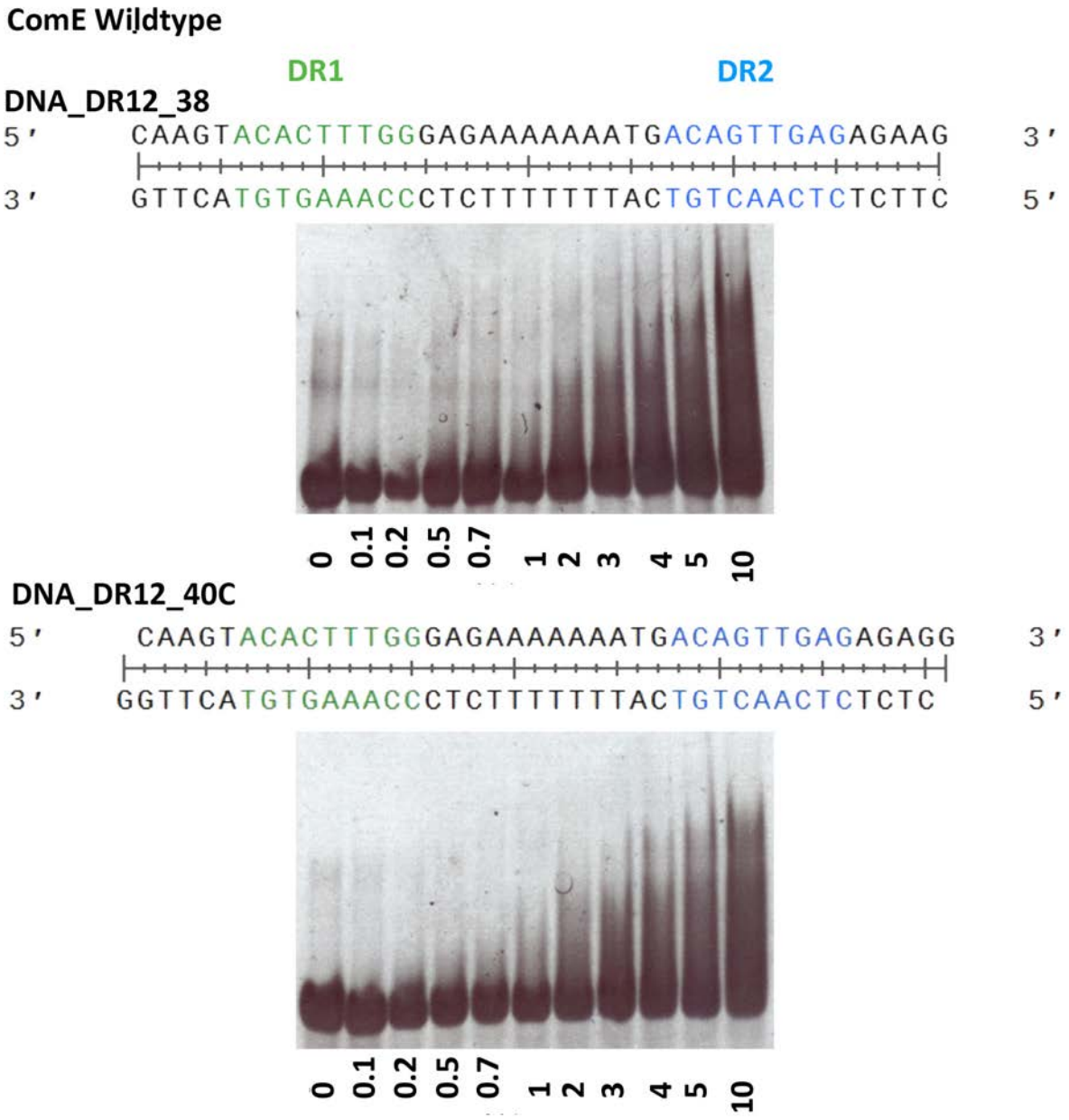


Figure 4.40 EMSA of "DNA_DR12_38" and "DNA_DR12_40C" with ComE^{Wildtype}

4.29. Redesigned oligonucleotide complex formation

Complex formation analysis was performed via SEC with the new oligonucleotides and ComE^{Wildtype}, which is shown in Figure 4.41.

protein-DNA complex was incubated using a 2:1 (protein: DNA) ratio for three hours before injecting the sample into the SEC.

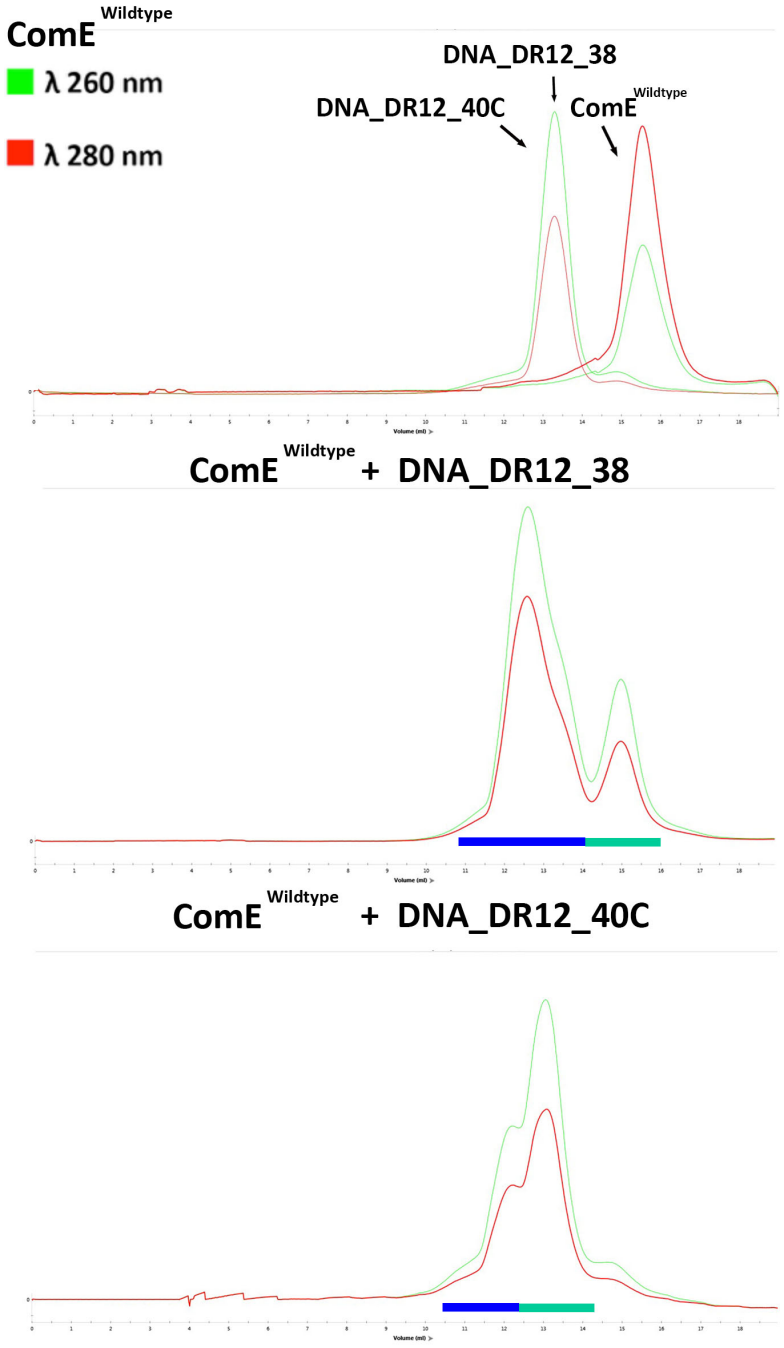


Figure 4.41 ComE wildtype protein-DNA complex trials with DNA_DR12_38 and DNA_DR12_40C

First chromatogram corresponds to the individual samples analysis. Both oligos when loaded into the SEC eluted at the same time.

We could visualize the proteins involved by staining the SDS-PAGE with “one step blue” and by staining them with Syber we could verify the presence of DNA as shown in figure 4.42.

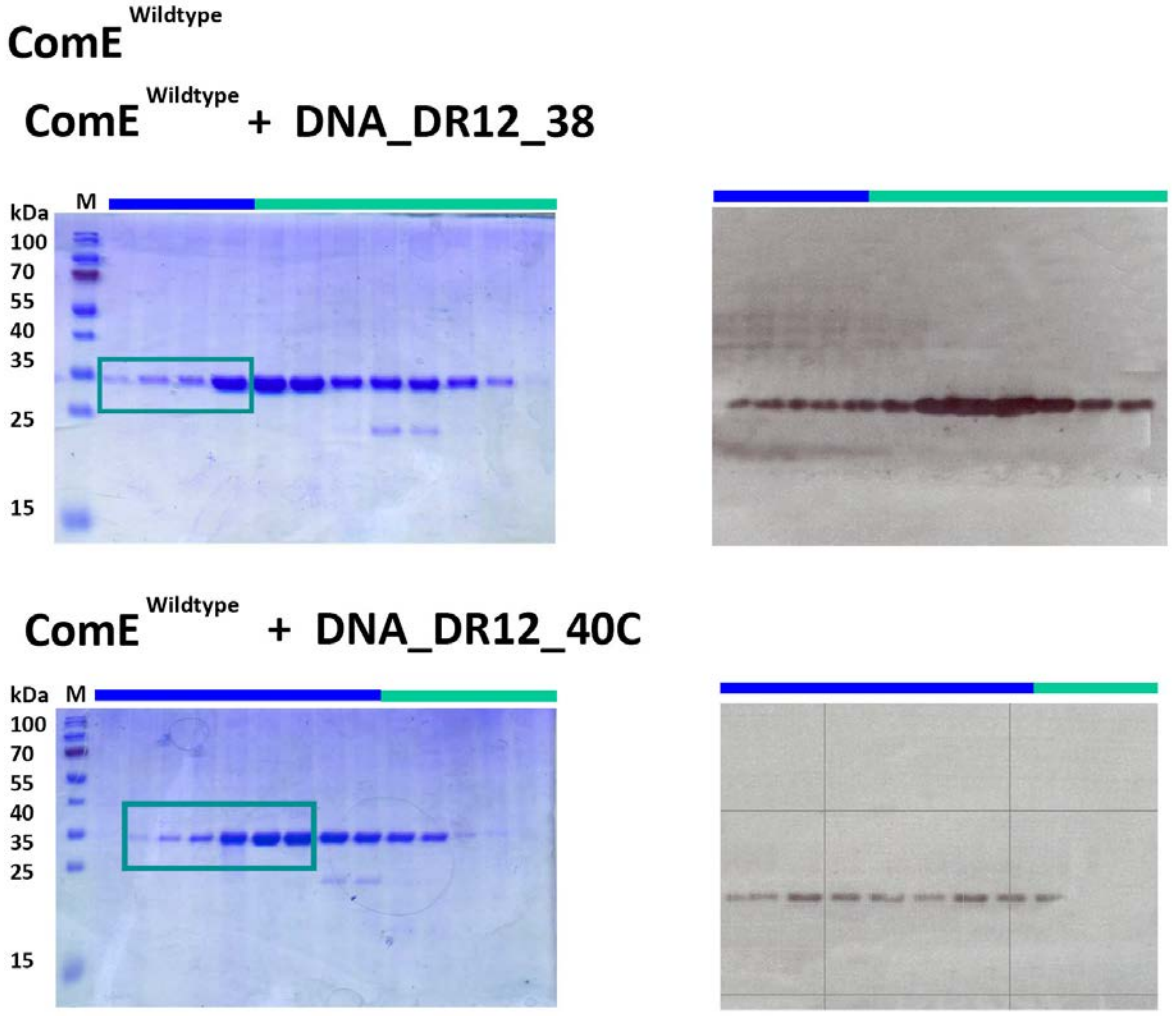


Figure 4.42 ComE^{Wildtype} protein-DNA complex trials with DNA_DR12_38 and DNA_DR12_40C SDS PAGE

Stained with one step blue on the left and Syber on the right.

Many general crystallization screening plates were set up at different conditions, no improvement appeared with the new oligonucleotides.

5. Conclusions ComD-ComE

The conclusions of this section of the thesis are the following:

1) ComD and ComE can be expressed and purified. The ComD-ComE complex binding seems not to be strong enough to endure the elution in a double SEC in tandem, suggesting that a number of factors could destabilize this complex.

2) ComD and ComE pure samples were obtained, and complexed with oligonucleotides containing both Direct Repeats from the *comCDE* promoter. The best complexes were obtained with oligonucleotides containing direct repeat 1 or both repeats in a 40 bp oligonucleotide. However, we were not able to optimize any condition and obtain crystals that diffracted at an acceptable resolution.

3) The MALS experiment suggests that at high concentrations of ComE, the dimer state has a tendency of appearing, which would explain the variability in the detections of the MW of the elution peak.

4) Crystals obtained were subject to X-ray analysis using Synchrotron radiation. Although diffraction was observed, the best resolution was not good enough for solving the structure. The variability of the protein species (oligomers) depending on the concentration suggests that a good quality diffracting crystal could be difficult to obtain. The lack of a complete homogeneity does not allow high quality-viable crystals to form.

5) Mutants seem to successfully mimic the nonphosphorylatable and the phosphorylated states, although surprisingly, the wildtype has more affinity to the DNA_DR12_40A oligonucleotide that we designed. This could be because there might be a different mechanism involved in this particular TCS and/or the glutamine might not imitate the phosphorylated state properly.

6) We theorize that there might be a mechanism during expression that phosphorylates the wildtype, which would explain why the binding of the wildtype is slightly higher than the one of the phosphorylating mimicking mutant.

Part II

ToxR-*toxT* interaction

Project II ToxR-*toxT* interaction

The second project focuses on the different structure interactions with DNA that arise after mutating residues that our collaborator observed cause a decrease in promoter activation

. The **Introduction** focuses in a one-component signaling system and summarizes the importance of studying cholera and understanding how to avoid and control it.

. The **Objectives** section enumerates the goals of this project, which includes obtaining the final crystal structure of three mutants.

. The **Materials and methods** section describes techniques, materials and instruments used throughout the realization of the thesis. It is divided in seven parts, the first three deal with the background and the formation of mutants that downregulate the activation of the promoter, the next two are related to the formation and crystallization of the new protein-DNA complexes and the three describe the steps needed to solve the structures.

. The **Results and discussion** chapter describes the process in which experimental data was obtained and interprets it. We highlight the changes that occur in the final structure and how these relate to the diminished activation of the *toxT* promoter.

. Finally, the **Conclusions** section recaps the main findings of this part of the PhD thesis.

6. Introduction ToxR project

In this part of the thesis we will elaborate on a one-component system; ToxR and its host, *Vibrio cholerae*.

6.1. ToxR as a one-component bacterial signal transduction system

One-component systems are comprised by a protein that contains both the sensory domain and an output domain as shown in Figure 6.1. There is more known about the two-component systems in comparison to the one-component systems, although one-component systems are more common in prokaryotes (Ulrich *et* al*.,* 2005). ToxR is an example of a one-component signal transduction system.

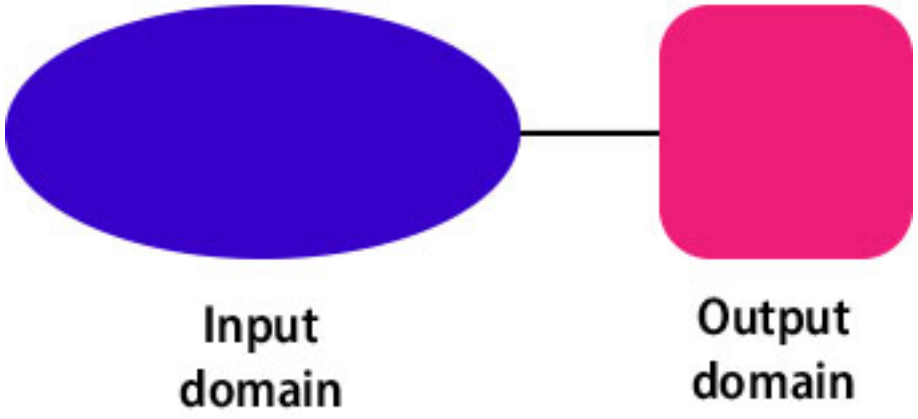


Figure 6.1 One-component system scheme

After sensing the environmental signal, the input domain produces the activation of the output domain without any chemical alteration.

Most one-component systems have a Helix-Turn-Helix (HTH)-DNA binding domain, and ToxR is no different (Buchner *et al.,* 2015). Although these HTH-DNA binding domains are not unique to the one-component systems, they do offer a simpler solution for signal transduction (Kolb *et al.,* 1993).

There are more than twenty known families of one-component systems in prokaryotes. They are defined by distinct conserved motifs and by amino-acid conservation in their DNA-binding domains and are shown in Table 6.1 (Cuthbertson, 2013).

Table 6.1 Examples of known families of one-component signal transduction systems
(Extracted from Cuthbertson, 2013).

One-component system	Defining features	Reference
AraC/XlyS	Involved in regulating pathways for the catabolism of various sugars, primarily transcriptional activators, C-terminal DNA-binding domain	Martin R.G. *et al.* 2001
ArgR	Involved in regulating amino acid metabolism, typically function as transcriptional repressors, N-terminal DNA-binding domain	Cherney L.T. *et al.* 2008
ArsR/SmtB	Involved in regulating metal homeostasis, primarily transcriptional repressors, DNA-binding domain located near the center of the protein	Busenlehner L.S. *et al.* 2003
AsnC/Lrp	Involved in regulating amino acid metabolism, function as both transcriptional activators and repressors, N-terminal DNA-binding domain	Yokoyama K. *et al.* 2006
Crp/Fnr	Involved in regulating many cellular processes, may function as activators and repressors, C-terminal DNA-binding domain	Körner H. *et al.* 2003
DeoR	Involved in regulating sugar metabolism, typically function as repressors, N-terminal DNA-binding domain	Zeng G. *et al.* 1996
DtxR	Involved in regulating metal homeostasis, primarily transcriptional repressors, N-terminal DNA-binding domain	Pennella M.A. *et al.* 2005
Fur	Involved in regulating metal homeostasis, primarily transcriptional repressors, N-terminal DNA-binding domain	Pennella M.A. *et al.* 2005
GntR	Involved in regulating numerous cellular processes, typically function as transcriptional repressors, N-terminal DNA-binding domain	Rigali S. *et al.* 2002
IclR	Involved in regulating carbon metabolism, function as both transcriptional activators and repressors, N-terminal DNA-binding domain	Molina-Henares A.J. *et al.* 2006
LacI	Involved in regulating carbon metabolism, typically function as transcriptional repressors, N-terminal DNA-binding domain	Swint-Kruse L. *et al.* 2009
LuxR	Involved in regulating quorum sensing, typically function as activators, C-terminal DNA-binding domain	Chen J. *et al.* 2011
LysR	Involved in regulating many cellular processes, function as both activators and repressors, N-terminal DNA-binding domain	Maddocks S.E. *et al.* 2008
MarR	Involved in regulating antibiotic resistance, typically function as transcriptional repressors, DNA-binding domain located near the center of the protein	Wilkinson S.P. *et al.* 2006
MerR	Involved in regulating metal homeostasis, typically function as transcriptional repressors, N-terminal DNA-binding domain	Hobman J.L. *et al.* 2005
MetJ	Involved in regulating many cellular processes, typically function as transcriptional repressors, N-terminal DNA-binding domain	Schreiter E.R. *et al.* 2007
ModE	Involved in regulating metal homeostasis, function as both transcriptional activators and repressors, N-terminal DNA-binding domain	Self W.T. *et al.* 1999
PadR	Poorly characterized family, N-terminal DNA-binding domain	Huillet E. *et al.* 2006
TetR	Involved in regulating antibiotic resistance, typically function as repressors, N-terminal DNA-binding domain	Ramos J.L. *et al.* 2005
ToxR	Involved in regulating various cellular processes, N-terminal DNA-binding domain, one transmembrane helix	Buchner S. *et al.* 2015
Xre	Involved in regulating various cellular processes, typically function as transcriptional repressors, N-terminal DNA-binding domain	McDonnell G.E. *et al.* 1994

Interestingly enough, the number of proteins from these families that are found in eukaryotes is minimum while the vast majority is found in bacterial genomes, and the reason for this lack of one-component regulators detected in eukaryotic genomes still has not been discovered (Cashin P *et al.,* 2006). This search was done against the protein database using Pfam version 19.0 (Bateman *et al.,* 2004).

It is worth mentioning that although it is often referenced to a response to an extracellular signal, many bacterial signal systems act intracellularly (Cashin *et al.,* 2006). This was confirmed by using the dense alignment surface method (Cserzo *et al*., 2004), they uncovered that most (97%) of one-component regulators with the HTH-binding domain lack transmembrane regions, predicting them to be cytoplasmic proteins (Ulrich *et al.,* 2005).

The ToxR family possesses a transmembrane domain between the output domain and the input domain. They have a C-terminal periplasmic sensing domain, a single transmembrane helix and an N-terminal cytoplasmatic winged-HTH-DNA binding domain.

6.2. Definition of cholera

Cholera is an acute diarrheal infection provoked by the ingestion of food or water contaminated with the gram-negative bacterium *Vibrio cholerae* of the O1 or the O139 serogroup and it's incubation period fluctuates from two hours up to five days. This disease is endemic in more than fifty countries as well as responsible for large epidemics worldwide (WHO, 2017), although it is rare in industrialized nations. Worldwide, registered cases of the disease have increased constantly since 2005, and the disease still proliferates in many places like Haiti, Southeast Asia and Africa (Mintz, *et al.,* 2013). Although cholera can be easily prevented as it depends on having access to safe water and efficiently treated, it can be life threatening. Patient mortality can be decreased greatly when they are treated with intense fluid replacement, and diarrhea can be reduced when treated with antibiotics (Harris et *al.,* 2012).

6.3. History of cholera

There is evidence that cholera has been an undesirable visitor for human civilization for at least one thousand years. There are records that it has had an impact in society, specifically the Bengal society as there is a goddess of cholera that asked for redemption in exchange of

protection of villages from the disease called Oola Beebee (Morris, 2011).

In 1817, the disease disseminated from India, and since then we have identified 7 cholera pandemics throughout the world (Mukandavire *et al.,* 2001). In 1854, a British physician who is now considered one of the fathers of epidemiology, called John Snow suggested that infected stools could contaminate drinkable water supplies. To test this, he conducted what could be considered his pioneering epidemiologic research as he created a map that illustrates (Figure 6.2) the proximity of the cluster of cholera cases and drew up the relationship between them (Johnson, 2006).

Figure 6.2 John Snow map of the location of the cholera cases and the water pipes in the London epidemic of 1854
(Retrieved from Johnson, 2006).

In 1884, Robert Koch isolated *V. cholerae* and described the disease (Lakhtakia, 2013), this was not the first time this happened, because

in 1854, Filippo Pacini performed the first observations and deductions of *V. cholerae* and although they were path breaking, they were ignored at the time (Howard-Jones, 1984).

6.4. Cholera worldwide

Cholera cases are seldom detected or registered, unless they are part of an outbreak or a pandemic. In 2017, 34 member states of the WHO reported a total of 227,391 cases and 5,654 deaths; an increase of 45% in the number of cases and 33% in the number of deaths from 2016. Yemen, which suffered from an outbreak, reported 84% of all cases and 41% of the fatalities (WHO, 2018a). Approximately 1.3 billion people are at risk for cholera in endemic countries, and it is estimated that 2,866 million cholera cases occur annually in said countries, and around 95,000 result in death. It is estimated that around half of cases and deaths occur in children under 6 years of age (WHO 2018b). In 2015, 69 countries were considered endemic and the cases were divided as described in Figure 6.3 (Ali *et al.,* 2015).

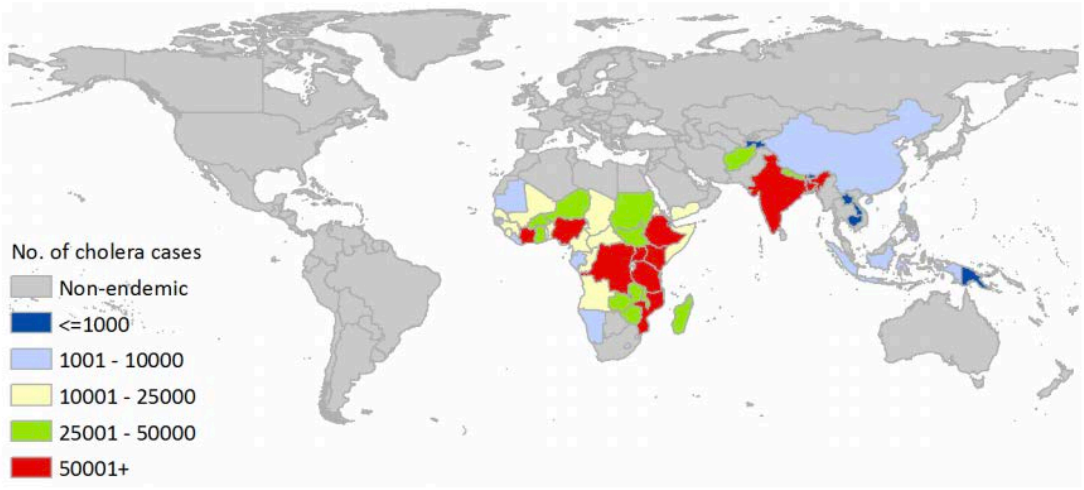


Figure 6.3 Number of cholera cases in endemic countries in 2014
(Image retrieved from Ali *et al.,* 2015).

Cholera is an opportunistic disease that thrives because of the continued destruction or deterioration of infrastructure services caused by conflict or natural disasters, which also lead to the contamination of drinking water sources (WHO, 2018b).

6.5. The pathogen "*Vibrio cholerae*"

V. cholerae is a gram-negative bacterium shown in Figure 6.4 (Kirn *et al.,* 2000). There are more than 200 serogroups. Epidemic cholera is only caused by serogroups O1 and O139. O1 is divided into the classical and El Tor biotype, the latter causing the most recent pandemic whilst the fifth and sixth pandemic was caused by the classical biotype (Harris *et al.,* 2012).

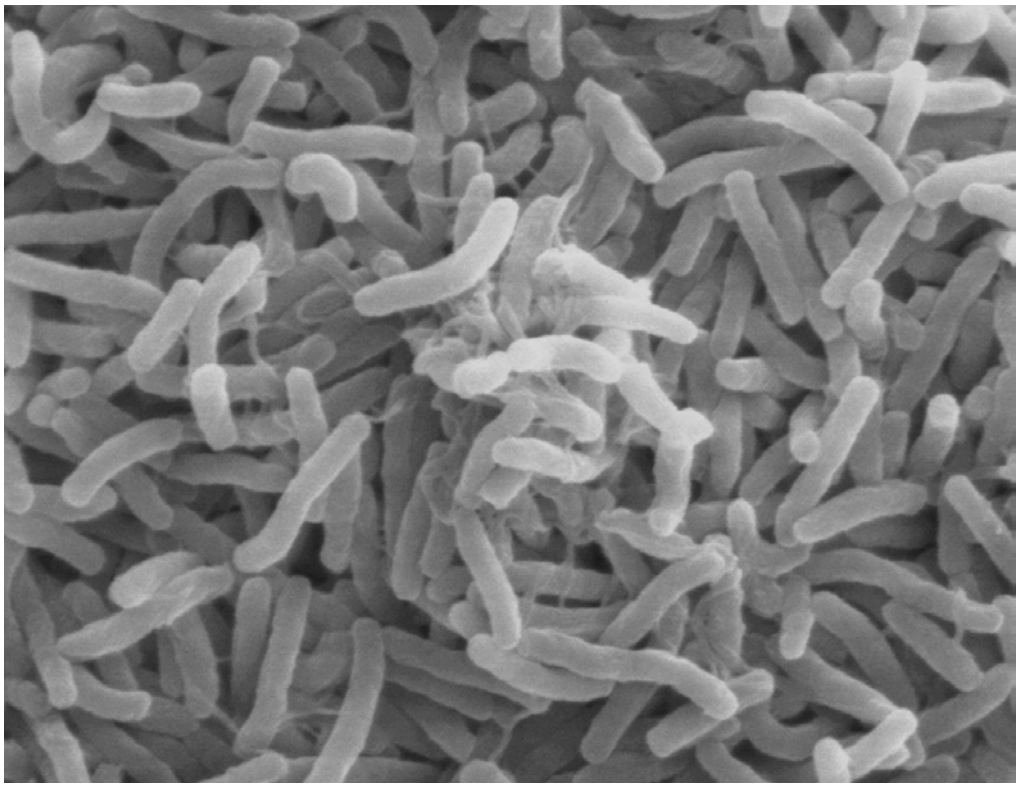


Figure 6.4 Scanning Electron Microscope (SEM) image of *Vibrio cholerae* bacteria
(Image extracted from Harris *et al.,* 2012).

It is a common and natural inhabitant of the aquatic environments that exist, both pandemic and non-pandemic strains. Its distribution depends on water salinity and temperature; as it can go undetected during cold weather and its environmental counts can increase during warmer periods (Morris, 2011). It is a survivor, some strains can adopt a rougher phenotype by producing high amounts of an exo-polysaccharide, which forms a biofilm that protects the bacterium

against UV light, chlorine and other standard disinfectants (Morris *et al.,* 1996). It is capable of staying viable for 50 days in a warm and low salinity environment (Harris *et al.,* 2012).

During its life cycle it transfers from an aquatic environment into host environments. When *V. cholerae* is ingested, it moves to the small intestine to colonize it. Colonization occurs after TCP expression, which aids and protects the bacteria by forming *V. cholerae* biofilm (Taylor *et al.,* 1987). When the environment is not ideal, the bacterium can switch into a non-culturable state that remains viable if ingested (Morris *et al.,* 1996).

6.6. *Vibrio cholerae* process

In the 1860's it was thought to be transferred from person to person, this conception changed in the following years when the attention focused on the environmental factors that triggered the disease, it is now known that it is caused by consuming food or water with contaminated material (Morris, 2011).

Most cholera bacteria are killed by gastric acid when they are ingested, yet the ones that survive are capable of producing CTX and TcpP and colonizing the small intestine. The mechanism of the cholera Toxin is illustrated in Figure 6.5. CTX is a toxin protein that consists of an A subunit and five B subunits. The five B subunits attach to the ganglioside receptor, and then the A subunit enters the cell membrane where it increases the adenylate cyclase activity and thus the intracellular cyclic AMP, generating a loss of nutrients in the form of secretory diarrhea (Gill *et al.,* 1978).

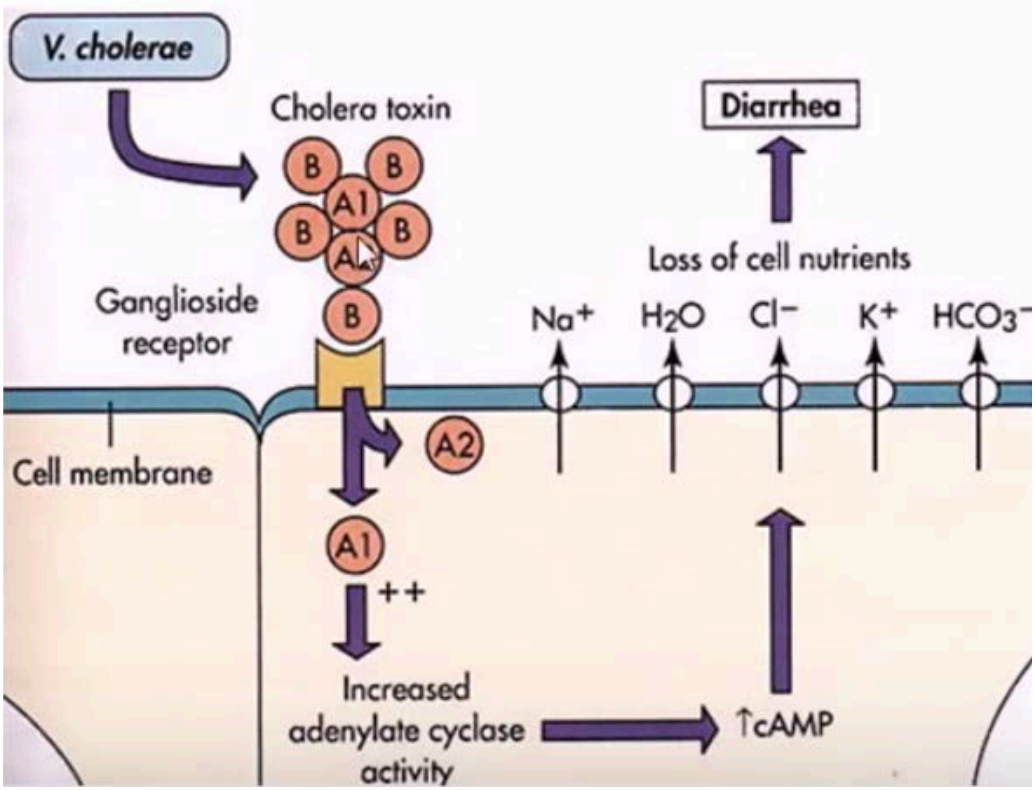


Figure 6.5 Cholera toxin, mode of action
(Extracted from Gill *et al.,* 1978).

6.7. ToxR, TcpP and CTX activation and transcription

The expression of CT and TCP is controlled by the activation of the *toxT* regulator, whose expression is dependent on two transcriptional factors (ToxR and TcpP) (Hase & Mekalanos, 1998). TCP is a protein that aids *V. cholerae* in the colonization of the intestine, by bringing bacteria in micro colonies, concentrating their secreted CTX and protecting them from the host defenses (Taylor *et al.,* 1987). CTX is an exotoxin protein that is responsible for the secretory diarrhea and vomiting (Gill *et al.,* 1978).

TcpP and ToxR are transmembrane transcription factor proteins with a cytoplasmic domain that belongs to the w-HTH family of transcription factors (Martínez-Hackert & Stock, 1997b) and to achieve complete transcriptional activation of the *toxT* promoter, cooperative binding of both transcription factors is required (Goss *et al*., 2010; Krukonis *et al.,* 2000).

6.8. The signaling cascade that activates ToxR and TcpP

V. cholerae has to travel from an aquatic environment to the human intestine facing dramatic changes of temperature, salinity amongst other environmental factors that stimulate the activation of the signaling cascade that culminates in the activation of ToxR and TcpP which are two membrane proteins that activate the master virulence regulator *toxT* (Herrington et *al.,* 1988). ToxR and TcpP result in the expression of ToxT, which activates the expression of two virulence factors: The cholera toxin (CTX) and the toxin co-regulated pilus (TCP). ToxT expression is regulated by a membrane-related mechanism controlled by TcpP and ToxR, which are bitopic proteins that collaborate in the activation of *toxT* transcription as well of other regulatory genes while remaining inside the membrane (Matson *et al.,* 2007; Miller *et al.,* 1989).

Both, TcpP and ToxR have a similar conformation; they have an N-terminal cytoplasmic winged HTH-DNA binding domain homologous to the OmpR/PhoB family of transcriptional activators (Martínez-Hackert *et al.,* 1997a), a transmembrane domain and a C-terminal periplasmic domain.

Apparently, TcpP has an important role for transcription activation binding to a region near the -35 position in the *toxT* promoter, whilst ToxR seems to play an auxiliary role binding to a distal region near the -80 position (Krukonis *et al.,* 2000). However, in the *ompU* promoter only ToxT binds near the -35 sequence, acting by itself to activate transcription. *ToxR* and TcpP operate alongside membrane found periplasmic effector proteins, ToxS and TcpH, respectively (DiRita *et al.,* 1991; Häse *et al.,* 1998). In the absence of TcpH, TcpP is degraded (Matson *et* al., 2005), so TcpH apparently protects TcpP's periplasmic domain from proteolytic action (Beck *et al.,* 2004). It has been proposed that ToxS aids the dimerization of ToxR and facilitates transcriptional activation (DiRita *et al*., 1991; Pfau *et al.,* 1998). Until recently, it was uncertain what ToxS did; ToxR undergoes regulated intramembrane proteolysis during late stationary phase at alkaline pH

in response to nutrient limitation. According to experiments performed by Almagro-Moreno, ToxR is stabilized and protected from premature proteolysis by ToxS (Almagro-Moreno et *al.,* 2015). ToxR transcriptional activity is induced in media rich in Asparagine, Arginine, Glutamic Acid and Serine; and it is worth mentioning that both ToxR and TcpP are regulated by external stimuli (Mey *et al.,* 2012).

Histone-like nucleoid structuring proteins (H-NS) silence the expression of many environmentally regulated genes, like the *toxT* gene, during growth under non-permissive conditions. H-NS apparently binds to curved DNA, which is the type of DNA where promoters are commonly found (Rimsky *et al.,* 2001). These transcription factors bind to the *toxT* promoter when *TcpPH* and *ToxRS* are induced suggesting that the H-NS can repress transcription by reducing the rate of open complex formation at the promoter (Nye *et al.,* 2000). High-resolution microscopy proposes that H-NS sequesters bound DNA into compact clusters in *E.coli*, and this might explain why in *V. cholerae* it limits ToxR access to genomic locations (Kazi *et al.,* 2016).

The role in controlling the cholera toxin expression of the ToxR regulon is shown in Figure 6.6.

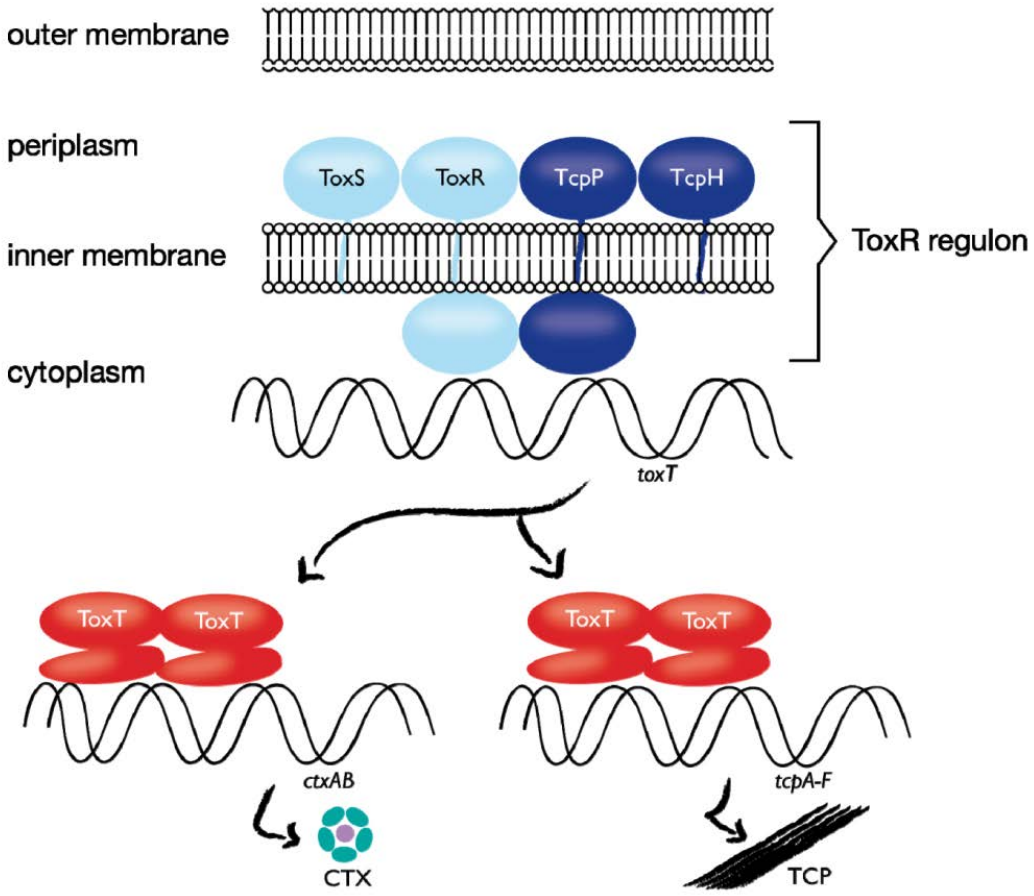


Figure 6.6 The ToxR regulon controls the expression of the cholera toxin (CTX)

TcpP and ToxR bind to the *toxT* promoter and the transcription of *toxT* is activated. ToxS and TcpH help stabilize ToxR and TcpP, respectively. ToxT activates *ctxAB* and *tcpA-F* transcription (extracted from Haas *et al.*, 2014).

ToxR and ToxS form a transcriptional activator complex called ToxRS while TcpP and TcpH form TcpPH. ToxRS has been characterized as an integral protein complex (DiRita *et al.,* 1996), yet it's expression can also be activated by a wing-helix DNA-binding protein called AphB (Xu *et al.,* 2010). This protein can activate the TcpPH operon and commence the virulence cascade, demonstrating that internal signals can also activate the virulence cascade (Kovacikova *et al.,* 2002). Figure 6.7 illustrates the ToxR regulon gene activation in *V. cholerae* (Amin Marashi *et al.,* 2013).

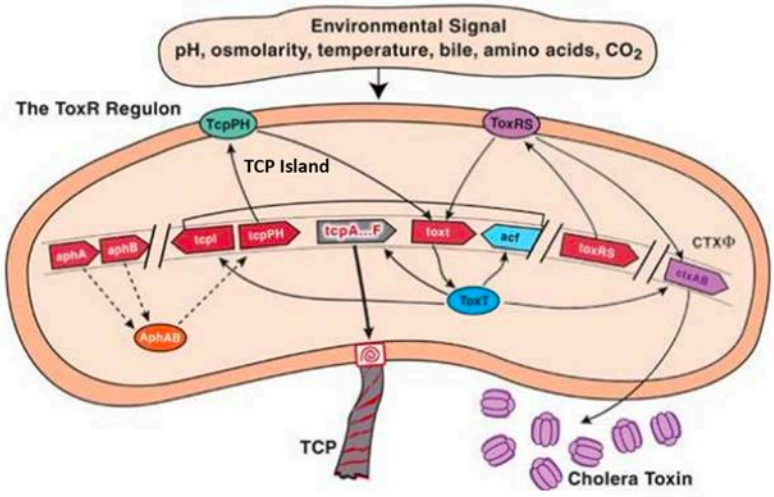


Figure 6.7 *Vibrio cholerae* ToxR regulon gene activation
(Extracted from Amin Marashi *et al.,* 2013).

6.9. The *toxT* promoter

Through DNase I footprinting, the ToxR binding site within the *toxT* promoter was defined. This analysis extended from -104 to -68 (Krukonis *et al.,* 2000). This is shown in Figure 6.8.

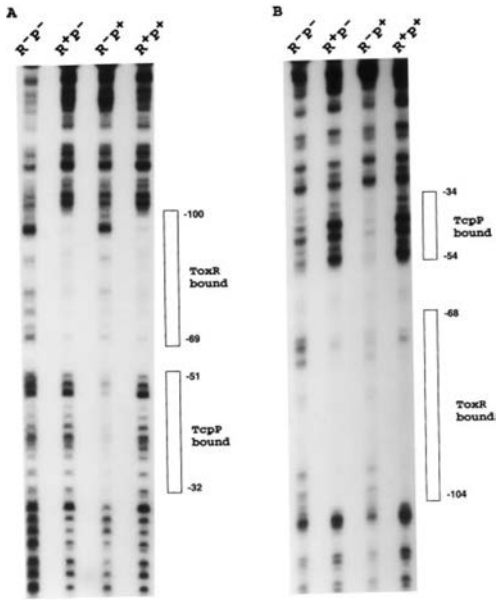


Figure 6.8 *toxT* promoter DNase I footprinting analysis
V.cholerae membrane extracted from strains expressing TcpP (P) and/or ToxR (R) and mixed with the *toxT* promoter in which the top(A)/bottom (B) stand was end labeled (Krukonis *et* al., 2000).

In Figure 6.9, it is possible to visualize the *toxT* promoter's DNA sequence from a *V.cholerae* classical strain O395. It shows the ToxR and TcpP binding sites as well as the Direct Repeats required for ToxR binding.

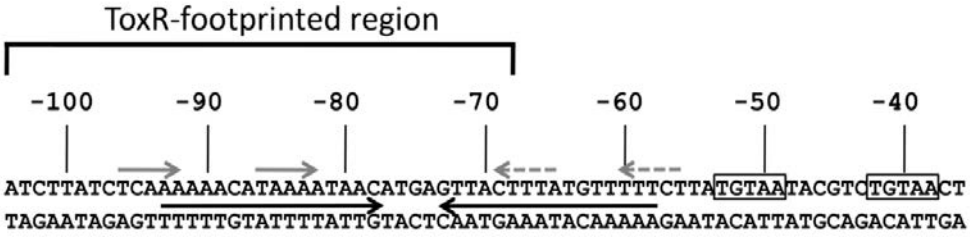


Figure 6.9 *V. cholerae* classical strain O395 DNA sequence of the proximal region of the toxT promoter

Solid grey arrows represent the Direct Repeat required for ToxR binding (5'-TNAAA-N5-TNAAA-3'). Solid black converging arrows show an inverted repeat sequence. Dashed grey arrows indicate degenerate ToxR binding site. The boxed sections indicate the TcpP binding site (Goss *et al.,* 2013).

TcpP can activate transcription of the *toxT* promoter in the absence of ToxR when it is overexpressed, yet ToxR cannot, even though ToxR can bind to the promoter when there is no TcpP. There for, it has been suggested that ToxR functions as a co-activator to the *toxT* promoter (Krukonis *et al.,* 2003).

6.10. The ToxR-*toxT* promoter interactions

In 2016, Simone Pieretti, a PhD student in our group solved the ToxR-*toxT* promoter complex crystal structure it is illustrated in Figure 6.10.

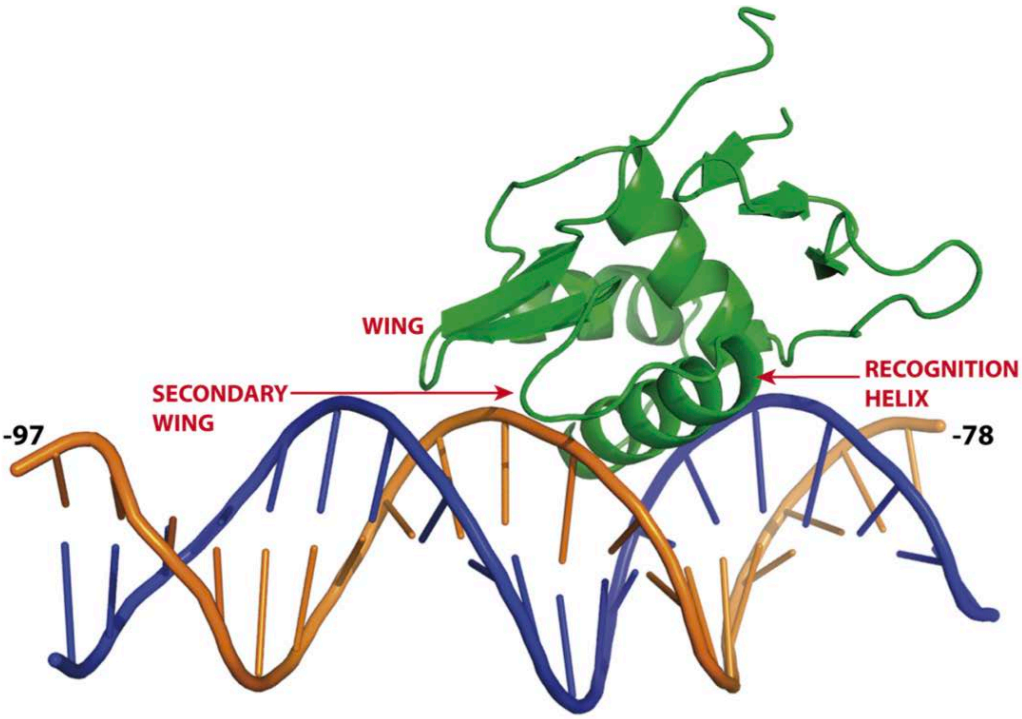


Figure 6.10 Pieretti's solved structure diffracted up to 2.07 Å
(Pieretti, 2016)

Pieretti described interactions between ToxR and the *toxT* promoter in his PhD thesis and Figure 6.11 illustrates them (Pieretti, 2016).

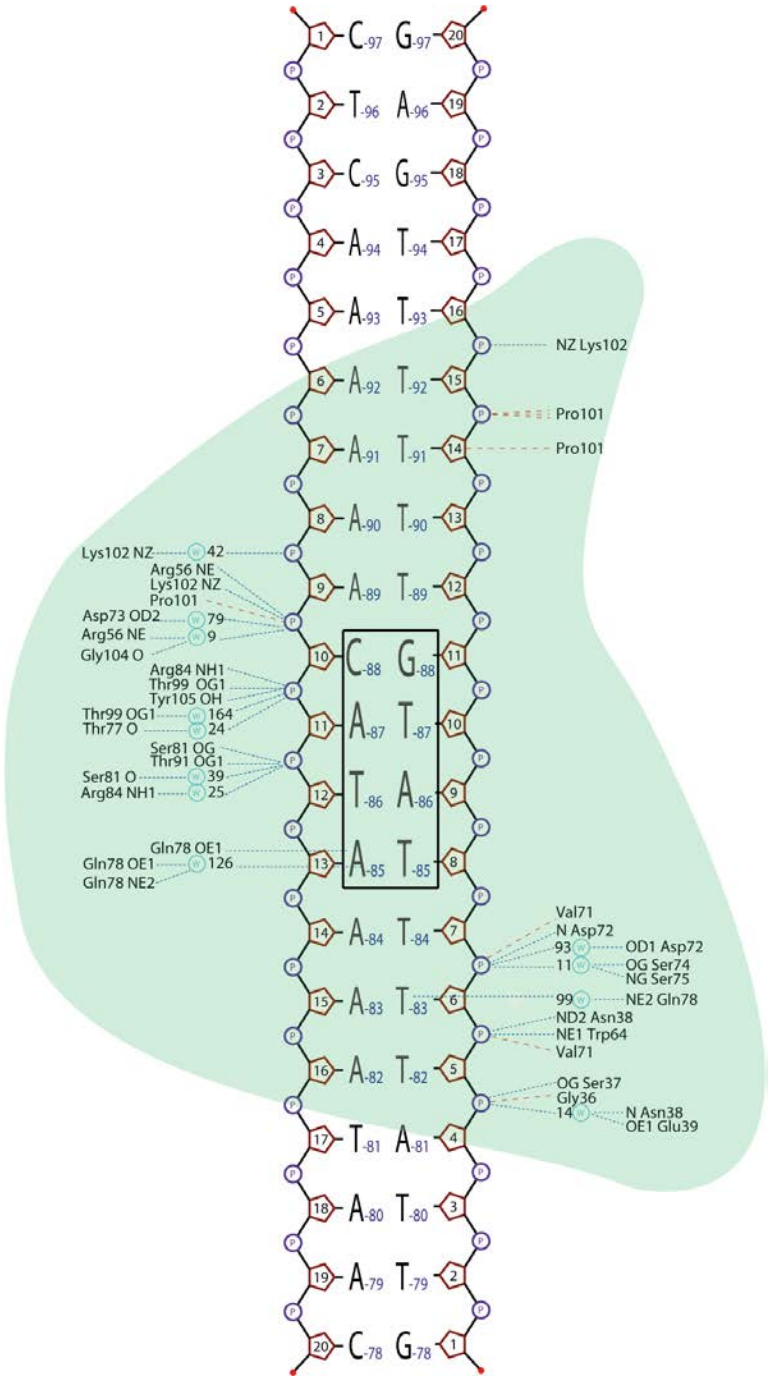


Figure 6.11 Interaction scheme of ToxR-DBD with DNA in the ToxRDNA20 structure

The CATA box and the ToxR binding site are highlighted (extracted from Pieretti, 2016).

Pieretti described the ToxR binding site and proposed an element referred to as “A-tract-CATA” element, which is characterized by an

AT-rich region sequence at the 5’ end followed by a CATA/CATG/TGTA box, where the last two bases (TA or TG) are required for protein binding.

ToxR DNA binding domain displays a heart-shaped α/β fold structure and is formed by a N-term five-stranded mixed β-sheet composed by four antiparallel β strands (β1-β4) and a fifth β strand parallel to β1 (β8); all of this, followed by a three-helix bundle (α1-α3) and a C-terminal β-hairpin (β6-β7) as shown in Figure 6.12 (Pieretti, 2016). α1-α3 helices compose the modified H-T-H DNA-binding motif, where α3 is the recognition helix and the α2-α3 connector replaces the turn. β5 is a short β strand positioned between α1 and α2 that interacts with the C-terminal hairpin forming a three-stranded antiparallel β-sheet. Recognition of the DNA minor groove is performed by the wing, which is formed by the turn between the two β strands of the C-term β-hairpin. ToxR-DBD is a winged-helix protein, because of its folding type; yet it possesses a secondary wing when compared with other proteins from the same family of the pdb database.

A central hydrophobic core is formed and is responsible for maintaining the fold of the protein, this is because of the residues Ile42, Leu43, Trp 44, Leu45 and Leu46 which aid the helix α1 as it connects the N-terminal β-sheet with the H-T-H motif and the β-hairpin.

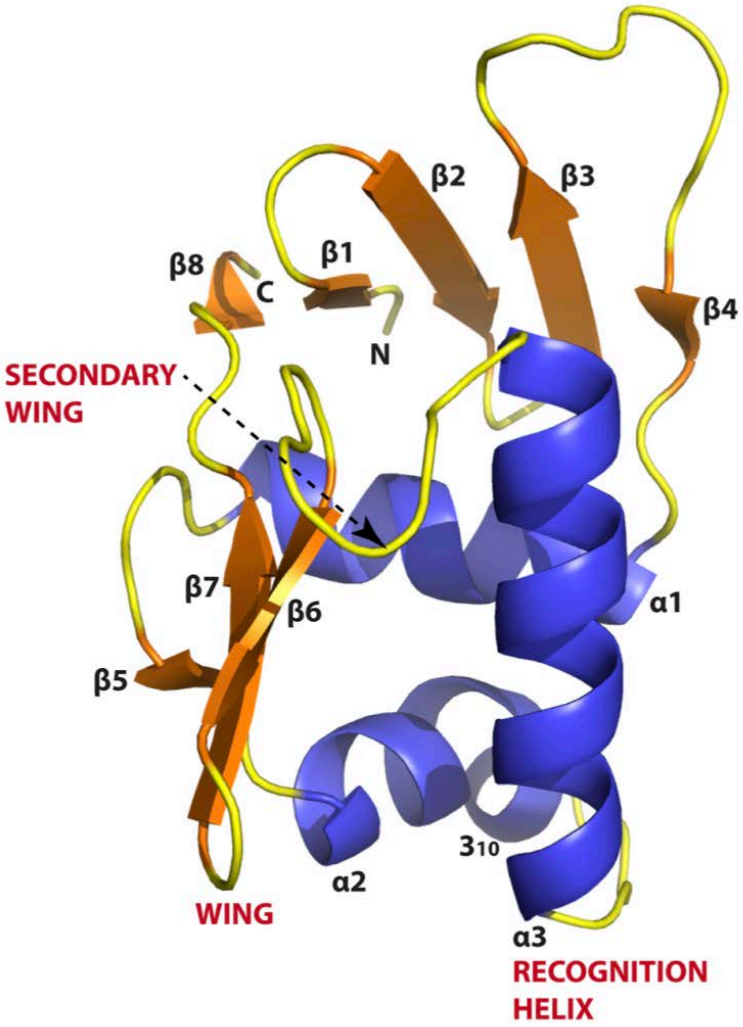


Figure 6.12 Structure ToxR-DBD

(Extracted from Pieretti, 2016).

Our lab is interested in observing the consequences of inserting a mutation that would affect the interactions reported by Pieretti and analyze when a decrease in activation occurs.

7. Objectives ToxR^{X} project

The objectives of this section were:

1. To solve the 3D high-resolution structure of the ToxR^{Q78A}-DNA, ToxR^{S81A}-DNA and ToxR^{P101A}-DNA, protein-DNA complexes using X-ray crystallography.

2. To analyze the ToxR^{Q78A}-DNA, ToxR^{S81A}-DNA and ToxR^{P101A}-DNA complexes structure and compare its features with the ToxR^{Wildtype}-DNA, in order to explain the decrease in the activation of the *toxT* promoter.

8. General Methods for the ToxR project

This section will focus on the methods used for experimentation in the project regarding ToxR mutants and *toxT* promoter activation.

8.1. ToxR mutants

As described in Simone Pieretti's thesis, ToxR DNA 20 bp oligonucleotide was designed and ToxR was purified to create a protein-DNA complex. Given that several amino acids of ToxR have interactions with the DNA, several mutants were produced by our collaborator Prof. Eric Krukonis to perform experiments that demonstrate that the activation of the *toxT* promoter is inhibited by the addition of single point mutations. Out of these, 3 were selected (Q78A, S81A and P101A) to attempt to crystalize with the previously used oligonucleotide shown in Figure 8.1.

5'-CTCAAAAAACATAAAATAAC-3'
3'-GAGTTTTTTGTATTTTATTG-5'

Figure 8.1 Sequence of DNA20

8.2. Production and purification of Q78A, S81A and P101A ToxR mutants

Protein purification was achieved through a multi-step chromatographic protocol using ÄKTA Purifier systems (GE Healthcare) or NGC Chromatography Systems (Bio-Rad) using the columns below, following manufacturer instructions (GE Healthcare):

- Histrap HP (5 ml) (Immobilized metal ion affinity chromatography (IMAC)):

Column was washed with 5 CV of H_2O, then washed with 5 CVs of buffer A, another 5 of buffer, and finally 10 CVs of buffer A. Protein sample was loaded into the system and then column was washed with 10 CVs of buffer A, after this, protein was eluted using a linear

gradient of 15 CV of buffer B. A 5 CV wash of buffer B was performed to ensure that all protein was released from the column.

- Superdex 75 gl (10/300) increase (Gel Filtration):

Sample was introduced and eluted using 1.2 CV of buffer AH. MW standards were used to analyze the eluted peak according to manufacturer instructions (GE Healthcare).

- Protease treatment:

Protease treatment for $ToxR^{Q78A}$, $ToxR^{s81A}$ and $ToxR^{P101A}$ was performed by measuring the total yield of protein after the first step of purification (Histrap) and then adding 1:100 Carboxyl peptidase A (IRB Protein Expression facility). Sample was left o/n at 20 ºC for 3 hours. Then the sample was loaded again into the column to do a reverse Histrap and the flowthrough concentrated to use to form DNA-protein complexes

Protein purity was checked by the SDS-PAGE gel of each chromatography and by analyzing the gel filtration chromatogram.

8.3. Protein-DNA complexes using Q78A, S81A and P101A ToxR mutants

Pure samples for each of the mutants was obtained and our goal was to create complexes with DNA from the *toxT* promoter to get crystals.

8.3.1. Formation of protein-DNA complexes

Pure protein samples were incubated for three hours with DNA with a 20 % protein excess ratio. Sample was then loaded into a Superdex 200 SEC column using buffer AH.

8.4. Crystallization of protein-DNA complexes

Several PAC crystallization screenings were set up using the Crystal Phoenix robot. When small crystals started to appear optimization crystallization drops were set up now using 24-well hanging drop

plates with 1 ml reservoirs and observed until the appearance of crystals.

8.4.1. Crystal mounting and freezing

By using nylon cryo-loops protein crystals were retrieved from the crystal drops. Fished crystals were cryo-protected by soaking them in reservoir solutions that contained a cryo-protectant.

8.4.2. Protein-DNA complexes with ToxR mutants' crystals data collection

X-ray data collection was performed at the beamline XALOC at the ALBA synchrotron in Cerdanyola del Vallès, Spain. Many crystals were tested and data-sets for each mutant complexed with DNA were obtained.

8.5. Data processing

Reflection spots need to be indexed, while crystal and detector parameters are refined and diffraction peaks integrated. It is necessary to create a relative scale between measurements; frames are merged and parameters are refined using the whole dataset. Statistical analysis is performed of the reflections, evaluating the dataset.

Diffraction data-sets were indexed and integrated using XDS. The CCP4 suite of crystallographic programs (Potterton *et al., 2003*) was then used to scale, reduce and merge the data.

8.5.1. Phasing

The amplitude of structure factors (F) is calculated during data processing as well as the phases of the reflections as they are both required to accomplish building a model. There are five commonly used basic phasing strategies; first three are experimental techniques while the latter two are performed is *in silico*:

- Single and Multiple Isomorphous Replacement (SIR and MIR)
- Single- and Multi-wavelength Anomalous Dispersion (SAD and MAD)
- A combination of both, named SIR and MIR with Anomalous Scattering (SIRAS and MIRAS)
- Molecular Replacement (MR)
- Direct methods.

In this project, MR was applied (Argos, P. & Rossman, M. G, 1980):

By using a similar model structure (sequence identity above 25%) for the calculation of the initial phases. Using the experimental data and the model, a Patterson map of interatomic vectors is calculated. By using both maps, rotating and translating them to be able to locate the probe model with respect to the origin of a new unit cell. By using the resulting location, initial phases are calculated (Taylor, 2010).

Molecular replacement was used to solve the structure using Phaser (McCoy et *al.,* 2007).

8.6. Refinement

In order to reduce the difference from the observed and the calculated amplitudes, changes are applied to the coordinates (*x*, *y* and *z*) and temperature *B*-factors from atoms of the model. This is an iterative process of manual fitting and building, geometric constraints, electron density map calculation from the experimental model and automatic optimization from the X-ray data. *R*-work and *R*-free values are used as guides. *R*-work value between 20% and 30% is accepted, yet as the resolution improves, the value of *R*-work is expected to decrease. *R*-free, usually higher than *R*-work (differences above 5% may indicate over-refinement or errors) characterizes the quality of the fit, as it is an independent measure and it originates from a test-dataset of reflections that are left out in the refinement. Density maps are also observed for refinement (Sum electron density maps ($2F_{obs}$-F_{calc}) and Difference density maps (F_{obs}-F_{calc})), one should appear like the model and the latter one indicating misplaced or missing atoms.

Multiple rounds of refinement were performed using Refmac (Murshudov *et al*., 2011) as well as manual model building using Coot (Emsley *et al.*, 2004) followed by further Refmac refinement. Phenix (Adams *et al.,* 2010) was also used with refinement purposes by correcting the rotamer outliers.

8.7. Validation and analysis

Using Coot, Ramachandran plots were generated and used to validate the final model. Using Nucplot (Luscombe *et al.,* 1997) protein:DNA interactions were analyzed and compared between wildtype and mutants. Using web-DNA (S *et al.*, 2019) and DNAproDB (Sagendorf *et al.,* 2017), DNA conformations were analyzed of all the structures, as well as precise conformations and approximations of the minor and major groove widths as well as other important DNA parameters. Lastly, the final model was validated using the pdb validation services (Berman *et al.,* 2003). Figures of the final structure were created utilizing Pymol (Delano, 2002).

9. ToxR Results and Discussion

This section focuses on the results and discussion of the experiments performed in order to solve the structure of three mutants of ToxR.

9.1. Construct design

The analysis of the sequence of ToxR will be presented in this part of the results and discussion, as well as a prediction of the secondary structure of the protein to have a better understanding of its structure.

9.2. ToxR sequence

Vibrio cholerae ToxR full length (Uniprot reference: W0B3Z6) is composed by a cytoplasmic domain (1-170), a transmembrane helix (171-186) and a periplasmic domain (187-282). Figure 9.1

```
        10         20         30         40         50         60
MSHIGTKFIL AEKFTFDPLS NTLIDKEDSE EIIRLGSNES RILWLLAQRP NEVISRNDLH

        70         80         90        100        110        120
DFVWREQGFE VDDSSLTQAI STLRKMLKDS TKSPQYVKTV PKRGYQLIAR VETVEEEMAR

       130        140        150        160        170        180
ESEAAHDISQ PESVNEYAES SSVPSSATVV NTPQPANVVA NKSAPNLGNR LFILIAVLLP

       190        200        210        220        230        240
LAVLLLTNPS QSSFKPLTVV DGVAVNMPNN HPDLSNWLPS IELCVKKYNE KHTGGLKPIE

       250        260        270        280
VIATGGQNNQ LTLNYIHSPE VSGENITLRI VANPKDAING CE
```

Figure 9.1 Sequence of full-length ToxR

9.3. ToxR secondary structure prediction

In order to reach a better understanding of the folding of ToxR, a prediction of the secondary structure of ToxR was done using the Psipred server as shown in Figure 9.2.

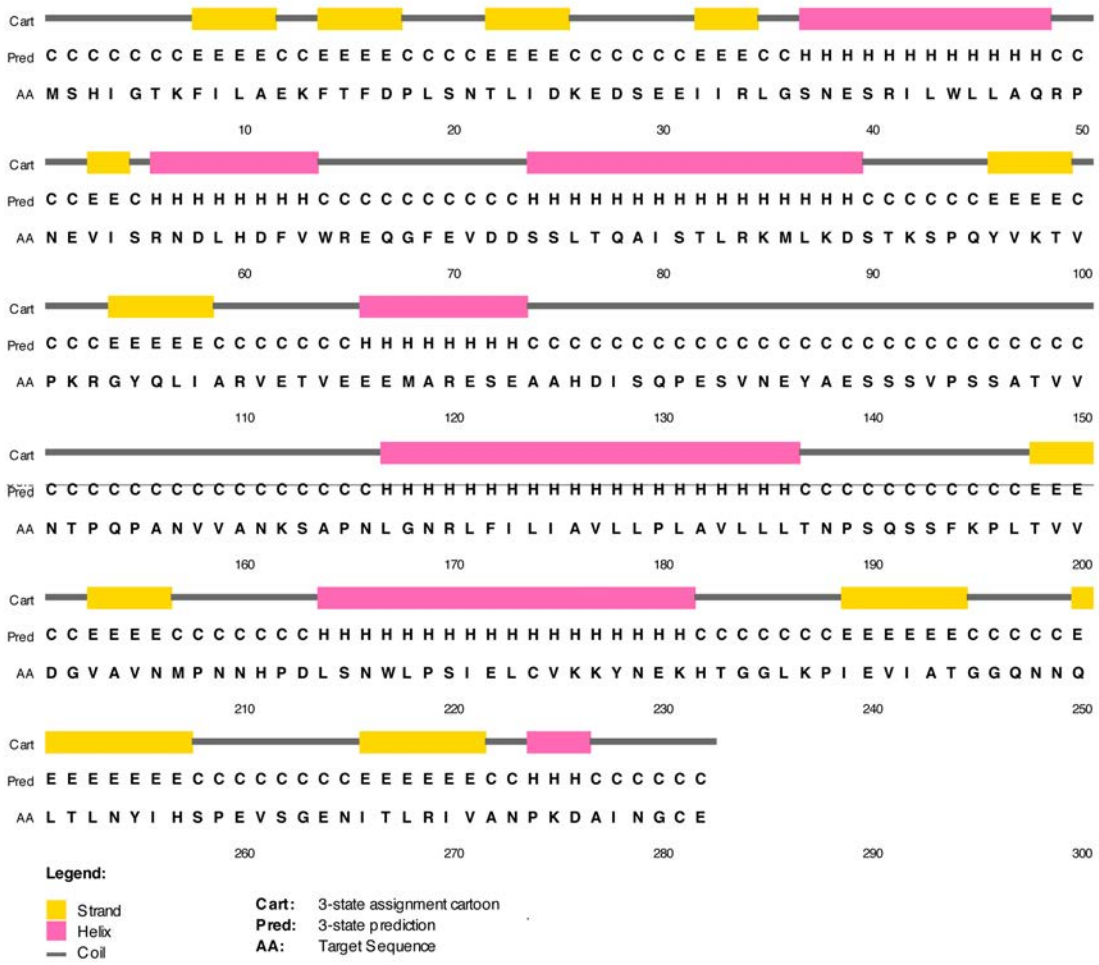


Figure 9.2 ToxR full-length secondary structure prediction

9.4. ToxR-DBD construct

A segment of the cytoplasmic domain, ToxR 6-114, which we identify as ToxR-DBD (for “DNA-Binding Domain”) was expressed and purified in our lab from a vector provided by the group of our collaborator Prof. Eric Krukonis at the University of Detroit Mercy and its crystal structure was solved in our lab by Simone Pieretti. Figure 9.3

```
         10          20          30          40          50          60
MSHIGTKFIL AEKFTFDPLS NTLIDKEDSE EIIRLGSNES RILWLLAQRP NEVISRNDLH

         70          80          90         100         110
DFVWREQGFE VDDSSLTQAI STLRKMLKDS TKSPQYVKTV PKRGYQLIAR VETV
```

Figure 9.3 Sequence of ToxR construct (6-114)

9.5. ToxR-DBD and *toxT* promoter interactions

Pieretti described interactions between ToxR and the *toxT* promoter in his PhD thesis, which are shown in figure 9.4. We wanted to see the effects of the introduction of point mutations in ToxR on *toxT* activation.

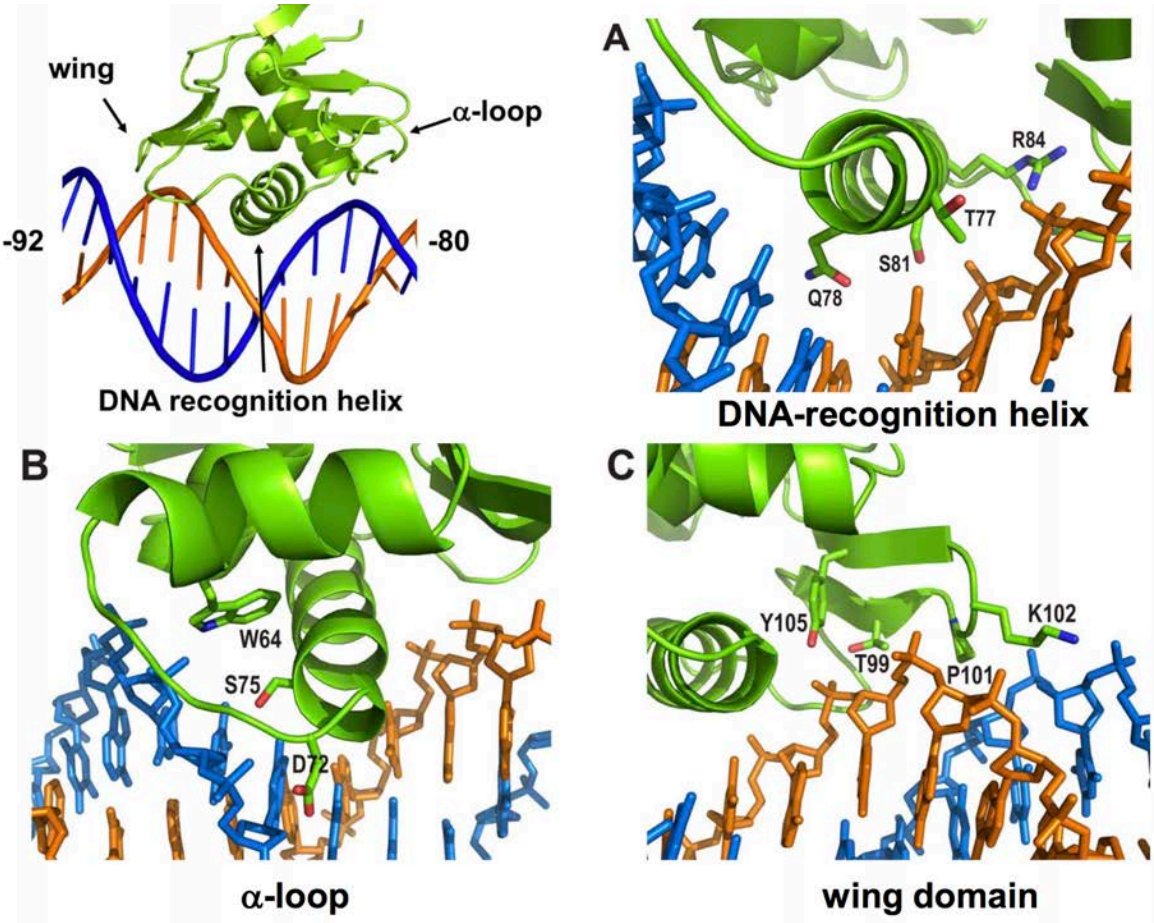


Figure 9.4 ToxR side chain interactions with the *toxT* promoter

(Extracted from Pieretti, 2016).

9.6. ToxR-DBD mutations.

Our collaborator's lab produced 15 mutants upon the information provided by our group and tested the effects of the altered side chains on *toxT* activation (Figure 9.5). With these results, our lab selected three of the mutants that resulted in reduced *toxT* activation to express, purify and solve their crystal structure and analyze the structural changes that these mutations generate.

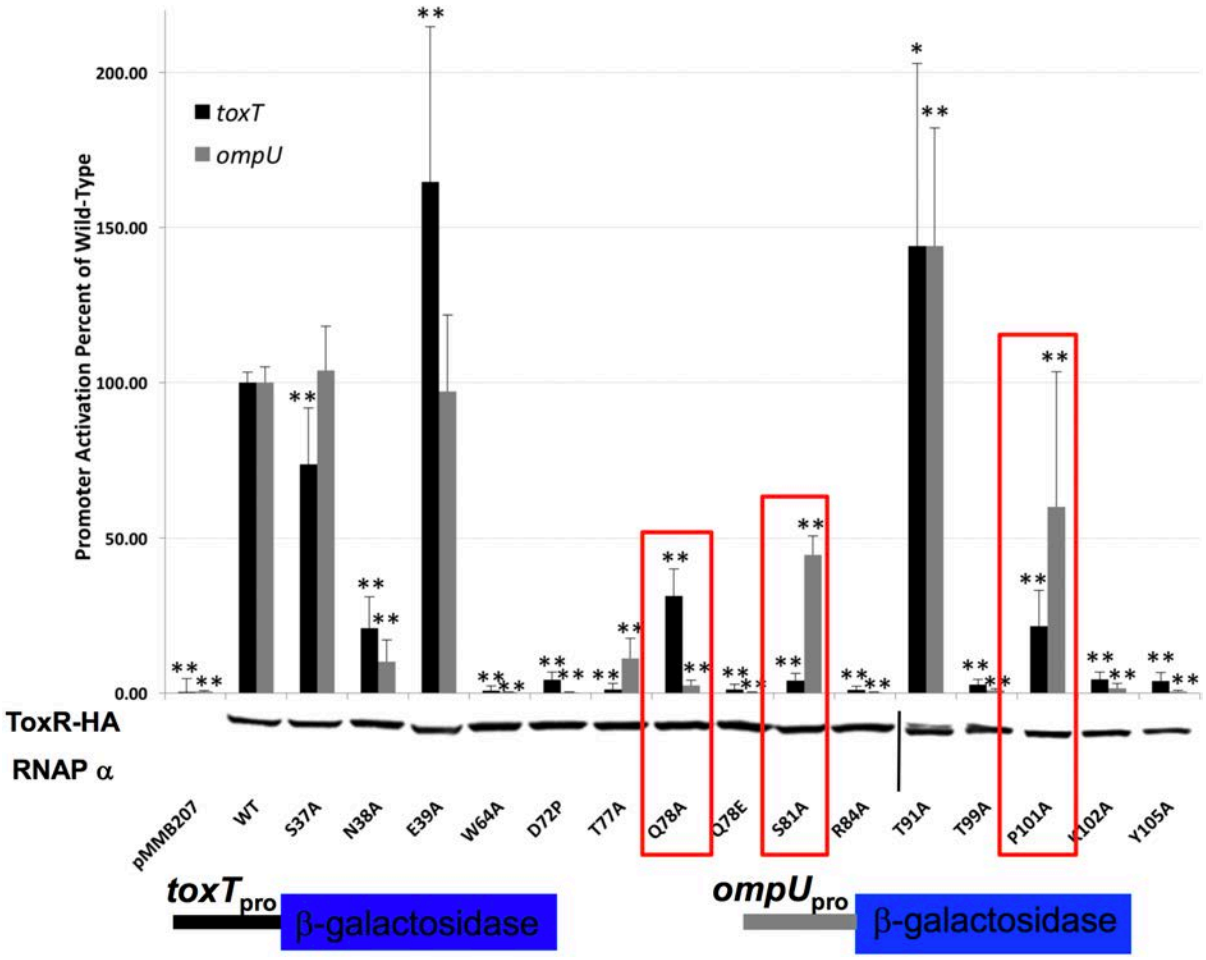


Figure 9.5 Effects of altered side chains on *toxT* and *ompU* activation (Krukonis, 2019).

9.7. ToxR^{X} sample preparation

Mutants that were selected were expressed by our lab in *E. coli* BL21(DE3) T1^{R} and are shown in Figure 9.6.

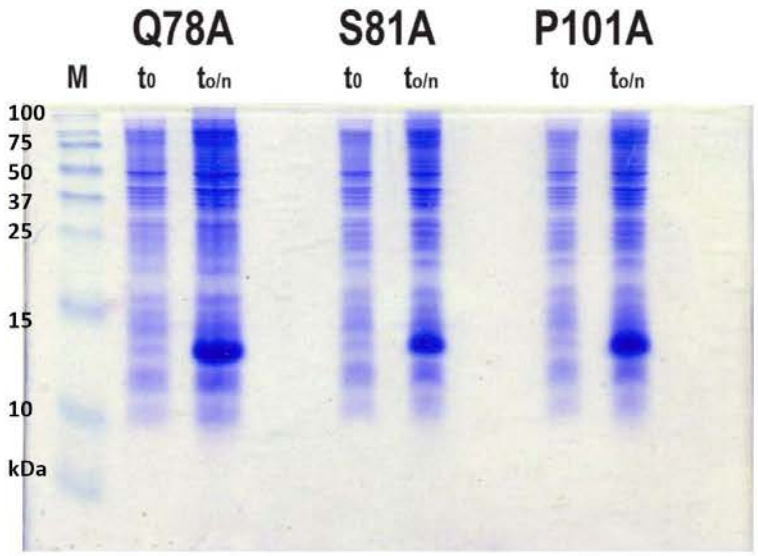


Figure 9.6 SDS-PAGE of over-expression of ToxR^{Q78A}, ToxR^{S81A} and ToxR^{P101A}

In our lab, we purified the selected mutants through a nickel-affinity chromatography followed by a SEC and finally, after removing the His-tag at the C-terminus with carboxyl peptidase A, an inverse nickel-affinity chromatography was performed.

We produced and purified ToxR^{Q78A}, ToxR^{S81A} and ToxR^{P101A}. They all were purified using the same purification scheme and had a similar behavior (chromatography profiles and final purity) and an example is shown in Figure 9.7.

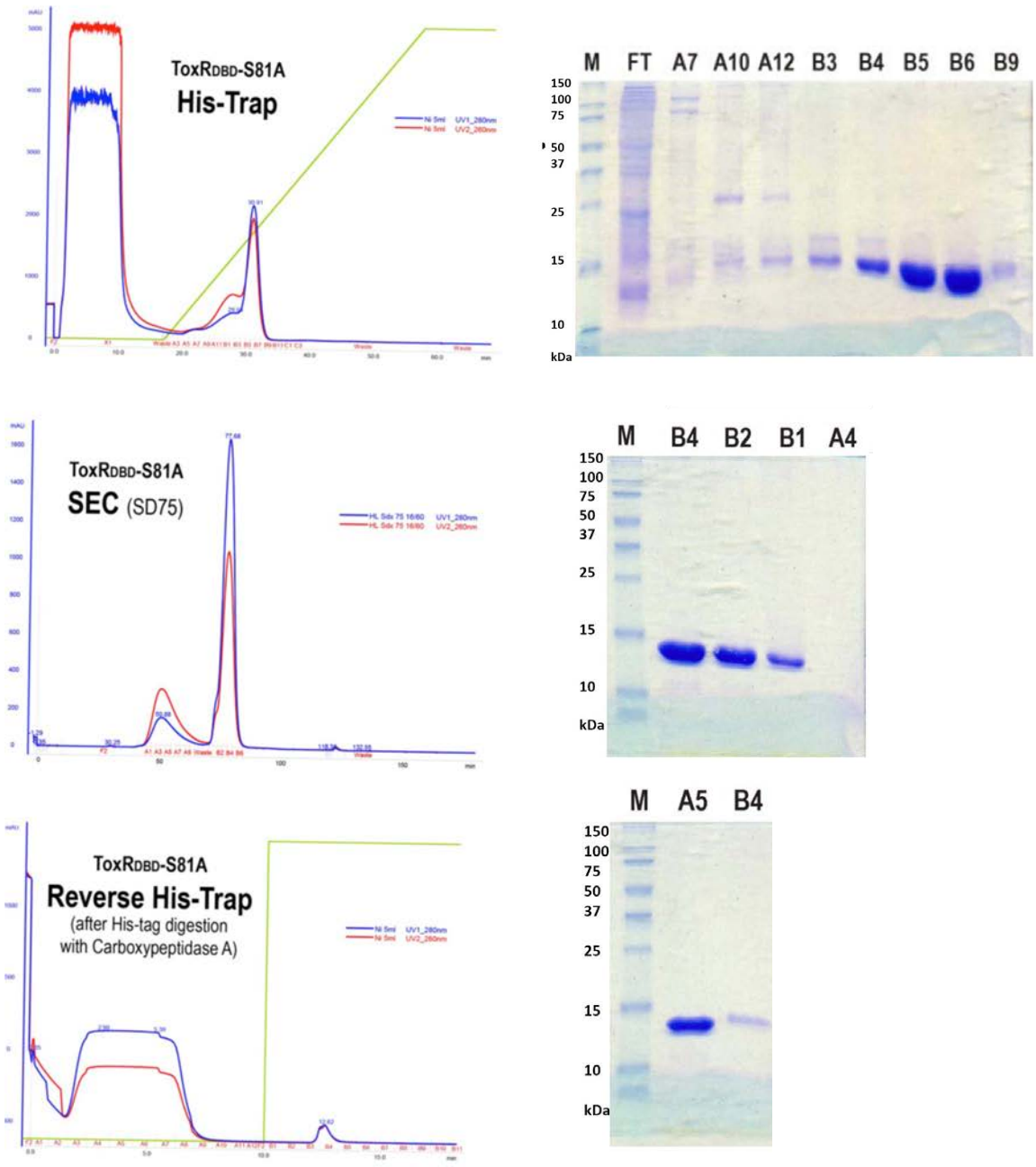


Figure 9.7 Purification scheme of ToxR^{S81}A

Chromatogram is on the left and SDS-PAGE on the right. All three mutants were purified in the same way.

9.8. ToxR^{X} in complex with a 20-bp oligonucleotide

We utilized the same 20-bp oligonucleotide that was used when solving the crystal structure of ToxR. It was designed as a result of

experiments performed in the lab of our collaborator Dr. Krukonis. Through footprinting analysis, it was shown that the ToxR binding site was from -104 to -68. This 20-bp oligonucleotide contains the specific sequence of the *toxT* promoter from C-97 to C-78 shown in Figure 9.8.

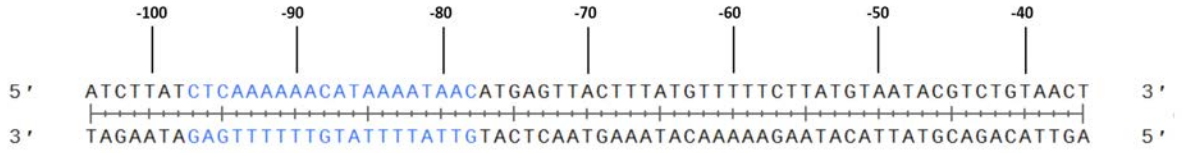


Figure 9.8 DNA sequence of the *toxT* promoter

20-bp section used for experimentation is highlighted in blue.

9.9. Complex formation

We incubated for three hours the purified mutants and the DNA with a 20 % protein excess ratio and the mixture sample was loaded into a superdex 200 SEC column. The protein-DNA complex eluted as a simple peak for all three mutants. They all behaved in the same way; they have a sharp peak profile indicating a proper protein-DNA binding as well as the conserved specificity of the protein for the sequence.

9.10. ToxR^{x} – *toxT* 20-bp oligonucleotide crystallization

After obtaining the complex of the three mutants with the 20-bp oligonucleotide, several different general crystallization-screening conditions were set up. These screenings were optimized and finally crystals for each of these mutants were obtained and tested for data collection.

9.11. Data collection and structure determination

Data sets were obtained at the Alba Synchrotron BL13-XALOC beamline. The best crystal for each mutant is described as follows.

9.11.1. ToxR^{Q78A}

Many crystals were tested at the synchrotron and the best crystal for the ToxR^{Q78A}-DNA complex diffracted at up to 2.5 Å (diffraction pattern shown in Figure 8.9).

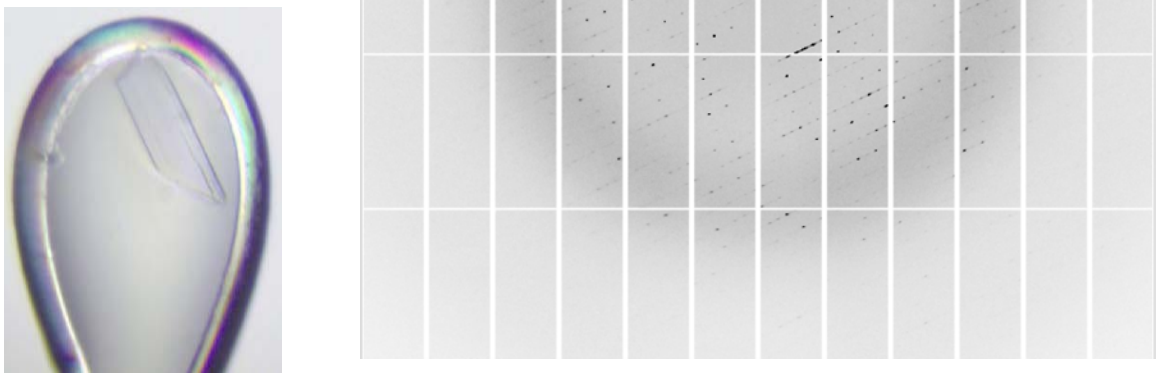


Figure 9.9 Crystal and diffraction pattern of ToxR^{Q78A}-DNA20

Crystal diffracted up to 2.5 Å.

This crystal's data collection and processing statistics are detailed in Table 9.1.

Table 9.1 Data collection and processing statistics for ToxR^{Q78A}DNA20

Parameters	Value
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	0.979
Space group	$P2_1$
Cell dimensions	a=49.17 Å, b=88.28 Å, c=88.88 Å α=90.00°, β=95.04°, γ=90.00°
Number of unique reflections	23,523
Resolution range (Å)- (overall/last shell)	88.54-2.66/2.66-2.55
R_{merge} (overall/last shell)	0.113/0.905
Multiplicity (overall/last shell)	2.7/2.2
Completeness (%) (overall/last shell)	95.0/86.1
<I>/ σ <I> (overall/last shell)	2.7/0.6
Wilson B-factor (Å^2)	45.90

Pieretti's structure was used as a model to solve the structures via molecular replacement using Phaser. We obtained a solution with a TFZ score of 44.8 and an overall LLG of 4381, with a $P2_1$ space group. We used Coot to manually refine the model as well as consecutive Refmac restrained refinement cycles. Rfactor and Rfree converged to

30% and 37%, respectively, after the addition of solvent molecules. These high values may be explained by the existence of a Translational Non-crystallographic pseudo symmetry at coordinate (0.243, 0.5, 0.5) at a 62.85 Å distance from the origin and with a p_value (Height) of 5.88 e^{-05} and a 51.1% height from the origin, as indicated by Phenix's Xtriage. This was also suggested by Phaser, when doing the molecular replacement. To verify the existence of this pseudo-origin, a Patterson-Harker map was created with the CCP4i server as illustrated in Figure 9.10 (Potterton *et al.,* 2003).

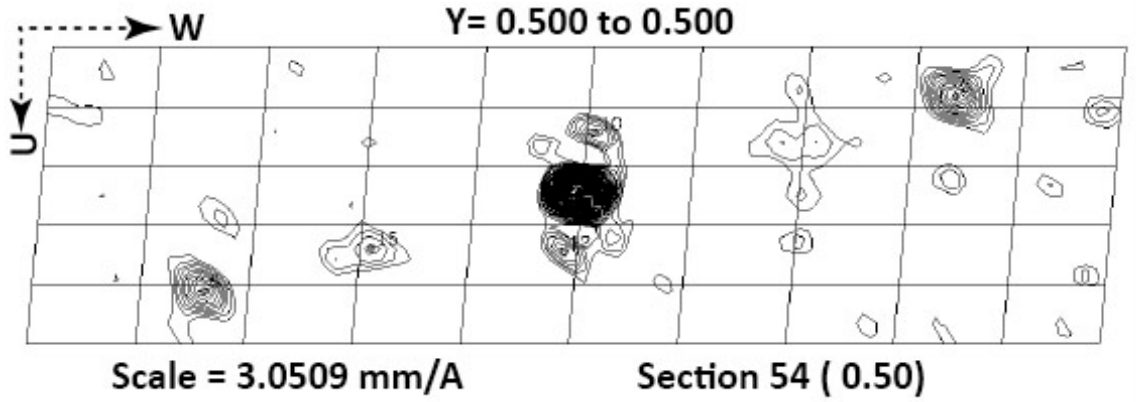


Figure 9.10 Patterson-Harker map of the pseudo-origin

In addition, a self-rotation function was calculated with MolRep as illustrated in Figure 9.11 (Vagin and Teplyakov, 2010). The pseudo-origin in the Patterson map explains why even if the model agrees with the density, the R values still do not decrease below 0.30. The projection at chi=180° shows the two-fold axes. The crystallographic axis is shown along Y and from the structure we know that there is a Non-Crystallographic Symmetry (NCS) axis, which is parallel to Y and is superposed in the self-rotation to the crystallographic axis (Figure 9.12a). The second NCS two-fold axis appears at theta=0, phi=70°, chi=180° which relates the two proteins to each other and runs approximately along the DNA axis (Figure 9.12b).

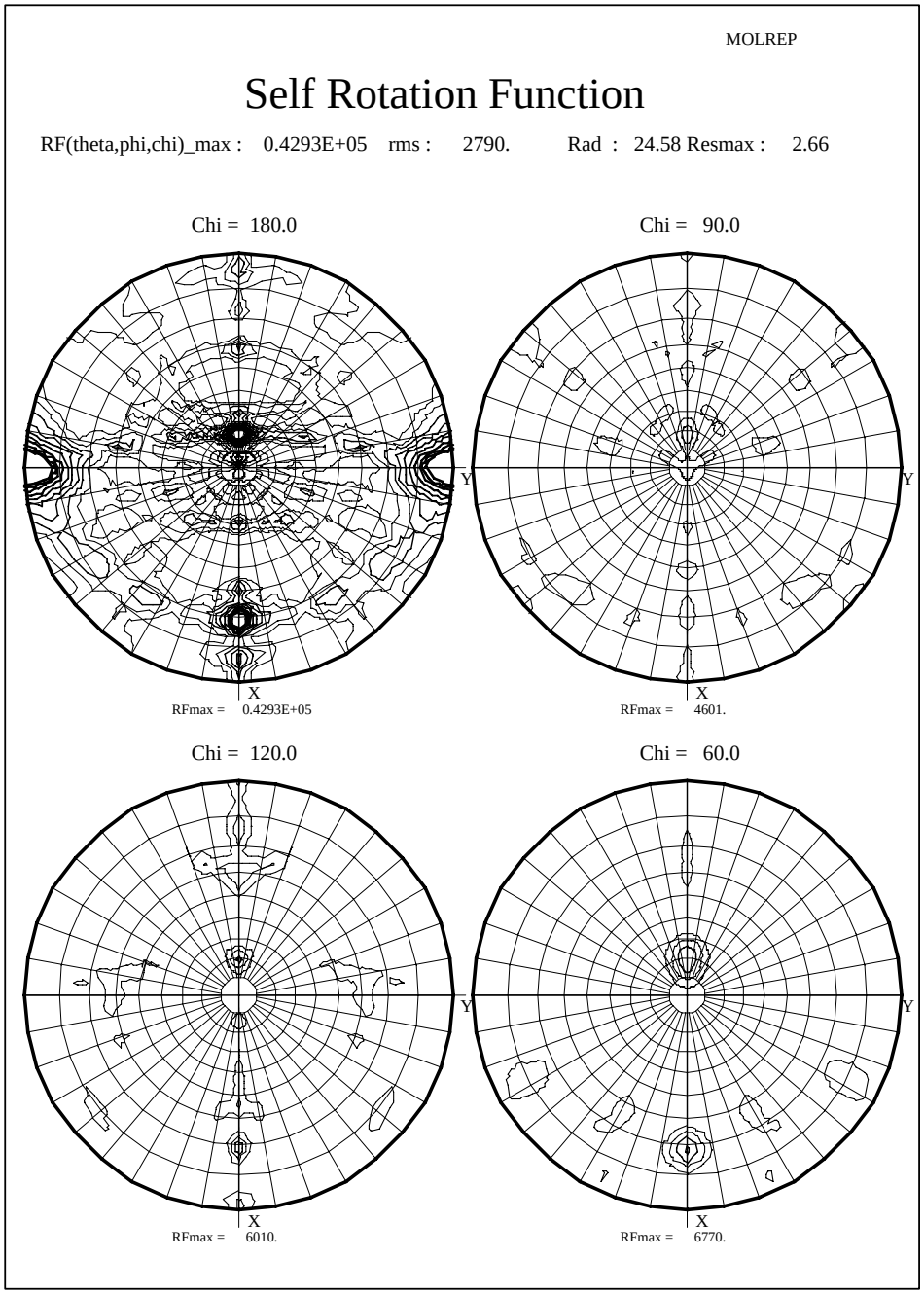


Figure 9.11 Rotation function depicting the pseudo symmetry

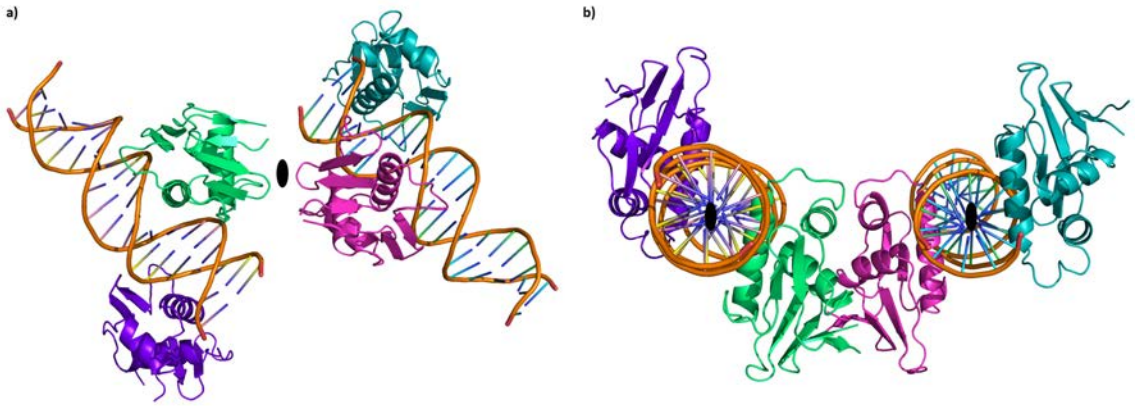


Figure 9.12 NCS two-fold axes

a) Illustrates the NCS two-fold axis that is parallel to Y. b) Illustrates the NCS two-foild axes that run along the DNA axes, these axes are only symmetric for the proteins.

9.11.1.1. ToxR^{Q78A} Translational non crystallographic pseudo symmetry correction trial

After detecting the existence of non crystallographic pseudo symmetry, we attempted to correct the space group by using Zanuda from the CCP4 server, which uses an array of refinements in compatible space groups with our own unit-cell parameters and then it choses the model with the highest symmetry space group from a subset of possible models that have the top refinement statistics as shown in Figure 9.12.

```
coordinates       TOXRQ78A_27sept3137.pdb
data              TOXRQ78A_27sept3137.mtz
readability test  passed (Refmac_5.8.0189)
resolution        2.549
spacegroup        P 1 21 1
cell              49.166 88.276 88.881 90.00 95.04 90.00

Step 1.
R-factors for the starting model.
Transformation into a supergroup.

current time:                                        Sep 27 17:19 BST
expected end of job (rough estimate):                Sep 27 17:23 BST

-------------------------------------------------------------------
| Subgroup | Spacegroup | R.m.s.d. |   Refinement in tested group   |
|          |            | from the |--------------------------------|
|   Ref    |            | starting |  Rigid   |     Restrained      |
|          |            | model, A |----------|---------------------|
|          |            |          |    R     |    R     |  R-free  |
|----------|------------|----------|----------|----------|----------|
| >>   2   | P 1 21 1   |  0.0002  |    --    |  0.4256  |  0.4426  |
|      2   | P 1 21 1   |  0.0002  |    --    |    --    |    --    |
^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

Step 2.
Refinements in subgroups.
There are 2 subgroups to test.

current time:                                        Sep 27 17:19 BST
expected end of job:                                 Sep 27 17:33 BST

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
| >>   2   | P 1 21 1   |  0.0002  |    --    |  0.4256  |  0.4426  |
-------------------------------------------------------------------
|      1   | P 1        |  1.7592  |  0.6297  |  0.5051  |  0.5303  |
|      2   | P 1 21 1   |  2.0567  |  0.6261  |  0.5072  |  0.5363  |
-------------------------------------------------------------------
| <<   1   | P 1        |  1.7592  |  0.6297  |  0.5051  |  0.5303  |
^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

Step 3.
Refinement of the best model.
Candidate symmetry elements are added one by one.

current time:                                        Sep 27 17:25 BST
expected end of job:                                 Sep 27 17:37 BST

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
| >>   1   | P 1        |  1.7592  |  0.6297  |  0.5051  |  0.5303  |
-------------------------------------------------------------------
|      1   | P 1        |  1.9617  |  0.6362  |  0.5039  |  0.5584  |
|      2   | P 1 21 1   |  2.0372  |    --    |  0.4927  |  0.5688  |
|      5   | P 21 21 2  |  5.3014  |    --    |  0.4455  |  0.4891  |
-------------------------------------------------------------------
| <<   5   | P 21 21 2  |  5.3014  |    --    |  0.4455  |  0.4891  |
-------------------------------------------------------------------

R-factor in the original subgroup is NOT the best.
The original spacegroup assignment seems to be incorrect.

end of job:                                          Sep 27 17:30 BST
```

Figure 9.12 Zanuda log

The model and map obtained from Zanuda positioned only two ToxR^{Q78A} molecules in the asymmetric unit; subsequently we attempted to place one 20 bp double stranded DNA molecule once

using Phaser and another time using Coot, but both trials were unsuccessful. The resulting models could not fit the Zanuda-selected space group and had high clash scores.

Interestingly, during data processing we selected the $P2_1$ monoclinic (mP) fitting instead of the $P2_12_12$ Orthoromic (oP) because the values obtained (92.2) were quite high in comparison to the value for the $P2_1$ (11.5). This part of the data processing is presented in Figure 9.13.

```
SPACE GROUP AND CELL PARAMETERS ARE UNKNOWN
AUTOINDEXING IS BASED ON   3560 OUT OF    3560 SPOTS

***** INDEXING OF OBSERVED SPOTS IN SPACE GROUP #   1 *****
     3544 OUT OF      3560 SPOTS INDEXED.
        0 REJECTED REFLECTIONS (REASON: OVERLAP)
       16 REJECTED REFLECTIONS (REASON: TOO FAR FROM IDEAL POSITION)
EXPECTED ERROR IN SPINDLE  POSITION     0.601 DEGREES
EXPECTED ERROR IN DETECTOR POSITION      1.09 PIXELS

***** DIFFRACTION PARAMETERS USED AT START OF INTEGRATION *****

REFINED VALUES OF DIFFRACTION PARAMETERS DERIVED FROM      3544 INDEXED SPOTS
REFINED PARAMETERS:   BEAM ORIENTATION CELL AXIS
STANDARD DEVIATION OF SPOT    POSITION (PIXELS)     1.09
STANDARD DEVIATION OF SPINDLE POSITION (DEGREES)    0.60
SPACE GROUP NUMBER      1
UNIT CELL PARAMETERS     49.529    88.829    89.347  89.778  84.868  89.833
REC. CELL PARAMETERS   0.020272  0.011258  0.011237  90.208  95.132  90.147
COORDINATES OF UNIT CELL A-AXIS    44.440     3.502    21.587
COORDINATES OF UNIT CELL B-AXIS   -38.764     3.323    79.856
COORDINATES OF UNIT CELL C-AXIS    11.234   -88.126     9.507
CRYSTAL MOSAICITY (DEGREES)     0.200
LAB COORDINATES OF ROTATION AXIS  0.999999 -0.001084  0.000118
DIRECT BEAM COORDINATES (REC. ANGSTROEM)   0.000088 -0.000468  1.021179
DETECTOR COORDINATES (PIXELS) OF DIRECT BEAM    1229.42   1268.28
DETECTOR ORIGIN (PIXELS) AT                     1229.19   1269.53
CRYSTAL TO DETECTOR DISTANCE (mm)       468.86
LAB COORDINATES OF DETECTOR X-AXIS  1.000000  0.000000  0.000000
LAB COORDINATES OF DETECTOR Y-AXIS  0.000000  1.000000  0.000000


 LATTICE-  BRAVAIS-   QUALITY  UNIT CELL CONSTANTS (ANGSTROEM & DEGREES)
 CHARACTER  LATTICE     OF FIT      a      b      c   alpha  beta gamma

    31        aP          0.0      49.5   88.8   89.3  89.8  84.9  89.8
    44        aP          2.6      49.5   88.8   89.3  90.2  95.1  89.8
    33        mP         11.5      49.5   88.8   89.3  90.2  95.1  89.8
    34        mP         30.3      49.5   88.8   89.3  90.2  95.1  89.8
    35        mP         85.9      88.8   49.5   89.3  95.1  90.2  89.8
    32        oP         92.2      49.5   88.8   89.3  90.2  95.1  89.8
    20        mC         96.9     126.2  125.7   49.5  86.5  93.7  90.3
    25        mC        103.1     126.2  125.7   49.5  93.5  93.7  89.7
    23        oC        104.7     125.7  126.2   49.5  93.7  93.5  89.7

BRAVAIS-            POSSIBLE SPACE-GROUPS FOR PROTEIN CRYSTALS
  TYPE                    [SPACE GROUP NUMBER,SYMBOL]
  aP      [1,P1]
  mP      [3,P2] [4,P2(1)]
 mC,mI    [5,C2]
  oP      [16,P222] [17,P222(1)] [18,P2(1)2(1)2] [19,P2(1)2(1)2(1)]
```

Figure 9.13 Space-group selection ToxR^{Q78A} during data processing

9.11.1.2. ToxR^{Q78A} final model

The geometry of the final model was validated using the wwPDB validation service (an extract is shown as Figure 9.14). The final model refinement statistics are shown in Table 9.2.

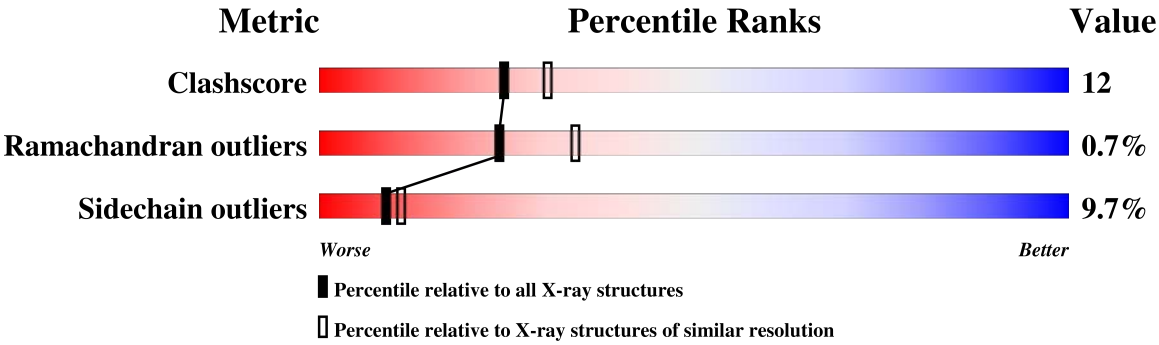


Figure 9.14 Excerpt from the wwPDB validation service report for the ToxR^{Q78A}-DNA20 crystal structure.

Table 9.2 Refinement statistics for ToxR^{Q78A}-DNA20

Parameter	Value
Resolution range (Å)	88.54-2.55
Number of protein/DNA atoms	5,137
Number of solvent molecules	28
RMSD for bonded angles (1) (°)	1.86
RMSD for bond lengths (1) (Å)	0.0119
Rfactor/Rfree	0.30/0.37
Average B-factor (Å^2)	43.6
Ramachandran favored (%)	90.63
Ramachandran outliers	0
Molprobility score	2.92

The final model presented a continuous and well-defined electron density for the structure as demonstrated in Figure 9.15.

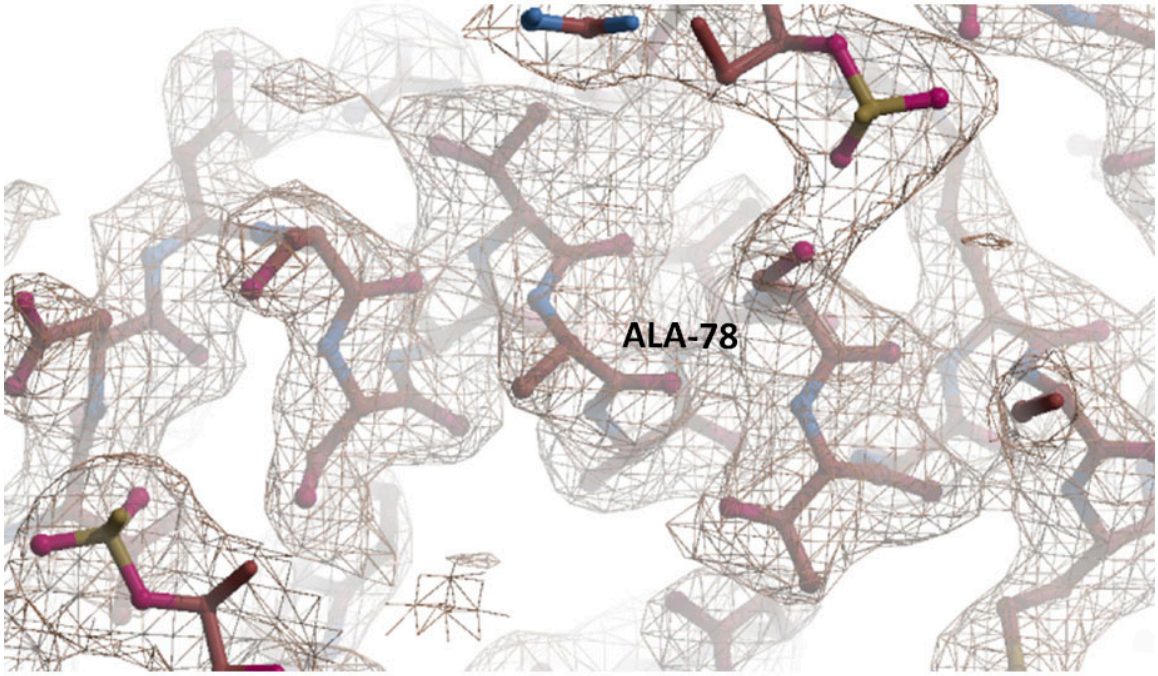


Figure 9.15 A fragment of the final ToxR^{Q78A}-DNA20 model fitted in the 2 F_o-F_c electron density map (contoured at 1.0 σ)

All the residues in the final model were located in the allowed regions of the Ramachandran plot shown in Figure 9.16.

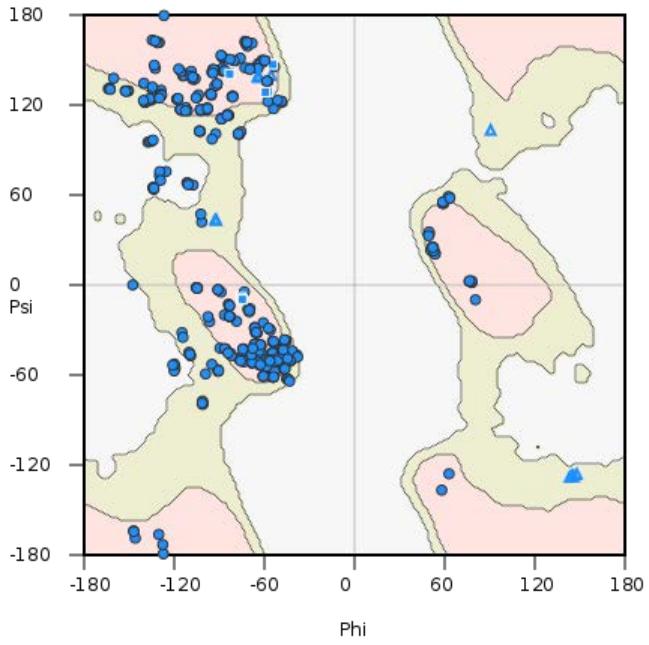


Figure 9.16 Ramachandran plot for ToxR^{Q78A}-DNA20 structure

The asymmetric unit is composed of two 20bp dsDNA and 4 ToxR^{Q78A} protein molecules, distributed in two proteins monomers interacting with each DNA molecule; this is illustrated in Figure 9.17. Interestingly enough, these monomers are bound in opposite directions. They are bound in tandem, one is bound at the CATA box while the other one one is bound further downstream in an “ATAA” repeat. This is a unique behavior, as there is no evidence of this happening before.

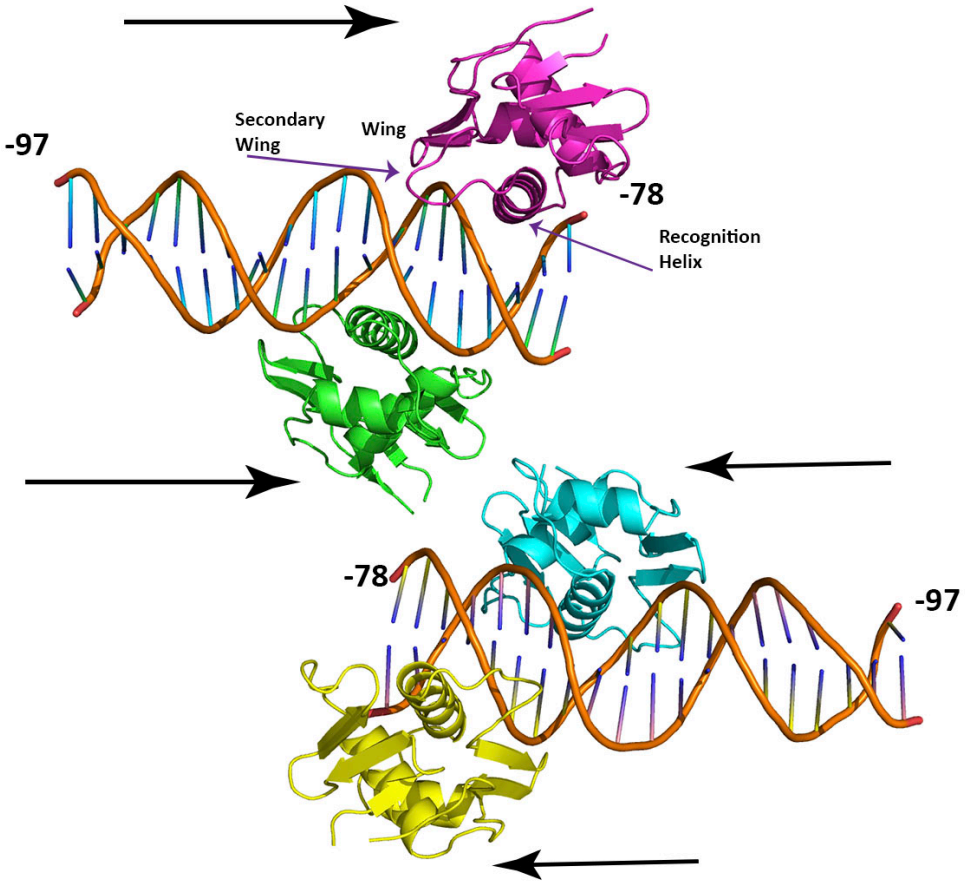


Figure 9.17 ToxR^{Q78A}-DNA20 structure

The four ToxR^{Q78A} molecules in the asymmetric unit bind the DNA using the same elements previously described: the wing domain, the recognition helix and a secondary wing as marked in Figure 9.17. Figure 9.18 illustrates by superposing the wildtype surface to the mutant it is possible to how see the loss of the interaction in the major groove when Q78 is lost, which is highlighted by the yellow circle.

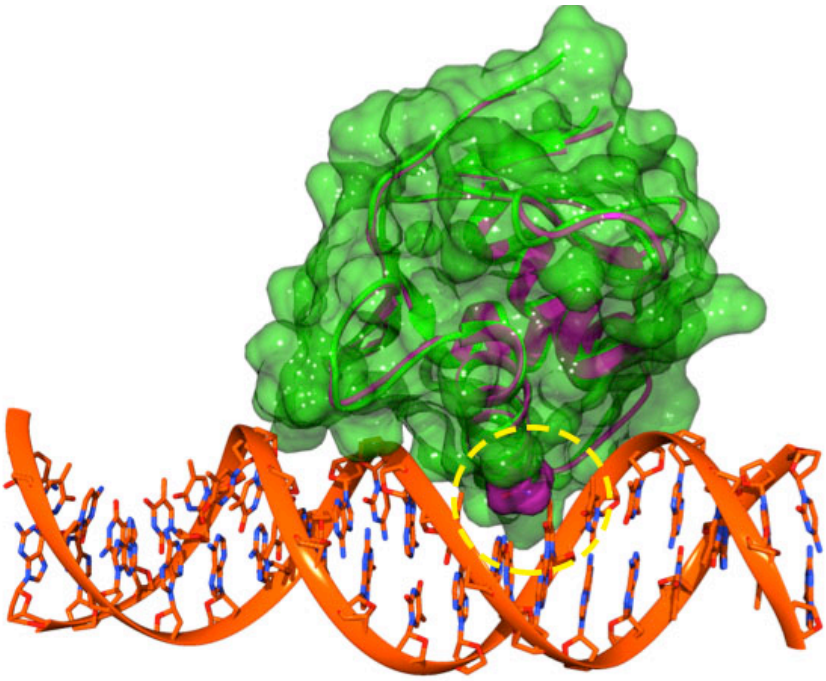


Figure 9.18 ToxR^{Q78A}-DNA20 vs ToxRWildtype-DNA20 surface difference
ToxR^{Wildtype} is shown in magenta, ToxR^{Q78A} is in lime green and the yellow dot circle surrounds the surface difference between them.

A representation of the electrostatic surface is displayed in Figure 9.19.

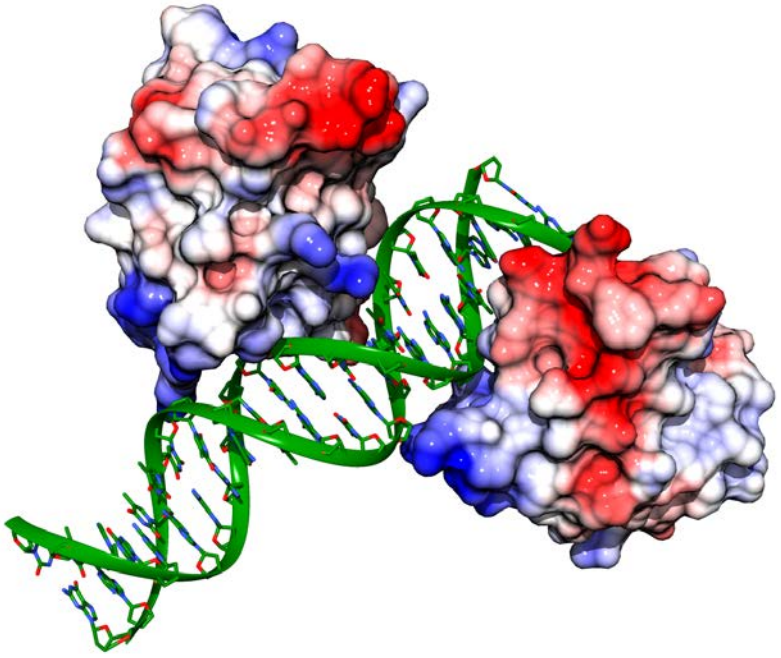


Figure 9.19 Electrostatic surface portrayal of the ToxR^{Q78A}-DNA20 structure

By superposing ToxR^{Wildtype} and ToxR^{Q78A} we can observe that they are very similar, yet, there is discrepancy in one of the monomer's α2-α3 loop, we believe this structural change is caused by the different packing contacts and is not related to the mutation. This discrepancy is illustrated in Figure 9.20.

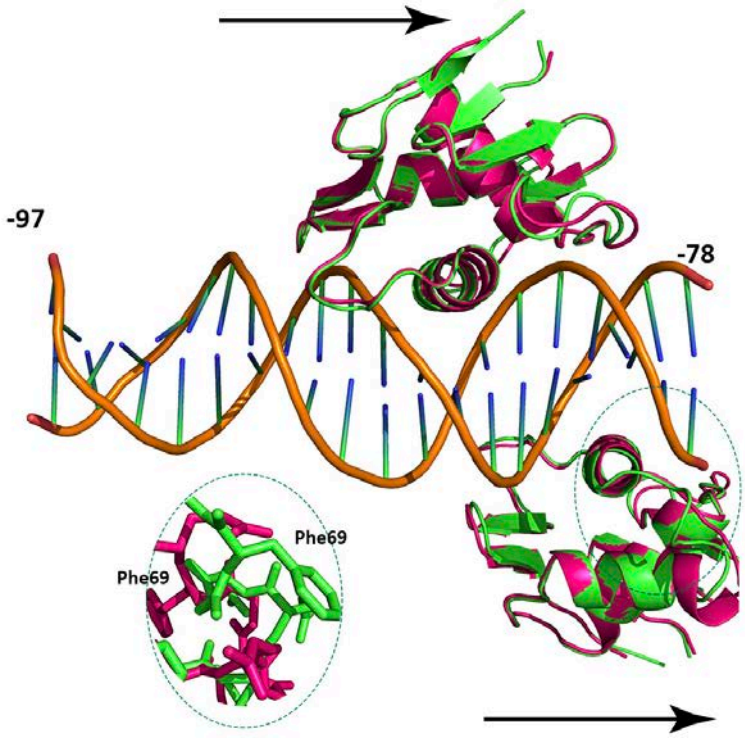


Figure 9.20 ToxR^{Wildtype}-DNA20 structure supperposed with theToxR^{Q78A}-DNA20 structure

ToxR^{Wildtype} is highlighted in magenta, ToxR^{Q78A-C} in lime green.

The r.m.s.d. between ToxR^{Wildtype} and ToxR^{Q78A} are low as shown in Table 9.3.

Table 9.3 r.m.s.d. values between ToxR^{Wildtype} and ToxR^{Q78A}

Element in ToxR^{Q78A} model (chain)	r.m.s.d. values (Å)
ToxR^{Q78A} Chain C	0.7064
ToxR^{Q78A} Chain F	0.7410
ToxR^{Q78A} Chain G	0.6772
ToxR^{Q78A} Chain H	0.8962

9.11.2. ToxR^{S81A}

Various crystals were tested at the synchrotron and the best of the ToxR^{S81A}-DNA complex diffracted at up to 2.95 Å (diffraction pattern shown in Figure 9.21).

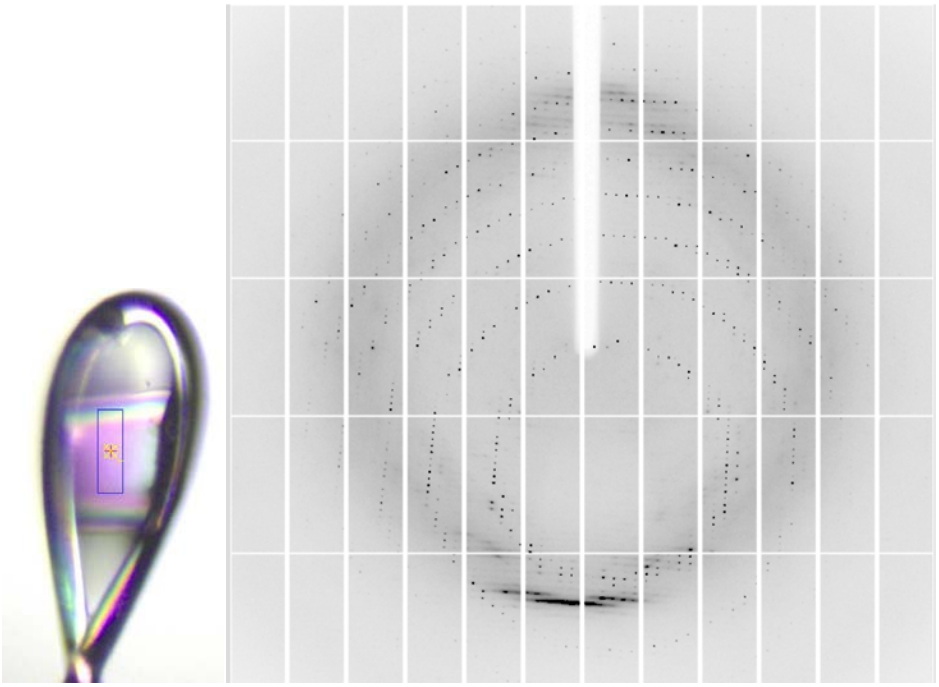


Figure 9.21 Crystal and diffraction pattern of ToxR^{S81A}-DNA20

Crystal diffracted up to 2.95 Å.

The data collection and processing statistics for the S81A mutant are specified in Table 9.4.

Table 9.4 Data collection and processing statistics ToxR^{S81A}-DNA20

Parameters	Value
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	0.979
Space group	$P222_1$
Cell dimensions	a=124.86 Å, b=77.00 Å, c=77.55 Å α=90.00°, β=90.00°, γ=90.00°
Number of unique reflections	29,361
Resolution range (Å)- (overall/last shell)	48.29-2.33/2.41-2.33
R_{merge} (overall/last shell)	0.065/0.811
Multiplicity (overall/last shell)	1.9/1.9
Completeness (%) (overall/last shell)	97.7/90.4
$\langle I \rangle / \sigma \langle I \rangle$ (overall/last shell)	8.2/0.7
Wilson B-factor (Å^2)	64.224

We used Phaser to solve the structure by molecular replacement and Pieretti's structure as a model. After analyzing the structure, we had to re-index the map as the Space Group was incorrect (P222), which was selected by the automatic ALBA processing. Systematic absence in one of the axes indicated a screw axis. After reindexing to $P222_1$, we tested Simone's model and obtained a solution with a refined TFZ

score of 35.4 and an Overall LLG of 1824. We used Coot to refine the model manually as well as successive Refmac cycles to achieve a restrained refinement. Rfactor and Rfree converged to 23% and 25% respectively. We used the wwPDB validation service to validate the geometry of the final model (an extract is shown as Figure 9.22). In Table 9.5 the final model refinement statistics are shown.

Table 9.5 Refinement statistics for ToxR^{S81A}-DNA20

Parameter	Value
Resolution range (Å)	48.29-2.33
Number of protein/DNA atoms	3,410
Number of solvent molecules	42
RMSD for bonded angles (1) (°)	2.04
RMSD for bond lengths (1) (Å)	0.0118
Rfactor/Rfree	0.23/0.25
Average B-factor (Å^2)	91.6
Ramachandran favored (%)	95.33
Ramachandran outliers	0
Molprobility score	2.33

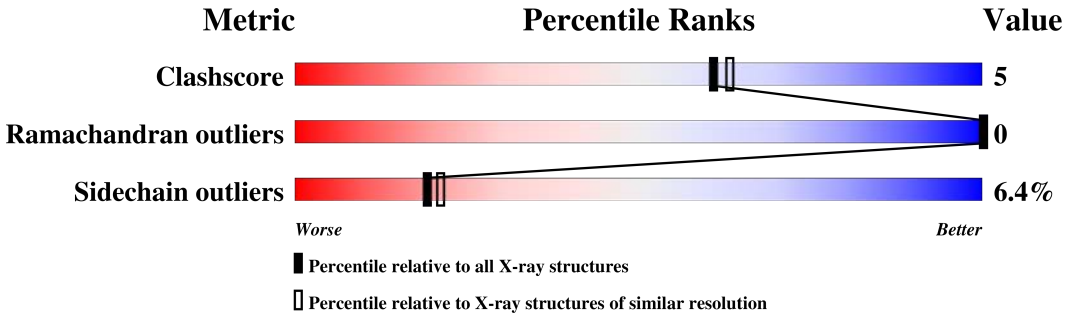


Figure 9.22 Fragment from the wwPDB validation service report for the ToxR^{S81A}-DNA20 crystal structure

In Figure 9.23, we illustrate how the final model presents a continuous and well-defined electron density for the structure.

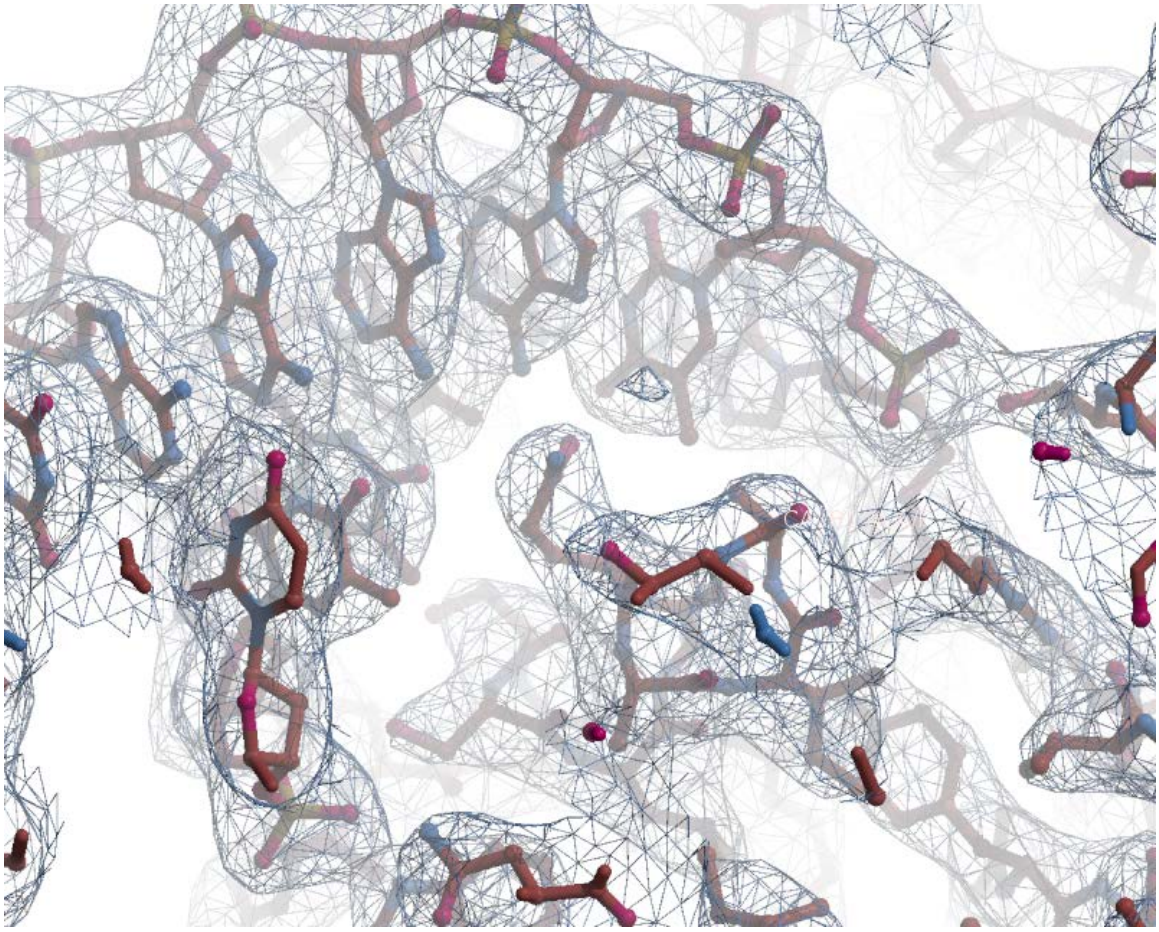


Figure 9.23 Section of the final ToxR^{S81A}-DNA20 model fitted in the 2 F_o-F_c electron density map (contoured at 1.0 σ)

In Figure 9.24, we illustrate how all the residues in the final model were located in the allowed regions of the Ramachandran plot. Which indicates the lack of areas in which the backbone torsion is deviating from what is expected in the solved structure.

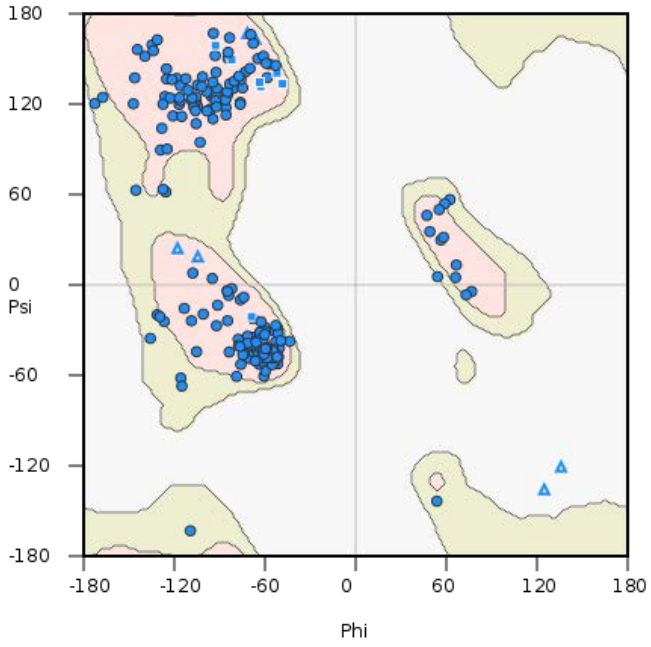


Figure 9.24 Ramachandran plot for ToxR^{S81A}-DNA20 structure

The ToxR^{S81A} final model is illustrated in Figures 9.25 and 9.26. Two 20bp dsDNA and 2 ToxR^{S81A} protein molecules compose the asymmetric unit. Figure 9.25 illustrates a cartoon representation of one a 20bp dsDNA molecule and a ToxR^{S81A} protein molecule and its domains. The electrostatic surface of ToxR^{S81A} is illustrated in Figure 9.26. After analyzing the structure, we know that there is a loss of interaction between ToxR^{S81A} and the major groove of the DNA.

9.11.2.1. ToxR^{S81A} final model

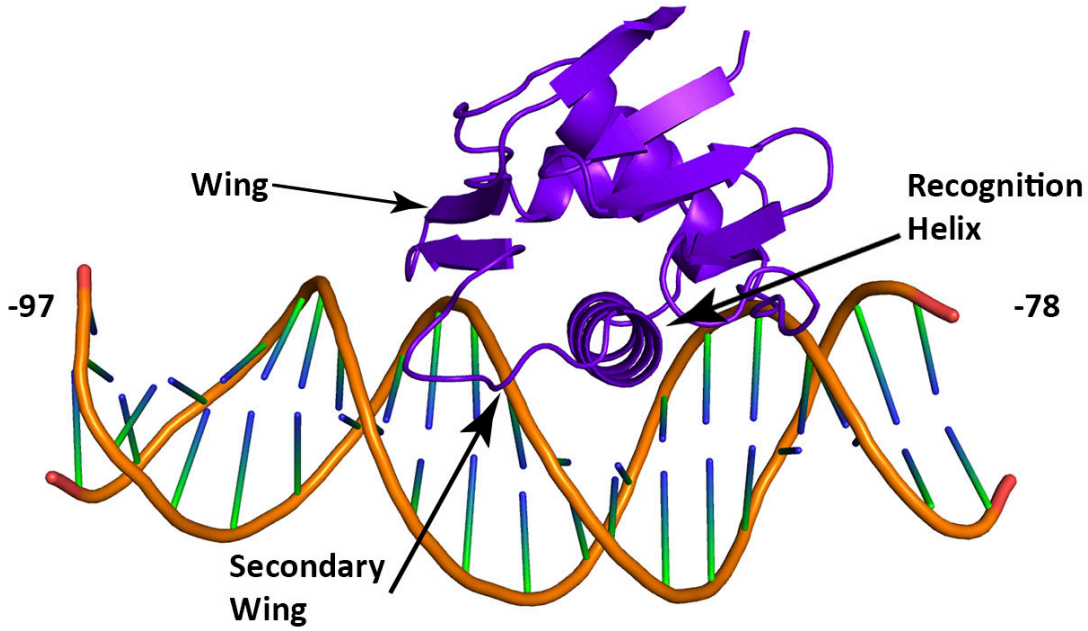


Figure 9.25 ToxR^{S81A}-DNA20 structure

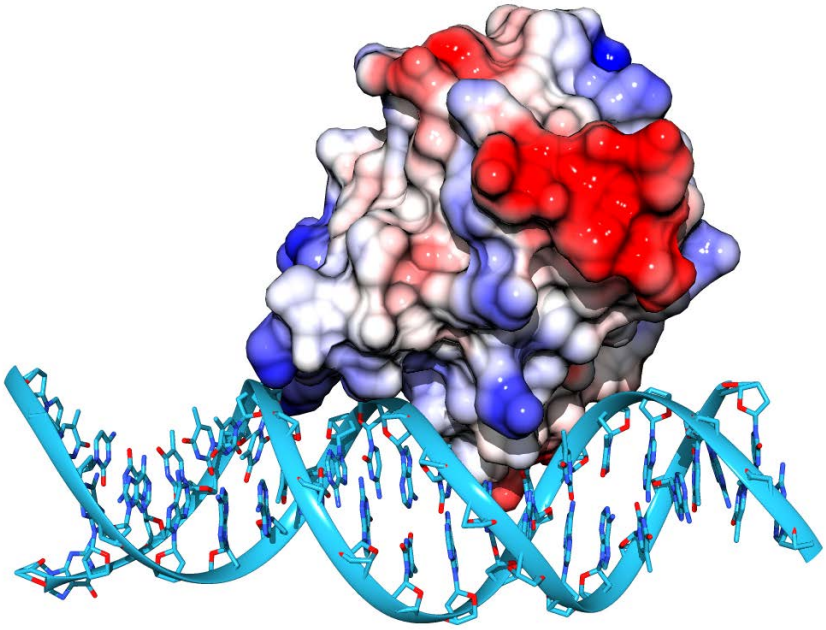


Figure 9.26 Electrostatic surface representation of the ToxR^{S81A}-DNA20 structure

Just like with the Q78A model, the r.m.s.d. between ToxR^{Wildtype} and ToxR^{S81A} are quite small as shown in Table 9.6.

Table 9.6 r.m.s.d. values between ToxR^{Wildtype} and ToxR^{S81A}

Element in ToxR^{S81A} model (chain)	r.m.s.d. values (Å)
ToxR^{S81A} Chain C	0.9134
ToxR^{S81A} Chain F	0.8493

9.11.3. ToxR^{P101A}

Several crystals were tested at the synchrotron and the best crystal for the ToxR^{P101A}-DNA complex diffracted at up to 2.95 Å (diffraction pattern shown in Figure 9.27).

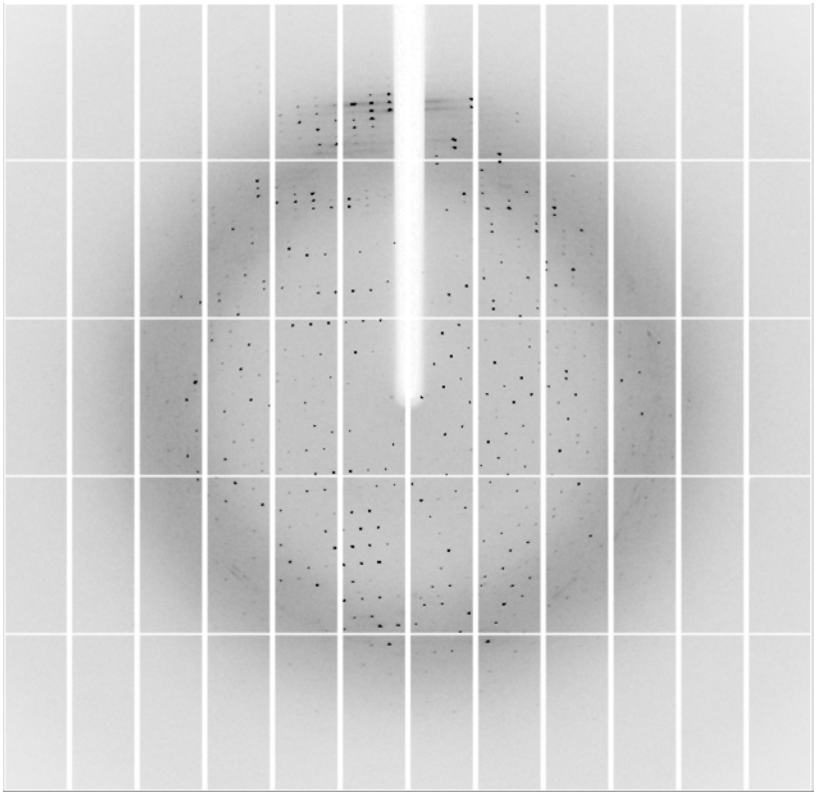


Figure 9.27 Diffraction pattern of ToxR^{P101A}-DNA20

Crystal diffracted up to 2.95 Å.

This crystal's data collection and processing statistics are detailed in Table 9.7.

Table 9.7 Data collection and processing statistics ToxR^{P101A}-DNA20

Parameters	Value
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	0.979
Space group	$P222_1$
Cell dimensions	a=125.14 Å, b=71.38 Å, c=78.82 Å α=90.00°, β=90.00°, γ=90.00°
Number of unique reflections	15,224
Resolution range (Å)- (overall/last shell)	66.69-2.95/3.13-2.95
R_{merge} (overall/last shell)	0.058/0.533
Multiplicity (overall/last shell)	1.8/1.9
Completeness (%) (overall/last shell)	98.66/99.3
<I>/ σ <I> (overall/last shell)	7.5/1.3
Wilson B-factor (Å^2)	71.36

Pieretti's structure was utilized as a model to solve the structures via molecular replacement using Phaser. We obtained a solution with a refined TFZ score of 32.1 and an Overall LLG of 1140.11, with a $P222_1$ space group, which is different from Pieretti's structure ($P2_1$). Coot was used to manually refine the model as well as consecutive cycles of

Refmac restrained refinement. Rfactor and Rfree converged to 21% and 25% respectively. The geometry of the final model was validated using the wwPDB validation service (an extract is shown as Figure 9.28). The final model refinement statistics are shown in Table 9.8.

Table 9.8 Refinement statistics for ToxR^{P101A}-DNA20

Parameter	Value
Resolution range (Å)	66.78-2.95
Number of protein/DNA atoms	3,410
Number of solvent molecules	15
RMSD for bonded angles (1) (°)	1.56
RMSD for bond lengths (1) (Å)	0.0101
Rfactor/Rfree	0.213/0.254
Average B-factor (Å^2)	96.8
Ramachandran favored (%)	87.85
Ramachandran outliers	0
Molprobility score	2.86

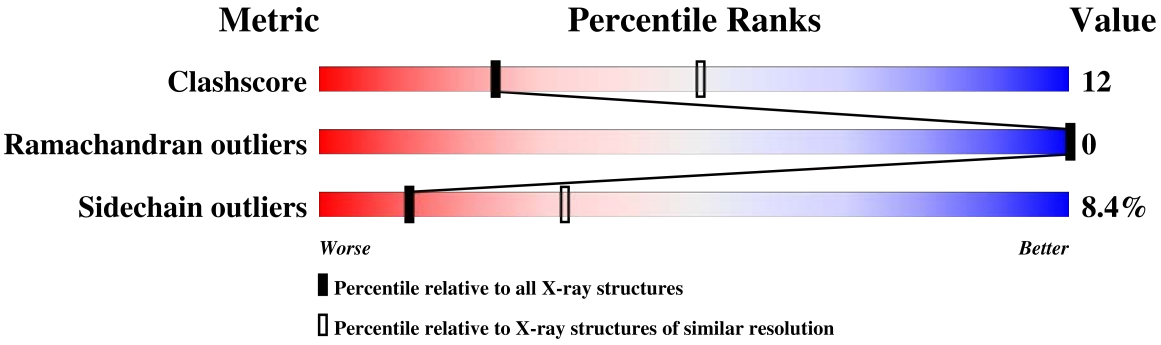


Figure 9.28 Excerpt from the wwPDB validation service report for the ToxR^{P101A}-DNA20 crystal structure

The final model presented a continuous and well-defined electron density for the structure, the area of the mutation is shown in Figure 9.29.

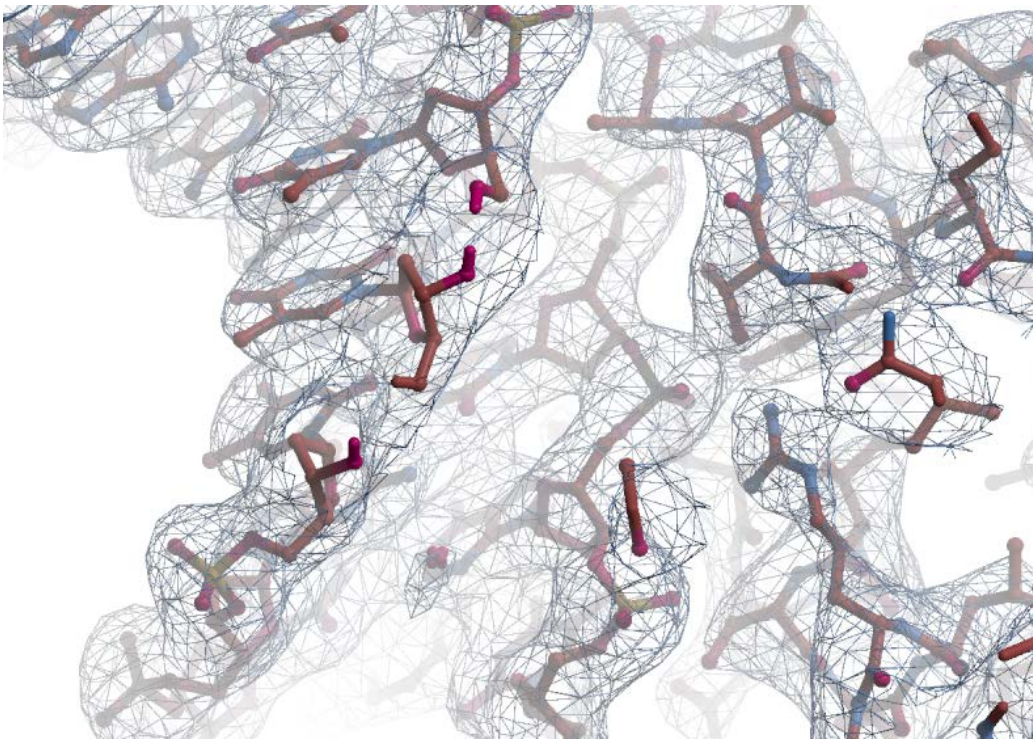


Figure 9.29 Fragment of the final ToxR^{P101A}-DNA20 model fitted in the 2 F_o-F_c electron density map (contoured at 1.0 σ)

All the residues in the final model were located in the allowed regions of the Ramachandran plot shown in Figure 9.30. This indicates that there are no areas where the backbone torsion deviates from what is expected in the solved structure.

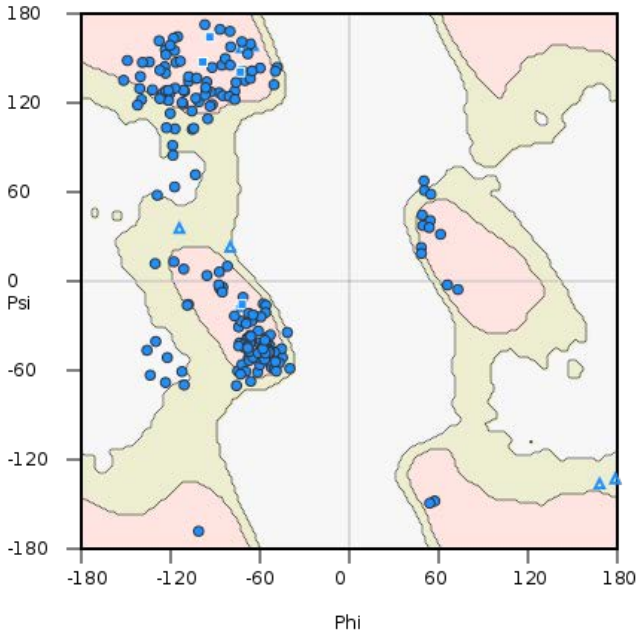


Figure 9.30 Ramachandran plot for ToxR^{P101A}-DNA20 structure

ToxR^{P101A} final model is illustrated in Figures 9.31 and 9.32. The asymmetric unit is composed of two 20bp dsDNA and 2 ToxR^{P101A} protein molecules. Figure 9.31 shows the cartoon representation of one a 20bp dsDNA molecule and a ToxR^{P101A} protein molecule and its domains. Figure 9.32 illustrates the electrostatic surface of ToxR^{P101A} and it is possible to visualize how P101A still enters the minor groove.

9.11.3.1. ToxR^{P101A} final model

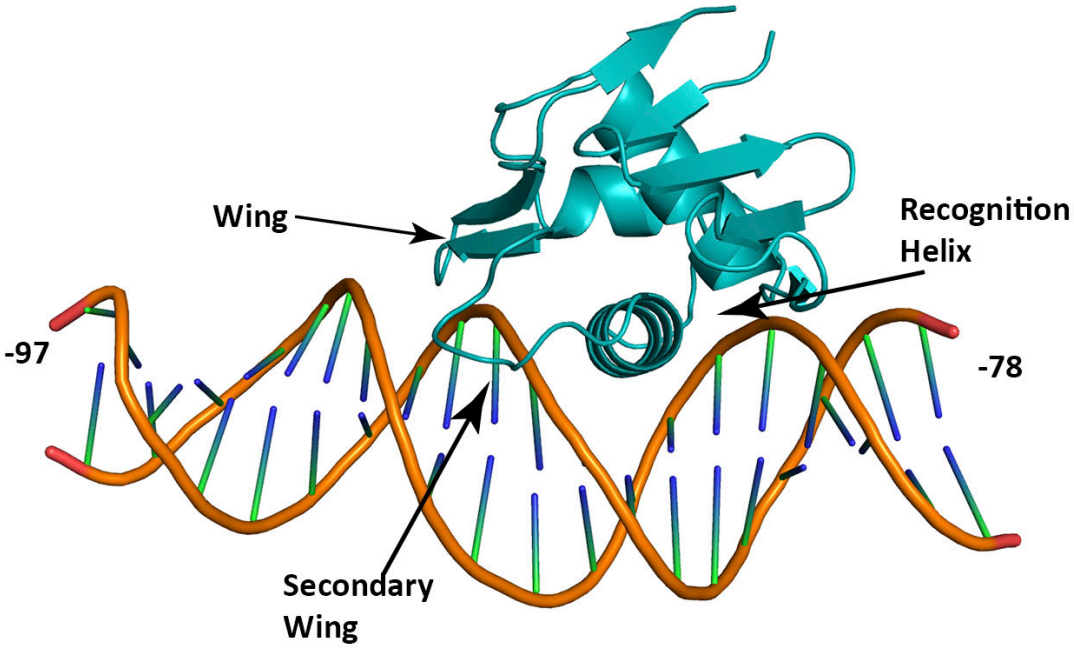


Figure 9.31 ToxR^{P101A}-DNA20 structure

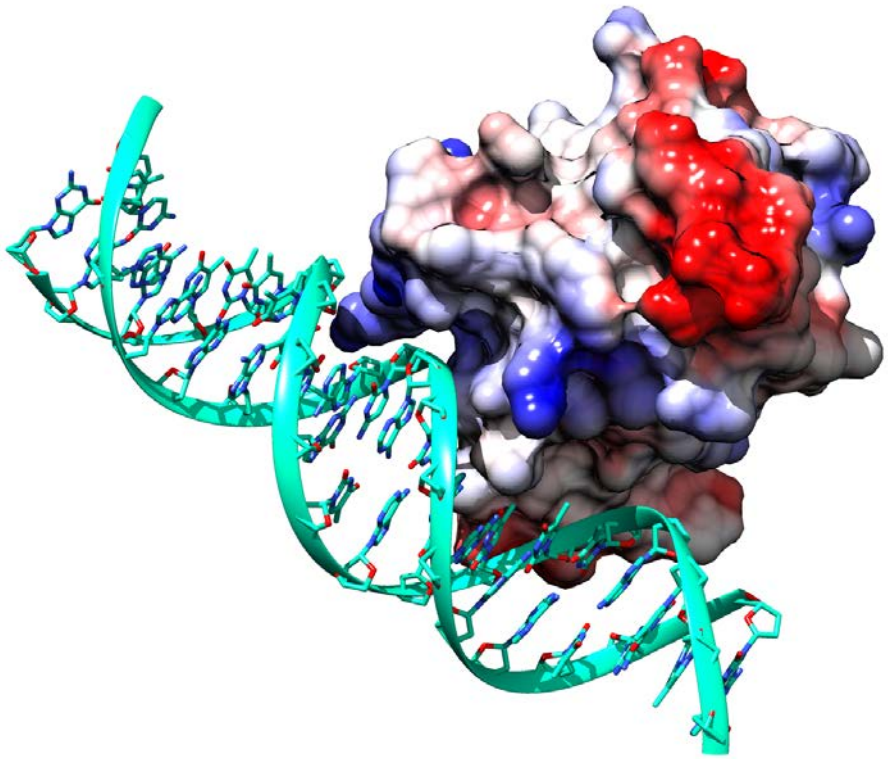


Figure 9.32 Electrostatic surface representation of the ToxR^{P101A}-DNA20 structure

As mentioned before, there is little change in the structure of the protein, using the cartoon representation, we can see that it conserves the secondary elements described by Pieretti. This is illustrated in Figure 9.33. It mantains the hydrophobic core formed by Ile42, Leu43, Trp 44, Leu45 and Leu46, plus Phe 69. This still sustains the fold of the protein and helps the helix α1 link the N-terminal β-sheet with the H-T-H motif and the β-hairpin.

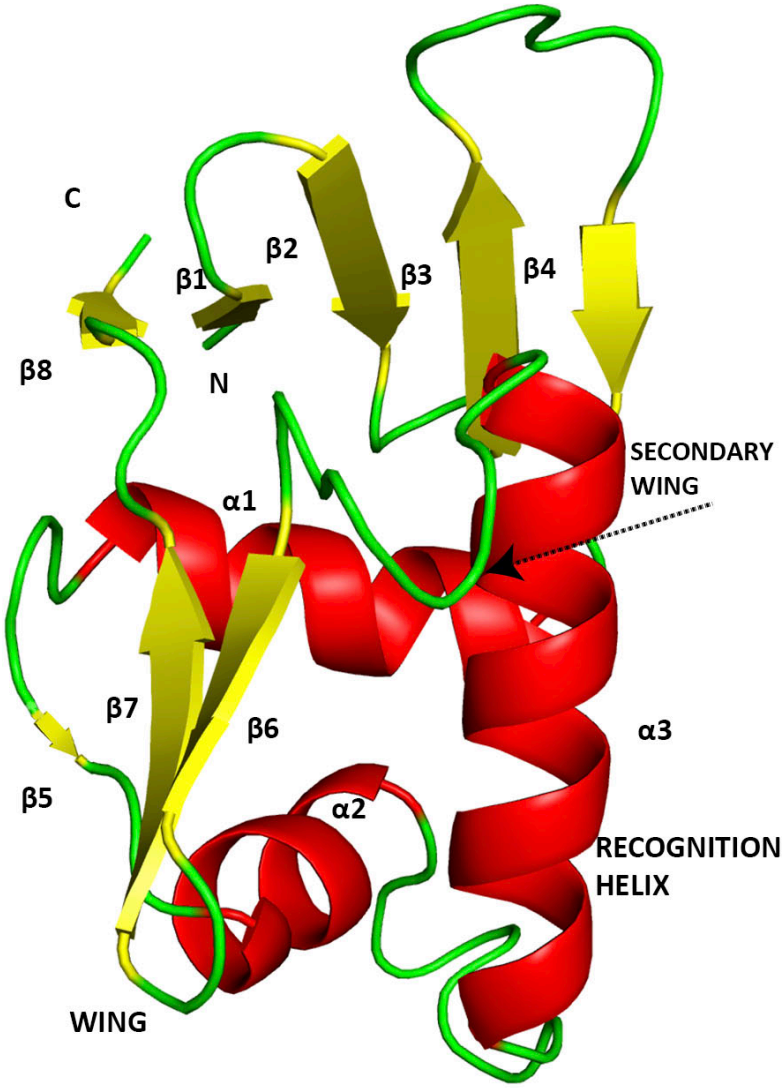


Figure 9.33 ToxR^{P101A}-DNA20 structure secondary elements

Once more, the r.m.s.d. between $ToxR^{Wildtype}$ and $ToxR^{P101A}$ are negligible as illustrated in Table 9.9.

Table 9.9 r.m.s.d. values between $ToxR^{Wildtype}$ and $ToxR^{P101A}$

Element in $ToxR^{P101A}$ model (chain)	r.m.s.d. values (Å)
$ToxR^{P101A}$ Chain C	0.9339
$ToxR^{P101A}$ Chain F	0.7961

9.12. Protein-DNA interactions

Using Nucplot (Luscome N.M. *et al.*, 1997) we created a plot of the protein-DNA interactions to compare the interactions of ToxR with the *toxT* promoter as well well as the changes that the mutations entail. These protein-DNA interactions are illustrated in Figures 9.34 for ToxR, 9.35 for $ToxR^{Q78A}$, 9.36 for $ToxR^{S81A}$ and 9.37 for $ToxR^{P101A}$.

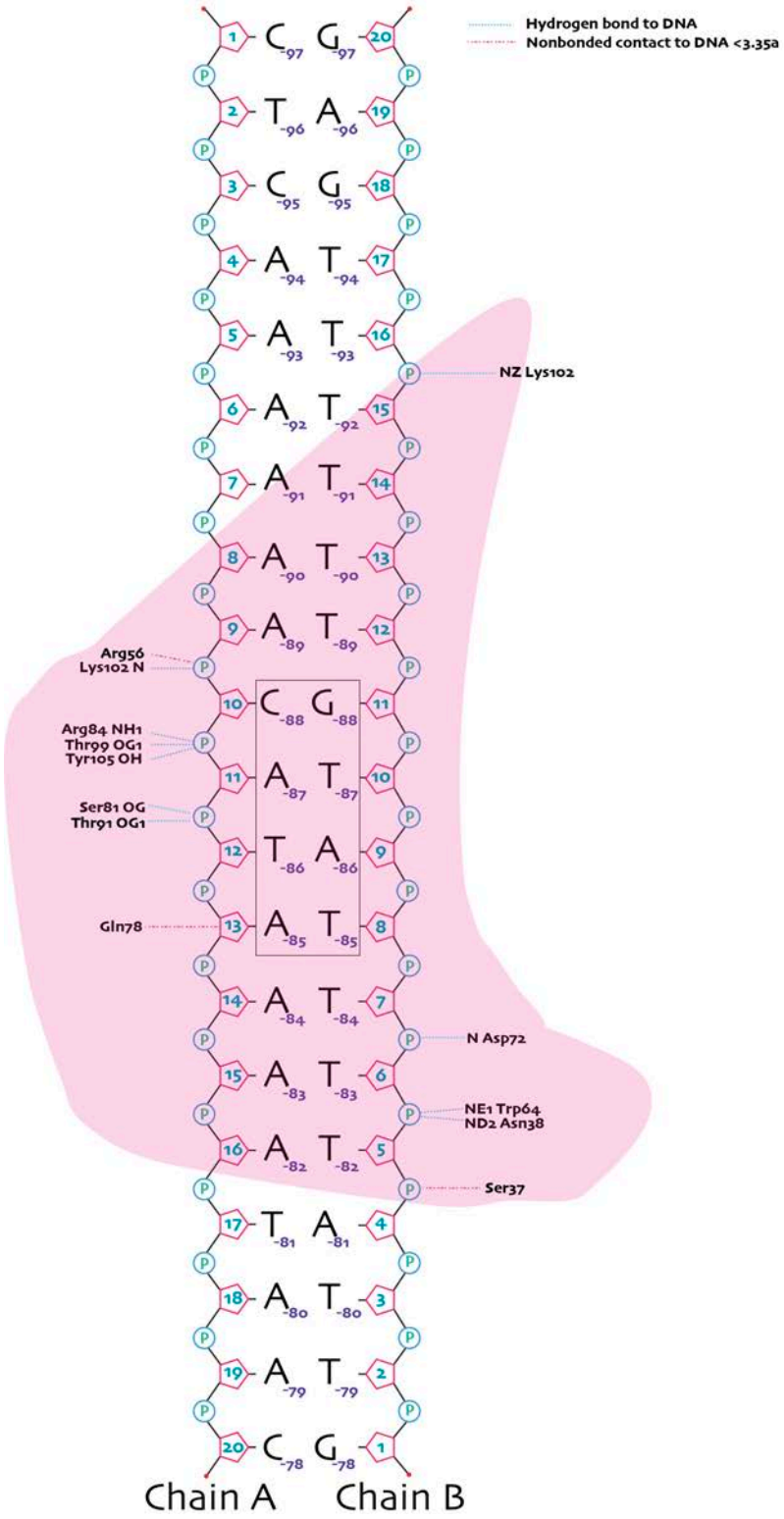


Figure 9.34 Protein-DNA interactions in the ToxR^{Wildtype}-DNA20 structure

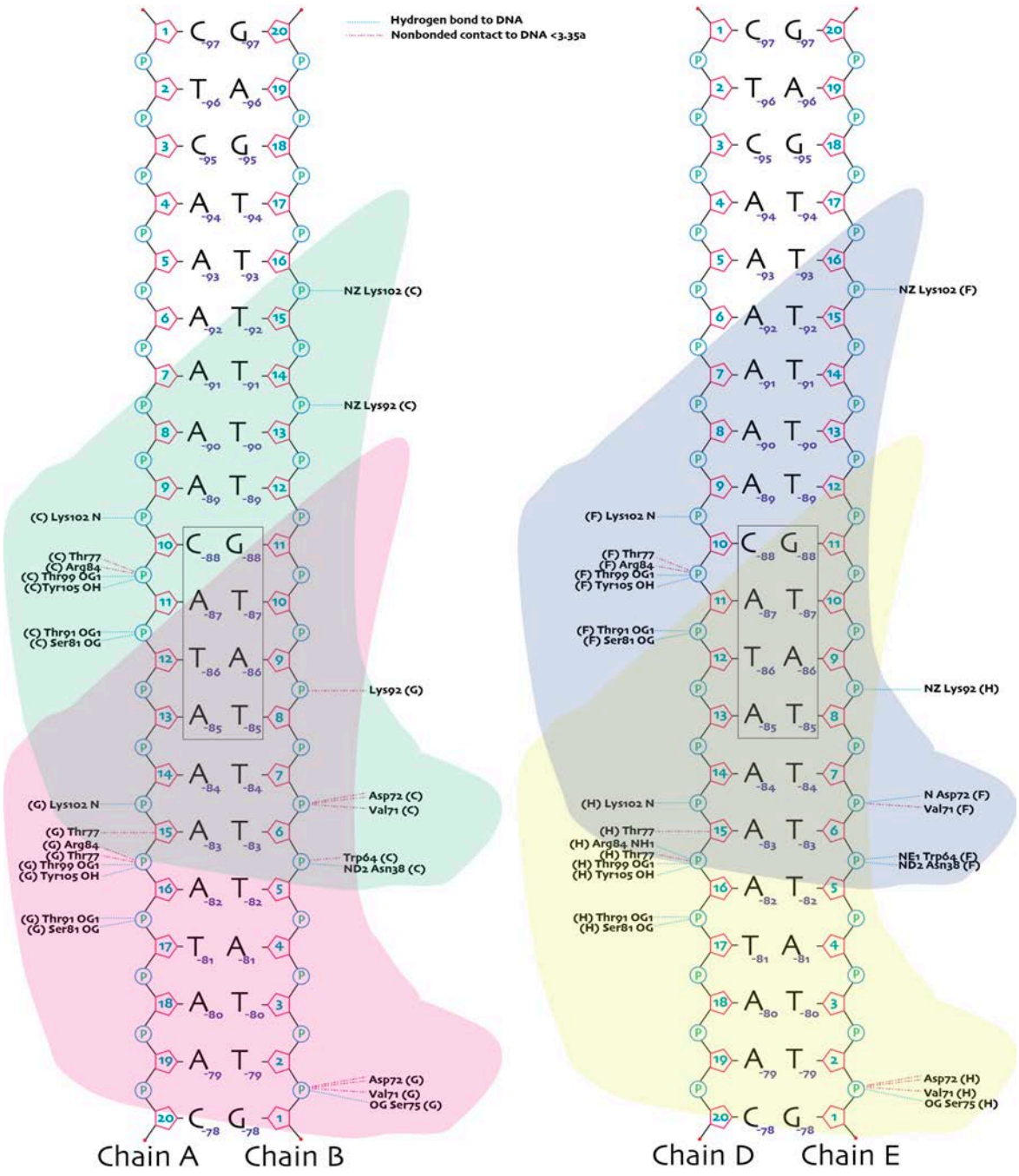


Figure 9.35 Protein-DNA interactions in the ToxR^{Q78A}-DNA20 structure

In the Q78A interaction scheme, we can observe that most of the interactions are conserved, yet at the first position of chain B and E respectively, we could use another basepair to decipher if it binds exactly in the same manner as the previous molecule as there is no phosphate to which Trp64 and Asn38 bind to.

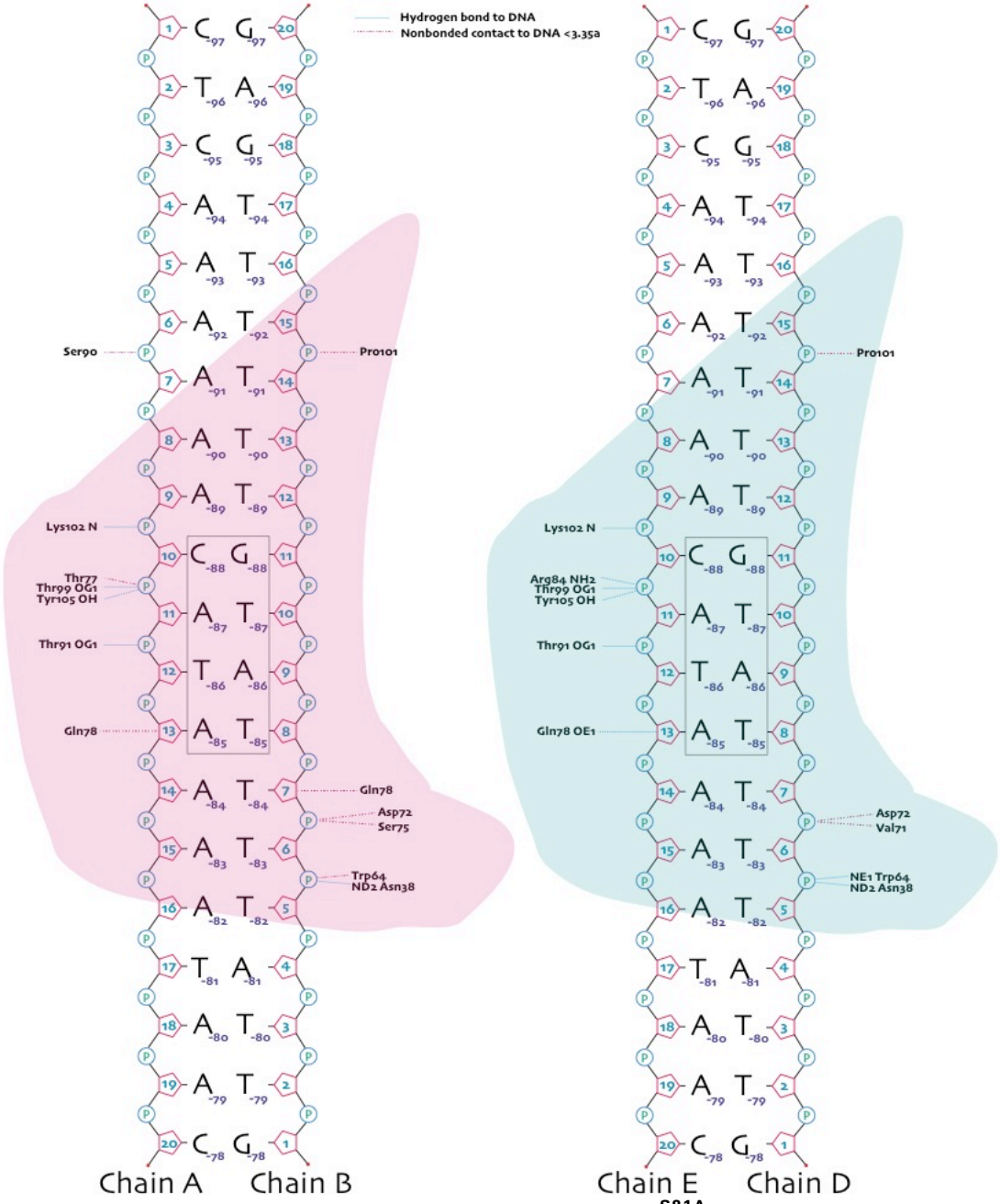


Figure 9.36 Protein-DNA interactions in the ToxR^{S81A}-DNA20 structure

The conservation of most of the interactions is evident in all of the mutant structures; it is important to notice that if these interactions are conserved, so we theorize that the decrease in binding should be because of defined interactions that the mutated residue does not have anymore.

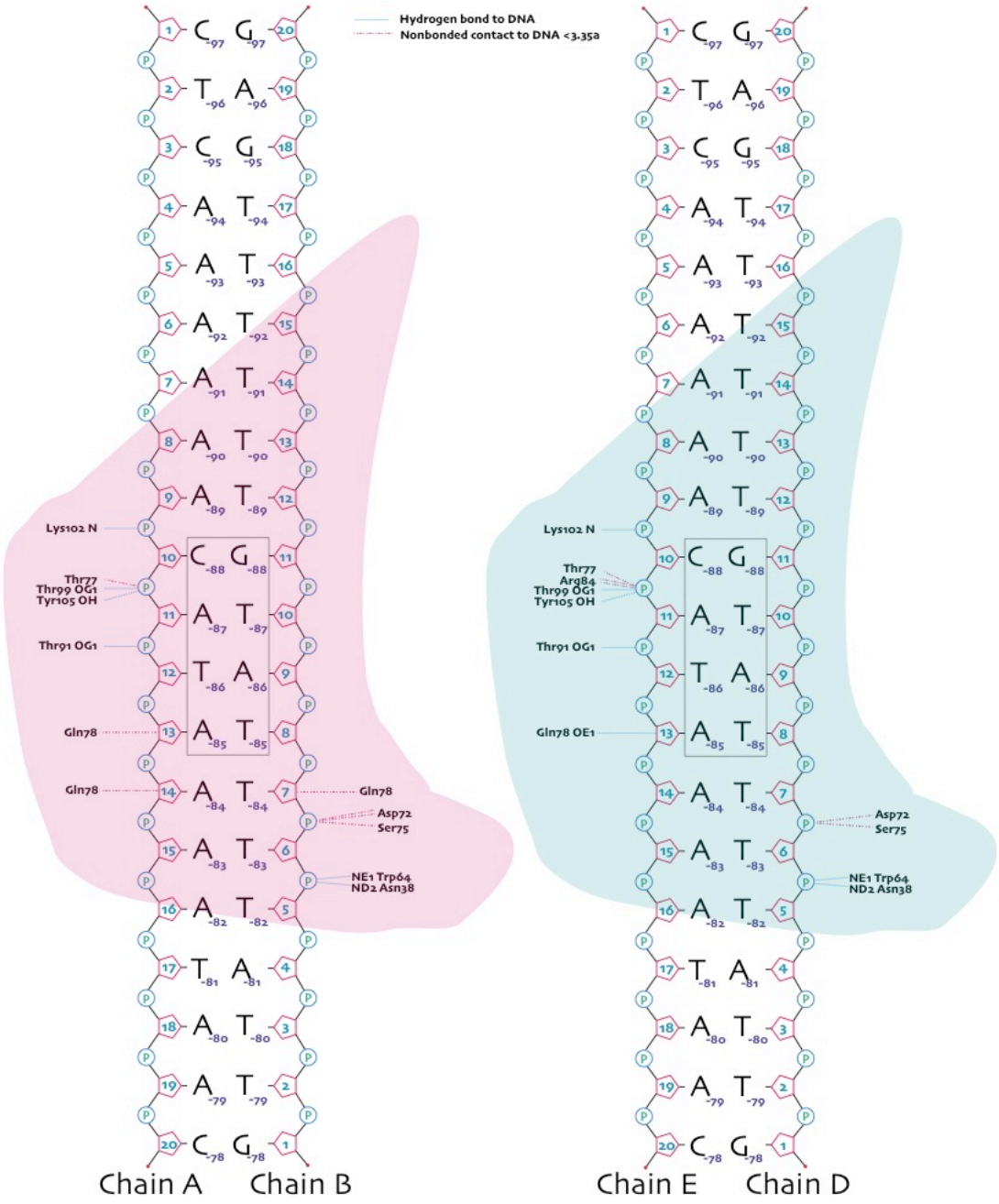


Figure 9.37 Protein-DNA interactions in the ToxR^{P101A}-DNA20 structure

The total of hydrogen bonds and nonbonded contacts observed are shown in Table 9.10. All, ToxR^{Wildtype}-DNA, ToxR^{S81A}-DNA and ToxR^{P101A}-DNA are composed of one protein bound to a molecule of DNA. Interestingly, the ToxR^{Q78A}-DNA crystal contains structures formed by two protein molecules and one DNA molecule. This has not

been observed with oligos this size; interestingly, the second protein molecule binds further down in a repeating "ATAA" segment in the same strand and direction of the promoter as shown in the schematic in Figure 9.38.

Table 9.10 Hydrogen bonds and nonbonded contacts between ToxR and DNA

	$ToxR^{Wildtype}$	$ToxR^{P101A\_1}$ (A-B chain)	$ToxR^{P101A\_2}$ (D-E chain)	$ToxR^{S81A\_1}$ (A-B chain)	$ToxR^{S81A\_1}$ (D-E chain)	$ToxR^{Q78A\_1}$ (A-B chain)	$ToxR^{Q78A\_2}$ (D-E chain)
First protein molecule (bonded at the CATA box)							
Hydrogen bonds	10	6	6	5	3	8	9
Nonbonded DNA contacts	3	7	7	8	8	4	3
	Second protein molecule (bonded at the ATAA repeating segment)						
Hydrogen bonds						6	7
Nonbonded DNA contacts						6	5

Figure 9.38 Schematic of protein-DNA binding in the ToxR-DNA20 structure

Schematic depicting the protein-DNA binding in the ToxR-DNA structures. ToxR is represented by the green structure with a yellow spiral used to depic the recognition helix and the red wing for the wing. First ToxR molecule binds in the same position in all four crystals structures ($ToxR^{Wildtype}$, $ToxR^{Q78A}$, $ToxR^{S81A}$, and $ToxR^{P101A}$). Semi-transparent ToxR protein molecule represents the second protein molecule that appears only in the $ToxR^{Q78A}$ crystal structure.

Most of the interactions are conserved in spite of the mutations; the specificity for the sequence is conserved and the DNA structure changes as will be explained in the following section. Even if the affinity and the activation are reduced, there is still binding at least in the high concentrated crystallization conditions. In the physiological conditions, the binding is not strong enough to enable a stable complex that allows the polymerase to be recruited.

9.13. DNA conformation

The DNA has a B-form straight conformation when interacting with $ToxR^{Wildtype}$. In this section we will enlist and compare some characteristics of the conformation of the DNA.

9.13.1. Minor groove analysis

The minor groove of the DNA with $ToxR^{Wildtype}$ narrows down up to 8.9 Å whilst for $ToxR^{Q78A}$, $ToxR^{S81A}$ and $ToxR^{P101A}$ up to 8.7 Å, 8.7 Å and 8.6 Å respectively. A deeper comparison of the behavior of the minor groove is done in this section.

9.13.1.1. Wildtype vs Q78A

The minor groove for the $ToxR^{Q78A}$ structure reaches the maximum width of 12.1 Å in the center of the 20 bp oligo used while in the wildtype structure this width is only 10.3 Å, and the maximum width (12.1 Å) is reached at the beginning of the oligo as illustrated in Figure 9.39. The residues that are interacting with the DNA molecule in the helix are displayed at the segment of the DNA in which they interact.

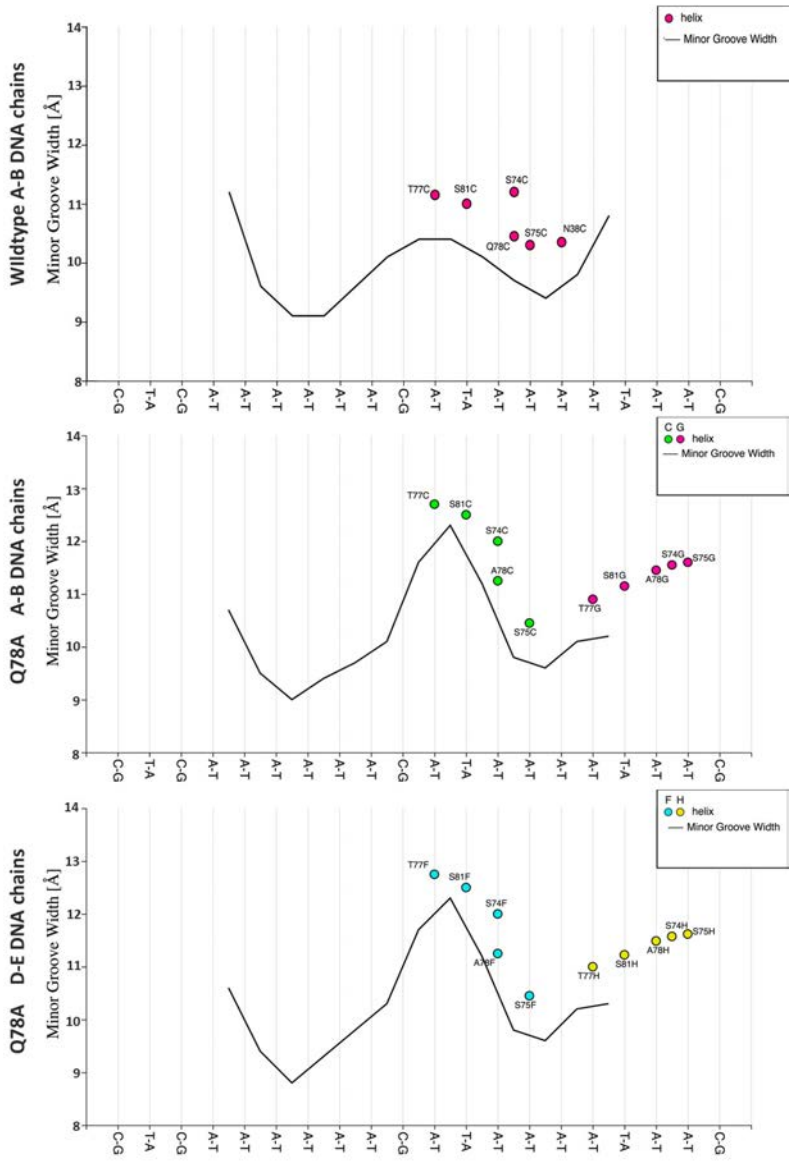


Figure 9.39 Minor groove width comparison between ToxR^{Wildtype} and ToxR^{Q78A}

Color code is kept as in Figure 9.35, green represents protein chain "C" while pink is "G" which interact with the A-B dsDNA chain and blue represents chain "F" and yellow is "H" which bind to D-E dsDNA chain.

9.13.1.2. Wildtype vs S81A

In the ToxR^{S81A} structure, the minor groove reaches its maximum width of of 11.9 Å at the center of the 20 bp sequence used while in the wildtype structure it is 10.3 Å, and the maximum width is

presented at the beginning of the promoter measuring 12.1 Å, as illustrated Figure 9.40.

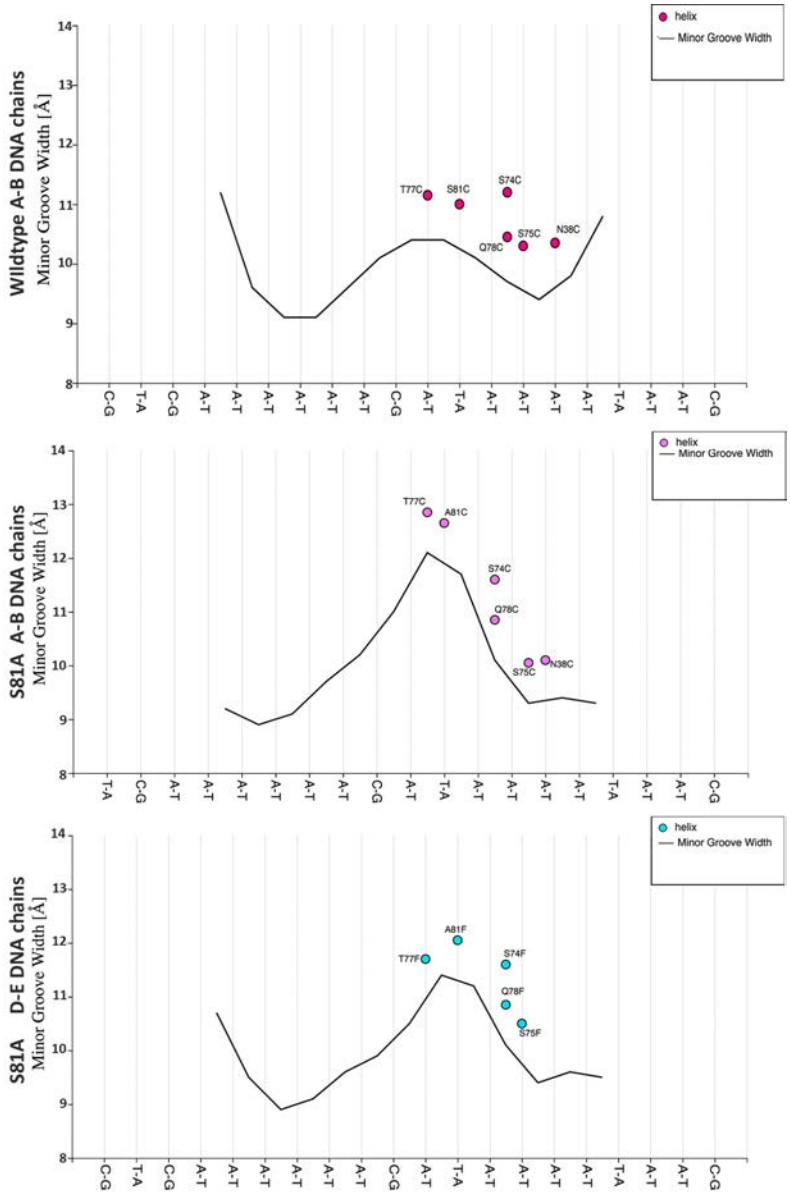


Figure 9.40 Minor groove width comparison between ToxR^{Wildtype} and ToxR^{S81A}

Residues in the helix that interact with the DNA molecule are presented in the DNA segment in which they interact.

9.13.1.3. Wildtype vs P101A

In this comparison, we can see that the minor groove reaches its biggest width at the center of the 20 bp sequence used and measures up to 13.1 Å as illustrated in Figure 9.41.

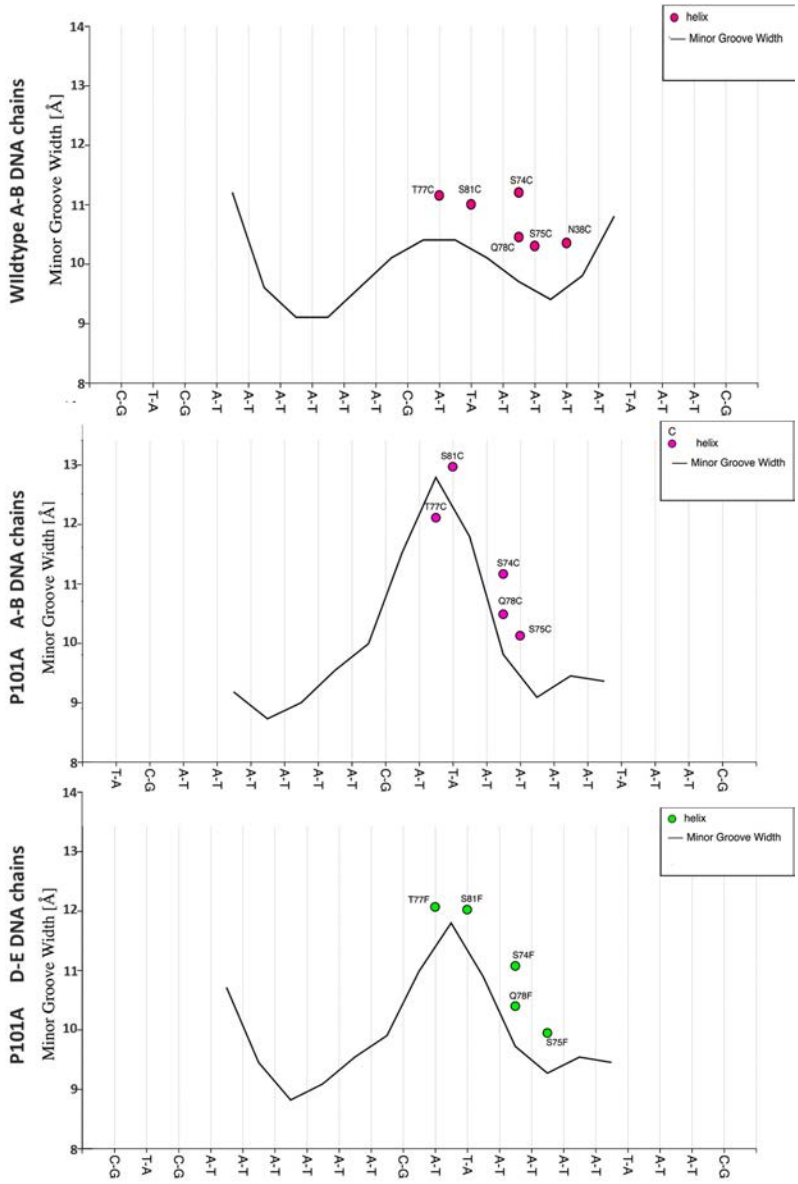


Figure 9.41 Minor groove width comparison between ToxR^{Wildtype} and ToxR^{P101A}

The residues that are interacting with the DNA molecule in the helix are displayed at the segment of the DNA in which they interact.

9.13.2. Major groove

The major groove of the DNA with ToxR^{Wildtype} measures up to 20 Å whilst for ToxR^{Q78A}, ToxR^{S81A} and ToxR^{P101A} up to 19.7 Å, 20.3 Å and 20.1 Å respectively. This is compared in more depth in this section.

9.13.2.1. Wildtype vs Q78A

There is a slight change in the major groove between the wildtype and Q78A mutant, yet there is the loss of the interaction of the Asn38 residue with the major groove, as well as the appearance of a second molecule in tandem, as illustrated in Figure 9.42. The newly appeared ToxR^{Q78A} molecule binds to an "ATAA" repeat that has a similar compression as the one presented in the CATA box for both (ToxR^{Wildtype} and ToxR^{Q78A}), which is approximately 17.4 Å for the CATA box and 17.2 Å for the "ATAA" repeat zone.

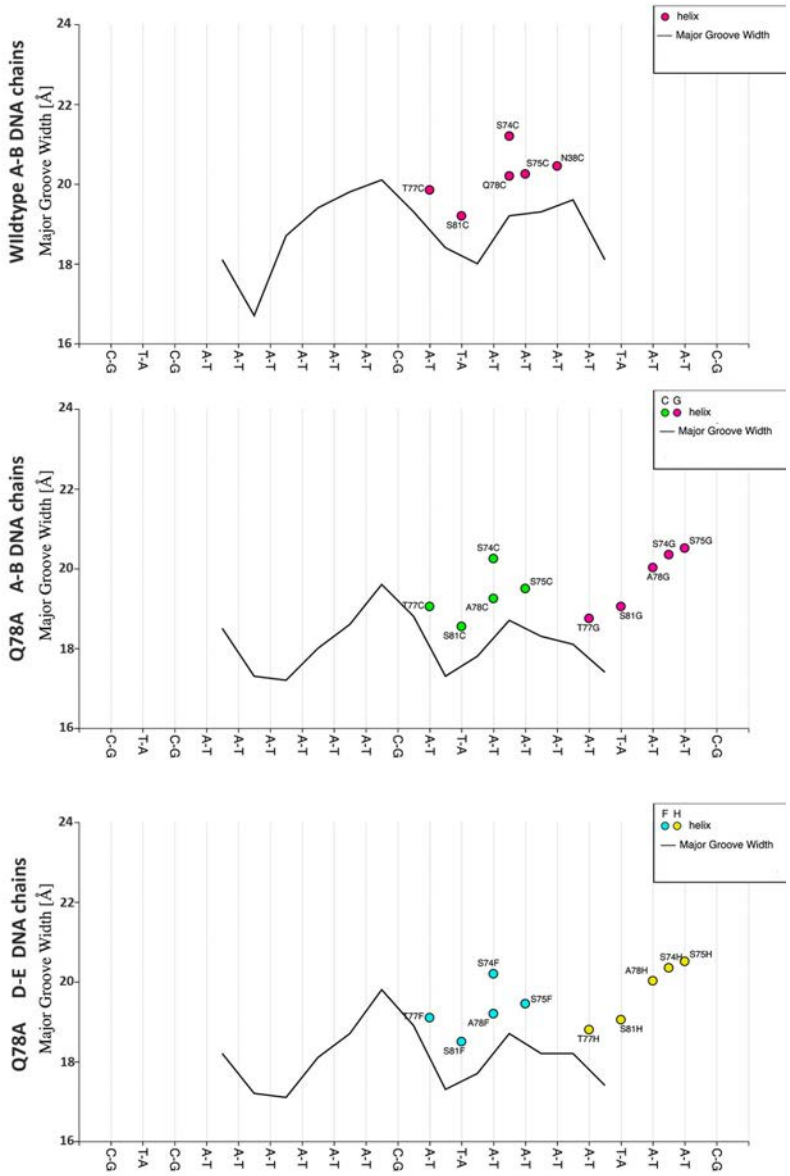


Figure 9.42 Major groove width comparison between ToxR^{Q78A} and ToxR^{Wildtype}

The residues that interact with the DNA molecule with the helix are displayed at the segment of the DNA in which they interact. Color code is kept as in Figure 9.35, green represents protein chain “C” while pink is “G” which interact with the A-B dsDNA chain and blue represents chain “F” and yellow is “H” which bind to D-E dsDNA chain.

9.13.2.2. Wildtype vs S81A

In Figure 9.43, we can see that in the ToxR^{S81A} structure the compression near the CATA box measures 17.5 Å; value that does not vary much from the wildtype. This indicates that even though affinity is decreased, there is still binding and stabilization of the compression by ToxR^{S81A}.

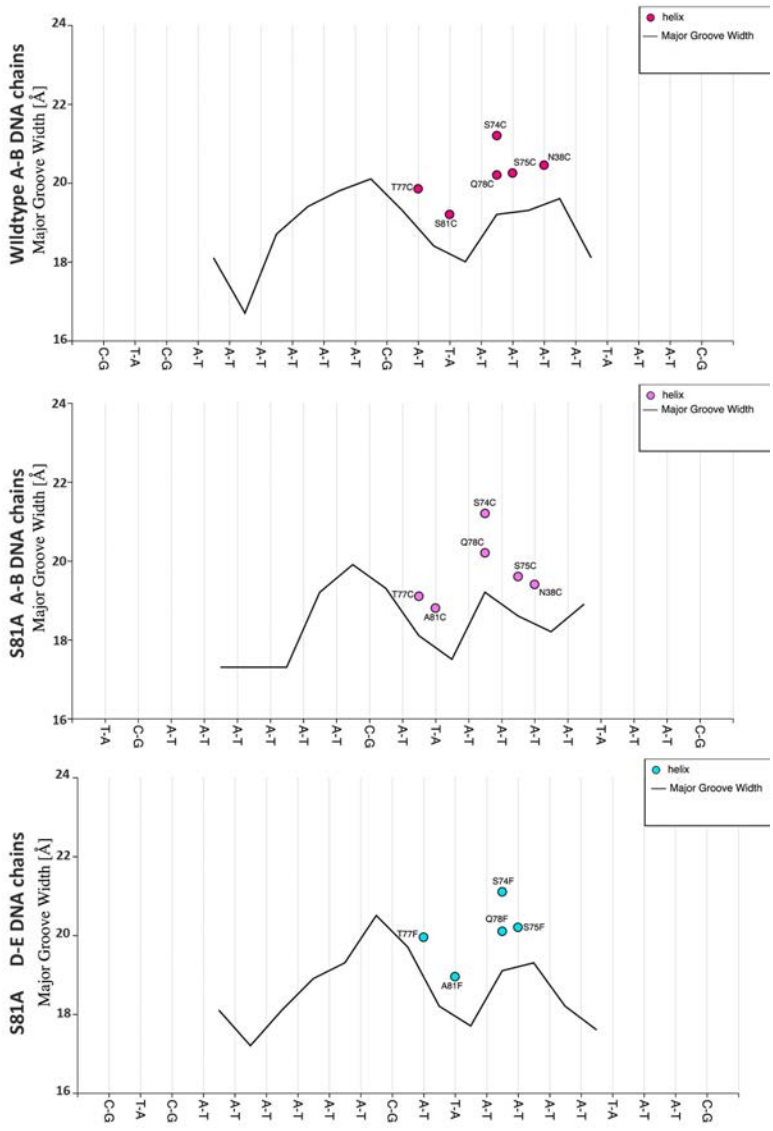


Figure 9.43 Major groove width comparison between ToxR^{S81A} and ToxR^{Wildtype}

Residues from the helix in the ToxR molecule interacting with the DNA are depicted as colored circles according to the helix they form part of.

9.13.2.3. Wildtype vs P101A

In Figure 9.44, we can see that the interaction between the Asn38 residue and the DNA is lost in the P101A mutant. The compression that occurs near the CATA box remains in a similar width (around 17.4 Å) suggesting that although affinity is reduced, the protein still binds to the 20 bp oligo, allowing ToxR^{P101A} to still contribute to stabilizing the compression.

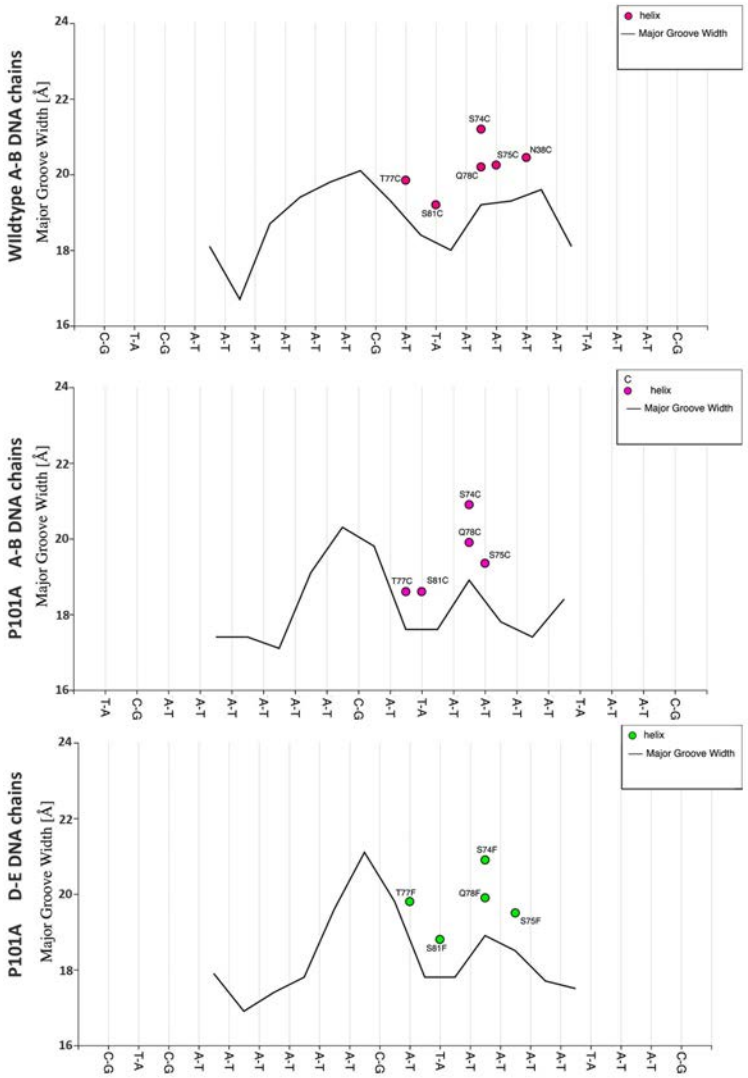


Figure 9.44 Major groove width comparison between ToxR^{P101A} and ToxR^{Wildtype}

Residues from the ToxR molecule that interact with the DNA are shown as colored circles depending on the helix they belong to.

9.14. DNA geometrical parameters

All the geometrical parameters of the DNA are presented in this section

9.14.1. DNA geometrical parameters for ToxR^{Q78A}

Table 9.11 Local base-pair parameters for ToxR^{Q78A}

Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
G-C	1.2	0.23	0.38	-1.96	-22.05	6.44
T-A	-0.22	-0.03	-0.15	9.33	-13.15	1.42
T-A	0.29	-0.2	-0.01	13.02	-13.92	-1.41
A-T	-0.11	-0.11	0.36	2.63	-6.84	6.02
T-A	0.25	-0.25	0.64	-0.75	-3.25	-0.5
T-A	0.34	0	0.89	-6.22	-9.36	3.04
T-A	0.36	-0.15	0.3	0.18	-7.92	-0.63
T-A	-0.41	-0.26	0	-3.68	-7.62	-4.02
A-T	0.22	-0.27	-0.28	-7.27	-2.43	2.7
T-A	0.13	-0.26	-0.04	-1.96	2.36	4.16
G-C	-0.34	-0.05	-0.31	-6.33	-15.94	-1.85
T-A	-0.32	-0.11	0.48	-0.97	-7.25	1.09
T-A	0.43	-0.46	0.25	4.13	-12.36	-1.01
T-A	-0.47	-0.39	-0.1	-1.63	-16.13	-3.69
T-A	-0.41	-0.4	0.37	-10.76	-16.33	0.62
T-A	-0.02	-0.06	0.19	-20.12	-4.96	-2.51
T-A	0.05	-0.14	-0.13	-14.21	-4.55	-4.44
G-C	0.11	-0.27	-0.67	-13.29	-7.49	-2.85
A-T	0.71	-0.2	-0.07	-3.25	-12.96	-4.43
G-C	0.25	0.1	-0.52	-7.86	-22.47	4.24
C-G	-0.37	0.09	-0.47	7.43	-20.43	4.27
T-A	-0.86	-0.13	-0.11	4.59	-12.61	-5.62
C-G	-0.22	-0.24	-0.72	15.13	-8.13	-3.97
A-T	-0.2	-0.08	-0.19	15.75	-5.72	-4.22
A-T	-0.18	0.07	0.1	21.48	-5	-1.86
A-T	0.32	-0.28	0.28	12.34	-16.99	0.5
A-T	0.47	-0.37	-0.2	2.53	-16.84	-3.62
A-T	-0.41	-0.51	0.13	-4	-11.92	-1.32
A-T	0.26	-0.19	0.4	0.89	-6.3	0.36
C-G	0.25	-0.12	-0.31	6.72	-14.79	-2.81
A-T	-0.31	-0.14	0.11	3.6	2.8	5.4
T-A	-0.35	-0.14	-0.15	9.8	-3.45	3.76
A-T	0.29	-0.15	0.16	6	-9.18	-4.21
A-T	-0.38	-0.07	0.44	0.97	-8.7	-2.04
A-T	-0.37	0.04	1.06	7.2	-9.82	2.47
A-T	-0.34	-0.18	0.81	2.31	-3.85	1.15
T-A	0.06	-0.05	0.44	-0.44	-7.62	8.33
A-T	-0.27	-0.31	0.01	-8.94	-14.22	0.57
A-T	0.17	-0.02	0.12	-9.26	-9.93	2.15
Ave.	-0.04	-0.15	0.09	0.71	-10.13	0.27
s.d.	0.44	0.17	0.40	8.80	6.05	3.79

Table 9.12 Local base-pair step parameters for ToxR^{Q78A}

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
GT/AC	-0.79	-1.4	2.92	1.13	-5.38	26.15
TT/AA	0.01	-0.09	3.13	0.93	-5.9	40.84
TA/TA	0.46	0.05	3.42	-1.19	0.73	38.14
AT/AT	-0.11	-0.86	3.37	-2.01	-3.53	33.78
TT/AA	0.38	-1.08	3.13	1.43	0.25	38.18
TT/AA	-0.26	-1	2.81	5.45	-1.68	31.59
TT/AA	-0.27	-0.56	3.15	-0.24	-0.05	35.1
TA/TA	0.11	0.78	3.48	-0.97	2.24	41.37
AT/AT	0.24	-0.25	3.25	-4.82	-2.08	29.18
TG/CA	-0.03	-0.76	3.26	4.25	4.01	36.51
GT/AC	0.08	-0.84	3.05	-5.99	0.31	35.39
TT/AA	-0.13	-0.37	3.13	1.08	-7.21	39.76
TT/AA	-0.3	-0.66	3.29	3.76	-3.45	32.31
TT/AA	-0.04	-0.57	3.44	-1.92	-2.8	35.97
TT/AA	-0.15	-0.44	3.41	3.21	-5.98	38.32
TT/AA	-0.12	-0.43	3.13	5.4	-1.99	37.28
TG/CA	1.02	0.88	3.22	5.4	7.48	37.92
GA/TC	-1.19	0.61	2.97	-8.95	3.31	33.87
AG/CT	0.56	-0.37	3.5	2.02	14.94	28.74
GC/GC	----	----	----	----	----	----
CT/AG	-0.61	-0.35	3.46	-1.79	14.2	27.53
TC/GA	1.22	0.63	2.95	9.36	3.48	33.79
CA/TG	-1	0.87	3.21	-5.53	7.33	38.04
AA/TT	0.14	-0.45	3.12	-5.35	-1.9	36.68
AA/TT	0.13	-0.45	3.39	-3.36	-5.67	38.57
AA/TT	0.04	-0.57	3.44	1.85	-2.84	36.52
AA/TT	0.28	-0.66	3.31	-3.98	-3.55	32.6
AA/TT	0.12	-0.39	3.14	-1.46	-7.15	39.52
AC/GT	-0.09	-0.86	3.06	5.32	0.47	35.1
CA/TG	0.13	-0.76	3.24	-5.03	3.79	35.69
AT/AT	-0.25	-0.23	3.22	4.93	-2.34	29.05
TA/TA	-0.18	0.8	3.46	0.69	2.12	41.83
AA/TT	0.22	-0.56	3.15	0.02	0.13	35.9
AA/TT	0.3	-0.98	2.79	-5.85	-1.79	31.89
AA/TT	-0.29	-1.08	3.11	-1.57	-0.11	37.59
AT/AT	0.14	-0.84	3.35	2.45	-3.79	33.7
TA/TA	-0.45	0.05	3.36	1.65	0.54	38.27
AA/TT	-0.09	-0.04	3.24	-3.5	-3.92	41.01
AC/GT	0.7	-1.42	2.79	-0.9	-4.38	25.17
Ave.	0	-0.39	3.21	-0.11	-0.32	35.23
s.d.	0.47	0.61	0.2	4.05	5.08	4.26

9.14.2. DNA geometrical parameters for ToxR^{S81A}

Table 9.13 Local base-pair parameters for ToxR^{S81A}

Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
G-C	-0.82	-0.11	-0.86	2.62	-10.61	9.26
T-A	-0.12	-0.21	-0.71	5.08	-26.68	11.21
T-A	-0.21	0.21	-0.31	2.85	-10.96	8.48
A-T	0.15	0.09	0.07	-6.69	-9.77	5.7
T-A	-0.16	-0.16	0.63	-2.77	-15.52	4.86
T-A	0	0.12	0.45	0.34	-15.95	6.21
T-A	0.39	-0.09	0.54	-0.3	-13.09	4.56
T-A	-0.23	-0.37	0.35	1.9	-14.7	3.79
A-T	0.25	-0.31	0.7	1.83	-14.18	2.11
T-A	-0.19	-0.36	0.55	-1.34	-5.08	0.54
G-C	-0.78	-0.32	0.1	-6.22	-7.33	-1.08
T-A	-0.07	-0.51	0.7	-12.34	-14.12	-4.6
T-A	-0.64	-0.19	0.38	-7.33	-18.9	-1.56
T-A	-0.71	-0.32	0.3	-8.59	-17.73	-2.58
T-A	-0.19	-0.05	0.2	-10.43	-18.42	5.05
T-A	-0.43	0.02	0.27	-5.05	-16.14	2.99
T-A	-0.68	-0.03	-0.13	-2.53	-9.78	-1.77
G-C	0.05	-0.13	-0.28	-1.06	-0.13	-2.64
A-T	0.78	-0.13	0.25	8.64	-6.31	-1.04
G-C	-1.03	-0.2	-0.39	-5.17	4.1	2.59
T-A	-0.44	-0.33	0.1	5.6	-8.1	-3.35
T-A	0.03	0.1	-0.08	5.14	-12.69	-0.27
A-T	0.25	-0.1	0.45	4.05	-7.92	2.64
T-A	0.06	-0.12	0.41	-3.7	-19.57	1.5
T-A	-0.09	-0.13	0.48	-0.44	-17.84	2.78
T-A	-0.03	0.05	0.08	3.48	-15.97	1.05
T-A	0.05	-0.35	-0.07	0.77	-12.18	2.9
A-T	-0.04	-0.08	0.54	-0.12	-14.66	0.8
T-A	-0.06	-0.33	0.59	-0.96	-11.75	2.31
G-C	-0.18	-0.13	0.33	-1.52	-6.6	6.02
T-A	-0.32	-0.09	0.88	-12.25	-11.37	-6.75
T-A	-0.59	-0.18	0.11	-2.35	-21.57	-5.82
T-A	-0.24	-0.28	0.28	-7.09	-20.41	-0.13
T-A	-0.05	-0.19	-0.08	-2.83	-21.22	1.7
T-A	0.04	-0.15	-0.38	-3.19	-19.22	7.84
T-A	0.2	-0.21	-0.22	-1.21	-11.6	-2.61
G-C	0.42	-0.09	-0.77	-11.17	-7.97	3.46
A-T	0.88	0.06	-0.14	-3.21	-13.37	-13.25
G-C	-1.63	-0.37	1.39	9.2	-18.42	12.16
Ave.	-0.16	-0.15	0.17	-1.75	-13.17	1.67
s.d.	0.47	0.16	0.46	5.36	6.06	5.02

Table 9.14 Local base-pair step parameters for ToxR^{S81A}

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
GT/AC	0.03	-0.07	2.98	-0.14	5.58	37.42
TT/AA	0.12	-0.03	3.35	-2.71	-1.12	36.45
TA/TA	-0.05	-0.16	3.54	-3.61	-1.25	39.61
AT/AT	-0.36	-0.65	3.17	-3.59	-6.29	32.65
TT/AA	0.2	-0.3	2.99	1.79	-5.99	38.76
TT/AA	0.14	-0.54	3.27	-1.24	-5.87	36.53
TT/AA	0.06	-0.46	3.09	5.24	-4.99	35.06
TA/TA	-0.23	0.06	3.13	-4.44	5.86	38.28
AT/AT	0.3	-0.75	3.45	0.62	0.08	29.41
TG/CA	0.04	-0.93	3.24	0.45	-0.16	35.55
GT/AC	-0.72	-1.2	3.48	-3.62	-0.57	36.76
TT/AA	-0.35	-0.91	2.89	0.07	-4.61	33.87
TT/AA	-0.31	-0.64	3.21	-0.69	-6.51	36.38
TT/AA	-0.07	-0.14	3.25	0.7	-3.39	38.49
TT/AA	0.13	0.32	3.16	-0.78	-10.59	39.43
TT/AA	0.05	-0.13	3.12	5.36	-4.13	33.75
TG/CA	1.21	1.17	3.43	5.21	1.05	39.34
GA/TC	-0.84	0.24	3.16	-5.49	2.59	39.29
AG/CT	----	----	----	----	----	----
GT/AC	-0.63	0.02	3.09	-6.16	4.54	34.54
TT/AA	0.04	-0.34	3.33	2.46	-6.73	36.84
TA/TA	0.05	-0.07	3.28	-5.43	-4.73	44.2
AT/AT	-0.19	-0.84	3.44	0.55	-3.63	30.2
TT/AA	0.02	-0.43	3.07	-0.83	-4.17	37.78
TT/AA	0.09	-0.56	3.08	2.72	-6.58	36.48
TT/AA	0.09	-0.41	3.38	1.43	-4	35.76
TA/TA	-0.07	-0.06	3.15	-4.34	7.84	38.36
AT/AT	0.36	-0.6	3.37	-1.4	-3.58	31.13
TG/CA	0.24	-1.17	3.06	0.68	-0.92	37.07
GT/AC	-1.11	-1.01	3.49	-4.46	2.13	33.65
TT/AA	-0.37	-1.1	2.78	3.49	-7.92	32.12
TT/AA	-0.09	-0.51	3.37	-4.35	-6.26	38.63
TT/AA	-0.22	-0.06	3.06	0.97	-3.25	39.23
TT/AA	0.01	0.14	3.23	2.05	-3.64	38.02
TT/AA	0.12	0	3.18	-3.08	0.29	39.49
TG/CA	1.01	0.79	3.7	7.3	5	35.44
GA/TC	-1.72	0.5	3.05	-8.05	0.66	38.37
AG/CT	1.93	-0.84	2.74	-8.13	4.53	26.35
Ave.	-0.03	-0.31	3.21	-0.85	-1.91	36.24
s.d.	0.6	0.54	0.21	3.78	4.44	3.43

9.14.3. DNA geometrical parameters for ToxR^{P101A}

Table 9.15 Local base-pair parameters for ToxR^{P101A}

Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
G-C	-0.16	0.28	-0.36	2.03	-3.32	22.01
T-A	-0.07	-0.27	-0.24	-3.56	-18.18	5.07
T-A	-0.67	-0.16	-0.44	4.72	-12.74	2.55
A-T	0.25	0.01	0.02	-1.23	-8.99	3.99
T-A	-0.23	-0.26	0.22	0.58	-19.87	-2.3
T-A	0.05	-0.27	0.29	-3.18	-12.2	-0.18
T-A	0.07	-0.11	0.26	-5.49	-9.76	2.35
T-A	-0.49	-0.26	0.2	-7.34	-7.53	0.49
A-T	0.36	-0.22	0.34	-4.37	-5.02	-4.12
T-A	0.34	-0.36	0.76	-0.75	-0.96	0.6
G-C	-0.25	-0.31	0.17	-0.08	-6.69	-3.67
T-A	-0.22	-0.34	0.37	-7.58	-10.32	-5.74
T-A	-0.09	-0.37	0.02	-1.77	-12.68	-2.99
T-A	-0.39	-0.34	-0.02	-4.42	-16.17	-0.28
T-A	-0.22	-0.22	-0.18	-6.21	-19.84	4.04
T-A	-0.35	-0.16	-0.17	-9.77	-15.62	3.13
T-A	-0.08	0	-0.21	-4.69	-13.96	7.41
G-C	-0.17	-0.2	0.11	3.43	3.43	1.08
A-T	0.78	-0.24	0.95	11.92	-6.13	3.97
G-C	-0.46	-0.22	-0.04	-6.77	2.2	2.2
T-A	0.51	-0.18	0.05	-1.78	-10.31	2.06
T-A	-0.5	-0.28	-0.34	8.4	-9.38	-2.87
A-T	0.11	-0.17	0.11	-0.09	-8.12	3.51
T-A	-0.28	-0.32	0.27	0.48	-19.03	-2.16
T-A	0.06	-0.3	0.27	-1.77	-11.92	1.27
T-A	0.03	-0.17	0.21	-3.98	-10.31	3.82
T-A	-0.59	-0.34	0.13	-6.18	-8.73	2.89
A-T	0.13	-0.22	0.42	-1.11	-6.84	-2.24
T-A	0.46	-0.31	0.73	-0.21	-0.79	-0.16
G-C	-0.1	-0.2	0.29	-0.04	-7.96	-0.44
T-A	-0.48	-0.19	0.41	-7.57	-10.94	-8.34
T-A	-0.12	-0.14	0.01	0.75	-13.85	-2.44
T-A	-0.61	-0.28	-0.09	-2.44	-19.73	-0.54
T-A	-0.3	-0.1	-0.31	-4.03	-22.11	7.36
T-A	-0.39	-0.07	-0.21	-6.6	-17.92	7.38
T-A	-0.1	-0.25	-0.42	1.29	-11.31	4.42
G-C	0.39	-0.1	-0.41	-1.8	-19.1	-3.65
A-T	0.84	0.11	0.58	5.54	-18.6	-15.25
G-C	-1.3	-1.18	0.2	14.44	-6.21	-38.2
Ave.	-0.11	-0.22	0.10	-1.31	-10.96	-0.1
s.d.	0.41	0.21	0.34	5.19	6.34	8.41

Table 9.16 Local base-pair step parameters for ToxR^{P101A}

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
GT/AC	-0.65	-0.45	3.46	0.5	-3.06	39.6
TT/AA	-0.02	-0.21	3.08	1.62	-2.92	30.38
TA/TA	0.13	0.15	3.44	-3.42	-1.64	45.04
AT/AT	-0.39	-0.74	3.24	-1.78	-3.4	30.91
TT/AA	0.27	-0.26	3.27	-0.25	-4.11	39.94
TT/AA	0.03	-0.51	3.24	1.96	-4.31	33.99
TT/AA	-0.12	-0.35	3.26	2.71	-5.33	34.66
TA/TA	0	0.23	3.14	0.3	5.01	41.08
AT/AT	0.88	-0.64	3.34	-3.63	2.87	31.29
TG/CA	0.14	-1.12	3.03	4.12	1.44	32.62
GT/AC	-0.82	-1.04	3.46	-2.61	3.15	35.62
TT/AA	-0.49	-0.85	2.98	-0.19	-4.23	37.72
TT/AA	-0.32	-0.46	3.25	-1.54	-4.38	34.57
TT/AA	-0.19	0.1	3.24	0.25	-2.79	39.08
TT/AA	0.01	0.32	3.32	0.22	-4.48	38.65
TT/AA	0.22	0.51	3.2	1.44	-1.87	37.6
TG/CA	0.81	1.46	3.26	-1.73	-2.79	35.84
GA/TC	-0.63	0.01	3.23	-6.17	-2.67	38.19
AG/CT	----	----	----	----	----	----
GT/AC	-0.25	0.23	3.32	-0.31	0.06	35.26
TT/AA	-0.33	-0.17	3.03	1.51	-4.59	28.61
TA/TA	0.23	0.11	3.54	-3.58	-1.8	43.77
AT/AT	-0.35	-0.7	3.27	-1.47	-2.7	31.72
TT/AA	0.28	-0.28	3.25	0.01	-3.91	39.78
TT/AA	0.03	-0.54	3.23	2.09	-4.1	33.67
TT/AA	-0.07	-0.36	3.25	2.93	-4.66	34.36
TA/TA	0.06	0.16	3.05	-0.87	4.06	42.18
AT/AT	0.71	-0.69	3.42	-2.95	1.42	33.21
TG/CA	0.3	-1.15	3.03	3.77	0.63	32.74
GT/AC	-0.98	-1.06	3.44	-2.26	2.59	32.46
TT/AA	-0.42	-0.86	2.92	-0.13	-6.1	38
TT/AA	-0.34	-0.48	3.21	-1.58	-4.82	33.48
TT/AA	-0.03	0.13	3.18	1.05	-1.43	40.44
TT/AA	0.05	0.32	3.29	-1.03	-3.41	37.71
TT/AA	0.19	0.3	3.1	2.03	-1.1	35.77
TG/CA	0.26	0.63	3.43	0.47	11.49	35.47
GA/TC	-0.99	-0.19	2.95	-10.55	-3.65	41.58
AG/CT	0.36	-1.01	3.22	2.36	3.16	19.74
AC/GT	0.7	-1.42	2.79	-0.9	-4.38	25.17
Ave.	-0.07	-0.26	3.23	-0.45	-1.47	35.86
s.d.	0.44	0.57	0.15	2.83	3.68	4.76

9.15. DNA crystal packing

Although the DNA has a normal B conformation and is almost straight, it has a peculiarity. We saw the appearance of non-Watson and Crick typical base pairing formed by the G1A, C20B, G1D and C20E base pairs. This is formed at the end of the two DNA molecules and is illustrated in Figure 9.45. In this image it is possible to see how 4 base pairs create hydrogen bonds that allow stabilizing the crystal structure while still having 4 bases in the same plane (Saenger W. *et al.,* 1984).

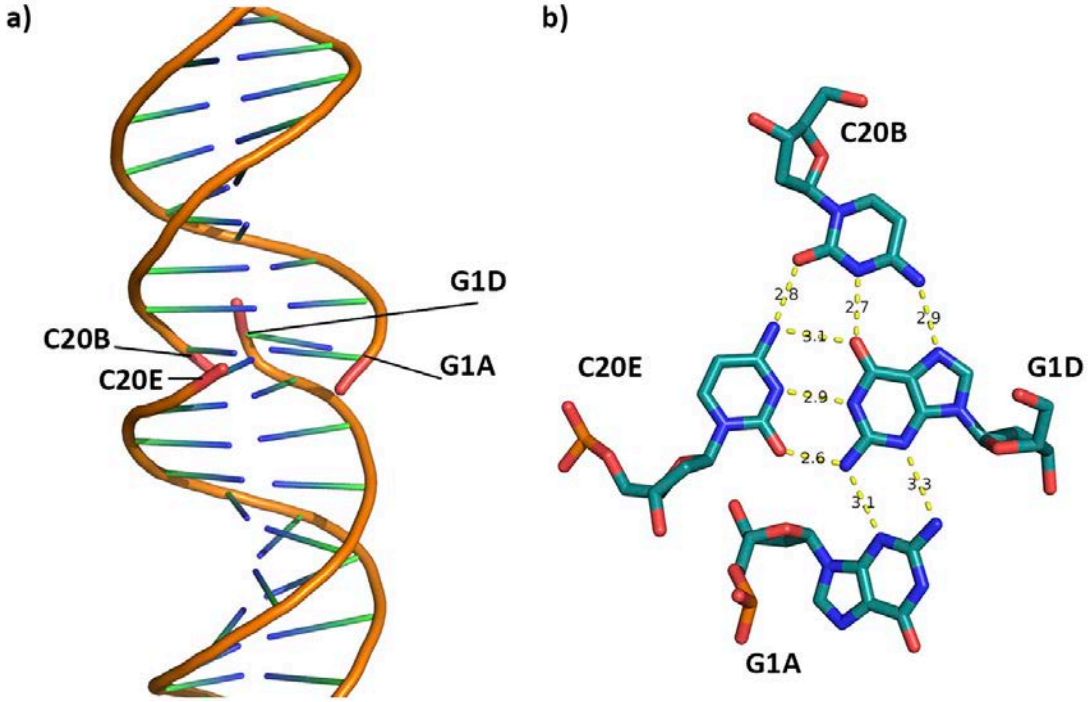


Figure 9.45 DNA crystal structure for ToxR^{S81A} and ToxR^{P101A}

a) Cartoon representation of the full 3-D structure of the two 20 bp *toxT* promoter oligonucleotide molecules inside the ToxR^{S81A} and the ToxR^{P101A} crystals, b) Stick representation of the 4 basepairs that interact of the two 20 bp oligos using hydrogen bonds with the distances labeled in Å.

The C20B is an imino tautomeric form of cytosine and it only occurs 0.1% of cases (Saenger W. *et al.,* 1984). The C20B requires an extra hydrogen atom to be able to create a hydrogen bond with the other base pairs. This is illustrated in Figure 9.46a. A second possibility is that the cytosine is protonated at N3 as shown in Figure 9.46b, this is unlikely to be this species because the pH of the crystal buffer was 7.3.

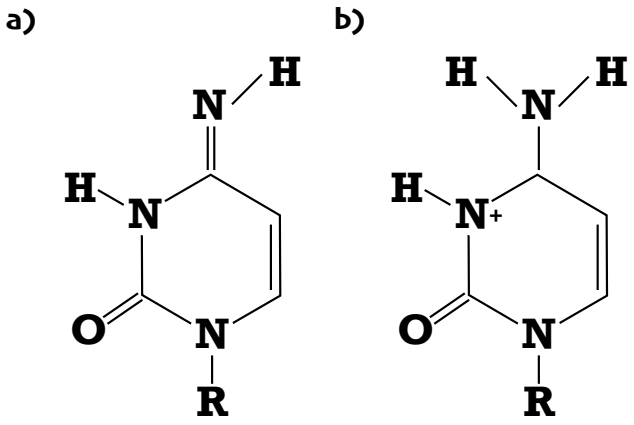


Figure 9.46 Schematic of the atomic structures of the imino tautomeric and protonated forms of cytosine

a) Imino tautomeric form of cytosine, b) Protonated form of cytosine.

This type of DNA stacking creates also an interesting "sandwich stacking" between the base of G20B and the sugars of A19B and the symmetrical of T2D as illustrated in Figure 9.47.

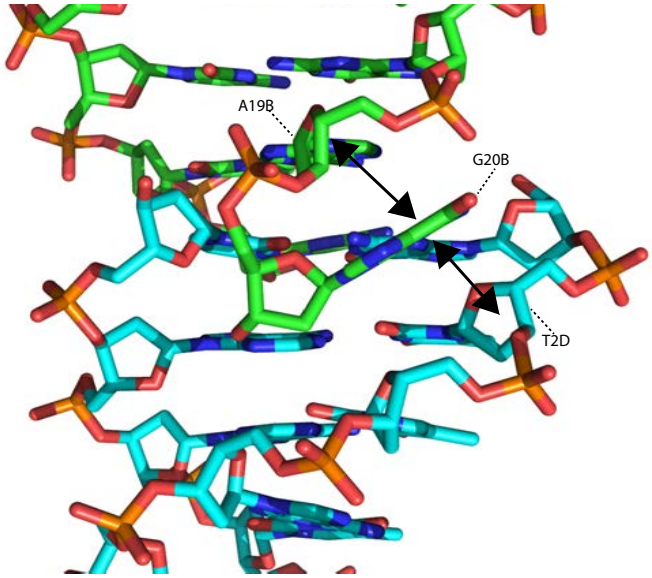


Figure 9.47 DNA sandwich stacking schematic for ToxR^{S81A} and ToxR^{P101A}

9.16. Interactions lost in mutation residues

In this section we will analyze in detail some of the structural differences that appear near the mutated residues.

9.16.1. Interactions lost with ToxR^{Q78A}

There are not many changes in the structure when mutating residue Gln78 into an alanine, yet activation of the promoter is greatly reduced. In Figure 9.48, we illustrate the atomic changes that mutating the residue 78 glutamine into an alanine entail. An interaction between OE1 Gln78 and adenine A13A disappears as well as one between NE2 Gln78 and T6B, leaving a distant interaction with Ala78. The distance between Ala78 and T6B increases from 4.85 Å to 6.18 Å, supposing a considerable change, as there is an interaction mediated by the water and two hydrogen bridges that disappears with the mutation, which are the only specific interactions with the complex, which only contacts the bases and not the sugar-phosphate backbone. Although the distance between Ala78 and A13A varies from 3.15 Å to 5.46 Å, the resulting change could be influencing the minor grooves increase in width. As it creates a peak near the CATA box, where this interaction fades, increasing from 10 Å to 12 Å.

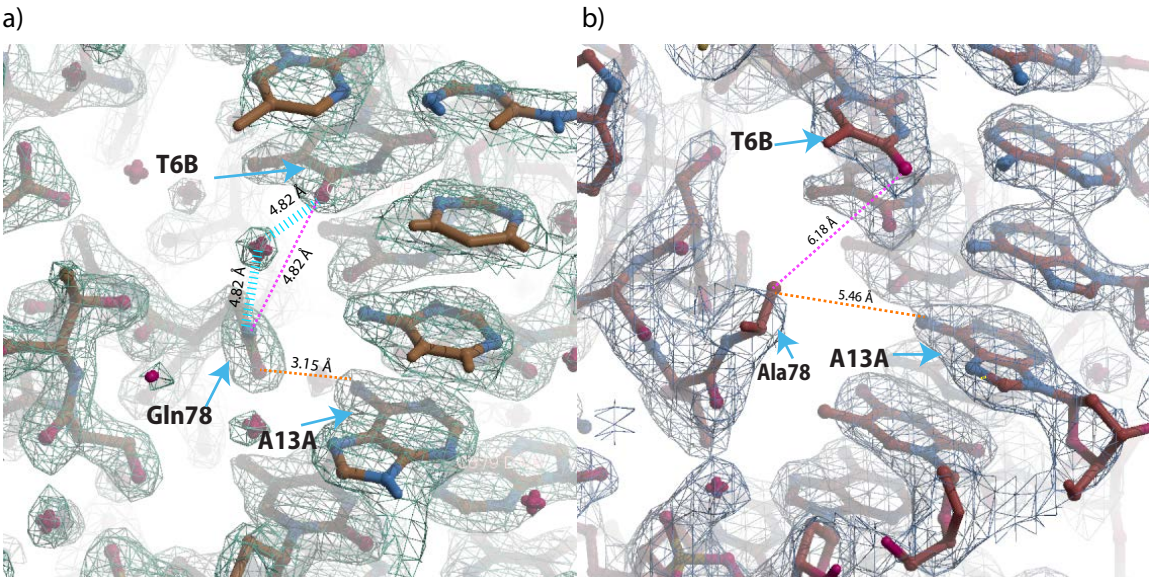


Figure 9.48 Comparative view of the structural changes observed in the ToxR^{Q78A}-DNA structure

a) Left image depicts the atomic interactions between Gln78 with T6B and A13A. b) Right image depicts the atomic interactions that were lost when mutating Gln78 into Ala78.

9.16.2. Interactions lost with ToxR^{S81A}

When mutating Ser81 to alanine, there is a considerable loss in the activation of the promoter, around 90%. In Figure 9.49, the atomic changes that arise from this mutation are illustrated; the interaction between Ser81 and T12A is lost and the distance between 12DT_A and Ala81 increases from 3.77 Å to 5.45 Å.

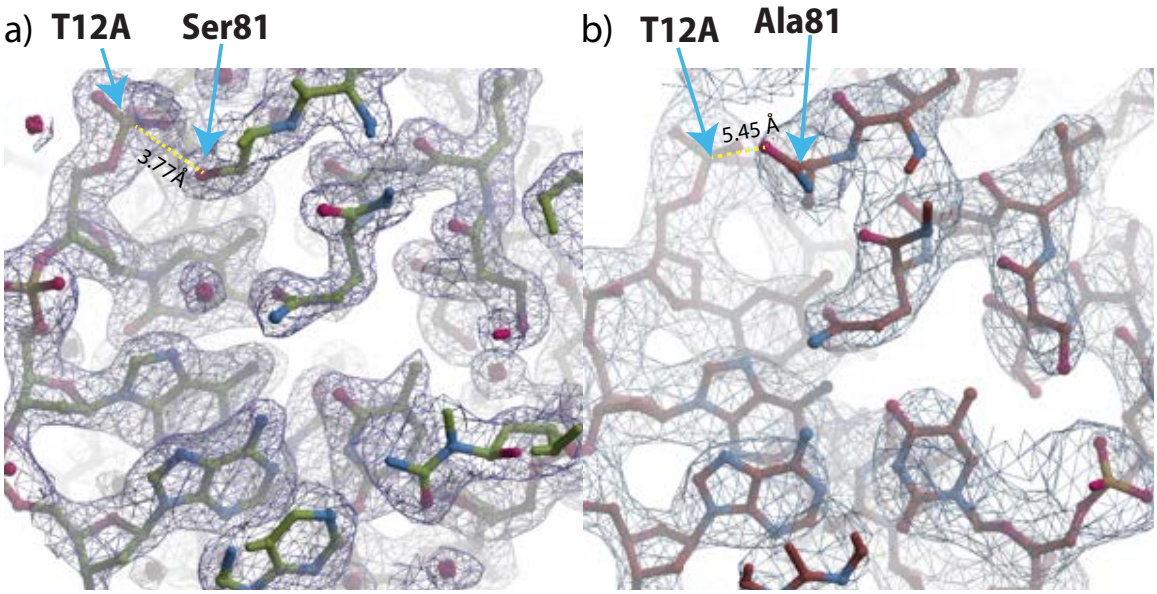


Figure 9.49 Comparative view of the structural changes observed in the ToxR^{S81A}-DNA structure

Left image depicts the atomic interactions between Ser81 with T12A. Right image depicts the atomic interactions that were lost when mutating Ser81 into Ala81.

9.16.3. Interactions lost with ToxR^{P101A}

By comparing the ToxR^{P101A}-DNA structure with that of the wildtype ToxR-DNA we notice a couple of significant conformational changes.

In Figure 9.50, we can appreciate a change in the ToxR^{P101A}-DNA structure, a conformational change in the α2-α3 loop, which might be a result of the different crystal packing as it is present in both molecules. The rest of the structure remains unchanged, yet approximately 75% of the promoter activation is lost.

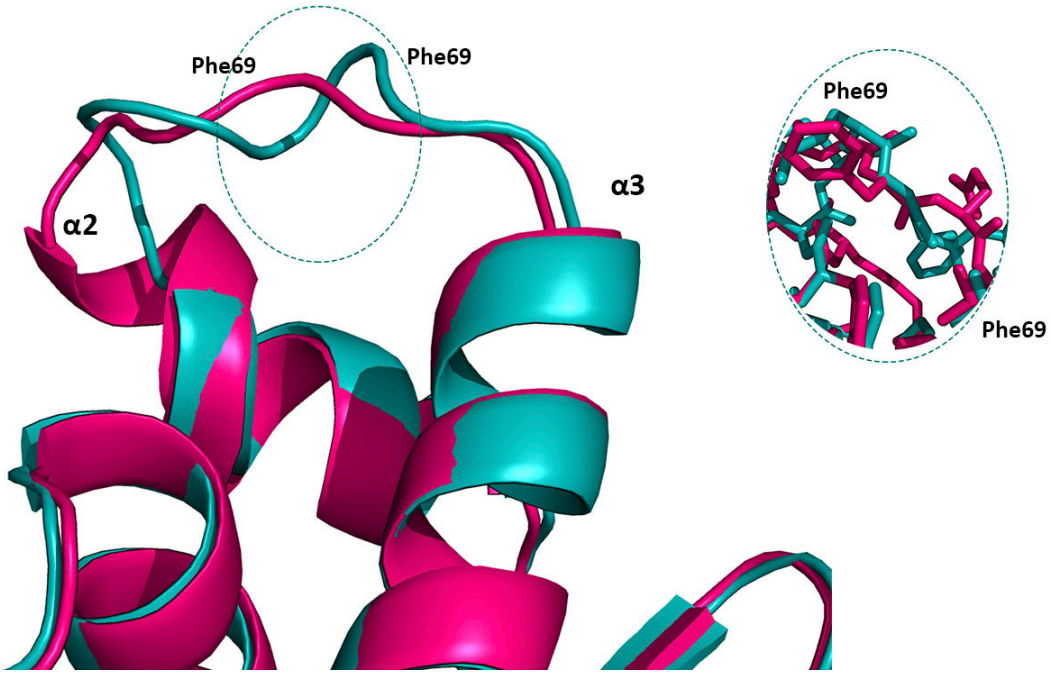


Figure 9.50 Cartoon representation of the structural modifications observed in the ToxR^{P101A} structure at the α2-α3 loop

Wildtype is shown in magenta and ToxR^{P101A} in teal. An outline of the differences using stick representation is also shown.

In this case, In Figure 9.51 we illustrate the hydrophobic interactions between Pro101 ring C and the sugar-CH_2 of both DNA chains. Pro101 is part of the wing and in the wildtype it enters and is "sandwiched" in the minor groove between sugars of the 20 bp segment of the promoter. By mutating the proline into an alanine, this interaction is completely lost. A bulkier hydrophobic C ring that interacts with the sugar phosphate backbone disappears and only an alanine remains, and although punctual atomic distances stay the same, there is a shift in the distances between grooves as shown in Figure 9.51c-d. The loss of the "sandwiching" of the proline results in the widenin of the groove when the hydrophobic interactions disappear leaving only a methyl group for a sidechain.

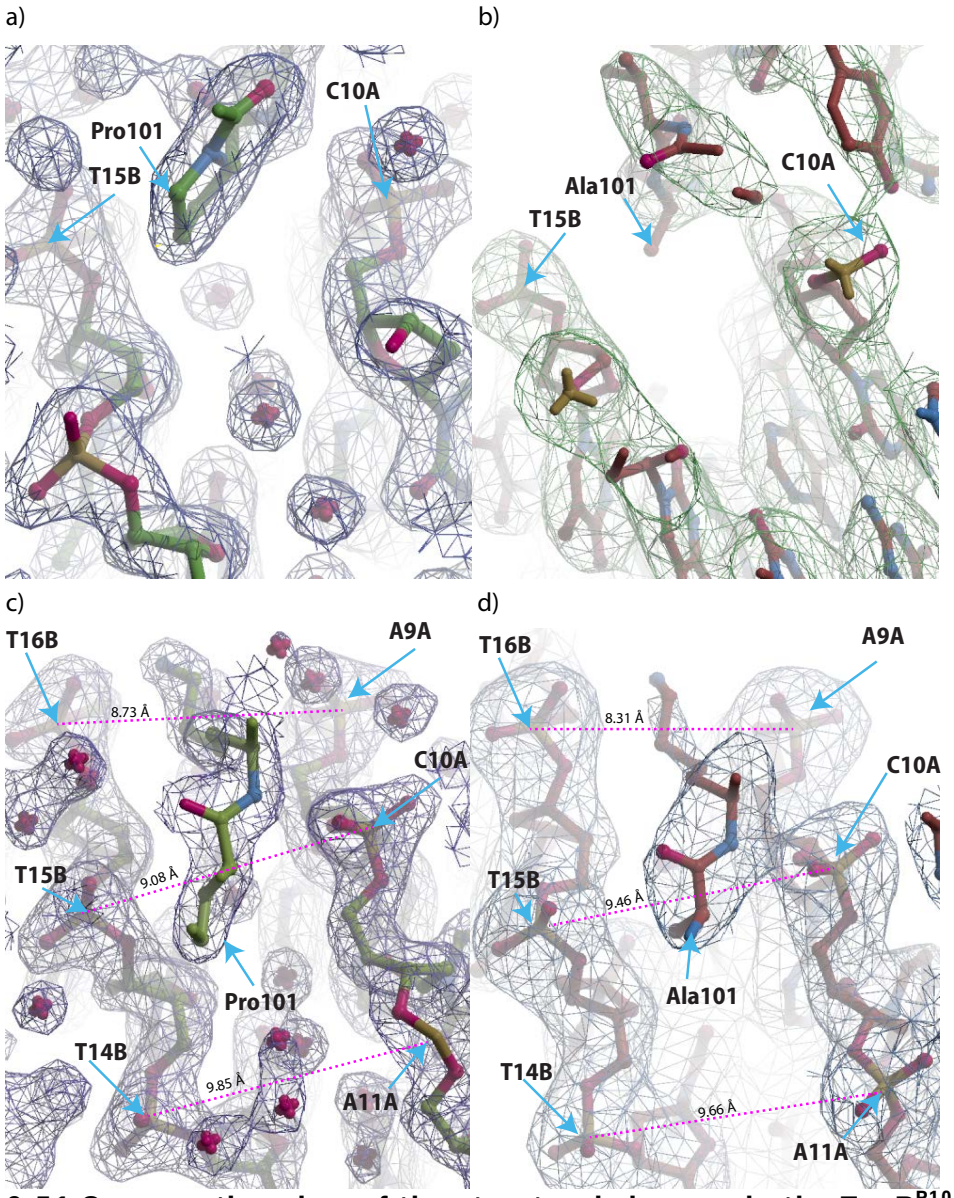


Figure 9.51 Comparative view of the structural changes in the ToxR^{P101A}-DNA structure

a) Top left image depicts the hydrophobic atomic interactions between Pro101 with C10A and T15B. b) Top right image depicts the atomic interactions that were lost when mutating Pro101 into Ala101. C) Bottom left image depicts the atomic distances between T16B-A9A, T15B-C10A and T14B-A11A in the wildtype structure. b) Bottom right image depicts the atomic distances between T16B-A9A, T15B-C10A and T14B-A11A in the P101A structure.

Although the atomic distance between the alanine and T15B and C10A does not radically change in the contact that remains, it is important to consider that an hydrophobic interaction completely disappears, which

could be the responsible for the peak in the minor groove that appears in that section at the CATA box that we mentioned earlier. Specially considering that the shift in the minor groove goes from 10 Å in the $ToxR^{Wildtype}$ structure to 13 Å in the $ToxR^{Wildtype}$ structure.

9. ToxR Conclusions

The conclusions of this section of this thesis are listed as follows:

1) We have solved and refined the crystal structures of three mutants complexed with 20 bp oligonucleotides of the promoter sequence; these mutants reduce the activation of the *toxT* promoter, and solving the structures gives us a better insight of how these mutations structurally affect the ToxR-*toxT* promoter complex.

2) All three mutants crystallize in the form of a protein-DNA complex, indicating that even thought that the activation is decreased, this mutations do not impair the binding in the high concentration conditions of crystallization drops. Even though some of these contacts disappear, the binding specificity to the sequence remains yet DNA shows a number of changes in its structure when crystallizing with these mutants.

3) The Patterson map clearly indicates that there is Translational Non Crystallographic Pseudo Symmetry in the Q78A crystal, reason why the R values will not decrease below 0.30, even if the model fits nicely with the density map. In this structure, we observed the loss of water mediated hydrogen bonds between the protein and the DNA 20 bp segment of the promoter and a direct hydrogen between the protein and a DNA adenine which ultimately result in an important de-activation of the promoter.

4) The S81A ToxR mutation almost completely reduces the activation of the *toxT* promoter. In this situation, there is a loss of hydrogen

bonds to the phosphate backbone of the DNA 20 bp promoter segment.

5) P101A ToxR mutation reduces the activation of the *toxT* promoter around 75%. P101A is part of the wing of ToxR and the fact that the mutation of this residue results in the inactivation of the promoter and that we can observe structural changes in the structure confirms that this motif is required for the activation of the *toxT* promoter as stated by our collaborator (Krukonis *et al*., 2019). The exchange of a proline to an alanine results in the loss of the “sandwiching” hydrophobic interactions between the proline and the sugar phosphate backbones at the DNA minor groove.

6) We observe that ToxR is capable of recognizing the CATA box, but can also bind with similar “boxes” as the $ToxR^{P101A}$ crystals present two molecules of protein bound to one of DNA, and the latter one is bound to an “ATAA” sequence.

7) $ToxR^{Q78A}$, $ToxR^{S81A}$ and $ToxR^{P101A}$ mutations, result in an increase in width of the minor groove. This could also decrease the affinity of the promoter for ToxR, resulting in the impossibility of recruiting the polymerase, which would translate into a decrease in promoter activation.

8) All together, even if the mutated proteins are able to bind to the correct sequence and mode of binding, affinity must be decreased, implying the existence of an unstable complex in physiological conditions, thus hindering the recruitment of the RNA polymerase for gene transcription.

10. Bibliography

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221. doi:10.1107/S0907444909052925

Ali, M., Lopez, A. L., You, Y. A., Kim, Y. E., Sah, B., Maskery, B., & Clemens, J. (2012). The global burden of cholera. *Bull World Health Organ, 90*(3), 209-218A. doi:10.2471/BLT.11.093427

Ali, M., Nelson, A. R., Lopez, A. L., & Sack, D. A. (2015). Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis, 9*(6), e0003832. doi:10.1371/journal.pntd.0003832

Almagro-Moreno, S., Root, M. Z., & Taylor, R. K. (2015). Role of ToxS in the proteolytic cascade of virulence regulator ToxR in Vibrio cholerae. *Mol Microbiol, 98*(5), 963-976. doi:10.1111/mmi.13170

Amin Marashi, S. M., Rajabnia, R., Imani Fooladi, A. A., Hojati, Z., Moghim, S., & Nasr Esfahani, B. (2013). Determination of ctxAB expression in Vibrio cholerae Classical and El Tor strains using Real-Time PCR. *Int J Mol Cell Med, 2*(1), 9-13.

Aminov, R. I. (2010). A brief history of the antibiotic era: lessons learned and challenges for the future. *Front Microbiol, 1*, 134. doi:10.3389/fmicb.2010.00134

Anthony J.F. Griffiths, U. o. B. C., Susan R. Wessler, U. o. C. R., Sean B. Carroll, H. H. M. I. U. o. W.-M., & John Doebley, U. o. W.-M. (2015). *Introduction to genetic analysis*: Eleventh edition. New York, NY : W.H. Freeman & Company.

Appelbaum, P. C. (1992). Antimicrobial resistance in Streptococcus pneumoniae: an overview. *Clin Infect Dis, 15*(1), 77-83. doi:10.1093/clinids/15.1.77

Argos, P., & Rossman, M. G. (1980). Molecular Replacement Method. In M. F. C. Ladd & R. A. Palmer (Eds.), *Theory and Practice of Direct Methods in Crystallography* (pp. 361-417). Boston, MA: Springer US.

Averhoff, B. (2004). DNA transport and natural transformation in mesophilic and thermophilic bacteria. *J Bioenerg Biomembr, 36*(1), 25-33.

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

Barrett, J. F., & Hoch, J. A. (1998). Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob Agents Chemother, 42*(7), 1529-1536.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., . . . Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res, 32*(Database issue), D138-141. doi:10.1093/nar/gkh121

Bauman, R. W., & Machunis-Masuoka, E. (2011). *Microbiology : with diseases by taxonomy*. San Francisco: Benjamin Cummings.

Beck, N. A., Krukonis, E. S., & DiRita, V. J. (2004). TcpH influences virulence gene expression in Vibrio cholerae by inhibiting degradation of the transcription activator TcpP. *J Bacteriol, 186*(24), 8309-8316. doi:10.1128/JB.186.24.8309-8316.2004

Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nat Struct Biol, 10*(12), 980. doi:10.1038/nsb1203-980

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., . . . Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res, 35*(6), e45. doi:10.1093/nar/gkm047

Bosch, F., & Rosich, L. (2008). The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology, 82*(3), 171-179. doi:10.1159/000149583

Boudes, M., Sanchez, D., Graille, M., van Tilbeurgh, H., Durand, D., & Quevillon-Cheruel, S. (2014). Structural insights into the dimerization of the response regulator ComE from Streptococcus pneumoniae. *Nucleic Acids Res, 42*(8), 5302-5313. doi:10.1093/nar/gku110

Brown, A. C., & Valiere, A. (2004). Probiotics and medical nutrition therapy. *Nutr Clin Care, 7*(2), 56-68.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res, 47*(W1), W402-W407. doi:10.1093/nar/gkz297

Buchner, S., Schlundt, A., Lassak, J., Sattler, M., & Jung, K. (2015). Structural and Functional Analysis of the Signal-Transducing Linker in the pH-Responsive One-Component System CadC of Escherichia coli. *J Mol Biol, 427*(15), 2548-2561. doi:10.1016/j.jmb.2015.05.001

Busenlehner, L. S., Pennella, M. A., & Giedroc, D. P. (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev, 27*(2-3), 131-143. doi:10.1016/S0168-6445(03)00054-8

Cashin, P., Goldsack, L., Hall, D., & O'Toole, R. (2006). Contrasting signal transduction mechanisms in bacterial and eukaryotic gene transcription. *FEMS Microbiol Lett, 261*(2), 155-164. doi:10.1111/j.1574-6968.2006.00295.x

Cassini, A., Högberg, L. D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G. S., . . . Group, B. o. A. C. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis, 19*(1), 56-66. doi:10.1016/S1473-3099(18)30605-4

CDC. (2017). *Streptococcus pneumoniae*. Retrieved from https://www.cdc.gov/pneumococcal/clinicians/streptococcus-pneumoniae.html

Chen, I., Christie, P. J., & Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science, 310*(5753), 1456-1460. doi:10.1126/science.1114021

Chen, I., & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nature Reviews Microbiology, 2*(3), 241-249. doi:10.1038/nrmicro844

Chen, J., & Xie, J. (2011). Role and regulation of bacterial LuxR-like regulators. *J Cell Biochem, 112*(10), 2694-2702. doi:10.1002/jcb.23219

Cherney, L. T., Cherney, M. M., Garen, C. R., Lu, G. J., & James, M. N. (2008). Crystal structure of the arginine repressor protein in complex with the DNA operator from Mycobacterium tuberculosis. *J Mol Biol, 384*(5), 1330-1340. doi:10.1016/j.jmb.2008.10.015

Clardy, J., Fischbach, M. A., & Currie, C. R. (2009). The natural history of antibiotics. *Curr Biol, 19*(11), R437-441. doi:10.1016/j.cub.2009.04.001

Claverys, J. P., & Havarstein, L. S. (2002). Extracellular-peptide control of competence for genetic transformation in Streptococcus pneumoniae. *Front Biosci, 7*, d1798-1814. doi:10.2741/claverys

Claverys, J. P., Martin, B., & Polard, P. (2009). The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol Rev, 33*(3), 643-656. doi:10.1111/j.1574-6976.2009.00164.x

Crawford, J. A., Krukonis, E. S., & DiRita, V. J. (2003). Membrane localization of the ToxR winged-helix domain is required for TcpP-mediated virulence gene activation in Vibrio cholerae. *Mol Microbiol, 47*(5), 1459-1473. doi:10.1046/j.1365-2958.2003.03398.x

Cserzo, M., Eisenhaber, F., Eisenhaber, B., & Simon, I. (2004). TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics, 20*(1), 136-137. doi:10.1093/bioinformatics/btg394

Cuthbertson, L., & Nodwell, J. R. (2013). The TetR family of regulators. *Microbiol Mol Biol Rev, 77*(3), 440-475. doi:10.1128/MMBR.00018-13

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography, 40*(1), 82-92.

DiRita, V. J., & Mekalanos, J. J. (1991). Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation. *Cell, 64*(1), 29-37. doi:10.1016/0092-8674(91)90206-e

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., . . . Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature, 477*, 457. doi:10.1038/nature10388

Egli, M. (2016). Diffraction Techniques in Structural Biology. *Curr Protoc Nucleic Acid Chem, 65*, 7.13.11-17.13.41. doi:10.1002/cpnc.4

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501. doi:10.1107/S0907444910007493

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571-607). Totowa, NJ: Humana Press.

Gest, H. (2004). The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec R Soc Lond, 58*(2), 187-201. doi:10.1098/rsnr.2004.0055

Gill, D. M., & Meren, R. (1978). ADP-ribosylation of membrane proteins catalyzed by cholera toxin: basis of the activation of adenylate cyclase. *Proc Natl Acad Sci U S A, 75*(7), 3050-3054. doi:10.1073/pnas.75.7.3050

Goss, T. J., Morgan, S. J., French, E. L., & Krukonis, E. S. (2013). ToxR recognizes a direct repeat element in the toxT, ompU, ompT, and ctxA

promoters of Vibrio cholerae to regulate transcription. *Infect Immun, 81*(3), 884-895. doi:10.1128/IAI.00889-12

Goulian, M. (2010). Two-component signaling circuit structure and properties. *Curr Opin Microbiol, 13*(2), 184-189. doi:10.1016/j.mib.2010.01.009

Haas, B. L., Matson, J. S., DiRita, V. J., & Biteen, J. S. (2014). Imaging live cells at the nanometer-scale with single-molecule microscopy: obstacles and achievements in experiment optimization for microbiology. *Molecules, 19*(8), 12116-12149. doi:10.3390/molecules190812116

Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T., & Calderwood, S. B. (2012). Cholera. *Lancet, 379*(9835), 2466-2476. doi:10.1016/S0140-6736(12)60436-X

Herrington, D. A., Hall, R. H., Losonsky, G., Mekalanos, J. J., Taylor, R. K., & Levine, M. M. (1988). Toxin, toxin-coregulated pili, and the toxR regulon are essential for Vibrio cholerae pathogenesis in humans. *J Exp Med, 168*(4), 1487-1492. doi:10.1084/jem.168.4.1487

Hilliard, J. J., Goldschmidt, R. M., Licata, L., Baum, E. Z., & Bush, K. (1999). Multiple mechanisms of action for inhibitors of histidine protein kinases from bacterial two-component systems. *Antimicrob Agents Chemother, 43*(7), 1693-1699.

Hobman, J. L., Wilkie, J., & Brown, N. L. (2005). A design for life: prokaryotic metal-binding MerR family regulators. *Biometals, 18*(4), 429-436. doi:10.1007/s10534-005-3717-7

Hoch, J. A. (2000). Two-component and phosphorelay signal transduction. *Curr Opin Microbiol, 3*(2), 165-170.

Howard-Jones, N. (1984). Robert Koch and the cholera vibrio: a centenary. *Br Med J (Clin Res Ed), 288*(6414), 379-381. doi:10.1136/bmj.288.6414.379

Huillet, E., Velge, P., Vallaeys, T., & Pardon, P. (2006). LadR, a new PadR-related transcriptional regulator from Listeria monocytogenes,

negatively regulates the expression of the multidrug efflux pump MdrL. *FEMS Microbiol Lett, 254*(1), 87-94. doi:10.1111/j.1574-6968.2005.00014.x

Häse, C. C., & Mekalanos, J. J. (1998). TcpP protein is a positive regulator of virulence gene expression in Vibrio cholerae. *Proc Natl Acad Sci U S A, 95*(2), 730-734. doi:10.1073/pnas.95.2.730

Jernberg, C., Löfmark, S., Edlund, C., & Jansson, J. K. (2010). Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology, 156*(Pt 11), 3216-3223. doi:10.1099/mic.0.040618-0

Johnsborg, O., Eldholm, V., & Håvarstein, L. S. (2007). Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol, 158*(10), 767-778. doi:10.1016/j.resmic.2007.09.004

Johnson, A., Alberts, B., Lewis, J., & al., e. (2002). Molecular Biology of the Cell. In G. Science (Ed.), *Introduction to Pathogens.*

Johnson, S. (2006). *The Ghost Map: The Story of London's Most Terrifying Epidemic--and how it Changed Science, Cities, and the Modern World*: Riverhead Books.

Kazi, M. I., Conrado, A. R., Mey, A. R., Payne, S. M., & Davies, B. W. (2016). ToxR Antagonizes H-NS Regulation of Horizontally Acquired Genes to Drive Host Colonization. *PLoS Pathog, 12*(4), e1005570. doi:10.1371/journal.ppat.1005570

Khurshid, S., Saridakis, E., Govada, L., & Chayen, N. E. (2014). Porous nucleating agents for protein crystallization. *Nat Protoc, 9*(7), 1621-1633. doi:10.1038/nprot.2014.109

Kirn, T. J., Lafferty, M. J., Sandoe, C. M., & Taylor, R. K. (2000). Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by Vibrio cholerae. *Mol Microbiol, 35*(4), 896-910. doi:10.1046/j.1365-2958.2000.01764.x

Kohanski, M. A., Dwyer, D. J., & Collins, J. J. (2010). How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol, 8*(6), 423-435. doi:10.1038/nrmicro2333

Kolb, A., Busby, S., Buc, H., Garges, S., & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem, 62*, 749-795. doi:10.1146/annurev.bi.62.070193.003533

Koonin, E. V., Makarova, K. S., & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol, 55*, 709-742. doi:10.1146/annurev.micro.55.1.709

Koropatkin, N. M., Cameron, E. A., & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol, 10*(5), 323-335. doi:10.1038/nrmicro2746

Kovacikova, G., & Skorupski, K. (2002). Regulation of virulence gene expression in Vibrio cholerae by quorum sensing: HapR functions at the aphA promoter. *Mol Microbiol, 46*(4), 1135-1147. doi:10.1046/j.1365-2958.2002.03229.x

Krukonis, E. S., & DiRita, V. J. (2003). DNA binding and ToxR responsiveness by the wing domain of TcpP, an activator of virulence gene expression in Vibrio cholerae. *Mol Cell, 12*(1), 157-165.

Krukonis, E. S., Yu, R. R., & Dirita, V. J. (2000). The Vibrio cholerae ToxR/TcpP/ToxT virulence cascade: distinct roles for two membrane-localized transcriptional activators on a single promoter. *Mol Microbiol, 38*(1), 67-84. doi:10.1046/j.1365-2958.2000.02111.x

Krüger, N. J., & Stingl, K. (2011). Two steps away from novelty--principles of bacterial DNA uptake. *Mol Microbiol, 80*(4), 860-867. doi:10.1111/j.1365-2958.2011.07647.x

Körner, H., Sofia, H. J., & Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev, 27*(5), 559-592. doi:10.1016/S0168-6445(03)00066-4

Lakhtakia, R. (2014). The Legacy of Robert Koch: Surmise, search, substantiate. *Sultan Qaboos Univ Med J, 14*(1), e37-41.

Laub, M. T., & Goulian, M. (2007). Specificity in Two-Component Signal Transduction Pathways. *Annual Review of Genetics, 41*(1), 121-145. doi:10.1146/annurev.genet.41.042007.170548

Lederberg, J., & Tatum, E. L. (1946). Gene Recombination in Escherichia Coli. *Nature, 158*(4016), 558-558. doi:10.1038/158558a0

Li, S., Olson, W. K., & Lu, X. J. (2019). Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res, 47*(W1), W26-W34. doi:10.1093/nar/gkz394

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

Lorenz, M. G., & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev, 58*(3), 563-602.

Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res, 25*(24), 4940-4945. doi:10.1093/nar/25.24.4940

Maddocks, S. E., & Oyston, P. C. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology, 154*(Pt 12), 3609-3623. doi:10.1099/mic.0.2008/022772-0

Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M., & Stahl, D. A. (2019). Brock biology of microorganisms.

Martin, R. G., & Rosner, J. L. (2001). The AraC transcriptional activators. *Curr Opin Microbiol, 4*(2), 132-137.

Martínez-Hackert, E., & Stock, A. M. (1997a). Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol, 269*(3), 301-312. doi:10.1006/jmbi.1997.1065

Martínez-Hackert, E., & Stock, A. M. (1997b). The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. *Structure, 5*(1), 109-124.

Matson, J. S., & DiRita, V. J. (2005). Degradation of the membrane-localized virulence activator TcpP by the YaeL protease in Vibrio cholerae. *Proc Natl Acad Sci U S A, 102*(45), 16403-16408. doi:10.1073/pnas.0505818102

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674. doi:10.1107/S0021889807021206

McDonnell, G. E., & McConnell, D. J. (1994). Overproduction, isolation, and DNA-binding characteristics of Xre, the repressor protein from the Bacillus subtilis defective prophage PBSX. *J Bacteriol, 176*(18), 5831-5834. doi:10.1128/jb.176.18.5831-5834.1994

Mey, A. R., Craig, S. A., & Payne, S. M. (2012). Effects of amino acid supplementation on porin expression and ToxR levels in Vibrio cholerae. *Infect Immun, 80*(2), 518-528. doi:10.1128/IAI.05851-11

Miller, V. L., DiRita, V. J., & Mekalanos, J. J. (1989). Identification of toxS, a regulatory gene whose product enhances toxR-mediated activation of the cholera toxin promoter. *J Bacteriol, 171*(3), 1288-1293. doi:10.1128/jb.171.3.1288-1293.1989

Mintz, E. D., & Tauxe, R. V. (2013). Cholera in Africa: a closer look and a time for action. *J Infect Dis, 208 Suppl 1*, S4-7. doi:10.1093/infdis/jit205

Molina-Henares, A. J., Krell, T., Eugenia Guazzaroni, M., Segura, A., & Ramos, J. L. (2006). Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol Rev, 30*(2), 157-186. doi:10.1111/j.1574-6976.2005.00008.x

Morris, J. G. (2011). Cholera--modern pandemic disease of ancient lineage. *Emerg Infect Dis, 17*(11), 2099-2104. doi:10.3201/eid1711.111109

Morris, J. G., Sztein, M. B., Rice, E. W., Nataro, J. P., Losonsky, G. A., Panigrahi, P., . . . Johnson, J. A. (1996). Vibrio cholerae O1 can assume a chlorine-resistant rugose survival form that is virulent for humans. *J Infect Dis, 174*(6), 1364-1368. doi:10.1093/infdis/174.6.1364

Mukandavire, Z., Liao, S., Wang, J., Gaff, H., Smith, D. L., & Morris, J. G. (2011). Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe. *Proc Natl Acad Sci U S A, 108*(21), 8767-8772. doi:10.1073/pnas.1019712108

Munita, J. M., & Arias, C. A. (2016). Mechanisms of Antibiotic Resistance. *Microbiol Spectr, 4*(2). doi:10.1128/microbiolspec.VMBF-0016-2015

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., . . . Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 355-367. doi:10.1107/S0907444911001314

Natividad, J. M., & Verdu, E. F. (2013). Modulation of intestinal barrier by intestinal microbiota: pathological and therapeutic implications. *Pharmacol Res, 69*(1), 42-51. doi:10.1016/j.phrs.2012.10.007

Nikolskaya, A. N., & Galperin, M. Y. (2002). A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res, 30*(11), 2453-2459. doi:10.1093/nar/30.11.2453

Nye, M. B., Pfau, J. D., Skorupski, K., & Taylor, R. K. (2000). Vibrio cholerae H-NS silences virulence gene expression at multiple steps in the ToxR regulatory cascade. *J Bacteriol, 182*(15), 4295-4303. doi:10.1128/jb.182.15.4295-4303.2000

Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature, 405*(6784), 299-304. doi:10.1038/35012500

Okada, A., Gotoh, Y., Watanabe, T., Furuta, E., Yamamoto, K., & Utsumi, R. (2007). Targeting two-component signal transduction: a

novel drug discovery system. *Methods Enzymol, 422*, 386-395. doi:10.1016/S0076-6879(06)22019-6

Ouwerkerk, J. P., de Vos, W. M., & Belzer, C. (2013). Glycobiome: bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol, 27*(1), 25-38. doi:10.1016/j.bpg.2013.03.001

Parkinson, J. S., & E.C., K. (1992). COMMUNICATION MODULES IN BACTERIAL SIGNALING PROTEINS. *Annual Review of Genetics, 26*(1), 71-112. doi:10.1146/annurev.ge.26.120192.000443

Patterson, M. J. (1996). Streptococcus. In S. Baron (Ed.), *Medical Microbiology*. Galveston (TX): University of Texas Medical Branch at Galveston

The University of Texas Medical Branch at Galveston.

Pennella, M. A., & Giedroc, D. P. (2005). Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. *Biometals, 18*(4), 413-428. doi:10.1007/s10534-005-3716-8

Peterson, R., & Packer, N. (2013). Dairy Australia Interim Final Report Phase 3 In (pp. 36).

Pfau, J. D., & Taylor, R. K. (1998). Mutations in toxR and toxS that separate transcriptional activation from DNA binding at the cholera toxin gene promoter. *J Bacteriol, 180*(17), 4724-4733.

Pieretti, S. (2016). *Thesis: toxT promoter recognition by ToxR transcription factor, a co-activator within the Vibrio cholerae virulence cascade.* Universitat de Barcelona, Barcelona, Spain.

Png, C. W., Lindén, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., . . . Florin, T. H. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol, 105*(11), 2420-2428. doi:10.1038/ajg.2010.281

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

Potterton, E., Briggs, P., Turkenburg, M., & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr, 59*(Pt 7), 1131-1137. doi:10.1107/s0907444903008126

Ramos, J. L., Martínez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., . . . Tobes, R. (2005). The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev, 69*(2), 326-356. doi:10.1128/MMBR.69.2.326-356.2005

Rigali, S., Derouaux, A., Giannotta, F., & Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem, 277*(15), 12507-12515. doi:10.1074/jbc.M110968200

Rimsky, S., Zuber, F., Buckle, M., & Buc, H. (2001). A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol, 42*(5), 1311-1323. doi:10.1046/j.1365-2958.2001.02706.x

Rivera-Espinoza, Y., & Gallardo-Navarro, Y. (2010). Non-dairy probiotic products. *Food Microbiol, 27*(1), 1-11. doi:10.1016/j.fm.2008.06.008

Roychoudhury, S., Zielinski, N. A., Ninfa, A. J., Allen, N. E., Jungheim, L. N., Nicas, T. I., & Chakrabarty, A. M. (1993). Inhibitors of two-component signal transduction systems: inhibition of alginate gene activation in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A, 90*(3), 965-969. doi:10.1073/pnas.90.3.965

Saenger, W. (1984). *Principles of nucleic acid structure*: Springer-Verlag.

Sagendorf, J. M., Berman, H. M., & Rohs, R. (2017). DNAproDB: an interactive tool for structural analysis of DNA–protein complexes. *Nucleic Acids Research, 45*(W1), W89-W97. doi:10.1093/nar/gkx272

Sanchez, D., Boudes, M., van Tilbeurgh, H., Durand, D., & Quevillon-Cheruel, S. (2015). Modeling the ComD/ComE/comcde interaction network using small angle X-ray scattering. *FEBS J, 282*(8), 1538-1553. doi:10.1111/febs.13240

Schreiter, E. R., & Drennan, C. L. (2007). Ribbon-helix-helix transcription factors: variations on a theme. *Nat Rev Microbiol, 5*(9), 710-720. doi:10.1038/nrmicro1717

Self, W. T., Grunden, A. M., Hasona, A., & Shanmugam, K. T. (1999). Transcriptional regulation of molybdoenzyme synthesis in Escherichia coli in response to molybdenum: ModE-molybdate, a repressor of the modABCD (molybdate transport) operon is a secondary transcriptional activator for the hyc and nar operons. *Microbiology, 145 (Pt 1)*, 41-55. doi:10.1099/13500872-145-1-41

Solomon, J. M., & Grossman, A. D. (1996). Who's competent and when: regulation of natural genetic competence in bacteria. *Trends in Genetics, 12*(4), 150-155. doi:https://doi.org/10.1016/0168-9525(96)10014-7

Stetefeld, J., McKenna, S. A., & Patel, T. R. (2016). Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys Rev, 8*(4), 409-427. doi:10.1007/s12551-016-0218-6

Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). Two-component signal transduction. *Annu Rev Biochem, 69*, 183-215. doi:10.1146/annurev.biochem.69.1.183

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif, 41*(1), 207-234.

Swint-Kruse, L., & Matthews, K. S. (2009). Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol, 12*(2), 129-137. doi:10.1016/j.mib.2009.01.009

Tang, Y. T., Gao, R., Havranek, J. J., Groisman, E. A., Stock, A. M., & Marshall, G. R. (2012). Inhibition of bacterial virulence: drug-like molecules targeting the Salmonella enterica PhoP response regulator. *Chem Biol Drug Des, 79*(6), 1007-1017. doi:10.1111/j.1747-0285.2012.01362.x

Taylor, G. L. (2010). Introduction to phasing. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 325-338. doi:10.1107/S0907444910006694

Taylor, R. K., Miller, V. L., Furlong, D. B., & Mekalanos, J. J. (1987). Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci U S A, 84*(9), 2833-2837. doi:10.1073/pnas.84.9.2833

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., . . . Fraser, C. M. (2001). Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. *Science, 293*(5529), 498-506. doi:10.1126/science.1061217

Tiwari, S., da Costa, M. P., Almeida, S., Hassan, S. S., Jamal, S. B., Oliveira, A., . . . Azevedo, V. (2014). C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds. *Integr Biol (Camb), 6*(11), 1088-1099. doi:10.1039/c4ib00140k

Tiwari, S., Jamal, S. B., Hassan, S. S., Carvalho, P. V. S. D., Almeida, S., Barh, D., . . . Azevedo, V. (2017). Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview. *Front Microbiol, 8*, 1878. doi:10.3389/fmicb.2017.01878

Ulijasz, A. T., & Weisblum, B. (1999). Dissecting the VanRS signal transduction pathway with specific inhibitors. *J Bacteriol, 181*(2), 627-631.

Ulrich, L. E., Koonin, E. V., & Zhulin, I. B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol, 13*(2), 52-56. doi:10.1016/j.tim.2004.12.006

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 1), 22-25. doi:10.1107/S0907444909042589

Waksman, S. A., & Woodruff, H. B. (1942). Streptothricin, a New Selective Bacteriostatic and Bactericidal Agent, Particularly Active Against Gram-Negative Bacteria. *Proceedings of the Society for*

Experimental Biology and Medicine, 49(2), 207-210. doi:10.3181/00379727-49-13515

Walker, L., Levine, H., & Jucker, M. (2006). Koch's postulates and infectious proteins. *Acta neuropathologica, 112*, 1-4. doi:10.1007/s00401-006-0072-x

Walsh, C. (2003). *Antibiotics: actions, origins, resistance*. Washington: American Society for Microbiology (ASM).

Ween, O., Gaustad, P., & Håvarstein, L. S. (1999). Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. *Mol Microbiol, 33*(4), 817-827. doi:10.1046/j.1365-2958.1999.01528.x

West, A. H., & Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci, 26*(6), 369-376.

Wilkinson, S. P., & Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol, 8*(1), 51-62.

Wilson, L. G. (1987). The Early recognition of streptococci as causes of disease. *Medical History, 31*(4), 403-414. doi:10.1017/S0025727300047268

Wolanin, P. M., Thomason, P. A., & Stock, J. B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol, 3*(10), REVIEWS3013. doi:10.1186/gb-2002-3-10-reviews3013

World Health Organization. (2014). Antimicrobial resistance. *Global report on surveillance 2014*, 257.

World Health Organization. (2018a). Cholera vaccine: WHO position paper, August 2017 - Recommendations. *Vaccine, 36*(24), 3418-3420. doi:10.1016/j.vaccine.2017.09.034

World Health Organization. (2018b). Weekly Epidemiological Record. *Weekly Epidemiological Record, 93*(38), 489-500. (Accession No. 27039411)

Worthington, R. J., Blackledge, M. S., & Melander, C. (2013). Small-molecule inhibition of bacterial two-component systems to combat antibiotic resistance and virulence. *Future Med Chem, 5*(11), 1265-1284. doi:10.4155/fmc.13.58

Wright, G. D. (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology, 5*, 175. doi:10.1038/nrmicro1614

Xu, Q., van Wezel, G. P., Chiu, H. J., Jaroszewski, L., Klock, H. E., Knuth, M. W., . . . Wilson, I. A. (2012). Structure of an MmyB-like regulator from C. aurantiacus, member of a new transcription factor family linked to antibiotic metabolism in actinomycetes. *PLoS One, 7*(7), e41359. doi:10.1371/journal.pone.0041359

Xu, X., Stern, A. M., Liu, Z., Kan, B., & Zhu, J. (2010). Virulence regulator AphB enhances toxR transcription in Vibrio cholerae. *BMC Microbiol, 10*, 3. doi:10.1186/1471-2180-10-3

Yokoyama, K., Ishijima, S. A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., . . . Suzuki, M. (2006). Feast/famine regulatory proteins (FFRPs): Escherichia coli Lrp, AsnC and related archaeal transcription factors. *FEMS Microbiol Rev, 30*(1), 89-108. doi:10.1111/j.1574-6976.2005.00005.x

Zeng, G., Ye, S., & Larson, T. J. (1996). Repressor for the sn-glycerol 3-phosphate regulon of Escherichia coli K-12: primary structure and identification of the DNA-binding domain. *J Bacteriol, 178*(24), 7080-7089. doi:10.1128/jb.178.24.7080-7089.1996

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., . . . Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A, 106*(7), 2365-2370. doi:10.1073/pnas.0812600106

Zinder, N. D., & Lederberg, J. (1952). Genetic exchange in Salmonella. *Journal of bacteriology, 64*(5), 679-699.

Zschiedrich, C. P., Keidel, V., & Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. *J Mol Biol, 428*(19), 3752-3775. doi:10.1016/j.jmb.2016.08.003