UNIVERSITAT DE BARCELONA

toxT promoter recognition by ToxR transcription factor, a co-activator within the *Vibrio cholerae* virulence cascade

Simone Pieretti

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

toxT promoter recognition by ToxR transcription factor, a co-activator within the *Vibrio cholerae* virulence cascade

PhD thesis of Simone Pieretti

Barcelona 2016

UNIVERSITAT DE BARCELONA

Doctoral Programme in Biotechnology

toxT promoter recognition by ToxR transcription factor, a co-activator within the *Vibrio cholerae* virulence cascade

IRB Barcelona, Institute for Research in Biomedicine

IBMB-CSIC, Molecular Biology Institute of Barcelona

PhD student
Simone Pieretti

Thesis directors: **Prof. Miquel Coll Capella**

Dr. Albert Canals Parera

Tutor: **Prof. Josefa Badia Palacín**

Barcelona 2016

This PhD thesis titled: "*toxT* promoter recognition by ToxR transcription factor, a co-activator within the *Vibrio cholerae* virulence cascade" 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 funded by the "La Caixa/IRB Barcelona International PhD Fellowship Programme".

Ai miei genitori,

a mio fratello Luca,

a Raffa.

“The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.”

Stephen Hawking

Index

Abstract

Cholera is an acute diarrheal infection caused by the bacterium *Vibrio cholerae.* An estimated 2.9 million of cases and about 100000 cholera deaths occur annually all over the world. Upon ingestion of the *V. cholerae*, the bacterium travels to small intestine where it colonizes and produces the cholera toxin. Cholera toxin raises intracellular cyclic AMP and leads to chloride secretion and the subsequent secretory diarrhoea. Cholera toxin production is regulated through the master virulence regulator ToxT. Trascriptional activation of *toxT* is activated by membrane-localized ToxR, in association with another membrane protein named TcpP. Both ToxR and TcpP work as a two-component regulatory system merged in single proteins: they receive an external signal through its periplasmic C-terminal domain and bind to the *toxT* promoter by their cytoplasmic N-terminal domains. This project thesis aims at characterizing part of the system by studying ToxR-DNA complexes, since two ToxR molecules are supposed to bind the promoter to recruit TcpP and hence the RNA polymerase for transcription activation. Using X-ray crystallography, we have solved the structure of three complexes of the ToxR DNA binding domain with 20-bp, 40-bp and 25-bp oligonucleotides at 2.0 Å, 2.6 Å and 3.2 Å resolution, respectively. According to the three structures, ToxR is able to bind to an extensive region of the *toxT* promoter that goes from the position -97 to the position -45. Considering an integrated model of the three structures, there are four ToxR molecules binding the *toxT* promoter: two molecules bind the DNA in tandem, one molecule binds the ToxR degenerate box and the last one is binding what is supposed to be the TcpP binding site. The structure determination of the three complexes is important to define with more accuracy the ToxR binding site in the *toxT* promoter. This site is characterized by eleven bases with a high A-T rich region sequence followed by a CATA/CATG/TGTA box, where the last two bases perform direct and specific contacts with the protein. In the three structures, ToxR shows a winged helix-turn-helix (w-HTH) fold. The wing interacts with the minor groove in an A-T rich region sequence while the recognition helix enters in the major groove at the region with a sequence

corresponding to a CATA box. We have compared ToxR with the other w-HTH family proteins and we have found a new structural element, the secondary wing, which displays interactions with the DNA. We have analyzed the protein-DNA contacts in the three structures, and also the protein-protein interactions in the ToxR-DNA40 structure, thus validating the data published on ToxR defective mutations. Finally, we put forward a model of *toxT* promoter activation at molecular level, based on our crystal structures and on what is known in literature and from our collaborators. We propose that ToxR acts as co-activator in the first steps of *toxT* transcription activation at different levels. First, it would capture the DNA and hold it close to the cytoplasmic membrane, since both ToxR and TcpP are membrane proteins. Second, it would play a key-role in relieving H-NS from the *toxT* promoter: H-NS binds the DNA and transcription is repressed, but ToxR is able to replace it in a region that goes from -97 to -45. Third, ToxR would not be recruiting the RNA polymerase directly, but creating the suitable conditions for the action of TcpP. ToxR recruits TcpP probably through a hand-holding mechanism since one of the ToxR binding site is very close to the TcpP's binding site.

List of figures

List of tables

List of abbreviations

Å	Angstrom
AMP	Adenosine monophosphate
bp	Base pairs
CTX	Cholera toxin
CV	Column volumes
Da	Dalton
ddH2O	double-distilled water
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
EDTA	Etylendiaminetetraacetate
GMP	Guanosine monophosphate
H-NS	Histone-like nucleoid structuring protein
µ	micro
M	molar
m	milli
min	Minute
n	nano
N-terminal	Amino-terminal
PAGE	Polyacrylamide gel
PDB	Protein Data Bank
$PhoB^{E}$	PhoB DNA binding effector domain
$PhoB^{R}$	PhoB receiver domain
Py	Pyrimidine
Pu	Purine
RMSD	Root mean square deviation
rpm	Revolutions per minute
s	Seconds
SEC	Size exclusion chromatography

TCP	toxin co-regulated pilus
TM	Transmembrane
ToxR-DBD	ToxR DNA binding domain
UV	Ultraviolet
V	volt
V. cholerae	*Vibrio cholerae*
[v/v]	volume per volume
WHO	World Health Organization
w-HTH	winged helix-turn-helix
[w/v]	weight per volume

Chapter 1: Introduction

1.1 Background

1.1.1 What is cholera?

Cholera is an acute diarrheal infection caused by the ingestion of food or water contaminated with the gram-negative bacterium *Vibrio cholerae* of the O1 or O139 serogroup. The disease is endemic in more than 50 countries and is also responsible for large epidemics worldwide. Patients with cholera are treated with vigorous fluid replacement, which can decrease mortality from more than 50% to less than 0.2%. Antibiotic medication reduces volume and duration of diarrhea by 50%. Safe water and a better sanitation are tow key-factors to prevent cholera (Harris *et al.* 2003). Oral cholera vaccines are increasingly used as an additional tool to control cholera outbreaks, although they are not fully effective (WHO 2010).

1.1.2 History of cholera

Cholera was described for the first time in Sanskrit in the 5th century B.C. and has persisted on the Indian subcontinent for centuries. The disease spread out of the Indian subcontinent in 1817, and between 1817 and 1923 six cholera pandemics occurred all over the world. In 1854, London physician John Snow proposed that cholera was a transmittable disease and that stool contained infectious material that could contaminate drinking water supplies. He devised a dot map to illustrate the cluster of cholera cases around the water pumps and used statistics to illustrate the connection between the quality of the water source and cholera cases (Figure 1) (Johnson 2006).

Figure 1. Colour version of John Snow's map showing the clusters of cholera cases in the London epidemic of 1854 (Johnson 2006).

In 1854, Filippo Pacini found comma-shaped organisms in cholera stools using a microscope. Finally, in 1884, Robert Koch was able to isolate a *V. cholerae* pure culture. From 1961, the seventh pandemic broke out causing 3-5 millon cases each year and killing 120000 people thereafter (Greenough *et al*. 2004; Morris *et al*. 2003; Sack *et al*. 2004).

1.1.3 The global burden of cholera

Most cholera infections are not detected, but in recent years large cholera outbreaks, such as those that occurred in Haiti (Chao *et al.* 2011), Vietnam (WHO 2008) and Zimbabwe (Zimbabwe: OCHA Cholera Update Situation Report No.22 15 Jul 2009) have occurred. It is estimated that around 1.4 billion people are at risk for cholera in endemic countries. 2.8 million cholera cases are found to occur annually in these countries and

about 8700 cholera cases occur in non-endemic countries. The incidence is calculated to be highest in children less than 5 years of age. Every year, around 91000 people die of cholera in endemic countries and 2500 people die of the disease in non-endemic countries (Ali *et al.* 2012). Countries were classified as being cholera endemics, cholera non-endemics, or cholera-free depending on the incident rate over an established threshold in at least three of five years (2008-2012) (Figure 2). The most recent estimation indicated 2.9 million cases and 95000 deaths in 69 endemic countries, with the majority of the cases in Sub-Saharian Africa (Ali *et al.* 2015).

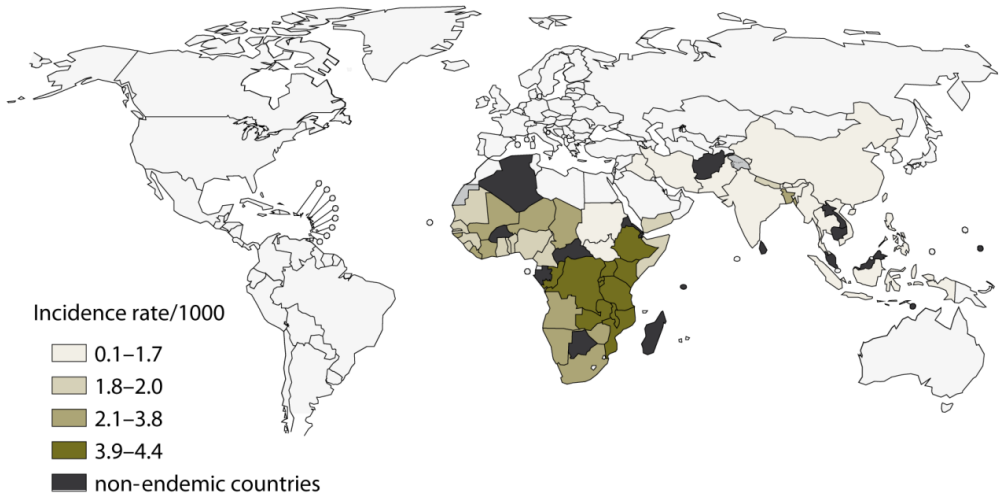

Figure 2. Geographical patterns of the estimated cholera incidences all over the world (Ali M. *et al.* 2012).

1.1.4 The causative agent of cholera: *Vibrio cholerae*

V. cholerae is a Gram-negative, comma-shaped bacterium (Figure 3) , classified into more than 200 serogroups based on the O antigen of the lipopolysaccharide. Among these, only O1 and O139 serogroups are able to cause epidemic cholera. O1 is further

classified into two biotypes, classical and El Tor. Although classical *V. cholerae* O1 caused the fifth and the sixth pandemics, the seventh pandemic is attributed to El Tor biotype, which has currently replaced the classical biotype (Harris *et al*. 2003). *V. cholerae* contains two circular chromosomes: the first one is the largest and houses genes for toxicity and important cellular functions, whereas the second contains essential genes for metabolism (Trucksis *et al.* 1998). Toxigenic *V. cholerae* has a pathogenicity island, which encodes *toxT*, the master virulence regulator. It also encodes genes required for the toxin co-regulated pilus (TCP), which plays a key role in microcolony formation and intestinal colonization (Manning *et al.* 1997). One of the transcription factors required for *toxT* activation, TcpP, is encoded on the *Vibrio* pathogenicity island (VPI-1), while the other, ToxR, is not (Karaolis *et al.* 1998).

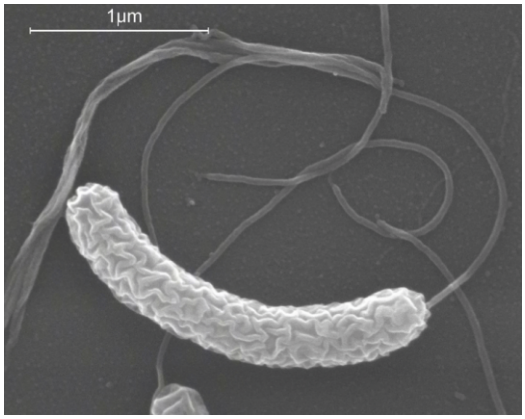

Figure 3. Scanning electron microscope (SEM) of *Vibrio cholerae* bacterium (Dartmouth Electron Microscope Facility).

1.1.5 *Vibrio cholerae*, transition from an aquatic organism to pathogen

V. cholerae cycle moves to and from aquatic and host environments. In aquatic environments such as estuaries, it is possible to find pandemic and non-pandemic strains. *V. cholerae* is able to remain viable about 50 days in a suitable water

environment characterized by low salt concentration and a warm temperature (Harris *et al*. 2003). In freshwater, an increased concentration of organic nutrient and divalent cations can compensate the lack of salt. If the water environment is not appropriate for *V. cholera,* the bacterium can survive passing into a viable but non-culturable stage that keeps virulence if ingested (Colwell *et al.* 1996). In aquatic environments, *V. cholerae* is mainly found on biofilms in contact with plankton and copepods since they are chitin-rich organisms (Reidl *et al.* 2002). If *V. cholerae* is ingested it moves to the small intestine and can colonize it. The colonization can occur after TCP expression, thus allowing the formation of *V. cholerae* microcolonies (Taylor *et al*. 1987).

1.1.6 Pathogenesis and pathophysiology of *V. cholerae*

Upon ingestion of *V. cholerae*, most of the bacteria are killed by gastric acid. However, surviving organisms are then able to colonize the small intestine and produce cholera toxin, which is the major virulence factor for pathogenic strains.

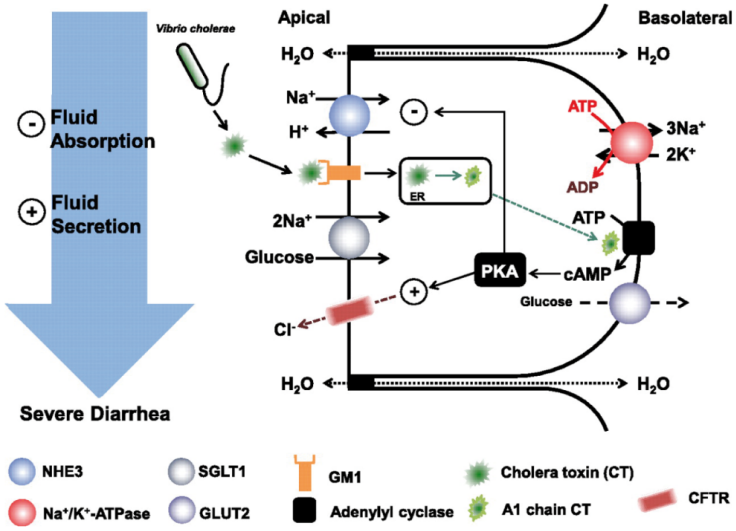

Figure 4. Mechanism of action of cholera toxin (Girardi *et al.* 2012).

Cholera toxin (CT) is an exotoxin protein consisting of one A subunit associated with five B subunits. The B subunit pentamer attaches to the ganglioside GM_1, on eukaryotic cells, and the A subunit is translocated intracellularly, where it acts enzymatically to activate adenylate cyclase and increase intracellular cyclic AMP concentration, resulting in chloride secretion through the apical chloride channel and secretory diarrhoea (Gill 1975; Gill *et al*. 1978) (Figure 4).

1.2 Environmental signalling and virulence cascade in *V. cholerae*

When *V. cholerae* moves from the aquatic environment to the human intestine, it has to survive dramatic changes of temperature, salinity, nutrient availability and other factors. These stimuli have a strong contribution to the virulence cascade regulation since they are essential in the activation of the signaling cascade in different steps. This signaling cascade culminates in ToxR and TcpP activation.

ToxR and TcpP are two membrane proteins that are able to activate *toxT* expression. As mentioned, *toxT* is the master virulence regulator, which is responsible for the expression of toxin co-regulated pilus (TCP) and the cholera toxin (CTX). ToxR and TcpP are co-transcribed with ToxS and TcpH, respectively (Miller *et al.* 1987). ToxS and TcpH are cytoplasmic activators that can modulate the activation of ToxR and TcpP inducing dimerization and enhancing their stability (Beck *et al*. 2004). The transcriptional activator complexes formed by ToxR and ToxS, and by TcpP and TcpH are named ToxRS and TcpPH, respectively. The activities of ToxR and TcpP are also regulated by environmental stimuli. ToxR transcriptional activation activity, and possibly *toxR* transcription, is induced in media that contain the amino acids asparagine, arginine, glutamic acid, and serine (Mey *et al*. 2012). Although ToxRS has been described as a constitutively expressed protein complex (DiRita *et al*. 1996), its expression may also be activated by AphB (Xu *et al.* 2010). AphB is a wing-helix DNA binding protein, which initiates the expression of the virulence cascade by activating the TcpPH operon.

Another factor which has an influence on the virulence cascade is the quorum sensing through several endogenous systems (Rutherford *et al*. 2011; Kovacikova *et al.* 2002). Several studies show that cyclic di-GMP responds to the quorum sensing machinery through repression of HapR and the subsequent activation of AphA. (Srivastava *et al.* 2012). AphA is a winged-helix transcription factor, which enhances activation of TcpPH expression by AphB transcription factor (Kovacikova *et al.* 1999) (Figure 5). Moreover, *tcpP* transcription is activated at low temperature (Skorupski *et al*. 1999). Another transcription factor able to bind to the *tcpP* promoter and induce *tcpP* transcription is PhoBR at low phosphate levels (Pratt J.T. *et al.* 2010). Although many aspects of virulence cascade regulation still remain unclear, ToxR regulon is commonly used to

represent the transcriptional activation and repression cascades that regulate virulence in *V. cholerae* (Figure 5).

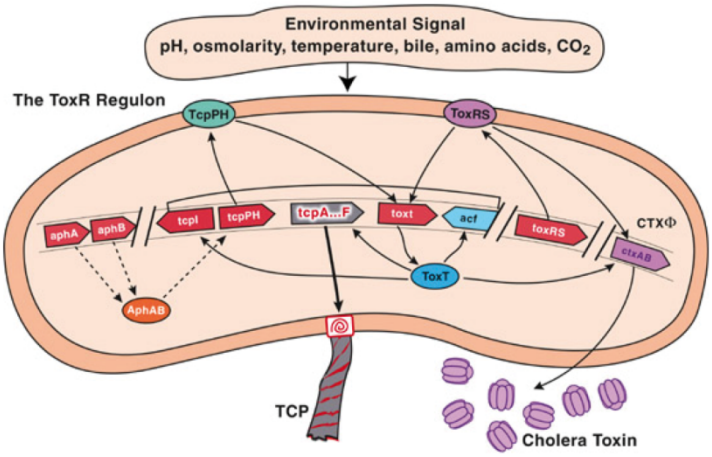

Figure 5. Virulence gene activation in *V. cholerae* (Amin Marashi *et al.* 2013).

1.3 One-component systems vs two-component systems

Bacteria use signal transduction systems to detect environment changes and react in response to them. These signal transduction systems can generally be divided into two groups: two-component systems and one-component systems (Laub *et al.* 2007).

Two-component signal transduction systems are composed of a sensor histidine kinase (HK) and a response regulator (RR) (Stock *et al.* 2000). The HK is able to autophosphorylate an inner domain. This event is followed by the phosphorylation of the receiver domain on the RR and the subsequent activation of the output domain, which can produce the appropriate physiological effect in the cell by controlling the expression of a specific gene (Figure 6).

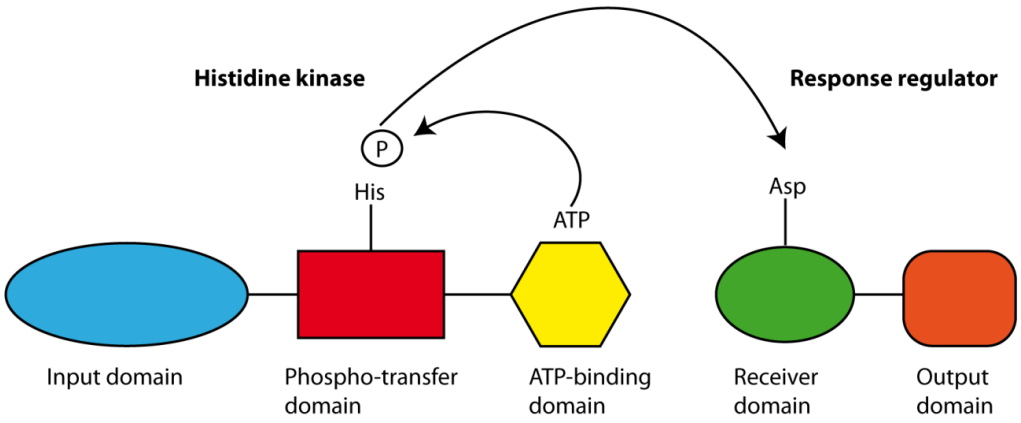

Figure 6. Two-component system scheme. Signal sensing by the input domain causes activation of the autokinase domain, which results in phosphorilation of a specific histidine residue in the phosphotransfer subdomain of the autokinase. The phosphoril group is then transferred to an aspartate residue in the receiver domain of the response regulator resulting in the modultation of a function.

Some HKs can display a dual function: they perform the phosphorylation and dephosphorylation of the RR according to the input sensory domain regulation. There is a subtype of a two-component system that is named the phospho-relay system. This system is characterized by the presence of a hybrid HK with an internal receiver domain and an intermediate histidine phosphotransfer protein, which is able to shuttle the phosphoryl group to the terminal RR (Figure 7). These additional components allow multiple signal integration (Varughese *et al.* 2002; Hoch *et al*. 2001).

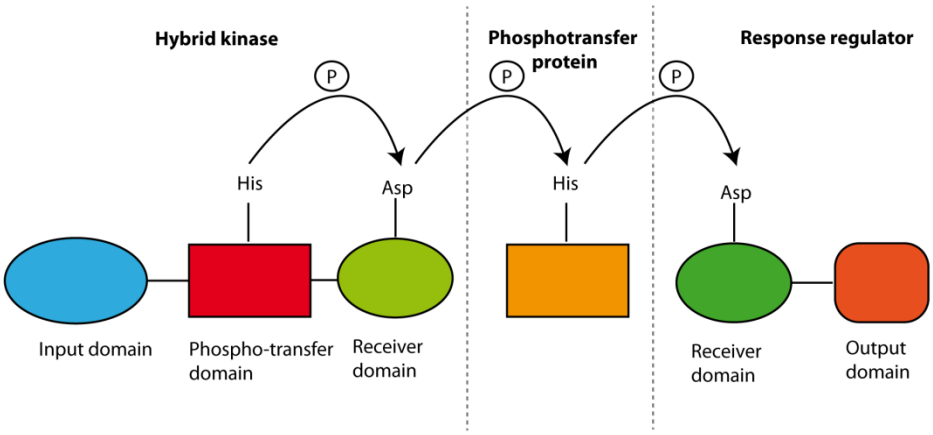

Figure 7. Phospho-relay system scheme. Signal sensing is followed by the auto-phosphoryation of the hybrid kinase, the phosphorylation of an internal receiver domain and the transfer of the phosphoryl group to the phosphotransfer protein, which is able to phosphorylate and activate the response regulator.

In one-component systems, one protein contains both the sensory, or input, domain and the output domain (Figure 8). The term two-component system is better known than the one-component system. However, one-component systems are more abundant in prokaryotes and show a greater diversity of domains than two-component systems (Ulrich *et al.* 2005).

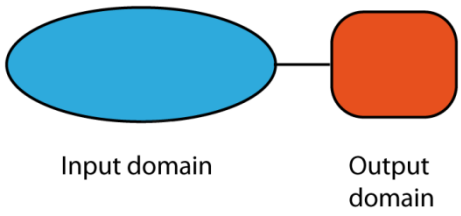

Figure 8. One-component system scheme. Signal sensing by the input domain produces an immediate activation of the output domain without any chemical modification.

There are at least 20 families of prokaryotic one-component systems that can be defined by amino acid conservation in their DNA-binding domains and are defined by different conserved motifs (Table 1) (Cuthbertson *et al*. 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

Table 1. Major families of one-component signal transduction systems (Cuthbertson *et al*. 2013)

The majority of one-component systems use a helix-turn-helix DNA binding domain, including ToxR family transcription factors (Buchner *et al.* 2015). Using the dense alignment surface method (Cserzo *et al*. 2004), it was found that 97% of the one-component regulators that contain an HTH domain do not have transmembrane regions and therefore are predicted to be cytosolic proteins (Ulrich *et al.* 2005). The peculiarity of ToxR family is to have a transmembrane region between the input domain and the output domain. ToxR family transcription factors are characterized by a C-terminal periplasmatic sensing domain, a single transmembrane helix and an N-terminal cytoplasmatic winged-helix-turn-helix DNA-binding domain (Figure 9). Signal transducton is mediated without chemical modification and is a direct form of bacterial stimulus-response mechanism.

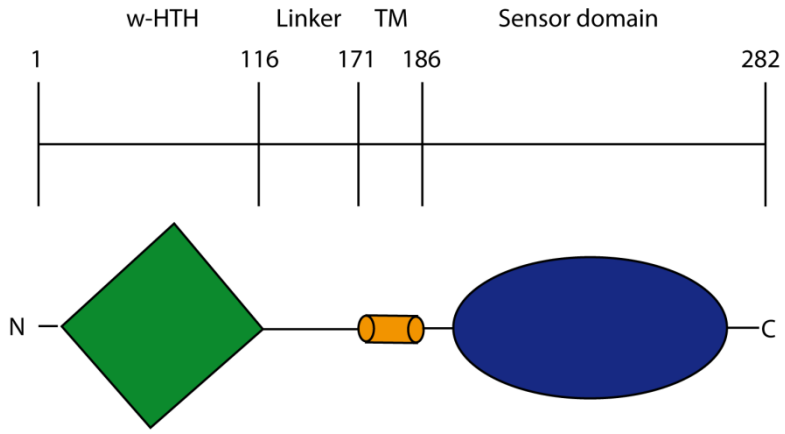

Figure 9. Domain architecture of *Vibrio cholerae* ToxR. It consists of an effector DNA binding domain with a helix-turn-helix motif, a flexible cytoplasmatic linker region, a single transmembrane helix and a C-terminal periplasmatic sensor domain.

1.3.1 PhoB: an example of a two-component system response regulator

PhoB is the response regulator of a two-component signaling transduction system and is able to activate nearly 40 genes in *E. coli* related to phosphate assimilation (Wanner *et*

al. 1996; Kim *et al*. 2000). In our laboratory, we determined the crystal structure of the two domains of PhoB: the receiver domain ($PhoB^R$) (Solà *et al*. 1999; Arribas-Bosacoma *et al.* 2007) and the DNA-binding effector domain ($PhoB^E$) (Blanco *et al*. 2002). The structure of $PhoB^E$ was first determined in the unbound state, in the absence of DNA (Blanco *et al*. 2002). As the cytoplasmatic domain of ToxR, $PhoB^E$ belongs to the OmpR-PhoB family of DNA binding effector domains and to the winged helix subfamily of helix-turn-helix proteins. $PhoB^E$ comprises an N-terminal four-stranded antiparallel β-sheet, three-α helix wrapped together and a C-terminal β-hairpin (Figure 10).

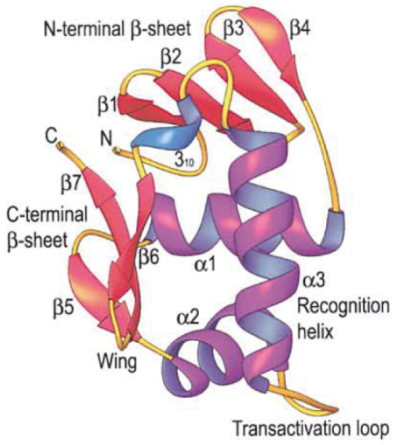

Figure 10. Overall structure of $PhoB^E$ (Blanco *et al.* 2002).

The structure of the $PhoB^E$, in complex with a 23-bp DNA fragment called *pho box*, shows a novel DNA tandem recognition behaviour, where two monomers are able to bind DNA in a head to tail conformation (Figure 11). $PhoB^E$ recognizes the DNA through two recognition elements: helix α3 that enters the major groove and a β-hairpin wing, which interacts with the minor groove through Arg219. This key-amino acid displays contacts with DNA sugar backbones. This structure fully describes the DNA binding of PhoB, but any notions of the σ^{70} activation of the RNA polymerase are not provided.

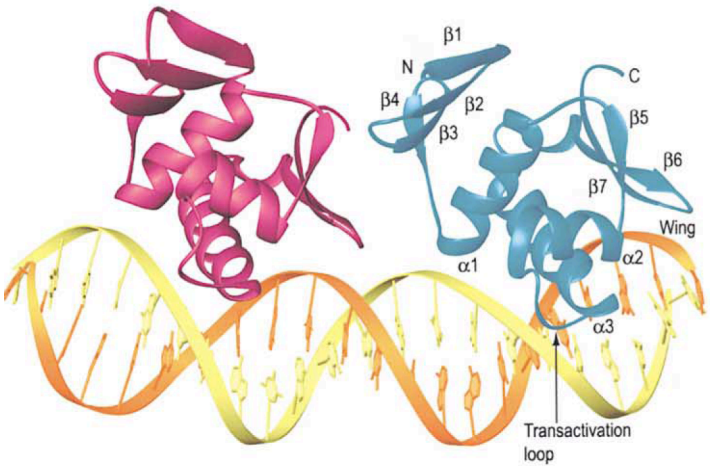

Figure 11. Ribbon representation of PhoB^{E} in complex with DNA (Blanco *et al.* 2002).

Another crystal structure, solved in our laboratory, indicates that PhoB can interact with σ_4 of the σ^{70} subunit of the *E. coli* RNA polymerase. This crystal structure includes PhoB^{E}, its target DNA and the σ_4 domain of the RNA polymerase σ^{70} subunit (Figure 12) (Blanco *et al*. 2012). The crystal structure of this ternary complex shows that RNAP σ^{70} subunit binds the -35 position in the *pho box* promoter when PhoB is bound. The regions used by PhoB^{E} to interact with σ^{70} factor are the transactivation loop and helix α2. On the other hand, the DNA recognition helix of σ_4 performs contacts with the major groove of DNA and PhoB. The interaction between PhoB^{E} and the σ_4 of the RNA polymerase σ^{70} subunit produces an enhancing effect on the RNAP folding. The function of PhoB is not only to recruit the polymerase, but also to remodel the σ_4 domain of σ^{70}. In fact, if we overpose the σ_4 of the holoenzyme with the σ_4 of this structure we can see a different orientation of the σ_4. The reorientation of the σ_4 involves the R3-4 linker and makes the RNA release possible, allowing a correct transcription process. These findings support the idea that the σ_4 domain acts as a versatile platform interacting with different transcription factors via specific amino acid residues. The σ_4 domain undergoes a remodeling process when it interacts with transcription factors or DNA-transcription factor complexes and, depending on this interaction, it can modulate the appropriate response (Canals *et al*. 2012). In conclusion, the PhoB example shows that the structural

characterization of transcriptional complexes can provide a mechanistic understanding of transcription in prokaryotes and represents a suitable model to study the ToxR DNA binding domain.

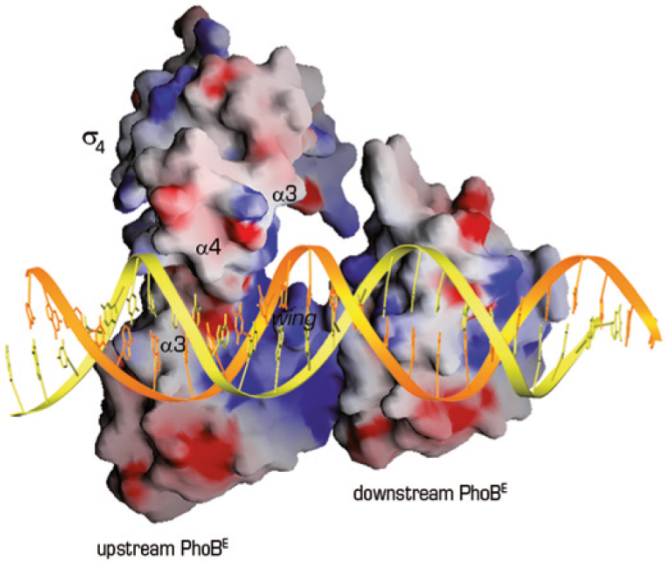

Figure 12. Overall structure of the transcription subcomplex constituted by PhoB^{E}, the σ_4 and the *pho box* promoter region of DNA (Canals *et al*. 2012).

1.4 ToxR and TcpP, two key proteins in the *V. cholerae* virulence cascade

The virulence cascade includes the activation of ToxR and TcpP and the subsequent expression of ToxT. Then, ToxT directly activates the expression of the two major virulence factors: the toxin co-regulated pilus (TCP), an essential intestinal colonization factor encoded on the Vibrio Pathogenicity Island, and the cholera toxin (CTX) encoded on the lysogenic CTX bacteriophage. ToxT expression is regulated by an unusual membrane-associated mechanism: the bitopic proteins ToxR and TcpP cooperate to activate *toxT* transcription while remaining localized to the inner membrane (Crawford *et al*. 2003; Matson *et al*. 2007). They both have an N-terminal cytoplasmatic DNA binding domain with a winged helix-turn-helix motif homologous to the OmpR/PhoB family of transcriptional activators (Martínez-Hackert *et al*. 1997a), a single transmembrane domain, and a C-terminal periplasmatic domain. Both ToxR and TcpP are required for *toxT* activation. TcpP seems to play the primary role binding a region around -40 in the promoter, while ToxR binds to a more distal region around -80 and appears to play an auxiliary role (Krukonis *et al.* 2000). Both TcpP and ToxR function together with membrane bound periplasmic effector proteins, TcpH and ToxS, respectively (DiRita *et al.* 1991; Häse *et al*. 1998). TcpH seems to influence the stability of TcpP by sheltering its periplasmatic domain from proteolytic action (Beck *et al.* 2004). In fact, in the absence of TcpH, TcpP is degraded by two sequential proteolytic events (Matson *et al.* 2005; Makinoshima *et al.* 2006). Since ToxR levels remain unchanged in the absence of ToxS, the functions of the membrane bound effector proteins for TcpP and ToxR do not seem to be equivalent. It has been proposed that ToxS may reinforce the dimerization of ToxR and facilitate transcriptional activation (DiRita *et al*. 1991; Pfau *et al.* 1998) (Figure 13).

An additional player that has to be taken into account is the histone-like nucleoid structuring protein H-NS. Actually, H-NS has the role in silencing the expression of a variety of environmentally regulated genes during growth under nonpermissive conditions and among them, the *toxT* gene. From a molecular point of view, it seems to bind curved DNA, which is commonly found at promoters (Rimsky *et al*. 2001). Upon

induction of TcpPH and ToxRS, these transcription factors bind to the *toxT* promoter, relieving repression by the H-NS, and activating *toxT* transcription (Nye *et al*. 2000).

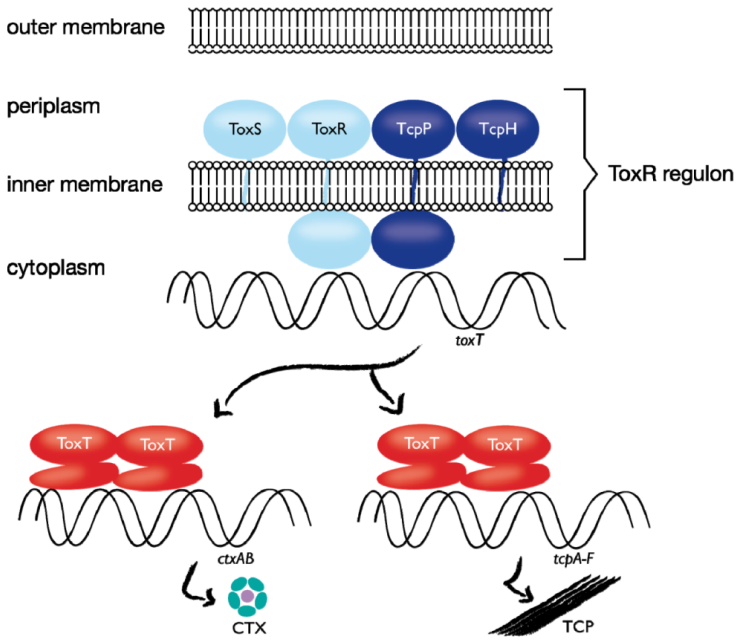

Figure 13. Role of the ToxR regulon in controlling toxin (CTX) expression. ToxR and TcpP bind the *toxT* promoter and activate the transcription of *toxT*. ToxS and TcpP contribute to the stability of ToxR and TcpP, respectively. ToxT activates *ctxAB* and *tcpA-F* transcription. (Haas *et al*. 2014).

1.4.1 Structure and function of ToxR and TcpP

ToxR and TcpP are both transmembrane transcription factors proteins with the cytoplasmic domain belonging to the winged helix-turn-helix (w-HTH) family of transcription factors (Martínez-Hackert and Stock 1997b). In prokaryotes, w-HTH transcription factors are part of a two-component regulatory system in which phosphorilation of an N-terminal regulatory domain of the response regulator protein leads to transcriptional activation of the w-HTH domain (Aravind *et al*. 2005). ToxR and TcpP are not endowed with an N-terminal regulatory domain, but they have a C-

terminal periplasmatic domain, with unknown homology. As mentioned, the periplasmic domain of TcpP regulates stability by interaction with TcpH. According to Crawford and his collaborators, the periplasmic domain of ToxR is involved in dimerization, but it is not required for transcriptional activation since truncations of ToxR missing the periplasmic domain are able to activate transcription (Crawford *et al.* 2003). Cooperative binding between two different w-HTH transcription factors is needed to activate the *toxT* promoter. TcpP binds to a proximal region on the *toxT* promoter, from -53 to -38 relative to the transcription start site (Goss *et al*. 2010) and ToxR binds further upstream (from -100 to -68) (Krukonis *et al.* 2000), but the binding of both transcription factors is required for complete transcriptional activation (Krukonis *et al.* 2000).

1.4.2 ToxR recognizes a direct repeat element in the *toxT p*romoter

The ToxR binding site within the *toxT* promoter was defined by DNase I footprinting analysis as extending from -104 to -68 (Krukonis *et al.* 2000) (Figures 14 and 15).

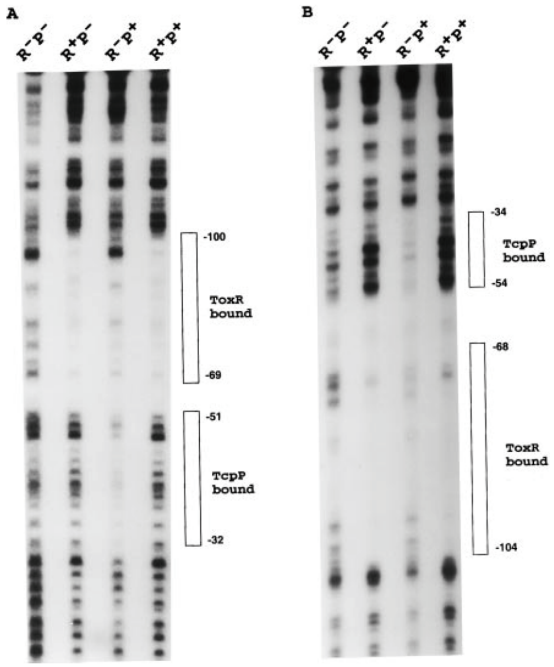

Figure 14. DNase I footprinting analysis of the *toxT* promoter. *V.cholerae* membrane extract derived from strains expressing ToxR (R) and/or TcpP (P) were mixed with the *toxT* promoter in which the top strand (A) or bottom strand (B) had been end labelled (Krukonis *et al.* 2000).

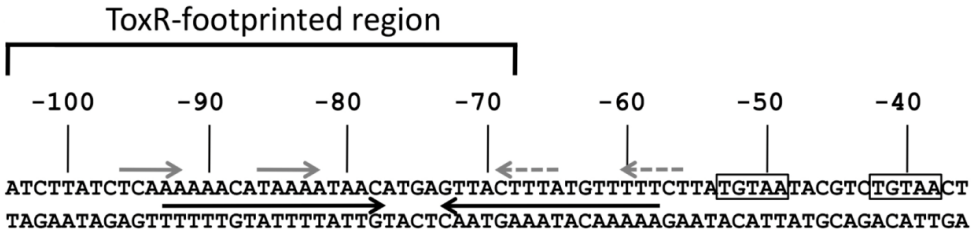

Figure 15. DNA sequence of the *V.cholerae* classical strain O395 proximal region of the *toxT* promoter. The solid gray arrows indicate the direct repeat motif important for ToxR binding (5'-TNAAA-N_5-TNAAA-3'). Dashed grey arrows indicate the degenerate ToxR binding site. The black convergent arrows indicate an inverted repeat sequence. The boxed nucleotides indicate the TcpP binding site. (Goss *et al.* 2013).

In order to identify the specific nucleotides within the ToxR region that are important for *toxT* activation, a library of *toxT* promoter derivatives with single transversions was constructed and tested. In total, 13 transversions reduced *toxT* promoter activity significatively and 12 of them affected nucleotides in the region from -92 to -82 (Figure 16). Of the transversions in the region from -81 to -57, only A(-74)C showed a significant reduction of the promoter activation (Figure 16). Based on these results, the 5'-CTNAAAAAANNTNAAA-3' nucleotide sequence seemed to be critical for ToxR-dependent *toxT* activation. In fact, the direct repeat motif 5'-TNAAA-N_5-TNAAA-3' would provide the binding site for two monomers, which would bind in a head-to-tail configuration to two 5'-TNAAA-3' (Figure 15). Once identified the ToxR binding site, an imperfect ToxR binding site was recognized from -69 to -56 on the opposite strand: 3'-ANAAA-N_4-TNAAG-5' (Figure 15). Gel-shift analysis indicated that ToxR is largely unable to bind this imperfect repeat element (Figure 17).

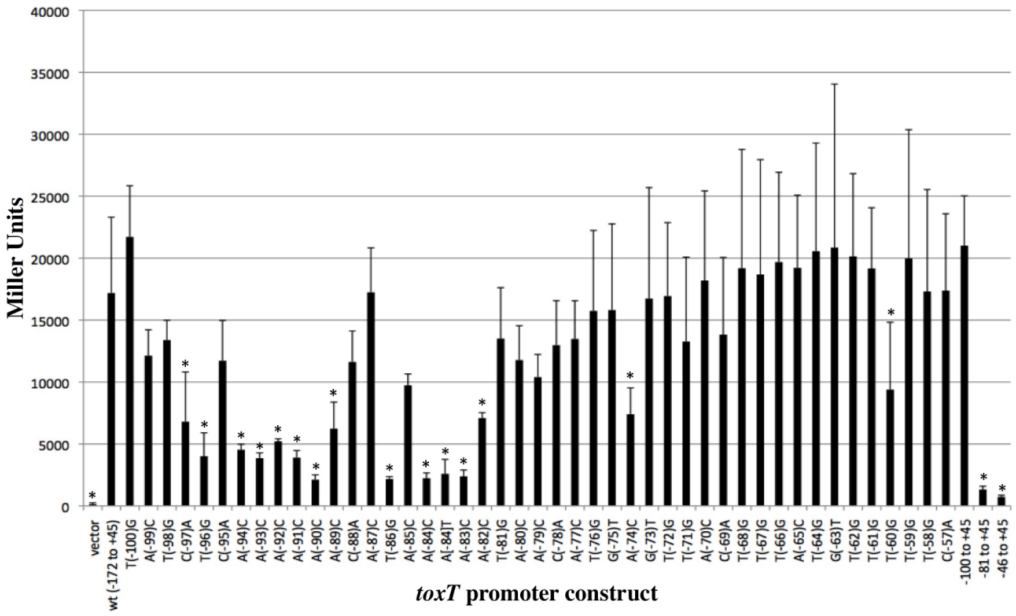

Figure 16. Effects of ToxR binding site mutations on *toxT* promoter. Each base transversion was tested measuring the residual *toxT-lacZ* activity (Goss *et al.* 2013).

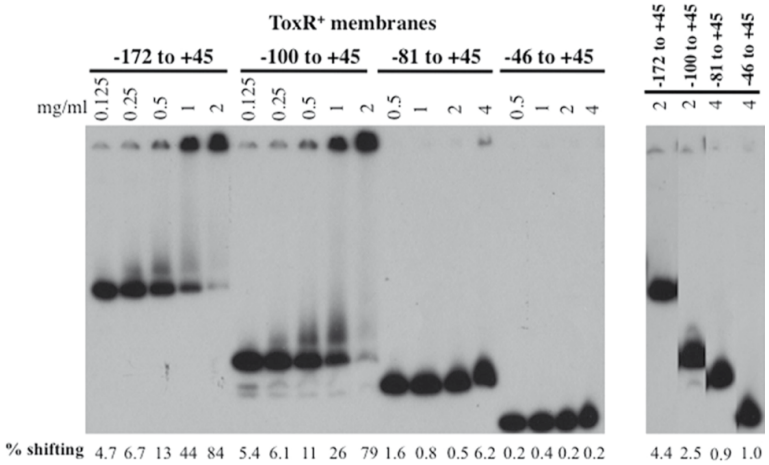

Figure 17. Electrophoretic shift assay of full length (-172 to +45), -100 to +45, -81 to +45, and -46 to + 45 *toxT* derivatives with increasing concentration of membranes containing ToxR. The degenerate ToxR binding site shows a weak ToxR binding capacity (Goss *et al.* 2013).

Moreover, ToxR failed to activate a shorter *toxT* promoter derivative containing the degenerate ToxR binding site from -69 to -56 (Figure 18).

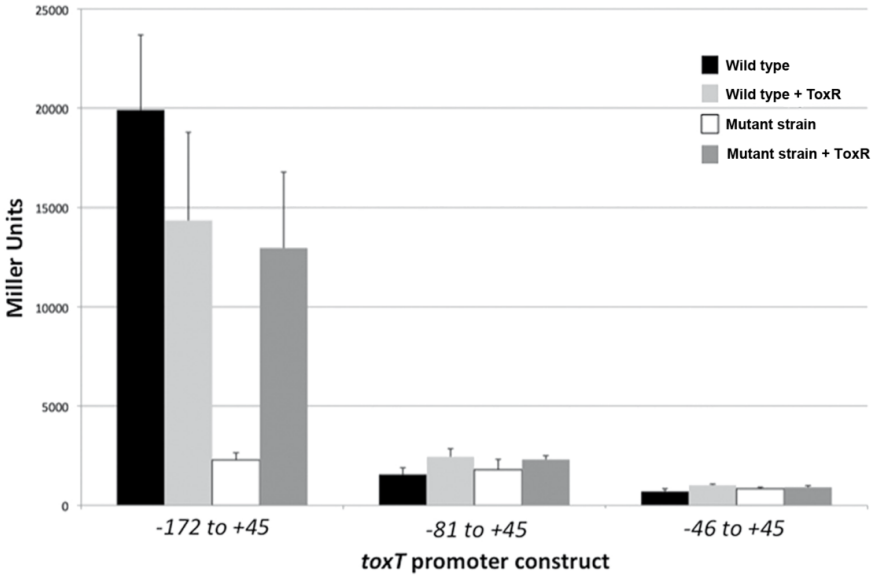

Figure 18. ToxR is not able to activate a *toxT-lacZ* truncated derivative containing the degenerate ToxR binding site from -69 to -56 (-81 to +45 promoter construct). The *toxT* promoter derivatives were tested for activation in wild type *V. cholerae* (O395) and *toxR* mutant strain with or without overexpression of ToxR from a plasmid (Goss *et al.* 2013).

This data indicated that ToxR binds the imperfect ToxR binding site poorly and that this specific DNA sequence does not contribute to ToxR-dependent *toxT* activation.

1.4.3 Models of transcriptional activation of *toxT* promoter

Even though ToxR and TcpP can act as direct activators of transcription, ToxR is thought to play a co-activation role on the *toxT* promoter, assisting TcpP in *toxT* activation. The TcpP binding site is located at -53 to -38 indicating that it is likely the transcription factor that interacts with RNAP and the direct activator of *toxT* transcription (Goss *et al.* 2010). Moreover, TcpP is able to activate transcription of the *toxT* promoter in the absence of ToxR if it is overexpressed from a plasmid (Häse *et al.* 1998; Krukonis *et al*. 2003). ToxR by itself is not able to activate expression of *toxT*, even when overexpressed, despite the fact that it is able to bind to the promoter in the absence of TcpP (Krukonis *et al*. 2003). Since TcpP binds weakly to the *toxT* promoter, it has been hypothesized that ToxR serves as a co-activator by enhancing TcpP recruitment to the promoter via ToxR-TcpP protein-protein interaction.

Three models have been proposed to explain the possible mechanism of ToxR as a co-activator of the *toxT* promoter. The first model is called the "hand-holding" model (Figure 19B). According to this model, ToxR and TcpP can interact and form a ternary complex with the DNA. ToxR would recruit TcpP and then stabilize the ternary complex on the DNA. The second model is called "catch and release" model (Figure 19C), and also is based on the ToxR-TcpP interaction for TcpP recruitment to the promoter. In spite of that, in this model the ToxR-TcpP interaction is not mantained after DNA binding allowing ToxR and TcpP to bind to the *toxT* promoter separately. The third model is named "promoter alteration" and encompasses other possible mechanisms of activation with the assumption that no ToxR-TcpP interaction is required for *toxT* activation. One possible mechanism is that ToxR would be able to enhance TcpP binding to the promoter introducing changes in the promoter architecture (Figure 19D). Another mechanism comprises the ability of ToxR to recruit the *toxT* promoter to a membrane proximal location (Figure 19E). This model is supported by single molecule tracking experiments using super-resolution microscope, which proved that the mobility of TcpP is higher in the presence of ToxR (Haas *et al*. 2015). Although these models are based on biochemical evidences, the coordination of TcpP and ToxR and their interactions with other components of the transcription complex have yet to be observed and the

mechanistic details of theses protein-protein and protein-DNA interactions remain unclear.

In our laboratory, we are currently working to put forward a molecular model for transcription activation of *V. cholerae* virulence genes. We aim to determine the structure of transcription initiation subcomplexes using X-ray crystallography and elucidate the molecular mechanisms by which *V. cholerae* produces its potent toxin. This could also provide avenues for developing novel therapeutic approaches.

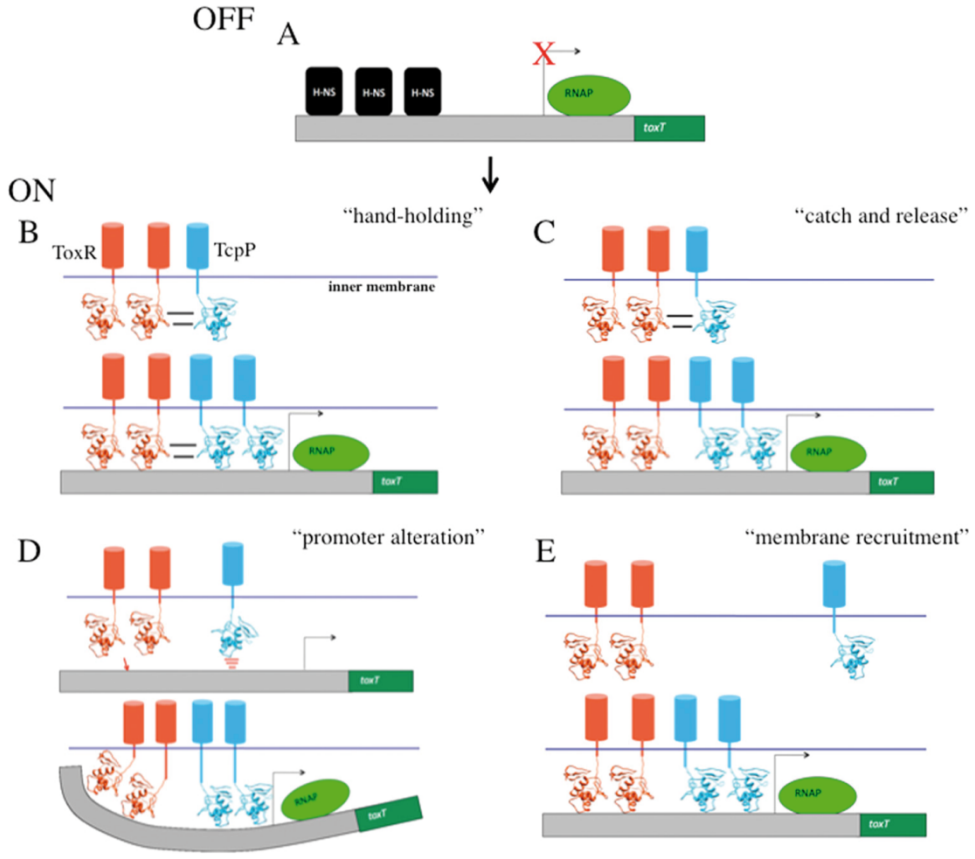

Figure 19. Models for the role of ToxR in TcpP-mediated *toxT* activation. (A) The *toxT* promoter is repressed by H-NS. (B) ToxR relieves H-NS repression and mantains the interaction with TcpP during transcription activation. (C) ToxR interacts with TcpP, but after H-NS displacement and upon DNA binding, ToxR releases TcpP, which can bind its binding site and activate transcription. (D) ToxR interaction with TcpP is not required. ToxR binds DNA and is able to displace H-NS and alter the architecture of *toxT* promoter facilitating TcpP binding. (E) Interaction between ToxR and TcpP is not required, ToxR only recruits the *toxT* promoter to the membrane and makes TcpP binding easier (Goss *et al.* 2013).

Chapter 2: Aim of the work

The general objective of this project thesis is to unveil the mode of action of ToxR as a co-activator of the *toxT* promoter.

The specific objectives are:

1) to determine the 3D structure of the ToxR DBD by X-ray crystallography and compare it with other DBDs of one- and two-component systems

2) to solve the structures of ToxR-DNA complexes to characterize the binding mode of the transcription factor to the *toxT* promoter

3) to put forward a model for activation of the *toxT* promoter.

In order to fulfill these objectives, the following work plan was established:

1. preparation of several DNA-ToxR complexes including DNA segments of the *toxT* promoter
2. crystallization of the DNA-protein complexes
3. data collection and processing
4. structure determination and refinement
5. building of a comprehensive model using the different complex structures solved.

Chapter 3: Materials and Methods

3.1 DNA oligonucleotide design

3.1.1 ToxRDNA20 20-bp DNA oligonucleotide design

According to the footprinting assays experiments previously discussed in the introduction chapter (Krukonis *et al.* 2000), the following DNA oligonucleotides were designed covering two putative ToxR binding sites present in the *toxT* promoter (5'-TNAAA-3'): 5'-CTCAAAAAACATAAAATAAC-3' and 5'-GTTATTTTATGTTTTTTGAG-3' from C-97 to C-78 (Figure 20).

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

Figure 20. Sequence of DNA20.

3.1.2 ToxRDNA40 40-bp DNA oligonucleotide design

The following DNA oligonucleotides were designed covering both the two putative ToxR binding sites and the degenerate ToxR binding site present in the *toxT* promoter (Goss *et al.* 2013): 5'-CCAAAAAACATAAAATAACATGAGTTACTTTATGTTTTTC-3' and 5'-GAAAAACATAAAGTAACTCATGTTATTTTATGTTTTTTGG-3' from T-96 to C-57 (Figure 21). We replaced the first thymine with a cytosine to have a GC base pair at both ends of the oligonucleotide.

```
5'-CCAAAAAACATAAAATAACATGAGTTACTTTATGTTTTTC-3'
3'-GGTTTTTTGTATTTTATTGTACTCAATGAAATACAAAAAG-5'
```

Figure 21. Sequence of DNA40.

3.1.3 ToxRDNA25 25-bp DNA oligonucleotide design

The following DNA oligonucleotides were designed covering the degenerate ToxR binding site and one TcpP binding site present in the *toxT* promoter: 5'-CTTTATGTTTTTCTTATGTAATACG-3' and 5'-CGTATTACATAAGAAAAACATAAAG-3' (Figure 22) from C-69 to G-45.

```
5'-CTTTATGTTTTTCTTATGTAATACG-3'
3'-GAAATACAAAAAGAATACATTATGC-5'
```

Figure 22. Sequence of DNA25.

3.2 Analysis of ToxR preparation

ToxR protein was prepared in the laboratory of Prof. Eric Kroukonis and delivered to us. Briefly, the protein was expressed in *E. coli* and purified through a nickel-affinity chromatography followed by a size-exclusion chromatography and finally an inverse nickel-affinity chromatography after His-tag removal. In our laboratory, we performed an analytical size-exclusion chromatography to ensure the quality of the purified ToxR sample. We used a GE Healthcare Superdex 200 10/300 column connected to an ÄKTA Explorer or Purifier Systems. We prepared Sdx200 buffer A containing 20 mM Tris, 1 mM EDTA, 125 mM NaCl at pH 7.6. We followed a standard protocol with a single step with Sdx200 buffer A (1,2 CV) at a flow of 0.4 mL/min.

3.3 Obtainment of ToxR-DNA complexes

3.3.1 Protein-DNA complex formation procedure

DNA oligonucleotides were diluted in ddH_20 to a final concentration of 2 mM. Complementary primers were mixed. The sample was denatured in a water bath at 80°C

for 30 min and annealed overnight by gradual cooling, after turning off the water bath. ToxR was slowly added to annealed DNA to get a 1:1 protein-DNA ratio and finally a 20% of DNA in excess was added. Complex formation was performed by incubating the protein-DNA mixture for 3 h at 4°C.

3.3.2 Purification of protein-DNA complexes

After incubating protein and DNA, complexes were purified by size-exclusion chromatography. Experiments were performed using a GE Healthcare Superdex 200 10/300 column and the Sdx200 buffer A (1,2 CV) with a flow of 0.4 mL/min. We checked the elution profile, in particular considering the ratio of the UV absorption at 260 and 280 nm. We compared the elution volume of the peak with the elution volume of the protein and DNA alone and we ensured that the DNA-protein complex peak eluted earlier, thus indicating the formation of the complex.

3.4 Structural characterization of ToxR-DNA complexes

3.4.1 Crystallization experiments

We performed crystallization experiments using the sitting drop vapor diffusion method. Initial screens were performed at the Automated Crystallography Platform (PAC) using 96-well MRC plates. Crystallization screenings were prepared using a Tecan Evo-1 robot (Tecan group, Switzerland) (Figure 23).

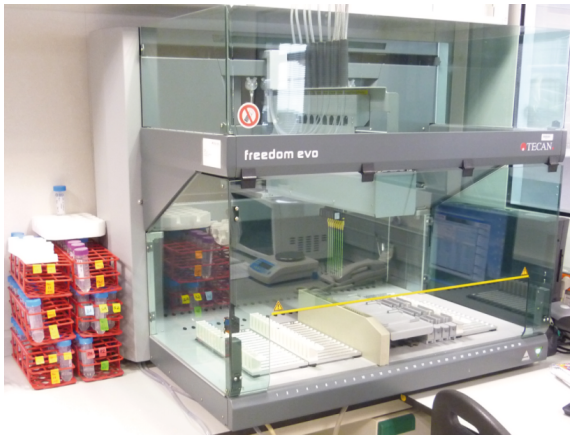

Figure 23. Tecan Evo-1 robot.

Crystallization drops were prepared using Cartesian Honeybee-X8 (Cartesian technologies, US) (Figure 24) and Phoenix RE (Rigaku Edition) (Art Robbins Instruments, US) robots (Figure 25), both providing micro-scale liquid handling for high-throughput screening of crystallization conditions.

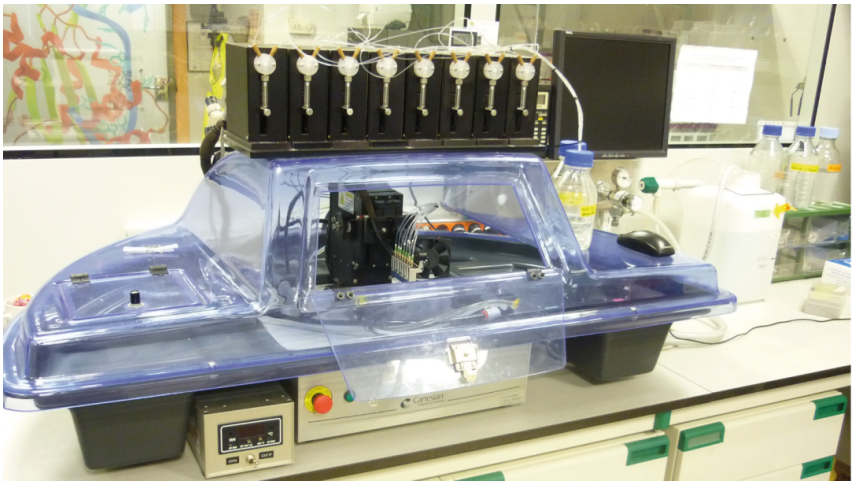

Figure 24. Cartesian Honeybee-X8 robot.

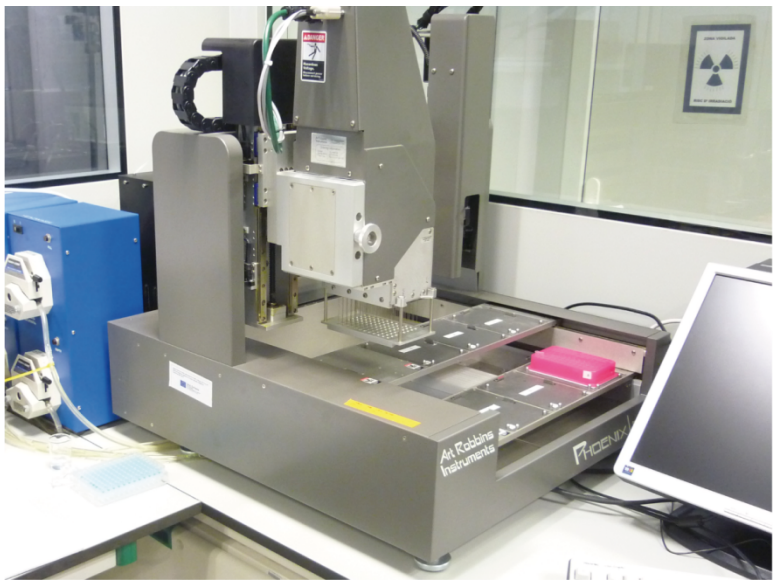

Figure 25. Phoenix RE robot.

We performed crystallization tests of the ToxR-DNA complexes using crystallization screens prepared at the PAC that are named PACs 1, 2, 9 and 10. PACs 1 and 2 comprise commercial screens from Hampton, while PACs 9 and 10 are compositions of conditions especially recommended for protein-DNA complexes (Tables 2, 3, 4 and 5). After setting up the 96-well MRC plates, we store them at 20°C and 4°C. Plate inspection was done automatically by a Crystal Farm 400 (Bruker Corporation, US) robot and using SMZ1000 (Nikon) optical microscopes. The best conditions were scaled up and optimized on 24-well Linbro plates. Protein crystals were fished with nylon cryoloops, and momentarily soaked in a solution containing the reservoir solution supplemented with a variable amount of cryoprotectant buffers. Cryoprotected crystals were flash frozen in liquid nitrogen.

Well	Num.	Salt	Buffer	Precipitant 1	Precipitant 2
A1	1	0.02 M Calcium chloride	0.1 M Sodium Acetate pH 4.6	30 %v/v MPD	
A2	2			0.4 M Na/K Tartrate	
A3	3			0.4 M Ammonium dihydrogen phosphate	
A4	4		0.1 M TRIS pH 8.5	2 M Ammonium sulfate	
A5	5	0.2 M Tri-sodium citrate	0.1 M Hepes pH 7.5	30 %v/v MPD	
A6	6	0.2 M Magnesium chloride	0.1 M TRIS pH 8.5	30 %w/v PEG 4000	
A7	7		0.1 M Cacodilate pH 6.5	1.4 M Sodium acetate	
A8	8	0.2 M Tri-sodium citrate	0.1 M Cacodilate pH 6.5	30 %v/v Isopropanol	
A9	9	0.2 M Ammonium acetate	0.1 M Tri-sodium citrate pH 5.6	30 %w/v PEG 4000	
A10	10	0.2 M Ammonium acetate	0.1 M Sodium Acetate pH 4.6	30 %w/v PEG 4000	
A11	11		0.1 M Tri-sodium citrate pH 5.6	1 M Ammonium dihydrogen phosphate	
A12	12	0.2 M Magnesium chloride	0.1 M HEPES pH 7.5	30 %v/v Isopropanol	
B1	13	0.2 M Tri-sodium citrate	0.1 M TRIS pH 8.5	30 %v/v PEG 400	
B2	14	0.2 M Calcium chloride	0.1 M Hepes pH 7.5	28 %v/v PEG 400	
B3	15	0.2 M Ammonium sulfate	0.1 M Cacodilate pH 6.5	30 %w/v PEG 8000	
B4	16		0.1 M HEPESpH 7.5	1.5 M Lithium sulfate anhydre	
B5	17	0.2 M Lithium sulfate anhydre	0.1 M TRIS pH 8.5	30 %w/v PEG 4000	
B6	18	0.2 M Mg acetate tetrahydrate	0.1 M Cacodilate pH 6.5	20 %w/v PEG 8000	
B7	19	0.2 M Ammonium acetate	0.1 M TRIS pH 8.5	30 %v/v Isopropanol	
B8	20	0.2 M Ammonium sulfate	0.1 M Sodium Acetate pH 4.6	25 %w/v PEG 4000	
B9	21	0.2 M Mg acetate tetrahydrate	0.1 M Cacodilate pH 6.5	30 %v/v MPD	
B10	22	0.2 Sodium acetate	0.1 M TRIS pH 8.5	30 %w/v PEG 4000	
B11	23	0.2 M Magnesium chloride	0.1 M HEPES pH 7.5	30 %v/v PEG 400	
B12	24	0.2 M Calcium chloride	0.1 M Sodium Acetate pH 4.6	20 %v/v Isopropanol	
C1	25		0.1 M Imidazole pH 6.5	1 M Sodium acetate	
C2	26	0.2 M Ammonium acetate	0.1 M Tri-sodium citrate pH 5.6	30 %v/v MPD	
C3	27	0.2 M Tri-sodium citrate	0.1 M HEPES pH 7.5	20 %v/v Isopropanol	
C4	28	0.2 Sodium acetate	0.1 M Cacodilate pH 6.5	30 %w/v PEG 8000	
C5	29		0.1 M HEPES pH 7.5	0.8 M Na/K Tartrate	
C6	30	0.2 M Ammonium sulfate		30 %w/v PEG 8000	
C7	31	0.2 M Ammonium sulfate		30 %w/v PEG 4000	
C8	32			2 M Ammonium sulfate	
C9	33			4 M Sodium formate	
C10	34		0.1 M Sodium Acetate pH 4.6	2 M Sodium formate	
C11	35		0.1 M Hepes pH 7.5	1.6 M Na/K phosphate	
C12	36		0.1 M TRIS pH 8.5	8 %w/v PEG 8000	
D1	37		0.1 M Sodium Acetate pH 4.6	8 %w/v PEG 4000	
D2	38		0.1 M Hepes pH 7.5	1.4 M Tri-sodium citrate	2 M Ammonium sulfate
D3	39		0.1 M Hepes pH 7.5	2 %v/v PEG 400	20 %w/v PEG 4000
D4	40		0.1 M Tri-sodium citrate pH 5.6	20 %v/v Isopropanol	20 %w/v PEG 4000
D5	41		0.1 M Hepes pH 7.5	10 %v/v Isopropanol	
D6	42	0.05 M Di- K hydrogen phosphate		20 %w/v PEG 8000	
D7	43			30 %w/v PEG 1500	
D8	44	0.2 M Zinc acetate	0.1 M Cacodilate pH 6.5	18 %w/v PEG 8000	
D9	45	0.2 M calcium acetate hydrate	0.1 M Cacodilate pH 6.5	18 %w/v PEG 8000	
D10	46		0.1 M Sodium Acetate pH 4.6	2 M Ammonium sulfate	
D11	47		0.1 M TRIS pH 8.5	2 M Ammonium dihydrogen phosphate	
D12	48	1 M Lithium sulfate anhydre		2 %w/v PEG 8000	
E1	49	2 M Sodium chloride		10 %w/v PEG 6000	0.01 M Mg chloride
E2	50	0.5 Sodium chloride		0.01 M cetil trimetil bromur ammonium	
E3	51			25 %v/v ethylene glycol	
E4	52			35 %v/v dioxane	
E5	53	2 M Ammonium sulfate		5 %v/v Isopropanol	
E6	54			1 M Imidazole pH 7.0	10 %w/v PEG 8000
E7	55			10 %w/v PEG 1000	
E8	56	1.5 M Sodium chloride		10 %v/v Ethanol	
E9	57		0.1 M Sodium Acetate pH 4.6	2 M Sodium chloride	
E10	58	0.2 M Sodium chloride	0.1 M Sodium Acetate pH 4.6	30 %v/v MPD	
E11	59	0.01 M Cobaltous chloride	0.1 M Sodium Acetate pH 4.6	1 M 1,6-Hexanediol	
E12	60	0.1 M Cadmium chloride	0.1 M Sodium Acetate pH 4.6	30 %v/v PEG 400	
F1	61	0.2 M Ammonium sulfate	0.1 M Sodium Acetate pH 4.6	30 %w/v PEG MME 2000	
F2	62	0.2 Potassium Sodium Tartrate	0.1 M Tri-sodium citrate pH 5.6	2 M Ammonium sulfate	
F3	63	0.5 M Ammonium sulfate	0.1 M Tri-sodium citrate pH 5.6	1 M lithium sulfate monohydrate	
F4	64	0.5 Sodium chloride	0.1 M Tri-sodium citrate pH 5.6	4 %v/v ethylene imine polymer	
F5	65		0.1 M Tri-sodium citrate pH 5.6	35 %v/v Tert-butanol	
F6	66	0.01 M Fe chloride hexahydrate	0.1 M Tri-sodium citrate pH 5.6	10 %v/v Jeffamine M-600	
F7	67		0.1 M Tri-sodium citrate pH 5.6	2.5 M 1,6-Hexanediol	
F8	68		0.1 M MES pH 6.5	1.6 M Magnesium sulfate heptahydrate	
F9	69	0.2 K/Na Tartrate	0.1 M MES pH 6.5	2 M Sodium chloride	
F10	70		0.1 M MES pH 6.5	12 %w/v PEG 20,000	
F11	71	1.6 M Ammonium sulfate	0.1 M MES pH 6.5	10 %v/v dioxane	
F12	72	0.05 M Cesium chloride	0.1 M MES pH 6.5	30 %v/v Jeffamine M-600	
G1	73	0.01 M Cobaltous chloride	0.1 M MES pH 6.5	1.8 M Ammonium sulfate	
G2	74	0.2 M Ammonium sulfate	0.1 M MES pH 6.5	30 %w/v PEG MME 5000	
G3	75	0.01 m Zinc sulfate heptahydrate	0.1 M MES pH 6.5	25 %v/v PEG MME 500	
G4	76			1.6 M Tri-sodium citrate pH 6.5	
G5	77	0.5 M Ammonium sulfate	0.1 M Hepes pH 7.5	30 %v/v MPD	5 %v/v MPD
G6	78		0.1 M Hepes pH 7.5	10 %w/v PEG 6000	
G7	79		0.1 M Hepes pH 7.5	20 %v/v Jeffamine M-600	
G8	80	0.1 m Sodium chloride	0.1 M Hepes pH 7.5	1.6 M Ammonium sulfate	
G9	81		0.1 M Hepes pH 7.5	2 M Sodium formate	
G10	82	0.05 M Cadmium sulfate	0.1 M Hepes pH 7.5	1 M Sodium acetate	
G11	83		0.1 M Hepes pH 7.5	70 %v/v MPD	
G12	84		0.1 M Hepes pH 7.5	4.3 M Sodium chloride	8 %v/v ethylene glycol
H1	85		0.1 M Hepes pH 7.5	10 %w/v PEG 8000	
H2	86		0.1 M Hepes pH 7.5	20 %w/v PEG 10,000	
H3	87	0.2 M Magnesium chloride	0.1 M TRIS pH 8.5	3.4 M 1,6-Hexanediol	
H4	88		0.1 M TRIS pH 8.5	25 %v/v Tert-butanol	
H5	89	0.01 M Nickel chloride	0.1 M TRIS pH 8.5	1 M lithium sulfate monohydrate	
H6	90	1.5 M Ammonium sulfate	0.1 M TRIS pH 8.5	12 %v/v glycerol	
H7	91	0.2 M NH4 dihydrogen phosphate	0.1 M TRIS pH 8.5	50 %v/v MPD	
H8	92		0.1 M TRIS pH 8.5	20 %v/v Ethanol	
H9	93	0.01 M Nickel chloride	0.1 M TRIS pH 8.5	20 %w/v PEG MME 2000	
H10	94	0.1 M Sodium chloride	0.1 M Bicine pH 9.0	30 %v/v PEG MME 500	
H11	95		0.1 M Bicine pH 9.0	2 M Magnesium chloride	10 %w/v PEG 20000
H12	96		0.1 M Bicine pH 9.0	2 %v/v dioxane	

Table 2. List of crystallization conditions of PAC1.

Well	Num.	Salt	Buffer	Precipitant 1	Precipitant 2
A1	1		0.1 M CHES 9.5	20 %w/v PEG 8000	
A2	2	0.2 M sodium chloride	0.1 M HEPES 7.5	10 %v/v Isopropanol	
A3	3		0.1 M CHES 9.5	15 %v/v Ethanol	
A4	4	0.2 M magnesium chloride	0.1 M imidazole 8	35 %v/v MPD	
A5	5		0.1 M CAPS 10.5	30 %v/v PEG 400	
A6	6		0.1 M Tri-sodium citrate 5.5	20 %w/v PEG 3000	
A7	7	0.2 M zinc acetate dihydrate	0.1 M MES 6	10 %w/v PEG 8000	
A8	8		0.1 M Tri-sodium citrate 5.5	2 M Ammonium sulfate	
A9	9		0.1 M Sodium acetate 4.5	1 M di-NH4 hydrogen phosphate	
A10	10		0.1 M TRIS 7	20 %w/v PEG MME 2000	
A11	11	0.2 M lithium sulfate anhydre	0.1 M MES 6	20 %v/v 1,4-butanediol	
A12	12	0.2 M calcium acetate hydrate	0.1 M imidazole 8	20 %w/v PEG 1000	
B1	13		0.1 M Cacodilate 6.5	1.26 M Ammonium sulfate	
B2	14		0.1 M Cacodilate 6.5	1 M sodium citrate	
B3	15	0.2 M lithium sulfate anhydre	0.1 M imidazole 8	10 %w/v PEG 3000	
B4	16		0.1 M Na/K phosphate 6.2	2.5 M sodium chloride	
B5	17	0.2 M lithium sulfate anhydre	0.1 M Sodium acetate 4.5	30 %w/v PEG 8000	
B6	18	0.2 M sodium chloride	0.1 M imidazole 8	1 M Na/K Tartrate	
B7	19		0.1 M TRIS 7	20 %w/v PEG 1000	
B8	20	0.2 M sodium chloride	0.1 M imidazole 8	0.4 M Na dihydrogen phosphate	1.6 M di-K hydrogen phosphate
B9	21		0.1 M HEPES 7.5	20 %w/v PEG 8000	
B10	22		0.1 M TRIS 8.5	10 %v/v Isopropanol	
B11	23	0.2 M magnesium chloride	0.1 M imidazole 8	15 %v/v Ethanol	
B12	24	0.2 M sodium chloride	0.1 M TRIS 7	35 %v/v MPD	
C1	25	0.2 M magnesium chloride	0.1 M TRIS 8.5	30 %v/v PEG 400	
C2	26		0.1 M CHES 9.5	10 %w/v PEG 3000	
C3	27	0.2 M lithium sulfate anhydre	0.1 M CAPS 10.5	1.2 M Na dihydrogen phosphate	0.8 M di-K hydrogen phosphate
C4	28	0.2 M sodium chloride	0.1 M HEPES 7.5	20 %w/v PEG 3000	
C5	29	0.2 M sodium chloride	0.1 M CHES 9.5	10 %w/v PEG 8000	
C6	30	0.2 M sodium chloride	0.1 M Sodium acetate 4.5	1.26 M Ammonium sulfate	
C7	31	0.2 M sodium chloride	0.1 M phosphate citrate 4.2	20 %w/v PEG 8000	
C8	32		0.1 M Na/K phosphate 6.2	10 %w/v PEG 3000	
C9	33	0.2 M lithium sulfate anhydre	0.1 M CAPS 10.5	2 M Ammonium sulfate	
C10	34		0.1 M imidazole 8	1 M di-NH4 hydrogen phosphate	
C11	35		0.1 M Sodium acetate 4.5	20 %v/v 1,4-butanediol	
C12	36		0.1 M imidazole 8	1 M Tri-sodium citrate	
D1	37		0.1 M imidazole 8	2.5 M sodium chloride	
D2	38	0.2 M lithium sulfate anhydre	0.1 M CHES 9.5	1 M Na/K Tartrate	
D3	39	0.2 M lithium sulfate anhydre	0.1 M phosphate citrate 4.2	20 %w/v PEG 1000	
D4	40	0.2 M calcium acetate hydrate	0.1 M MES 6	10 %v/v Isopropanol	
D5	41		0.1 M CHES 9.5	30 %w/v PEG 3000	
D6	42		0.1 M TRIS 7	15 %v/v ethanol	
D7	43		0.1 M Na/K phosphate 6.2	35 %v/v MPD	
D8	44	0.2 M calcium acetate hydrate	0.1 M Sodium acetate 4.5	30 %v/v PEG 400	
D9	45		0.1 M Sodium acetate 4.5	20 %w/v PEG 3000	
D10	46	0.2 M calcium acetate hydrate	0.1 M imidazole 8	10 %w/v PEG 8000	
D11	47	0.2 M lithium sulfate anhydre	0.1 M TRIS 8.5	1.26 M Ammonium sulfate	
D12	48	0.2 M zinc acetate dihydrate	0.1 M Sodium acetate 4.5	20 %w/v PEG 1000	
E1	49	0.2 M zinc acetate dihydrate	0.1 M Sodium acetate 4.5	10 %w/v PEG 3000	
E2	50	0.2 M lithium sulfate anhydre	0.1 M MES 6	35 %v/v MPD	
E3	51	0.2 M magnesium chloride	0.1 M Tris 8.5	20 %w/v PEG 8000	
E4	52	0.2 M sodium chloride	0.1 M Cacodilate 6.5	2 M ammonium sulfate	
E5	53	0.2 M sodium chloride	0.1 M HEPES 7.5	20 %v/v 1,4-butanediol	
E6	54	0.2 M lithium sulfate anhydre	0.1 M phosphate citrate 4.2	10 %v/v Isopropanol	
E7	55	0.2 M sodium chloride	0.1 M TRIS 7	30 %w/v PEG 3000	
E8	56	0.2 M sodium chloride	0.1 M Na/K phosphate 6.2	10 %w/v PEG 8000	
E9	57		0.1 M phosphate citrate 4.2	2 M ammonium sulfate	
E10	58		0.1 M TRIS 8.5	1 M di-NH4 hydrogen phosphate	
E11	59	0.2 M zinc acetate dihydrate	0.1 M Cacodilate 6.5	10 %v/v Isopropanol	
E12	60	0.2 M lithium sulfate anhydre	0.1 M Cacodilate 6.5	30 %v/v PEG 400	
F1	61	0.2 M lithium sulfate anhydre	0.1 M Tri-sodium citrate 5.5	15 %v/v ethanol	
F2	62	0.2 M sodium chloride	0.1 M Na/K phosphate 6.2	20 %w/v PEG 1000	
F3	63		0.1 M HEPES 7.5	1.26 M ammonium sulfate	
F4	64		0.1 M CHES 9.5	1 M Tri-sodium citrate	
F5	65	0.2 M magnesium chloride	0.1 M TRIS 7	2.5 M sodium chloride	
F6	66	0.2 M calcium acetate hydrate	0.1 M TRIS 7	20 %w/v PEG 3000	
F7	67		0.1 M phosphate citrate 4.2	1.6 M Na dihydrogen phosphate	0.4 M di-K hydrogen phosphate
F8	68	0.2 M zinc acetate dihydrate	0.1 M MES 6	15 %v/v ethanol	
F9	69		0.1 M Sodium acetate 4.5	35 %v/v MPD	
F10	70		0.1 M imidazole 8	10 %v/v Isopropanol	
F11	71	0.2 M magnesium chloride	0.1 M HEPES 7.5	15 %v/v ethanol	
F12	72	0.2 M sodium chloride	0.1 M imidazole 8	30 %w/v PEG 8000	
G1	73	0.2 M sodium chloride	0.1 M HEPES 7.5	35 %v/v MPD	
G2	74		0.1 M CHES 9.5	30 %v/v PEG 400	
G3	75	0.2 M magnesium chloride	0.1 M Cacodilate 6.5	10 %w/v PEG 3000	
G4	76	0.2 M calcium acetate hydrate	0.1 M MES 6	20 %w/v PEG 8000	
G5	77	0.2 M sodium chloride	0.1 M CHES 9.5	1.26 M ammonium sulfate	
G6	78	0.2 M zinc acetate dihydrate	0.1 M imidazole 8	20 %v/v 1,4-butanediol	
G7	79	0.2 M sodium chloride	0.1 M TRIS 7	1 M Tri-sodium citrate	
G8	80		0.1 M TRIS 8.5	20 %w/v PEG 1000	
G9	81	0.2 M sodium chloride	0.1 M Tri-sodium citrate 5.5	1 M di-NH4 hydrogen phosphate	
G10	82		0.1 M imidazole 8	10 %w/v PEG 8000	
G11	83		0.1 M Sodium acetate 4.5	0.8 M Na dihydrogen phosphate	1.2 M di-K hydrogen phosphate
G12	84	0.2 M sodium chloride	0.1 M phosphate citrate 4.2	10 %w/v PEG 3000	
H1	85	0.2 M lithium sulfate anhydre	0.1 M TRIS 7	1 M Na/K Tartrate	
H2	86	0.2 M lithium sulfate anhydre	0.1 M Sodium acetate 4.5	2.5 M sodium chloride	
H3	87	0.2 M sodium chloride	0.1 M CAPS 10.5	20 %w/v PEG 8000	
H4	88	0.2 M zinc acetate dihydrate	0.1 M imidazole 8	20 %w/v PEG 3000	
H5	89	0.2 M lithium sulfate anhydre	0.1 M TRIS 7	2 M ammonium sulfate	
H6	90	0.2 M sodium chloride	0.1 M HEPES 7.5	30 %v/v PEG 400	
H7	91	0.2 M magnesium chloride	0.1 M TRIS 7	10 %w/v PEG 8000	
H8	92	0.2 M magnesium chloride	0.1 M Cacodilate 6.5	20 %w/v PEG 1000	
H9	93		0.1 M MES 6	1.26 M ammonium sulfate	
H10	94	0.2 M sodium chloride	0.1 M imidazole 8	1 M di-NH4 hydrogen phosphate	
H11	95	0.2 M zinc acetate dihydrate	0.1 M imidazole 8	2.5 M sodium chloride	
H12	96		0.1 M MES 6	1 M Na/K Tartrate	

Table 3. List of crystallization conditions of PAC2.

Well	Num.	Salt 1	Salt 2	Buffer	Precipitant
A1	1	0.01 M magnesium chloride		0.05 M MES 5.6	2 M lithium sulfate monohydrate
A2	2	0.01 M Mg acetate tetrahydrate		0.05 M MES 5.6	2.5 M ammonium sulfate
A3	3	0.1 M Mg acetate tetrahydrate		0.05 M MES 5.6	20 %v/v MPD
A4	4	0.2 M potassium chloride	0.01 M Mg sulfate hydrate	0.05 M MES 5.6	10 %v/v PEG 400
A5	5	0.2 M potassium chloride	0.01 M Mg sulfate hydrate	0.05 M MES 5.6	5 %w/v PEG 8000
A6	6	0.1 M ammonium sulfate	0.01 M Mg sulfate hydrate	0.05 M MES 5.6	20 %w/v PEG 8000
A7	7	0.02 M magnesium chloride		0.05 M MES 6	15 %v/v Isopropanol
A8	8	0.1 M ammonium acetate	5 mM Mg sulfate hydrate	0.05 M MES 6	0.6 M sodium chloride
A9	9	0.1 M potassium chloride	0.01 M Mg sulfate hydrate	0.05 M MES 6	10 %v/v PEG 400
A10	10	5 mM Mg sulfate hydrate		0.05 M MES 6	5 %w/v PEG 4000
A11	11	0.01 M Mg chloride		0.05 M Cacodilate 6	1 M lithium sulfate monohydrate
A12	12	0.01 M Mg sulfate hydrate		0.05 M Cacodilate 6	1.8 M lithium sulfate monohydrate
B1	13	15 mM Mg acetate tetrahydrate		0.05 M Cacodilate 6	1.7 M ammonium sulfate
B2	14	0.1 M potassium chloride	25 mM Mg chloride	0.05 M Cacodilate 6	15 %v/v Isopropanol
B3	15	0.04 M magnesium chloride		0.05 M Cacodilate 6	5 %v/v MPD
B4	16	0.04 M Mg acetate tetrahydrate		0.05 M Cacodilate 6	30 %v/v MPD
B5	17	0.2 M potassium chloride	0.01 M calcium chloride	0.05 M Cacodilate 6	10 %w/v PEG 4000
B6	18	0.01 M Mg acetate tetrahydrate		0.05 M Cacodilate 6.5	1.3 M lithium sulfate monohydrate
B7	19	0.01 M Mg sulfate hydrate		0.05 M Cacodilate 6.5	2 M ammonium sulfate
B8	20	0.1 M ammonium acetate	15 mM Mg acetate tetrahydrate	0.05 M Cacodilate 6.5	10 %v/v Isopropanol
B9	21	0.2 M potassium chloride	5 mM Mg chloride	0.05 M Cacodilate 6.5	10 %w/v 1,6-hexanediol
B10	22	0.08 M Mg acetate tetrahydrate		0.05 M Cacodilate 6.5	15 %v/v PEG 400
B11	23	0.2 M potassium chloride	0.01 M Mg chloride	0.05 M Cacodilate 6.5	10 %w/v PEG 4000
B12	24	0.2 M ammonium acetate	0.01 M calcium chloride	0.05 M Cacodilate 6.5	10 %w/v PEG 4000
C1	25	0.08 M Mg acetate tetrahydrate		0.05 M Cacodilate 6.5	30 %w/v PEG 4000
C2	26	0.2 M potassium chloride	0.1 M Mg acetate tetrahydrate	0.05 M Cacodilate 6.5	10 %w/v PEG 8000
C3	27	0.2 M ammonium acetate	0.01 M Mg acetate tetrahydrate	0.05 M Cacodilate 6.5	30 %w/v PEG 8000
C4	28	0.05 M Mg sulfate hydrate		0.05 M HEPES 7	1.6 M lithium sulfate monohydrate
C5	29	0.01 M Mg chloride		0.05 M HEPES 7	4 M lithium chloride
C6	30	0.01 M Mg chloride		0.05 M HEPES 7	1.6 M ammonium sulfate
C7	31	5 mM Mg chloride		0.05 M HEPES 7	25 %v/v PEG MME 500
C8	32	0.2 M potassium chloride	0.01 M Mg chloride	0.05 M HEPES 7	20 %w/v 1,6-hexanediol
C9	33	0.2 M ammonium chloride	0.01 M Mg chloride	0.05 M HEPES 7	30 %w/v 1,6-hexanediol
C10	34	0.1 M potassium chloride	5 mM Mg sulfate hydrate	0.05 M HEPES 7	15 %v/v MPD
C11	35	0.1 M potassium chloride	0.01 M Mg chloride	0.05 M HEPES 7	5 %v/v PEG 400
C12	36	0.1 M potassium chloride	0.01 M calcium chloride	0.05 M HEPES 7	10 %v/v PEG 400
D1	37	0.2 M potassium chloride	25 mM Mg sulfate hydrate	0.05 M HEPES 7	20 %v/v PEG 200
D2	38	0.2 M ammonium acetate	0.15 M Mg acetate tetrahydrate	0.05 M HEPES 7	5 %w/v PEG 4000
D3	39	0.1 M ammonium acetate	0.02 M Mg chloride	0.05 M HEPES 7	5 %w/v PEG 8000
D4	40	0.01 M Mg chloride		0.05 M TRIS 7.5	1.6 M ammonium sulfate
D5	41	0.1 M potassium chloride	15 mM Mg chloride	0.05 M TRIS 7.5	10 %v/v PEG MME 500
D6	42	0.01 M Mg acetate tetrahydrate		0.05 M TRIS 7.5	5 %v/v Isopropanol
D7	43	0.05 M ammonium acetate	0.01 M Mg chloride	0.05 M TRIS 7.5	10 %v/v MPD
D8	44	0.2 M potassium chloride	0.05 M Mg chloride	0.05 M TRIS 7.5	10 %w/v PEG 4000
D9	45	25 mM Mg sulfate hydrate		0.05 M TRIS 8.5	1.8 M ammonium sulfate
D10	46	5 mM Mg sulfate hydrate		0.05 M TRIS 8.5	35 %w/v 1,6-hexanediol
D11	47	0.1 M potassium chloride	0.01 M Mg chloride	0.05 M TRIS 8.5	30 %v/v PEG 400
D12	48	0.2 M ammonium chloride	0.01 M calcium chloride	0.05 M TRIS 8.5	30 %w/v PEG 4000
E1	49			0.1 M Tri-Na citrate 5.6	10 %w/v PEG MME 5000
E2	50			0.1 M MES 6.5	10 %w/v PEG MME 5000
E3	51			0.1 M HEPES 7.5	10 %w/v PEG MME 5000
E4	52			0.1 M Tri-Na citrate 5.6	20 %w/v PEG MME 5000
E5	53			0.1 M MES 6.5	20 %w/v PEG MME 5000
E6	54			0.1 M HEPES 7.5	20 %w/v PEG MME 5000
E7	55			0.1 M Tri-Na citrate 5.6	10 %w/v PEG 3350
E8	56			0.1 M MES 6.5	10 %w/v PEG 3350
E9	57			0.1 M HEPES 7.5	10 %w/v PEG 3350
E10	58			0.1 M Tri-Na citrate 5.6	20 %w/v PEG 3350
E11	59			0.1 M MES 6.5	20 %w/v PEG 3350
E12	60			0.1 M HEPES 7.5	20 %w/v PEG 3350
F1	61			0.1 M Tri-Na citrate 5.6	15 %w/v PEG 400
F2	62			0.1 M MES 6.5	15 %w/v PEG 400
F3	63			0.1 M HEPES 7.5	15 %w/v PEG 400
F4	64			0.1 M Tri-Na citrate 5.6	30 %w/v PEG 400
F5	65			0.1 M MES 6.5	30 %w/v PEG 400
F6	66			0.1 M HEPES 7.5	30 %w/v PEG 400
F7	67			0.1 M Tri-Na citrate 5.6	15 %v/v MPD
F8	68			0.1 M MES 6.5	15 %v/v MPD
F9	69			0.1 M HEPES 7.5	15 %v/v MPD
F10	70			0.1 M Tri-Na citrate 5.6	25 %v/v MPD
F11	71			0.1 M MES 6.5	25 %v/v MPD
F12	72			0.1 M HEPES 7.5	25 %v/v MPD
G1	73			0.1 M Tri-Na citrate 5.6	10 %w/v PEG 6000
G2	74			0.1 M MES 6.5	10 %w/v PEG 6000
G3	75			0.1 M HEPES 7.5	10 %w/v PEG 6000
G4	76			0.1 M Tri-Na citrate 5.6	10 %w/v PEG 8000
G5	77			0.1 M MES 6.5	10 %w/v PEG 8000
G6	78			0.1 M HEPES 7.5	10 %w/v PEG 8000
G7	79			0.1 M TRIS 8	10 %w/v PEG MME 5000
G8	80			0.1 M TRIS 8	20 %w/v PEG MME 5000
G9	81			0.1 M TRIS 8	10 %w/v PEG 3350
G10	82			0.1 M TRIS 8	20 %w/v PEG 3350
G11	83			0.1 M TRIS 8	15 %w/v PEG 400
G12	84			0.1 M TRIS 8	30 %w/v PEG 400
H1	85			0.1 M TRIS 8	15 %v/v MPD
H2	86			0.1 M TRIS 8	25 %v/v MPD
H3	87			0.1 M TRIS 8	10 %w/v PEG 6000
H4	88			0.1 M TRIS 8	10 %w/v PEG 8000

Table 4. List of crystallization conditions of PAC9.

Well	Num.	Precipitant	Buffer	Additive 1	Additive 2
A1	1	1 M Lithium sulfate monohydrate	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A2	2	10 %v/v Butanol	0.05 M Tri-Na citrate 5.5		
A3	3	20 %v/v MPD	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A4	4	30 %v/v Glycerol	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A5	5	15 %v/v 2,3-Butanediol	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	0.15 M Sodium chloride
A6	6	25 %v/v DMSO	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A7	7	10 %v/v PEG 200	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A8	8	10 %v/v PEG 400	0.05 M Tri-Na citrate 5.5	0.2 M Potassium Chloride	
A9	9	10 %w/v PEG MME 500	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A10	10	10 %w/v PEG 1000	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	0.2 M Potassium Chloride
A11	11	5 %w/v PEG 8000	0.05 M Tri-Na citrate 5.5	5 mM Magnesium Chloride	
A12	12	10 %w/v PEG 8000	0.05 M Tri-Na citrate 5.5	0.1 M Magnesium Chloride	
B1	13	1.75 M Ammonium sulfate	0.05 M MES 6	0.01 M Magnesium Chloride	
B2	14	20 %v/v Ethanol	0.05 M MES 6	0.01 M Magnesium Chloride	
B3	15	10 %w/v PEG 3350	0.05 M MES 6	0.01 M Magnesium Chloride	0.2 M Potassium Chloride
B4	16	10 %v/v PEG 400	0.05 M MES 6	5 mM Barium Chloride	
B5	17	15 %w/v PEG MME 2000	0.05 M MES 6		
B6	18	20 %w/v PEG 8000	0.05 M MES 6		
B7	19	20 %v/v Propanediol	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	
B8	20	1.5 M Lithium sulfate monohydrate	0.05 M Cacodilate 6.5		
B9	21	10 %v/v Ethanol	0.05 M Cacodilate 6.5	5 mM Barium Chloride	
B10	22	20 %v/v PEG 200	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	
B11	23	15 %v/v Isopropanol	0.05 M Cacodilate 6.5	0.1 M Potassium Chloride	
B12	24	10 %v/v MPD	0.05 M Cacodilate 6.5	0.1 M Magnesium Chloride	
C1	25	20 %w/v PEG 3350	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	
C2	26	15 %v/v Glycerol	0.05 M Cacodilate 6.5	0.05 M Cesium Chloride	
C3	27	10 %v/v PEG 400	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	
C4	28	20 %v/v Ethylene glycol	0.05 M Cacodilate 6.5		
C5	29	10 %w/v PEG 1000	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	
C6	30	12.5 %w/v PEG MME 5000	0.05 M Cacodilate 6.5	0.01 M Magnesium Chloride	0.1 M Potassium Chloride
C7	31	1.5 M Sodium chloride	0.05 M Cacodilate 7	5 %v/v Ethylene glycol	
C8	32	20 %v/v Ethylene glycol	0.05 M Cacodilate 7	5 mM Barium Chloride	
C9	33	20 %v/v 2,3-Butanediol	0.05 M Cacodilate 7	0.01 M Magnesium Chloride	
C10	34	10 %w/v PEG MME 500	0.05 M Cacodilate 7	0.01 M Magnesium Chloride	
C11	35	30 %v/v MPD	0.05 M Cacodilate 7	0.1 M Magnesium Chloride	
C12	36	15 %v/v Isopropanol	0.05 M Cacodilate 7	0.01 M Magnesium Chloride	0.1 M Potassium Chloride
D1	37	30 %w/v PEG MME 350	0.05 M Cacodilate 7	0.01 M Magnesium Chloride	
D2	38	10 %v/v PEG 400	0.05 M Cacodilate 7	0.01 M Magnesium Chloride	0.1 M Potassium Chloride
D3	39	10 %w/v PEG 1000	0.05 M Cacodilate 7		
D4	40	20 %w/v PEG 3000	0.05 M Cacodilate 7	0.2 M Potassium Chloride	
D5	41	2.5 %w/v PEG 8000	0.05 M Cacodilate 7	0.4 M Potassium Chloride	
D6	42	10 %w/v PEG 8000	0.05 M Cacodilate 7	0.01 M Barium Chloride	
D7	43	15 %v/v Ethylene glycol	0.05 M TRIS 7.5		
D8	44	25 %w/v PEG MME 500	0.05 M TRIS 7.5		
D9	45	30 %v/v Propanediol	0.05 M TRIS 7.5		
D10	46	30 %v/v Glycerol	0.05 M TRIS 7.5	0.2 M Lithium chloride	
D11	47	15 %v/v Isopropanol	0.05 M TRIS 7.5	0.01 M Magnesium Chloride	0.2 M Potassium Chloride
D12	48	20 %v/v MPD	0.05 M TRIS 7.5	0.1 M Cesium Chloride	
E1	49	20 %v/v PEG 400	0.05 M TRIS 7.5	0.01 M Magnesium Chloride	0.1 M Potassium Chloride
E2	50	25 %w/v PEG 1000	0.05 M TRIS 7.5		
E3	51	40 %w/v PEG 6000	0.05 M TRIS 7.5	0.2 M Potassium Chloride	
E4	52	20 %w/v PEG MME 2000	0.05 M TRIS 7.5	0.01 M Magnesium Chloride	
E5	53	12.5 %w/v PEG 3350	0.05 M TRIS 7.5	0.2 M Potassium Chloride	
E6	54	10 %w/v PEG 8000	0.05 M TRIS 7.5	5 %v/v Ethylene glycol	
E7	55	20 %v/v DMSO	0.05 M TRIS 8		
E8	56	6 %v/v Glycerol	0.05 M TRIS 8	2.5 M Sodium chloride	
E9	57	20 %w/v PEG 3350	0.05 M TRIS 8	0.05 M Magnesium Chloride	
E10	58	10 %v/v MPD	0.05 M TRIS 8	0.2 M Magnesium Chloride	
E11	59	30 %v/v Ethylene glycol	0.05 M TRIS 8	5 mM Barium Chloride	
E12	60	10 %v/v PEG 200	0.05 M TRIS 8	0.2 M Magnesium Chloride	
F1	61	30 %v/v Tert-butanol	0.05 M TRIS 8	0.2 M lithium chloride	
F2	62	20 %v/v PEG 600	0.05 M TRIS 8	5 %v/v MPD	
F3	63	10 %w/v PEG 1000	0.05 M TRIS 8	0.01 M Magnesium Chloride	
F4	64	20 %w/v PEG 3350	0.05 M TRIS 8	0.01 M Magnesium Chloride	
F5	65	5 %w/v PEG MME 2000	0.05 M TRIS 8	10 %v/v Ethylene glycol	
F6	66	20 %w/v PEG 8000	0.05 M TRIS 8	0.1 M Cesium Chloride	
F7	67	25 %v/v 2,3-Butanediol	0.05 M TRIS 8	0.01 M Magnesium Chloride	
F8	68	1 M Lithium sulfate monohydrate	0.05 M TRIS 8	5 %v/v Ethylene glycol	
F9	69	30 %v/v Ethylene glycol	0.05 M TRIS 8	0.01 M Magnesium Chloride	
F10	70	12.5 %w/v PEG 3000	0.05 M TRIS 8	0.1 M Magnesium Chloride	
F11	71	25 %w/v PEG MME 5000	0.05 M TRIS 8	0.01 M Zinc Acetate	
F12	72	30 %v/v Glycerol	0.05 M TRIS 8	0.01 M Magnesium Chloride	
G1	73	15 %v/v Ethanol	0.05 M TRIS 8.5	6 %v/v Glycerol	
G2	74	20 %v/v MPD	0.05 M TRIS 8.5	0.2 M Potassium Chloride	
G3	75	20 %w/v PEG MME 2000	0.05 M TRIS 8.5	0.1 M Magnesium Chloride	
G4	76	35 %v/v PEG 400	0.05 M TRIS 8.5	0.01 M Magnesium Chloride	
G5	77	10 %v/v PEG 600	0.05 M TRIS 8.5	0.5 M Sodium chloride	
G6	78	20 %w/v PEG MME 350	0.05 M TRIS 8.5	0.05 M Cesium Chloride	
G7	79	10 %w/v PEG 1000	0.05 M TRIS 8.5	0.01 M Zinc Acetate	
G8	80	20 %v/v PEG 4000	0.05 M TRIS 8.5	0.2 M Potassium Chloride	
G9	81	15 %w/v PEG MME 2000	0.05 M TRIS 8.5	0.05 M Cesium Chloride	
G10	82	20 %w/v PEG 3350	0.05 M TRIS 8.5	0.01 M Magnesium Chloride	
G11	83	5 %w/v PEG MME 5000	0.05 M TRIS 8.5	0.01 M Magnesium Chloride	12 %v/v Glycerol
G12	84	5 %w/v PEG 8000	0.05 M TRIS 8.5		
H1	85	1.5 M Sodium chloride	0.05 M CHES 9	5 %v/v MPD	
H2	86	10 %w/v PEG MME 2000	0.05 M CHES 9	0.05 M Magnesium Chloride	
H3	87	10 %v/v 2,3-Butanediol	0.05 M CHES 9		
H4	88	12 %v/v Glycerol	0.05 M CHES 9	0.1 M Cesium Chloride	
H5	89	15 %v/v Isopropanol	0.05 M CHES 9	0.05 M Magnesium Chloride	
H6	90	20 %v/v MPD	0.05 M CHES 9	0.05 M Magnesium Chloride	
H7	91	25 %v/v PEG 200	0.05 M CHES 9		
H8	92	10 %w/v PEG MME 500	0.05 M CHES 9	0.5 M Magnesium Chloride	
H9	93	20 %w/v PEG MME 5000	0.05 M CHES 9	0.01 M Magnesium Chloride	
H10	94	25 %w/v PEG MME 350	0.05 M CHES 9	5 %v/v Isopropanol	
H11	95	10 %w/v PEG 8000	0.05 M CHES 9	6 %v/v Glycerol	
H12	96	12.5 %w/v PEG 8000	0.05 M CHES 9	0.2 M Magnesium Chloride	

Table 5. List of crystallization conditions of PAC10.

3.4.2 X-ray diffraction

All diffraction data were collected in ALBA synchrotron at BL13-XALOC beamline using native protein crystals. Located in Cerdanyola del Vallès near Barcelona, ALBA is a third generation 3-GeV synchrotron light facility (Figure 26). The BL13-XALOC beam has two different configurations: the unfocused one and the focused one. In the unfocused configuration, the beam is highly collimated (0.03 mrad) and hence it can be used for crystals with very large unit cells like protein-protein complexes and viruses. In the focused configuration, the mirrors along the beam path allow the focus of the beam on small crystals maintaining a small vertical divergence (0.1 mrad).

Figure 26. Aerial photo of ALBA synchrotron (http://www.lightsources.org).

The end-station of the beam is constituted by the detector, the diffractometer and the beam-conditioning elements (Figure 27). The detector is a fast read-out Dectris Pilatus 6M detector (a 360° 1.1 Å resolution dataset can be collected in 33 seconds). The diffractometer is a MD2M-Bruker diffractometer that is endowed with Mini-Kappa, a system that allows a specific re-orientation of the crystal. The BL13-XALOC beam

conditioning elements are located in front of the sample and are constituted by: 12 ESRF-type foil attenuators, DECTRIS position monitors, JJ-Xray ESRF type slits, a FPS400 CEDRAT fast shutter and a kapton window. The sample changer, an Irelec CATS (Cryogenic Automated Transfer System), which is suitable for both SPINE pins and crystallization plates, allows a high-throughput data collection. The sample storage DEWAR contains 9 cassettes with 10 samples each and hence has a maximum capacity of 90 samples. Diffraction characterization and data collection strategy were performed using EDNA program.

Figure 27. BL13-XALOC end-station.

3.4.3 Structural determination

Diffraction data were integrated with XDS (Kabsch 2010) and scaled and merged with SCALA (Evans 2006). We used the molecular replacement method to solve all the structures showed in this thesis manuscript. These calculations were carried out using Phaser (McCoy *et al.* 2007).

3.4.4 Refinement of the three-dimensional structure

We performed multiple rounds of refinement with Refmac (Murshudov *et al*. 2011) followed by manual rebuilding using Coot (Emsley *et al*. 2004).

3.4.5 Validation of the final model

We used Molprobity (Chen *et al.* 2010) to validate the final model of the three-dimensional structure. Ramachandran plots were generated using Procheck (Laskowski *et al*. 1993).

3.4.6 Representation and analysis of structural data

Structural similarity searches were carried out using the PDBefold server (Protein structure comparison service PDBeFold at European Bioinformatics Institute) (Krissinel 2007) (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver). Protein-protein interactions of ToxRDNA40 structure were detected and evaluated using PISA (Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute) (Krissinel and Henrick 2007) (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). ESPript 3.0 was used to represent multiple sequence alignment and secondary structure of ToxR-DBD (Robert and Gouet 2014) (http://espript.ibcp.fr). Protein-DNA interactions were analyzed with Nucplot (Luscombe *et al.* 1997). DNA conformations of the three structures were analyzed with 3DNA, which allowed us to get a precise estimation of major and minor grove widths and other DNA parameters (Lu and Olson 2003). All figures were made using Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

Chapter 4: Results

4.1 ToxR construct design

Vibrio cholerae ToxR 6-113 (Uniprot reference: W0B3Z6), which will be referred to as ToxR-DBD in this dissertation, was expressed and purified in the laboratory of Prof. Eric Krukonis, at the University of Detroit Mercy. The full-length protein is made of a cytoplasmic domain (1-170), a transmembrane helix (171-186) and a periplasmic domain (187-282) (Figure 28).

```
         10         20         30         40         50
MSHIGTKFIL AEKFTFDPLS NTLIDKEDSE EIIRLGSNES RILWLLAQRP
         60         70         80         90        100
NEVISRNDLH DFVWREQGFE VDDSSLTQAI STLRKMLKDS TKSPQYVKTV
        110        120        130        140        150
PKRGYQLIAR VETVEEEMAR ESEAAHDISQ PESVNEYAES SSVPSSATVV
        160        170        180        190        200
NTPQPANVVA NKSAPNLGNR LFILIAVLLP LAVLLLTNPS QSSFKPLTVV
        210        220        230        240        250
DGVAVNMPNN HPDLSNWLPS IELCVKKYNE KHTGGLKPIE VIATGGQNNQ
        260        270        280
LTLNYIHSPE VSGENITLRI VANPKDAING CE
```

Figure 28. Sequence of full-length ToxR.

Our protein construct is a portion of the cytoplasmatic domain that is supposed to include the DNA binding domain (Figure 29).

```
         10         20         30         40         50
      MKFIL AEKFTFDPLS NTLIDKEDSE EIIRLGSNES RILWLLAQRP
         60         70         80         90        100
NEVISRNDLH DFVWREQGFE VDDSSLTQAI STLRKMLKDS TKSPQYVKTV
        110
PKRGYQLIAR VET
```

Figure 29. Sequence of ToxR construct (6-113), which will be mentioned as ToxR-DBD in this thesis manuscript.

4.2 ToxR sample preparation

In our laboratory, we performed an analytical size-exclusion chromatography using a Superdex 200 (10-300) column. The protein showed a single and symmetrical peak. The protein, which has a molecular weight of 12943 Da, eluted as a dimer according to the column calibration with standards of known molecular weight (Figure 30).

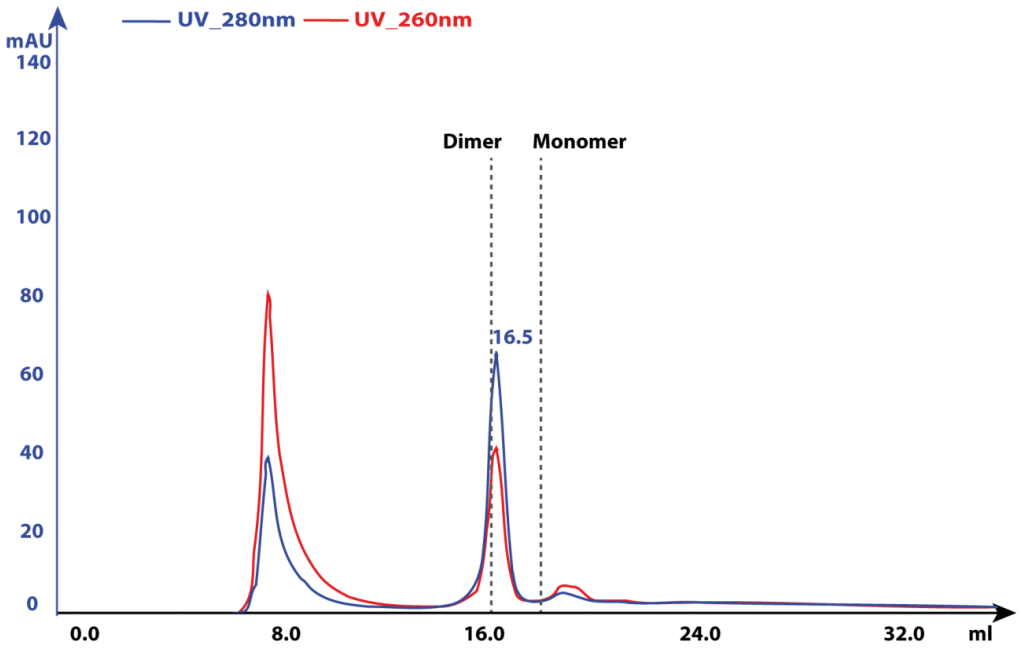

Figure 30. ToxR Superdex 200 (10-300) gel filtration and expected elution volumes for its dimeric and monomeric states (dashed lines).

4.3 ToxR in complex with a 20-bp oligonucleotide

We designed a 20-bp oligonucleotide taking advantage of the previous studies done at Eric Kroukonis' laboratory, in particular the footprinting experiments done with ToxR on the *toxT* promoter (see section 1.1.10). Since the footprinting analysis showed a ToxR binding site going from -104 to -68, we tried to get a complex of ToxR with a 20-bp oligonucleotide (DNA20) containing the specific sequence of the *toxT* promoter from C-97 to C-78 (Figure 31). We named this complex ToxRDNA20.

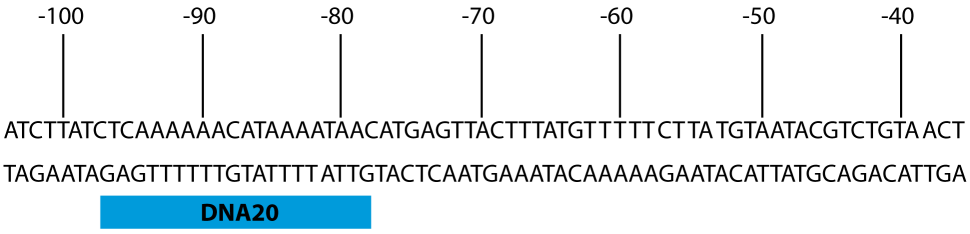

Figure 31. DNA20 sequence comprises the sequence of the *toxT* promoter from C-97 to C-78.

4.3.1 ToxRDNA20 complex formation

We incubated the protein and the DNA with a [1:1.2] ratio, and then we loaded the mixture into a Superdex 200 size-exclusion chromatography column. The ToxRDNA20 complex eluted as a single peak, showing a nice profile and confirming the specificity of the protein for the sequence (Figure 32).

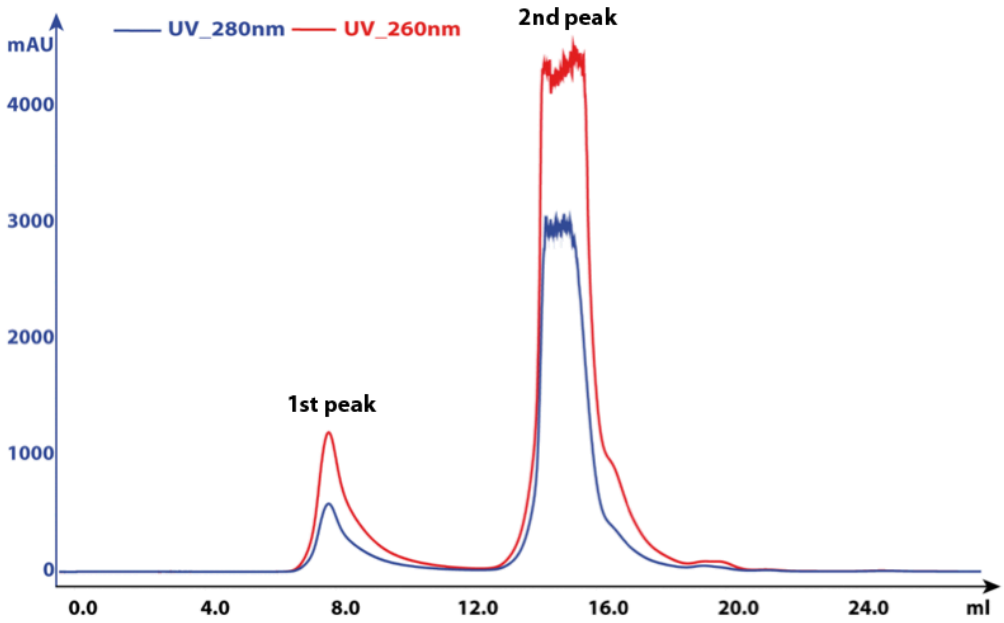

Figure 32. Peak of ToxRDNA20 complex after gel filtration using a Superdex 200 (10-300) column.

We verified by SDS-PAGE that all the fraction of the second peak corresponded to the ToxRDNA20 complex and not to DNA alone (Figure 33).

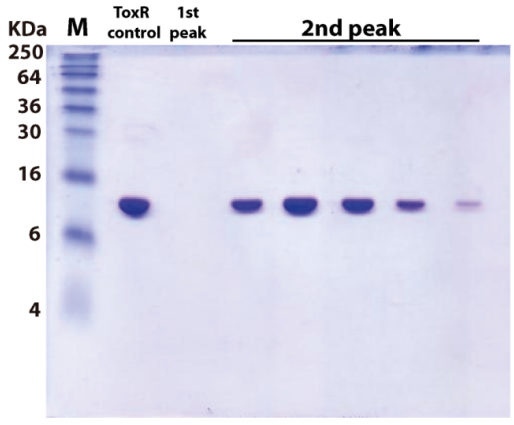

Figure 33. SDS-PAGE analysis of purified ToxRDNA20 complex.

4.3.2 Crystallization of the ToxRDNA20 complex

We obtained hits of interest from different crystallization screening conditions: PAC1-A10, PAC1-B11, PAC1-C6 and PAC10-G3 at 20°C. We successfully scaled up to microdrops the first condition (30% PEG4000, 0.1 M sodium acetate pH 4.6 and 0.2 M ammonium acetate) and were able to get elongated plate-shaped crystals after optimization of the crystallization condition: 30% PEG4000, 0.1 M sodium acetate pH 4, 0.15 M ammonium acetate (Figure 34).

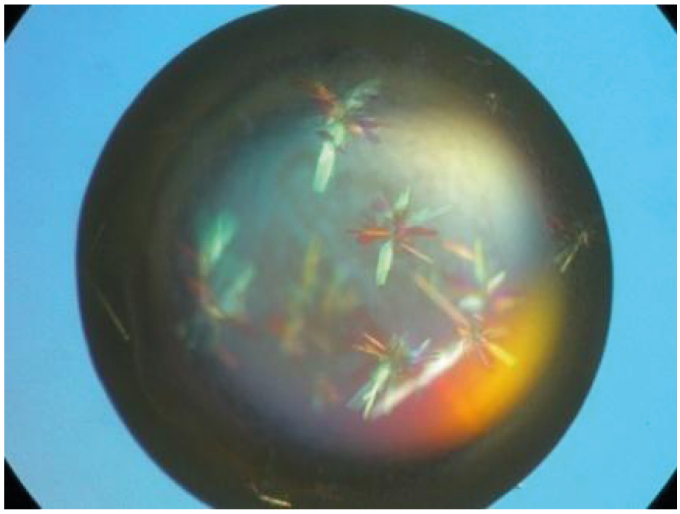

Figure 34. ToxRDNA20 crystals after optimization.

4.3.3 Data collection and structure determination of ToxRDNA20

In total, we tested thirty ToxRDNA20 crystals at Alba Synchrotron BL13-XALOC beamline and collected twenty data sets. The best crystal, measuring 0.27 x 0.08 x 0.02 mm^3, diffracted up to 2.1 Å (diffraction pattern shown in Figure 35).

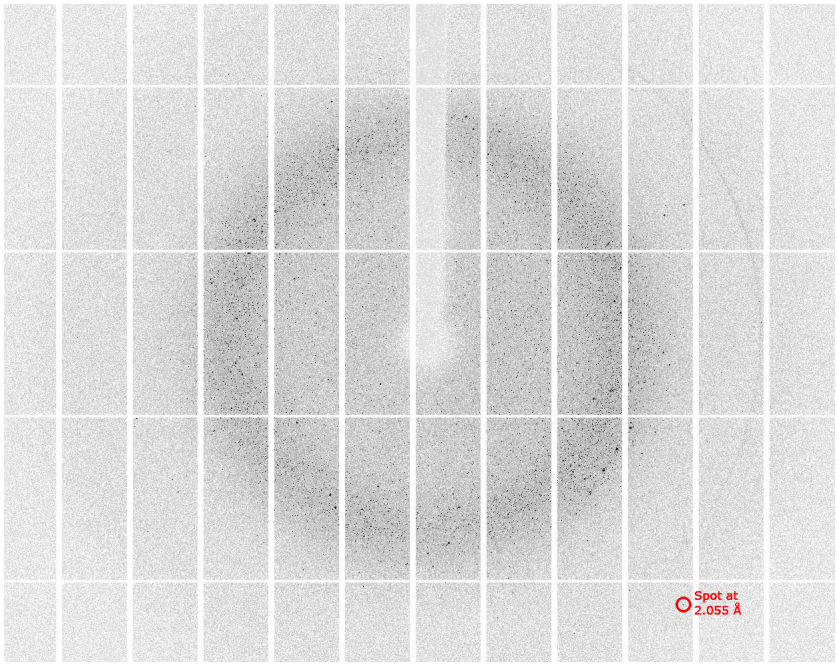

Figure 35. Diffraction pattern of a ToxRDNA20 crystal that diffracted up to 2.1 Å.

Since ToxR was homologous to the *E.coli* PhoB DNA binding domain (25.0% identity, 41.4% similarity) that had been previously characterized by our research group (Blanco *et al.* 2002), we aimed at solving the structure by molecular replacement. Data collection and processing statistics for this crystal are detailed in Tables 6 and 7. All the analyzed crystals had the same space group and similar cell dimensions.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
40.00	3	1	2	50.0%	718.9%	753.5%	3	-0.54	880.4%	0.0	0	0.000	0
10.00	515	144	149	96.6%	4.1%	3.8%	507	29.59	4.8%	99.5*	1	0.840	71
6.00	1951	500	500	100.0%	4.1%	4.2%	1944	27.65	4.8%	99.8*	3	0.873	295
4.00	6048	1513	1517	99.7%	5.1%	4.8%	6024	25.89	5.9%	99.6*	-1	0.851	950
3.00	12133	2882	2900	99.4%	7.0%	6.9%	12100	18.98	8.0%	99.7*	-1	0.814	1956
2.50	15489	3645	3652	99.8%	19.5%	19.3%	15440	9.31	22.1%	98.7*	-5	0.770	2459
2.30	10071	2450	2454	99.8%	32.9%	32.2%	10029	4.93	37.6%	95.2*	-1	0.758	1545
2.00	23839	5707	5727	99.7%	68.4%	68.0%	23732	2.39	78.1%	85.4*	-7	0.679	3705
1.90	12029	2798	2802	99.9%	143.2%	150.2%	11972	1.15	162.7%	62.0*	0	0.632	1933
1.80	12513	3428	3438	99.7%	238.7%	250.6%	12370	0.58	278.4%	29.3*	-3	0.587	1589
total	94591	23068	23141	99.7%	11.0%	10.9%	94121	7.67	12.6%	99.7*	-3	0.721	14503

Table 6. Data processing outputs for the best dataset of ToxRDNA20.

Parameters	**Value**
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	0.979
Space group	$P2_1$
Cell dimensions	36.50 Å 36.50 Å 36.50 Å 90.00° 92.58° 90.00°
Number of unique reflections	15359
Resolution range (Å)- (overall/last shell)	63.59-2.07/2.14-2.07
R_{merge} (%) (overall/last shell)	10.55/101.60
Multiplicity (overall/last shell)	7.8/7.4
Completeness (%) (overall/last shell)	99.82/99.74
<I>/σ<I> (overall/last shell)	16.32/2.71
Wilson B-factor ($Å^2$)	31.19

Table 7. Data collection and processing statistics for ToxRDNA20.

In Table 7, we report the R_{merge}, an indicator of data quality, which is defined by the following equation:

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{n} |I(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I(hkl)}$$

We determined the structure of ToxRDNA20 by molecular replacement method using Phaser (Mccoy *et al*., 2007). As searching models, we used the DNA binding domain of PhoB (monomer E) taken from the structure of PhoB in complex with DNA (pdb code:

1GXP) and an idealized double strand DNA of 10 bp. Calculations gave one solution with a high TFZ score (9.5), confirming $P2_1$ as the correct special group. We grew the DNA strands manually, and refined the model by successive cycles of manual building of the model using Coot and restrained refinement using Refmac. Rfactor and Rfree converged to 18% and 24%, respectively, after the addition of solvent molecules. ToxRDNA20 geometry was validated using Molprobity, which assigned to the structure an overall score of 1.49, an optimal result considering the maximal resolution of the structure (2.1 Å). Table 8 summarizes the refinement statistics.

Parameter	**Value**
Resolution range (Å)	63.59-2.07
Number of protein/DNA atoms	1700
Number of solvent molecules	177
RMSD for bonded angles $_{(1)}$ (°)	1.76
RMSD for bond lengths $_{(1)}$ (Å)	0.014
Rfactor/Rfree	0.1822/0.2446
Average B-factor (Å^2)	37.70
Ramachandran favored (%)	96.0
Ramachandran ouliers	0
Molprobity score	1.49

Table 8. Refinement statistics for ToxRDNA20. Rfactor= $\Sigma|Fobs - Fcalc| / \Sigma|Fobs|$. $_{(1)}$ RMSD from target values.

The final model showed a continuous and well-defined electron density for all the structure (Figure 36).

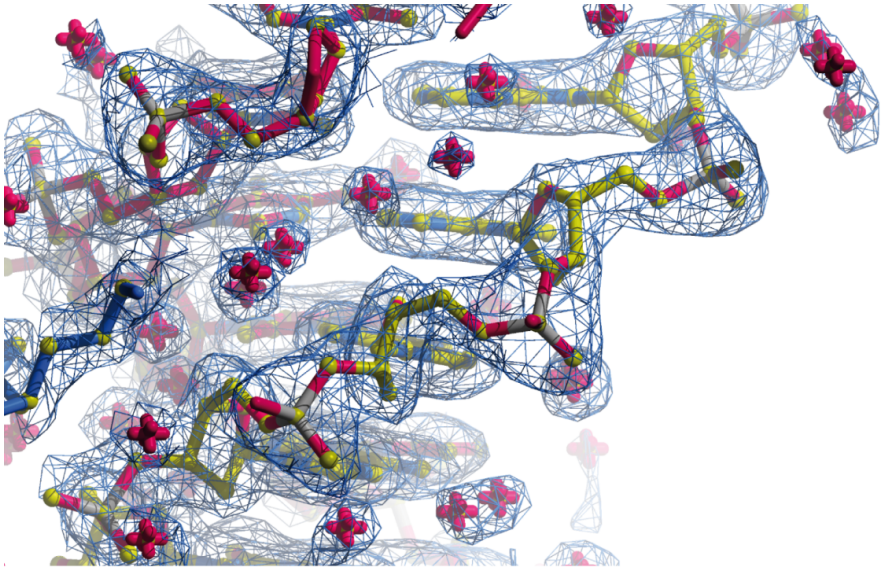

Figure 36. 2 F_o-F_c electron density map (contoured at 1.0 σ) with the final ToxRDNA20 model fitted in the density.

All the residues were located in the allowed regions of the Ramachandran plot (Figure 37).

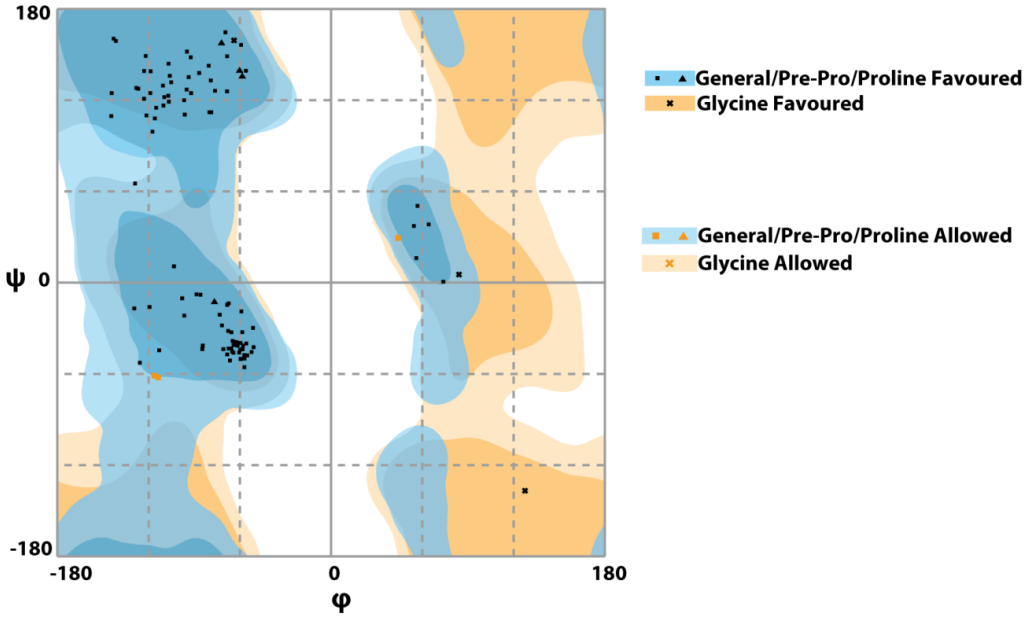

Figure 37. Ramachandran plot for ToxR-DBD in the ToxRDNA20 structure.

4.3.4 ToxR-DBD fold

ToxR DNA binding domain shows a heart-shaped α/β fold structure and is constituted by an N-term five-stranded mixed β-sheet formed by four antiparallel β strands (β1-β4) and one β strand parallel to β1 (β8), followed by a three-helix bundle (α1-α3) and a C-term β-hairpin (β6-β7) (Figure 38). α1-α3 helices form the modified helix-turn-helix DNA-binding motif, where α3 is the recognition helix and the α2-α3 connector replaces the turn. There is a short β strand (β5) located between α1 and α2 that interacts with the C-terminal hairpin forming a three-stranded antiparallel β-sheet. The turn between the two β strands of the C-term β-hairpin forms the wing, which is responsible for the recognition of the DNA minor groove. According to this folding type, ToxR-DBD can be classified as a winged-helix protein. However, when it is compared with the pdb structures of the other proteins from this family ToxR-DBD shows a unique turn between the recognition helix α3 and β4, which we called "secondary wing".

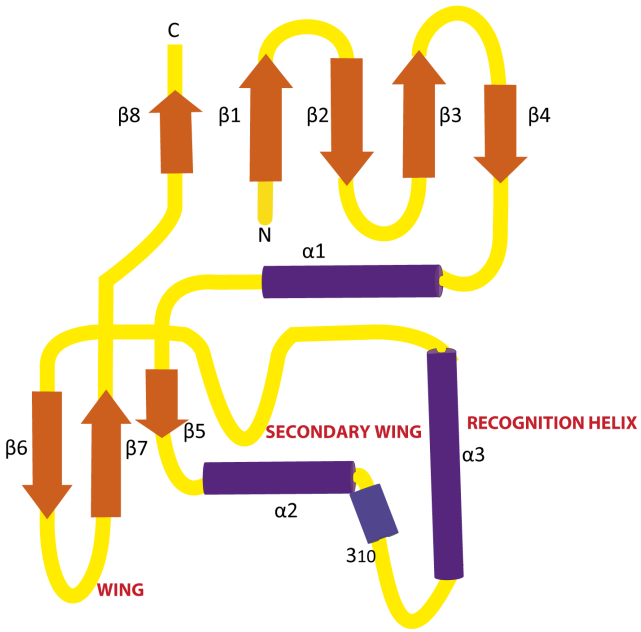

Figure 38. Topology diagram of ToxR-DBD.

In Figure 39, the overall structure of ToxR-DBD is shown using a cartoon representation and the secondary structure elements are highlighted. If we look at the tertiary structure of the protein, a central hydrophobic core is responsible for maintaining the fold of the protein. Helix α1 is connecting the N-terminal β-sheet with the helix-turn-helix motif and the β-hairpin by the residues Ile42, Leu43, Trp 44, Leu45 and Leu46.

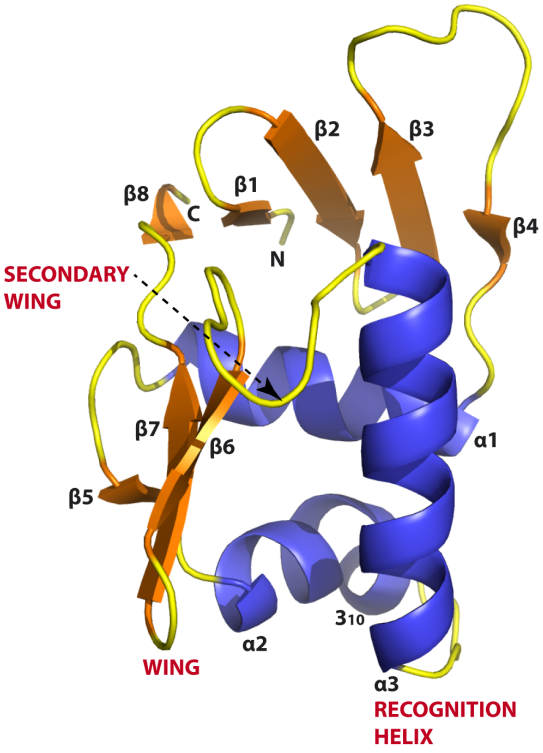

Figure 39. Overall structure of ToxR-DBD.

In Figure 40, we show the multiple sequence alignment of ToxR-DBD with the DNA binding domain of *E. coli* PhoB and *V. cholerae* TcpP elaborated using ESPript 3.0 program. The secondary structure elements present in ToxR-DBD are also indicated.

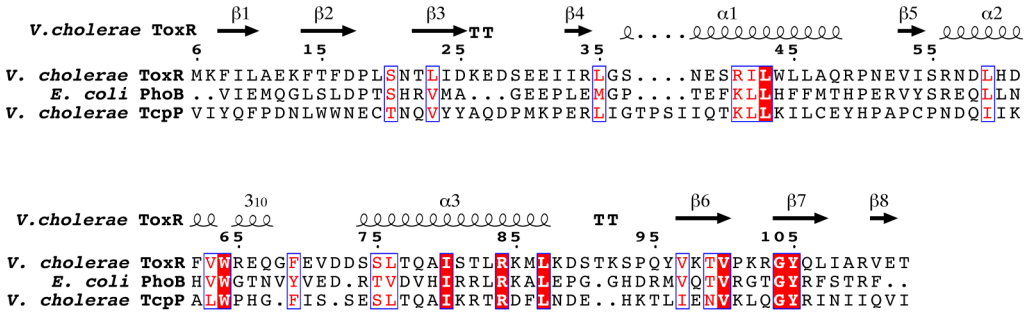

Figure 40. Multiple sequence alignment and secondary structure of ToxR-DBD formatted with ESPript 3.0.

In particular, when we compare ToxR-DBD to PhoB DNA binding domain, we can see that the two proteins share the most important structural and functional elements (the recognition helix and the wing) with the exception of the secondary wing (Figure 41). Differently to ToxR, in PhoB a 3_{10} helix is located between α3 and β6 instead of the secondary wing. Another difference between the two structures is that in ToxR-DBD we found an extra parallel β-strand (β8) in the C-term that is forming a wider β-sheet with the four β strands located in the N-term. In PhoB, this β-sheet is needed for contacting the receiver domain of PhoB but ToxR has no receiver domain. Possibly, the different orientation of the C-term and the presence of a large loop between β3 and β4 in ToxR-DBD can imply a diverse function of the N-term β-sheet in the two proteins. Moreover, in ToxR-DBD, α2 is more extended than in PhoB since it continues with a 3_{10} helix. This makes the α2-α3 loop shorter in ToxR-DBD. In PhoB, the α2-α3 loop, named transactivation loop, is essential for transcription since it interacts with the σ^{70} subunit of the *E. coli* RNA polymerase.

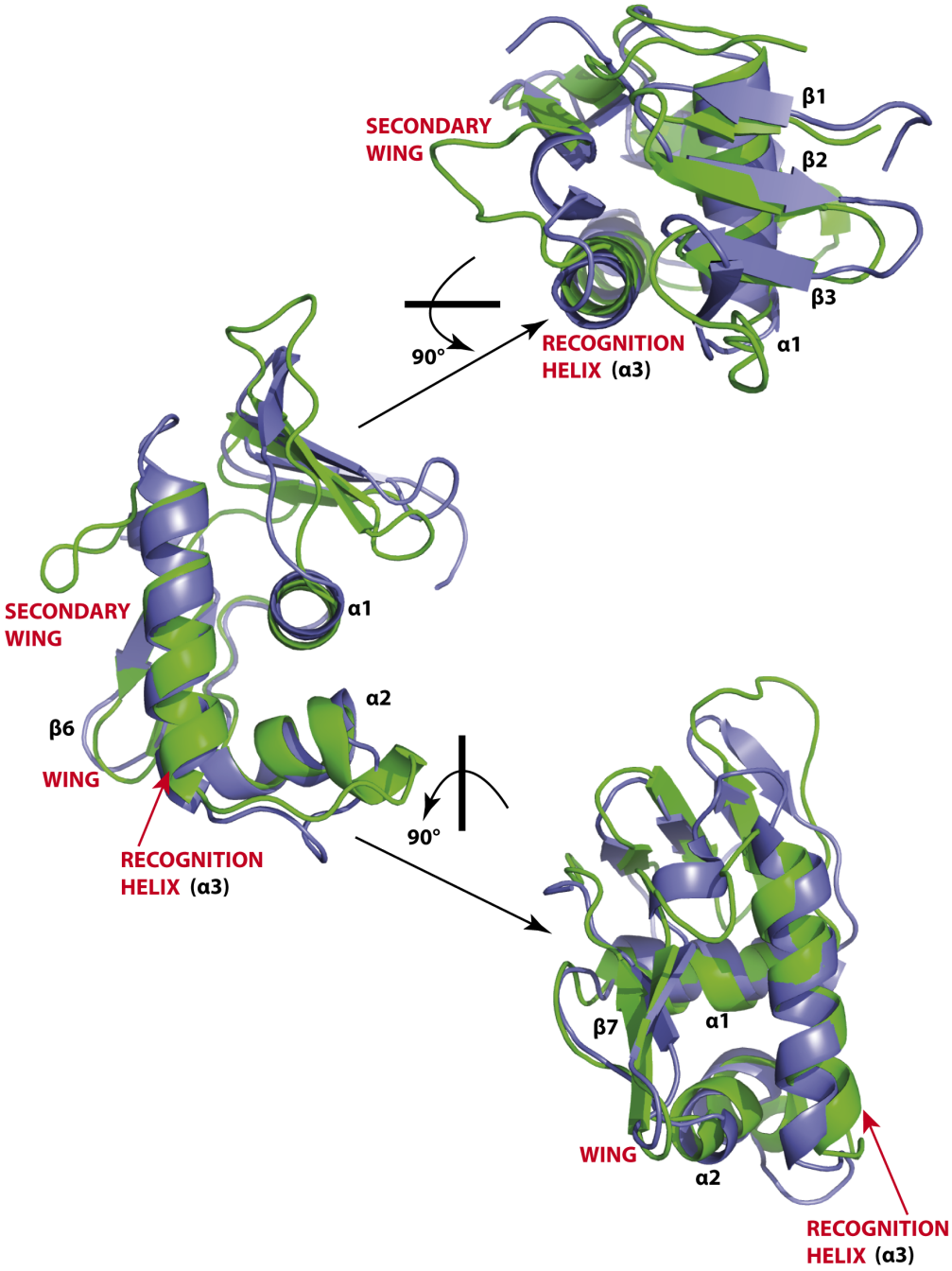

Figure 41. Superposition of ToxR-DBD (green) and the DNA binding domain of *E.coli* PhoB (blue).

4.3.5 Global description of the protein-DNA interactions in the ToxRDNA20 complex structure

In ToxRDNA20 one molecule of ToxR-DBD binds the DNA through three essential elements: the wing domain, the recognition helix and the secondary wing (Figures 42 and 43).

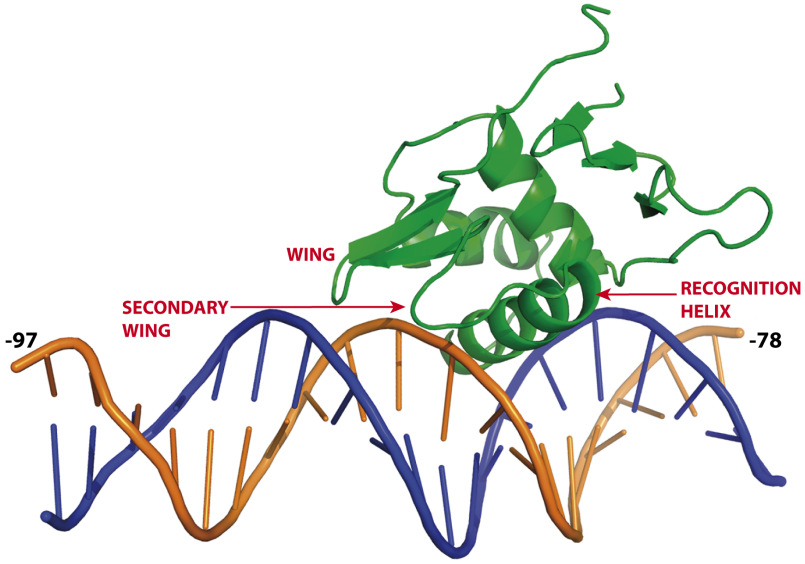

Figure 42. Overview of the structure of the ToxRDNA20 complex.

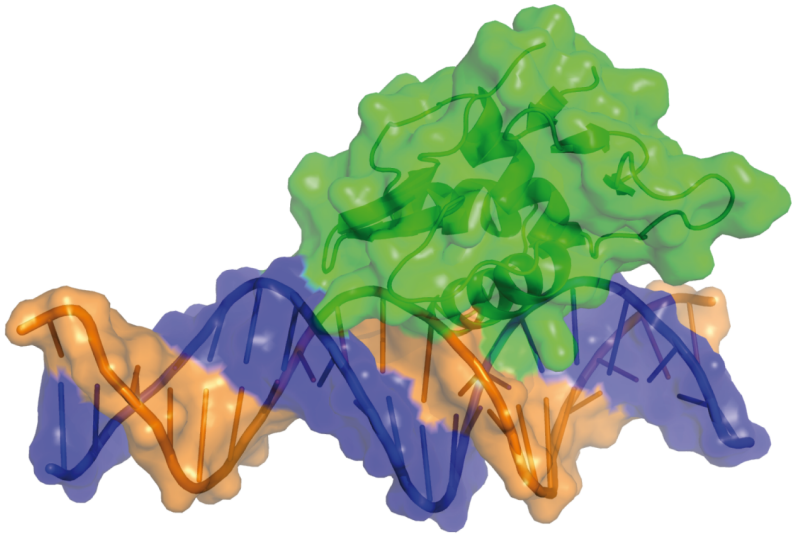

Figure 43. Surface representation of the ToxRDNA20 complex.

4.3.6 DNA-protein interactions

A plot of DNA-protein interactions calculated using Nucplot (Luscombe N.M. *et al.*, 1997) is shown in Figure 46. The wing displays important interactions with the minor groove through residues Thr99 (H-bond between γ-OH of Thr99 and the phosphate of A-87 of chain A and a Van der Waals' interaction between the γ-CH3 of Thr99 and the ribose of C-88 of DNA, chain A), Pro101 (Van der Waals' interaction between the side chain of Pro101 and the ribose of T-91, chain B) and Tyr105 (H-bond between η-OH of Tyr105 and the phosphate of A-87 of DNA, chain A). Remarkably, Lys102 forms an H-bond between the N of the main chain and the phosphate of C-88 (DNA, chain A) and a second H-bond between the N of the side chain and the phosphate of T-93 (chain B) (Figures 44 and 46).

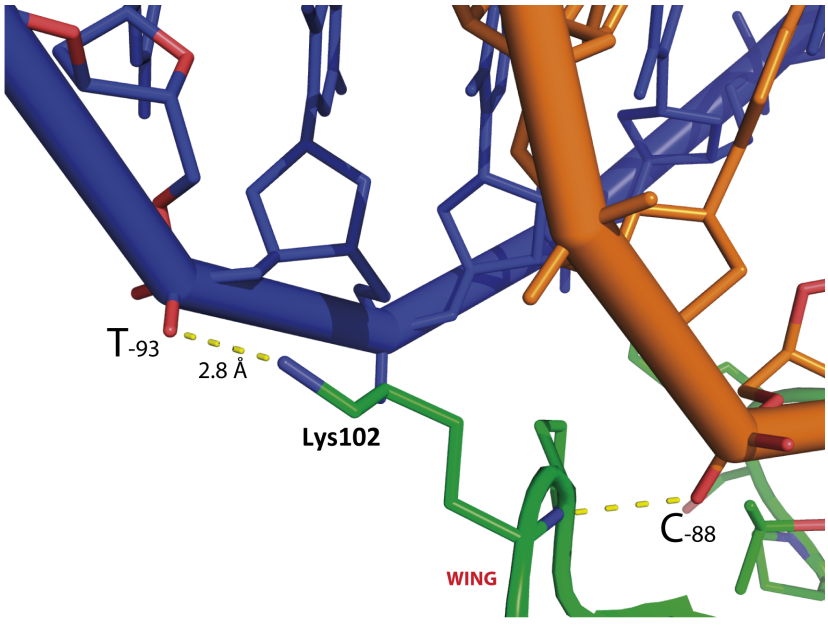

Figure 44. Molecular details of the contacts between Lys102 and DNA.

The majority of the interactions are between the recognition helix and four DNA bases, from C-88 to A-85, a region that we named the CATA box. The recognition helix interacts

with the major groove through eight residues: Asp72, Asp73, Ser74, Ser75, Thr77, Gln78, Ser81 and Arg84. Above all, Gln78 seems to play an essential role in the recognition of this specific DNA sequence since it is able to interact directly with the DNA bases (Figures 45 and 46). It forms an H-bond between the δ-C=O and the N6 of adenine A-85 (chain A). Moreover, it forms two water mediated H-bonds: a first H-bond between the ε-NH2 of the side chain and the N9 of adenine of A-85 (chain A) mediated by water W126 and a second H-bond between the ε-NH2 of the side chain and the O2 of the thymine of T-83 (chain B) mediated by water W99.

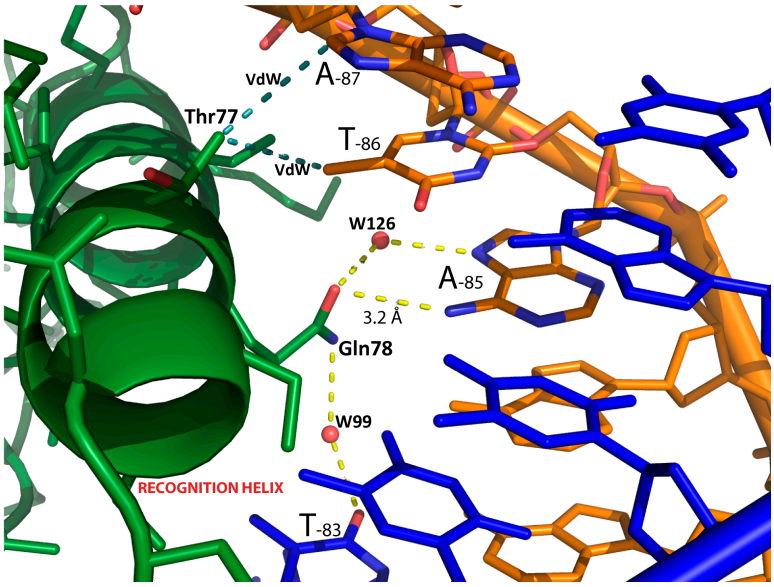

Figure 45. Molecular details of the contacts between two key residues of the recognition helix, Thr77 and Gln78, and the DNA.

In the recognition helix another key residue is Thr77, which is able to perform direct interactions with the DNA, displaying Van der Waals' interactions between the γ-CH3 and the adenine of A-87 (chain A) and thymine of T-86 (chain A). The secondary wing is also interacting with the major groove through the contacts between the γ-OH of Thr91 and the phosphate of T-86, chain A (H-bond) and between the γ-CH3 of Thr91 and the ribose of A-87, chain A (Van der Waals' interaction). Other residues that display interactions with DNA are Ser37, Asn38, Glu39, Trp64, Val71. Analysis of DNA-protein interaction allowed us to define the ToxR binding site that goes from A-92 to T-81.

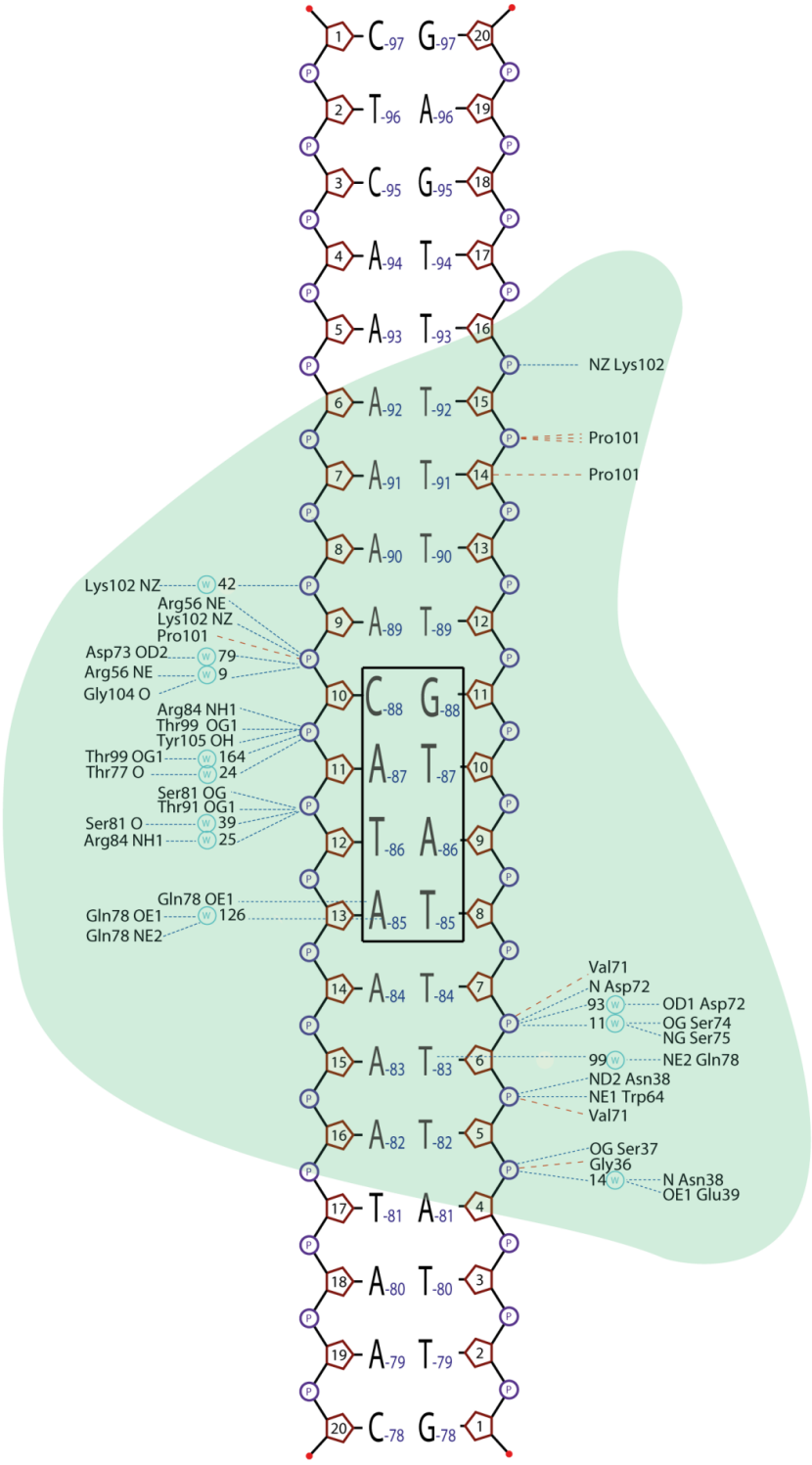

Figure 46. Interaction scheme of ToxR-DBD with DNA in the ToxRDNA20 structure. The CATA box and the ToxR binding site are highlighted.

4.3.7 DNA conformation

The DNA has a B conformation and is almost straight. It presents a slight curvature (4.9°) that is caused by the compression of the minor groove in a A-T rich sequence region from A-92 to A-89, resulting in the formation of a typical hydration spine which had been firstly described by Drew and Dickerson (Drew H.R *et al*. 1981) (Figure 36). The minor groove narrows up to 8.9 Å between these atoms (Figure 47) and allows the suitable fitting of Lys 102, which enters in the minor groove and displays important contacts with DNA (Figure 44) without affecting the spine of hydration. Moreover, there is a slight compression of the major groove that narrows up to 17.9 Å in the CATA box region (Figure 47). This region is where the recognition helix displays most of the contacts with DNA (Figure 45) and thus ToxR might contribute to stabilize the compression. All the geometrical parameters of the DNA are presented in Tables 9 and 10.

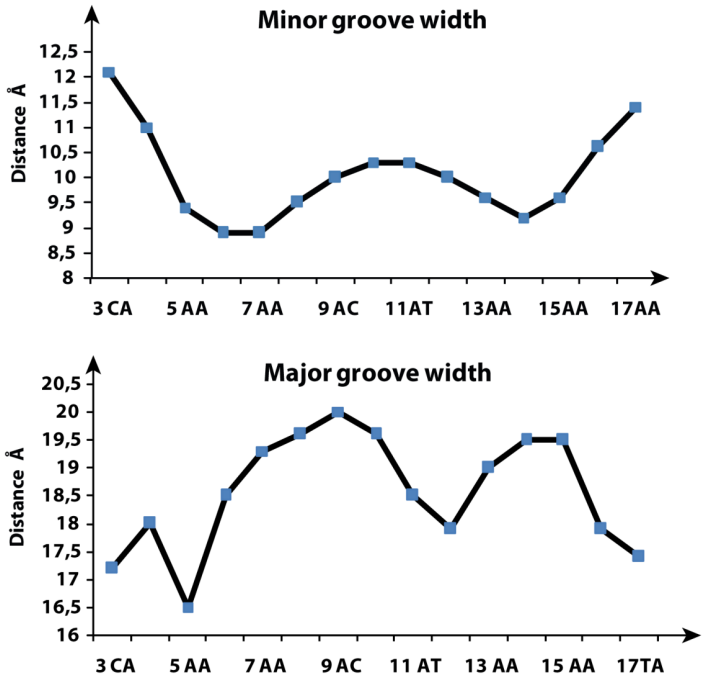

Figure 47. Minor groove and major groove width plots for the ToxRDNA20 structure. Minor and major groove widths are measured calculating the distance between phosphates. In the x-axis we indicate the number of the phosphate of the chain A and the corresponding DNA bases that are located before and after the phosphate.

bp	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
1 C-G	-0.21	-0.27	0.02	-4.77	-4.78	-4.33
2 T-A	-0.15	-0.28	0.33	-3.60	-11.18	-6.87
3 C-G	0.26	-0.26	-0.22	6.21	-4.88	-0.05
4 A-T	-0.20	-0.05	-0.22	5.88	-13.44	2.82
5 A-T	0.07	0.13	0.29	12.82	-14.64	3.21
6 A-T	-0.06	-0.04	0.25	6.48	-15.10	5.04
7 A-T	-0.13	-0.13	0.02	2.95	-13.37	0.65
8 A-T	-0.03	-0.07	0.11	3.54	-13.48	2.50
9 A-T	0.07	-0.01	0.03	4.58	-10.43	4.55
10 C-G	0.20	-0.14	-0.14	7.61	-8.34	-0.97
11 A-T	0.03	-0.10	-0.08	-5.31	-1.69	3.88
12 T-A	0.01	-0.14	-0.01	2.67	-5.98	-0.98
13 A-T	0.01	-0.08	0.05	-4.83	-13.91	5.06
14 A-T	0.20	-0.05	0.18	-0.81	-12.82	2.57
15 A-T	-0.05	0.00	0.05	-3.07	-12.45	5.50
16 A-T	-0.05	-0.18	0.17	-2.38	-12.45	5.20
17 T-A	-0.20	-0.15	0.11	-5.96	-13.54	3.00
18 A-T	0.21	-0.13	-0.16	-3.28	-7.40	-1.71
19 A-T	-0.13	-0.14	-0.12	6.46	-9.48	-7.93
20 C-G	0.33	-0.27	-0.22	3.25	-10.80	-8.16
Average	0.01	-0.12	0.02	1.42	-10.51	0.65
S.D.	0.16	0.10	0.17	5.37	3.81	4.45

Table 9. Local base-pair parameters for the ToxRDNA20 structure.

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
1 CT/AG	-0.61	0.39	3.44	-3.09	-5.54	39.91
2 TC/GA	1.14	1.38	3.18	5.66	-3.45	38.17
3 CA/TG	-0.86	1.09	3.42	-2.85	5.30	34.07
4 AA/TT	0.06	0.05	3.05	-4.15	0.14	36.17
5 AA/TT	0.18	0.05	3.38	-0.07	-4.18	38.88
6 AA/TT	-0.02	-0.42	3.36	1.98	-3.35	34.35
7 AA/TT	0.32	-0.44	3.23	0.61	-4.12	37.41
8 AA/TT	0.18	-0.59	3.21	0.34	-1.68	35.84
9 AC/GT	0.04	-0.82	3.18	-0.61	-0.07	35.03
10 CA/TG	-0.18	-0.95	3.50	-0.76	4.85	36.58
11 AT/AT	-0.58	-0.46	3.08	0.15	-2.69	31.25
12 TA/TA	0.12	-0.11	3.35	-0.13	1.80	40.52
13 AA/TT	0.15	-0.41	3.07	-0.06	-2.04	36.82
14 AA/TT	-0.05	-0.61	3.29	0.74	-1.40	31.92
15 AA/TT	-0.22	-0.43	3.14	-1.66	-1.52	36.32
16 AT/AT	-0.05	-0.60	3.34	-0.04	-0.69	32.01
17 TA/TA	0.35	0.47	3.29	3.31	2.23	40.08
18 AA/TT	0.02	-0.02	3.10	2.30	1.70	29.29
19 AC/GT	0.25	-0.22	3.33	1.87	-1.97	42.17
Average	0.01	-0.14	3.26	0.19	-0.88	36.15
S.D.	0.42	0.61	0.13	2.29	2.97	3.45

Table 10. Local base-pair step parameters for the ToxRDNA20 structure.

Notably, the DNA shows high values of propeller-twist in the A-T rich regions of the sequence of the *toxT* promoter.

4.4 ToxR in complex with a 40-bp oligonucleotide

In order to gain further information on the binding of ToxR to the *toxT* promoter, we aimed at obtaining a complex of ToxR-DBD with a 40-bp oligonucleotide (DNA40) containing the specific sequence of the *toxT* promoter from T-96 to C-57 (Figure 48). We substituted the first thymine with a cytosine to increase the stability of the ds DNA. We named this complex ToxRDNA40.

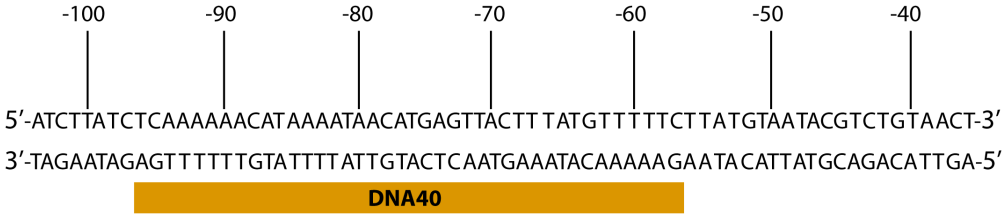

Figure 48. DNA40 sequence comprises the sequence of the *toxT* promoter from T-96 to C-57

4.4.1 ToxRDNA40 complex formation

We incubated the protein and the DNA with a 1:1.2 ratio, and then we loaded the reaction product into a Superdex 200 size-exclusion chromatography column. The ToxRDNA40 complex eluted showing a strong peak, confirming the specificity of the protein for the sequence. We verified by SDS-PAGE that all the fractions of the second peak corresponded to the ToxRDNA40 complex and not to DNA alone (Figure 49) and we collected the fractions 21-25.

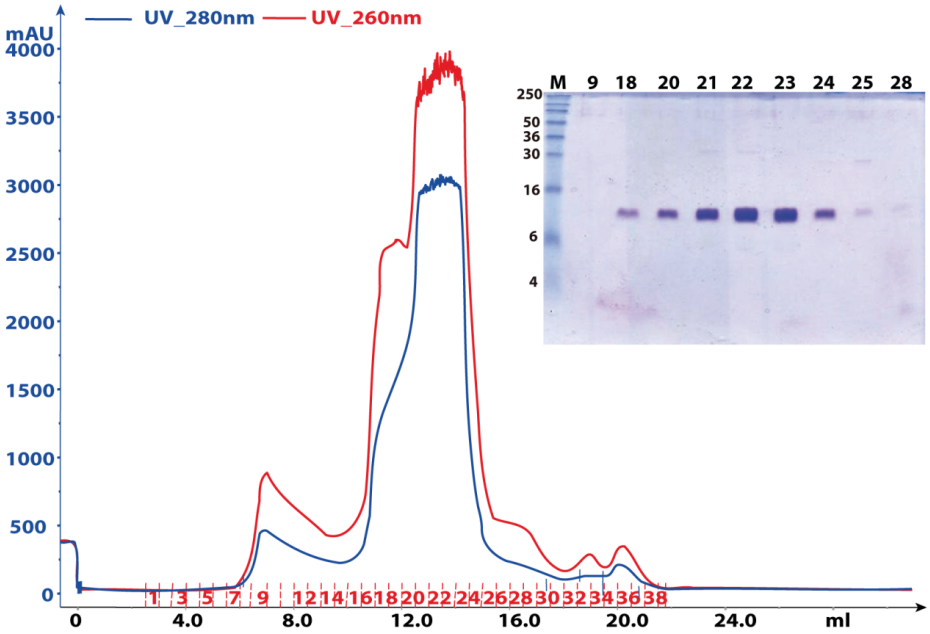

Figure 49. Peak of ToxRDNA40 complex after gel filtration using a Superdex 200 (10-300) column and the SDS-PAGE analysis of the purified complex.

4.4.2 Crystallization of the ToxRDNA40 complex

We obtained hits from different crystallization screening conditions in robot nanodrops and we successfully scaled up to microdrops the condition PAC1-B9. We were able to get hexagonal-shaped crystals after optimization of the crystallization condition (35% MPD, 0.3 M magnesium acetate, 0.1 M Hepes pH 7.0) (Figure 50).

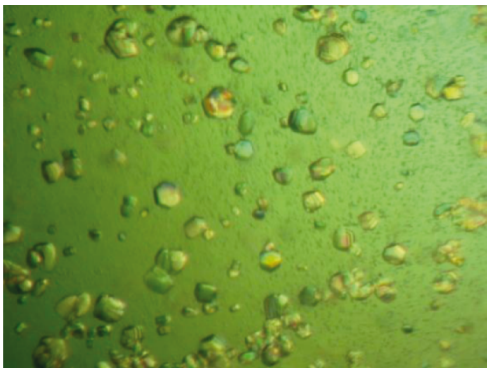

Figure 50. ToxRDNA40 optimized crystals.

We tested the ToxRDNA40 crystals at Alba Synchrotron BL13-XALOC beamline and we got the typical diffraction pattern of DNA fibers, indicating that these crystals were made of only DNA and not the ToxRDNA40 complex (Figure 51).

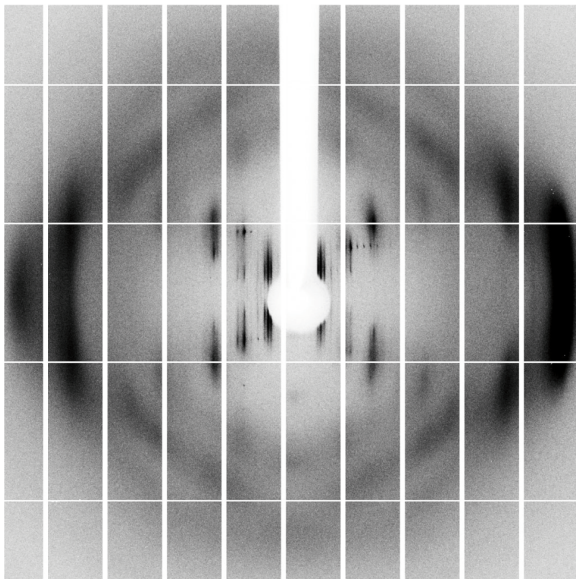

Figure 51. Diffraction pattern of a crystal grown from a ToxRDNA40 solution but diffracting as DNA fibers.

Then, we focused on another promising condition and scaled it up to microdrops. After optimization, we were able to get large bar-shaped crystals using the following crystallization condition: 32% PEG MME 500, 0.175 M sodium chloride, 0.1 M glycine pH 8.5 (Figure 52).

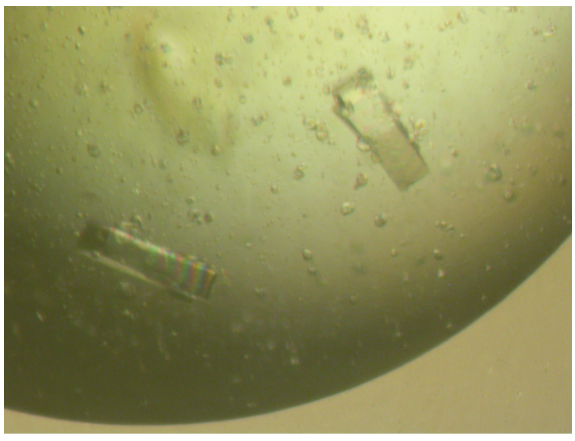

Figure 52. ToxRDNA40 crystals after optimization.

4.4.3 Data collection and structural determination of ToxRDNA40

In total, we tested twenty ToxRDNA40 crystals at Alba Synchrotron BL13-XALOC beamline and collected seven data sets. The best crystal measuring 0.32 x 0.09 x 0.06 mm^3, diffracted up to 2.6 Å resolution (diffraction pattern shown in Figure 53).

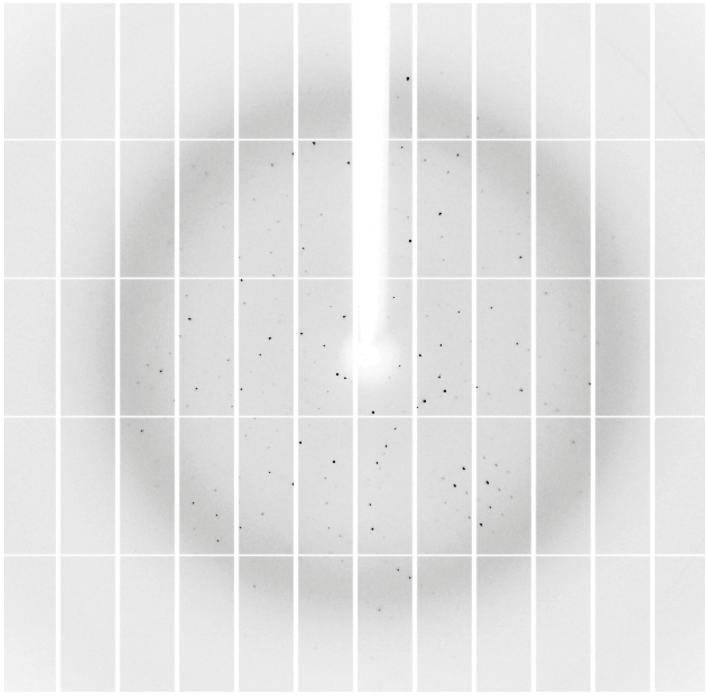

Figure 53. Diffraction pattern of a ToxRDNA40 crystal that diffracted up to 2.6 Å.

Data collection and processing statistics for this crystal are detailed in Tables 11 and 12. All the analyzed crystals belonged to the same space group and had similar cell dimensions.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
5.48	5738	3404	3733	91.2%	3.9%	3.5%	4646	19.18	5.6%	99.5*	2	0.856	687
3.87	10750	6350	6793	93.5%	6.3%	6.3%	8695	15.26	8.9%	98.3*	4	0.838	1334
3.16	13444	7987	8795	90.8%	13.5%	13.3%	10708	8.92	19.0%	96.0*	10*	0.837	1554
2.74	15204	9047	10406	86.9%	29.3%	28.2%	12059	2.89	41.1%	90.0*	41*	0.961	1685
2.45	18162	10799	11791	91.6%	71.2%	74.5%	14380	1.07	99.8%	56.4*	32*	0.805	2068
2.24	14139	10249	13038	78.6%	153.6%	176.1%	7634	0.37	215.4%	20.3*	22*	0.724	693
2.07	8017	6754	14263	47.4%	308.5%	389.1%	2525	0.12	436.2%	9.2	13	0.598	229
1.94	4153	3684	15131	24.3%	973.4%	1229.7%	938	0.04	1376.6%	0.3	-13	0.409	53
1.83	1078	1017	16241	6.3%	665.7%	779.1%	122	0.02	941.4%	-0.3	0	0.000	1
total	90685	59291	100191	59.2%	9.1%	9.1%	61707	4.65	12.9%	99.1*	25*	0.837	8304

Table 11. Data processing outputs for the best dataset of ToxRDNA40.

Parameters	**Value**
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	1.033
Space group	C2
Cell dimensions	225.10 Å 42.62 Å 63.18 Å 90.00° 105.74° 90.00°
Number of unique reflections	17004
Resolution range (Å)- (overall/last shell)	108.30-2.64/2.73-2.64
R_{merge} (%) (overall/last shell)	12.83/84.41
Multiplicity (overall/last shell)	3.0/3.0
Completeness (%) (overall/last shell)	97.74/98.00
<I>/σ<I> (overall/last shell)	8.86/2.43
Wilson B-factor ($Å^2$)	62.47

Table 12. Data collection and processing statistics for ToxRDNA40. R_{merge} is defined in section 4.3.3.

We determined the structure of ToxRDNA40 by the molecular replacement method using Phaser (Mccoy *et al*., 2007). As searching models, we used the structure of ToxR taken from the ToxRDNA20 complex and two molecules of idealized double strand DNA of 10 bp and 6 bp, respectively. Calculations gave one solution with a high TFZ score (9.6), confirming C2 as the correct special group. We grew the DNA strands manually, and refined the model by successive cycles of manual building of the model using Coot and restrained refinement using Refmac. No solvent molecules could be assigned to the model. Rfactor and Rfree converged to 22% and 30%. ToxRDNA20 geometry was validated using Molprobity, which assigned to the structure an overall score of 2.58. This

value was considered good enough taking into account the maximal resolution of the structure (2.6 Å). Table 13 summarizes the refinement statistics.

Parameter	**Value**
Resolution range (Å)	108.30-2.64
Number of protein/DNA atoms	4305
Number of solvent molecules	0
RMSD for bonded angles $_{(1)}$ (°)	1.96
RMSD for bond lengths $_{(1)}$ (Å)	0.019
Rfactor/Rfree	0.2230/0.2960
Average B-factor (Å^2)	60.10
Ramachandran favored (%)	86.0
Ramachandran ouliers	2.5
Molprobity score	2.58

Table 13. Refinement statistics for ToxRDNA40. Rfactor= Σ|*Fobs* - *Fcalc*| / Σ|*Fobs*|. $_{(1)}$ RMSD from target values.

The final model showed a continuous and quite well defined electron density for all the structure (Figure 54).

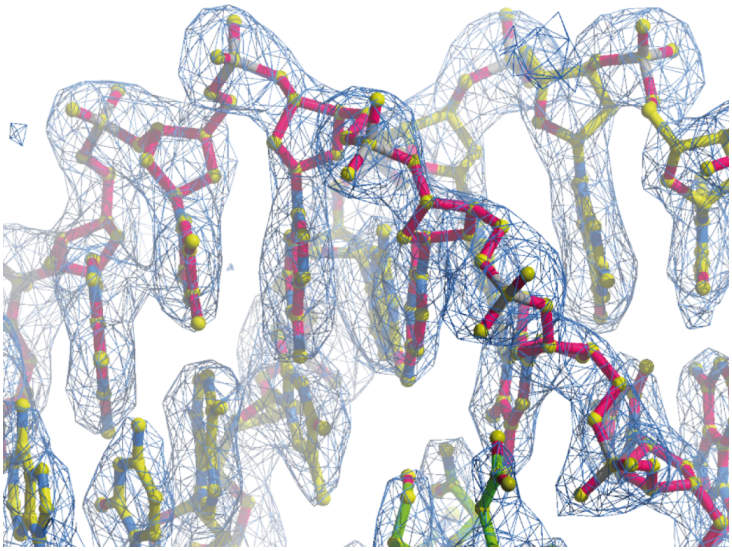

Figure 54. 2 F_o-F_c electron density map (contoured at 1.0 σ) with the final ToxRDNA40 model fitted in the density.

All the residues were located in allowed regions of Ramachandran plot with few exceptions (Figure 55).

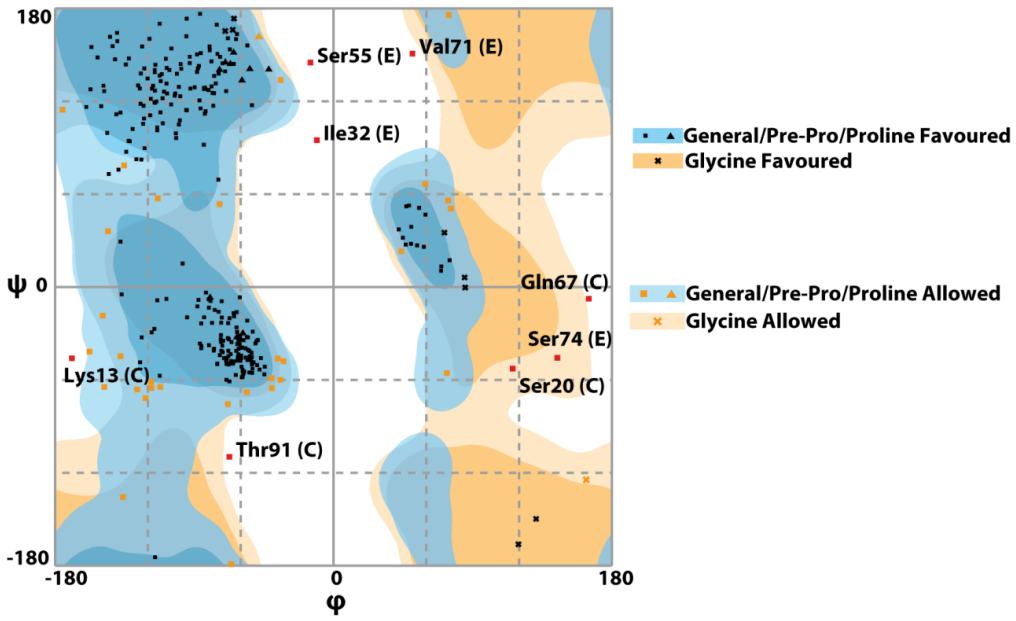

Figure 55. Ramachandran plot for the three ToxR molecules (monomers C, D and E) in the ToxRDNA40 structure.

4.4.4 Global description of the protein-DNA complex structure

In ToxRDNA40, three molecules of ToxR-DBD bind the DNA through the same elements analyzed in ToxRDNA20 structure: the wing, the recognition helix and the secondary wing (Figures 56 and 57). The peculiarity of this structure is that the upstream (C and D) two monomers are binding the DNA in tandem, while the third one (E, downstream) is oriented in the opposite direction (note the arrows on Figure 56). Monomer E is not aligned with monomers C and D because there is not a complete turn of helix between monomers D and E.

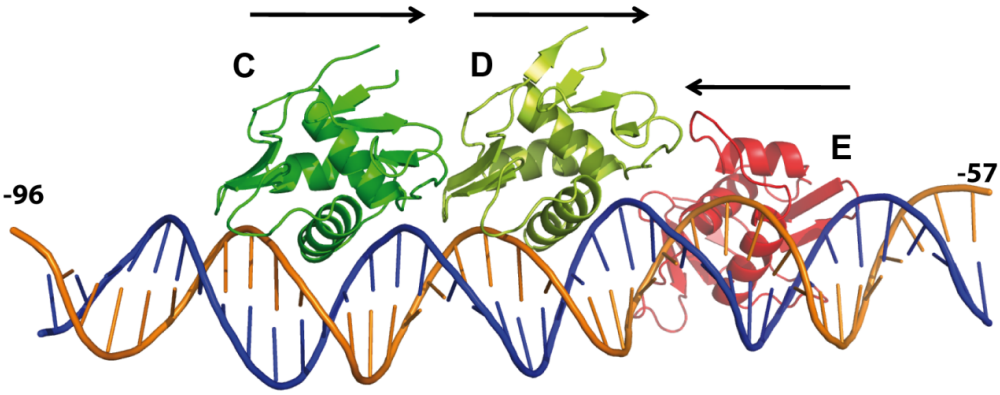

Figure 56. Overview of the structure of the ToxRDNA40 complex.

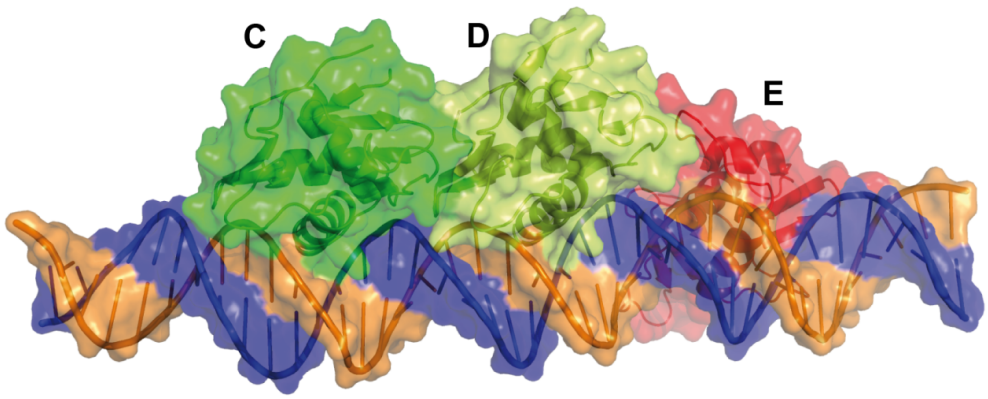

Figure 57. Surface representation of the ToxRDNA40 complex.

4.4.5 DNA-protein interaction

The three monomers maintained the same DNA-protein interactions, which are described for the ToxRDNA20 complex. There is a remarkable difference for monomer D. In this case, ToxR is binding a CATG box, and the key-residue of the recognition helix, Gln78, is making two direct contacts with the DNA bases (Figure 58). It forms a first H-bond between the ε-NH_2 of Gln78 and the O6 of guanine G-75 (chain A) and a second H-bond between the ε-NH_2 of Gln78 and the O4 of thymine T-74 (chain B).

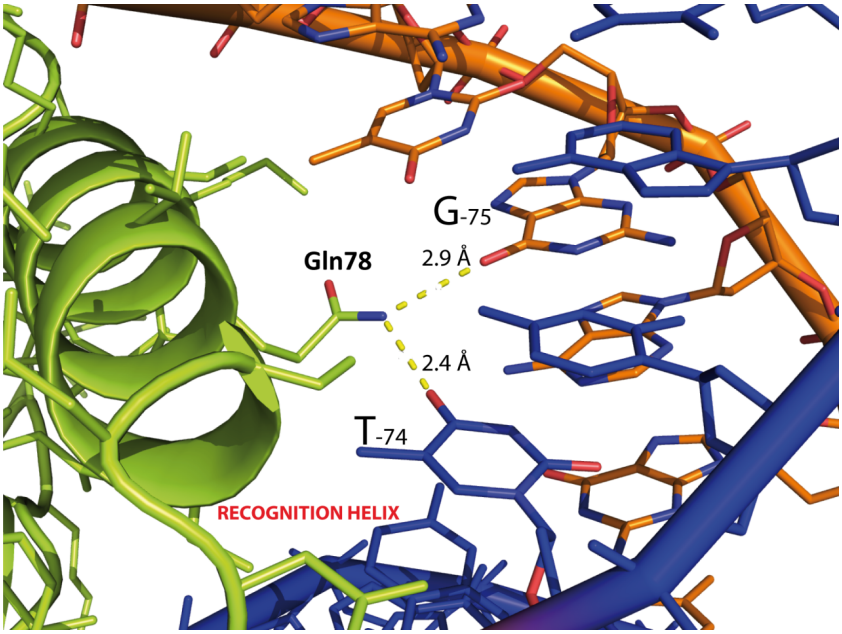

Figure 58. Molecular details of the contacts between Gln78 of monomer D and DNA.

Analysis of DNA-protein interaction allowed us to define the ToxR binding site for the three monomers: the binding site for monomer C goes from A-92 to T-81, the one for monomer D goes from A-83 to T-72 and finally the binding site for monomer E is from C-69 to T-58. As the binding sites of monomers C and D overlap, the two ToxR molecules are bound in tandem and interacting with each other. Figure 59 shows a plot of DNA-protein interactions, which was made by using Nucplot (Luscombe *et al.* 1997).

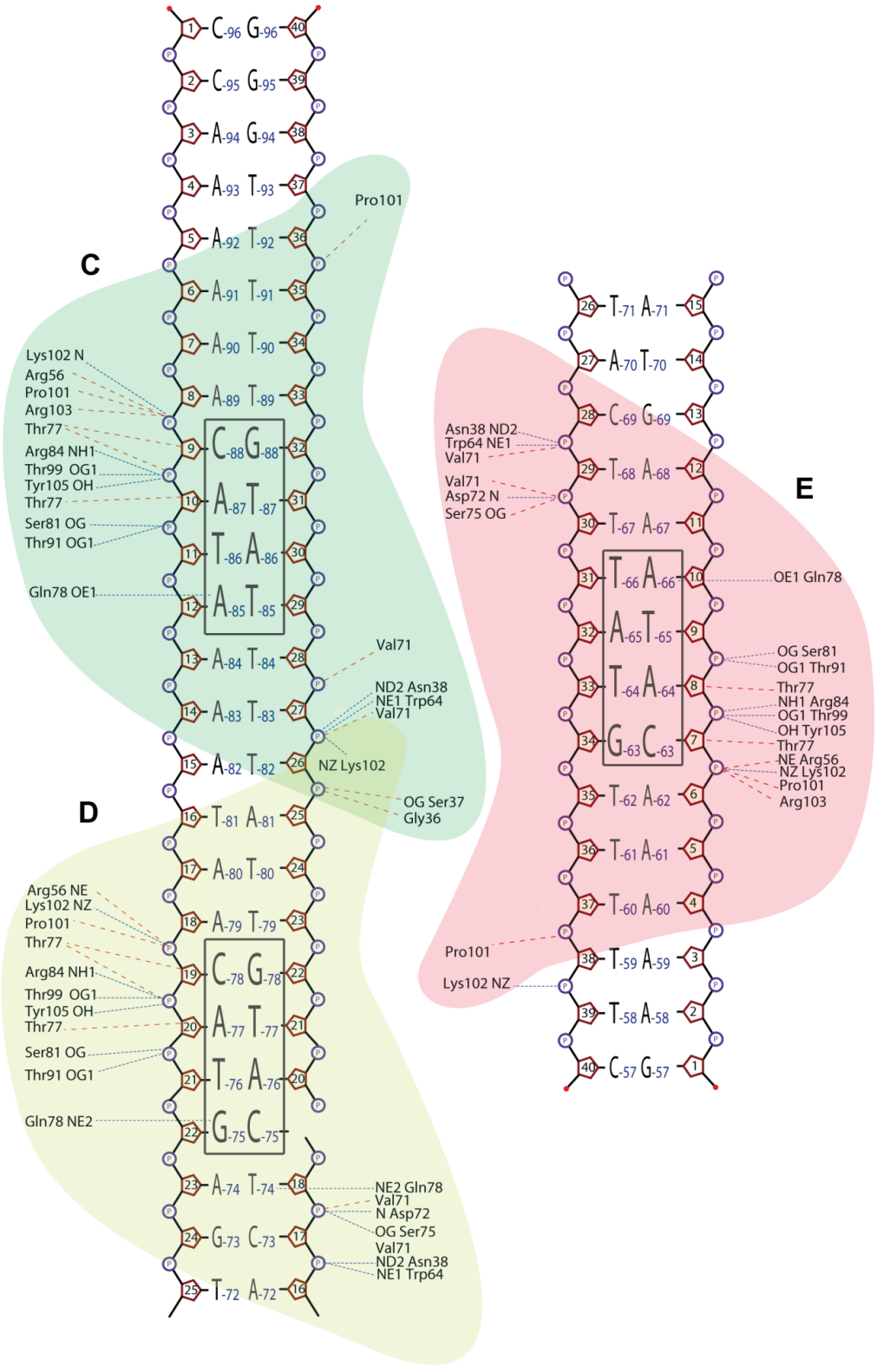

Figure 59. Interaction scheme of ToxR-DBD with DNA in the ToxRDNA40 structure. The CATA and CATG boxes and the ToxR binding sites are highlighted.

4.4.6 Protein-protein interactions

The two upstream monomers (C and D) are binding the DNA in tandem and are displaying several interactions in the interface area. We assessed the significance of these interactions by using the PISA software and we got a complex formation score of 0.061, which implied that this interface was not significant for dimer complex formation by itself, although it covers an area of 409.9 $Å^2$. In details, monomer C is interacting through the loop connecting β2 and β3, the loop connecting β4 and α1 and the α2-α3 connector, while monomer D makes contacts by the wing, the secondary wing and the loop connecting α1 and β5 (Figures 60 and 61). Monomer C displays H-bonds with monomer D through residues Arg34 (H-bond between the NH_2 of the side chain and the carbonyl oxygen of the main chain of Pro94) (Figure 62), Gln67 (H-bond between the ε-carbonyl and the η- NH_2 of Arg103) (Figure 63), Asn21 (H-bond between the δ-oxygen of the carbonyl group and the N of the side chain of Lys96), Leu19 (H-bond between the carbonyl oxygen of the main chain and the ε- NH_2 of Gln51) and Ser20 (H-bond between the γ-OH and the ε- NH_2 of Gln51 and H-bond between the γ-OH and the ε-carbonyl of Gln51) (Figure 64). Monomer C displays Van der Waals' interactions with monomer D through residues Ser20 (interaction with Ile108) and Arg34 (interactions with Ser93 and Pro94). Moreover, we found an electrostatic interaction between the side chain of Lys102 from monomer C and the phenyl ring of Phe69 from monomer D (Figure 63). These interactions may be important for tandem formation, although not enough to drive dimerization of the protein in the absence of DNA. Hence, we hypothesize two possible scenarios: a) the dimer observed in the ToxRDNA40 complex is the same dimer observed in solution without DNA (Figure 26), but it is weak complex interaction, so ToxR is able to bind DNA as a monomer; b) the dimer observed in the ToxRDNA40 complex is different from the dimer observed in solution without DNA. The eventual conformational changes from a dimeric form to another one would be probably due to the presence of DNA.

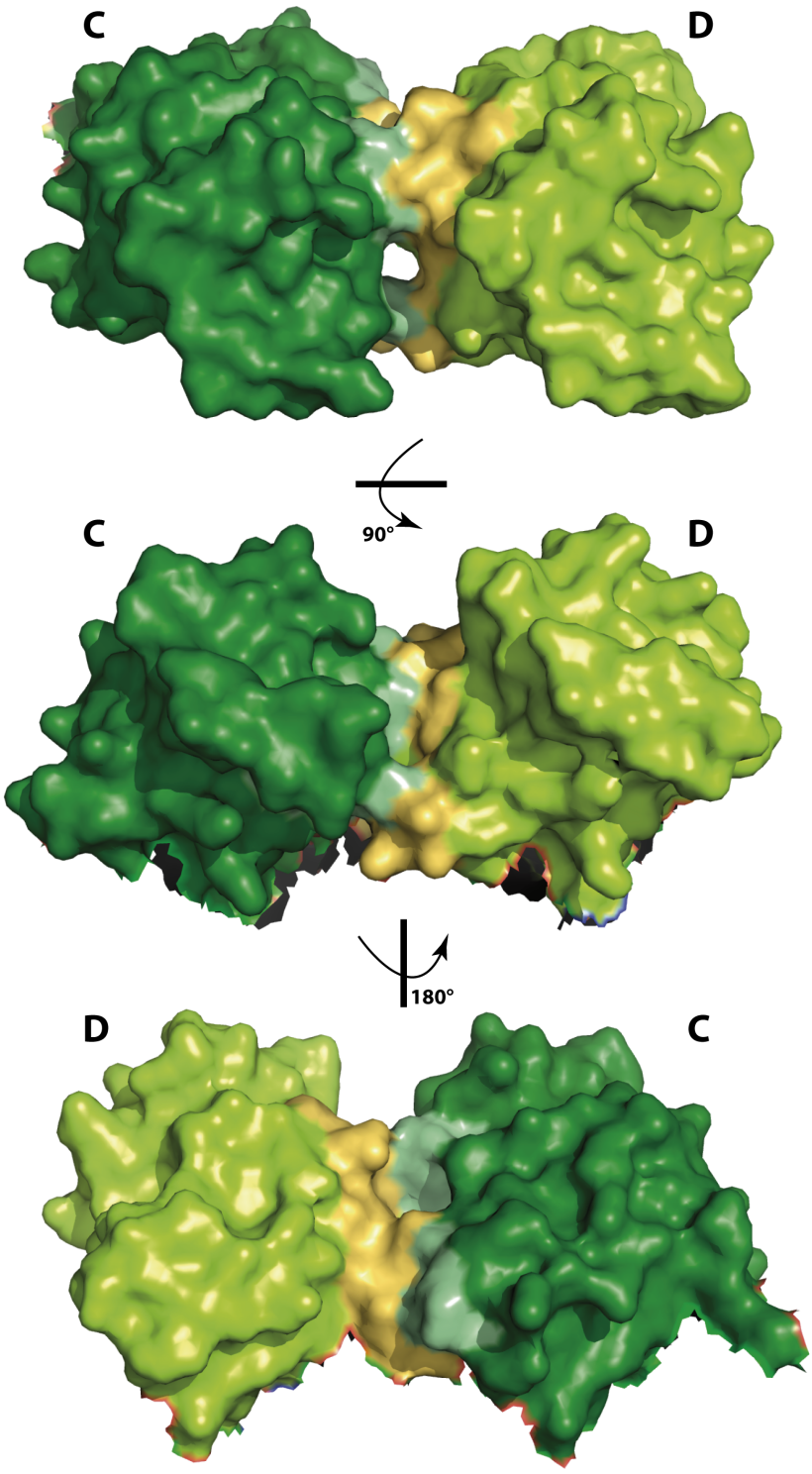

Figure 60. Top, front and rear views of the ToxR-DBD tandem interface. Monomers, A and B, are represented in dark and light green, respectively. The interface area is coloured in palegreen (monomer A) and yellow-orange (monomer B).

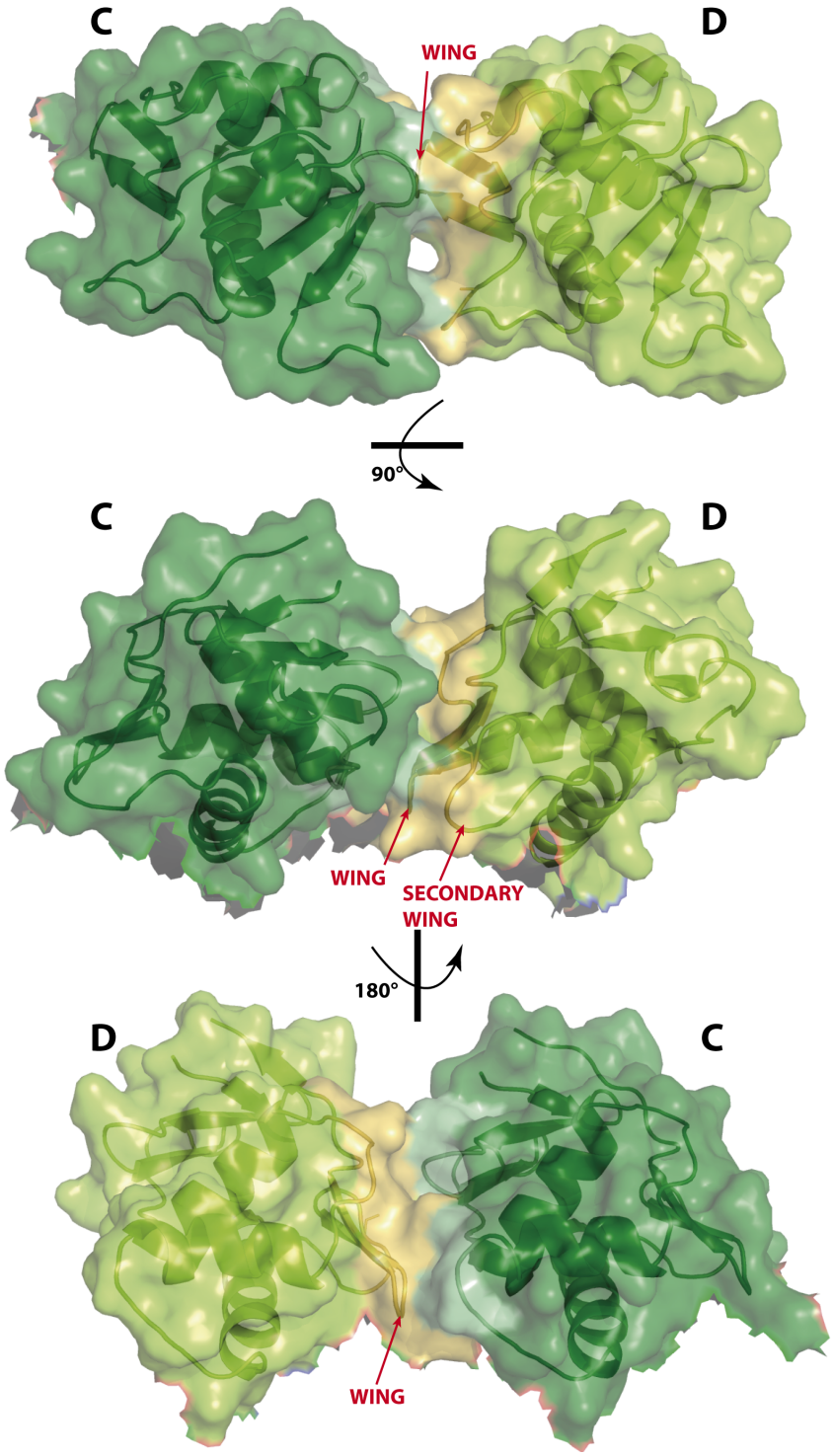

Figure 61. Cartoon and surface representations of ToxR-DBD tandem interface. The ToxR characteristic elements that are present in the interface area, are indicated.

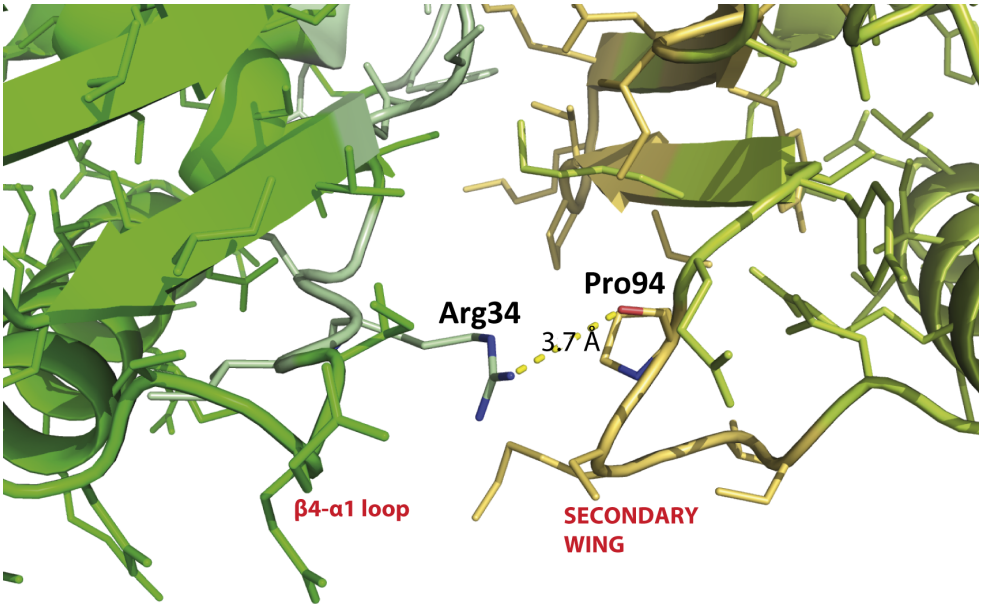

Figure 62. H-bond between Arg34 of monomer C and Pro94 of monomer D.

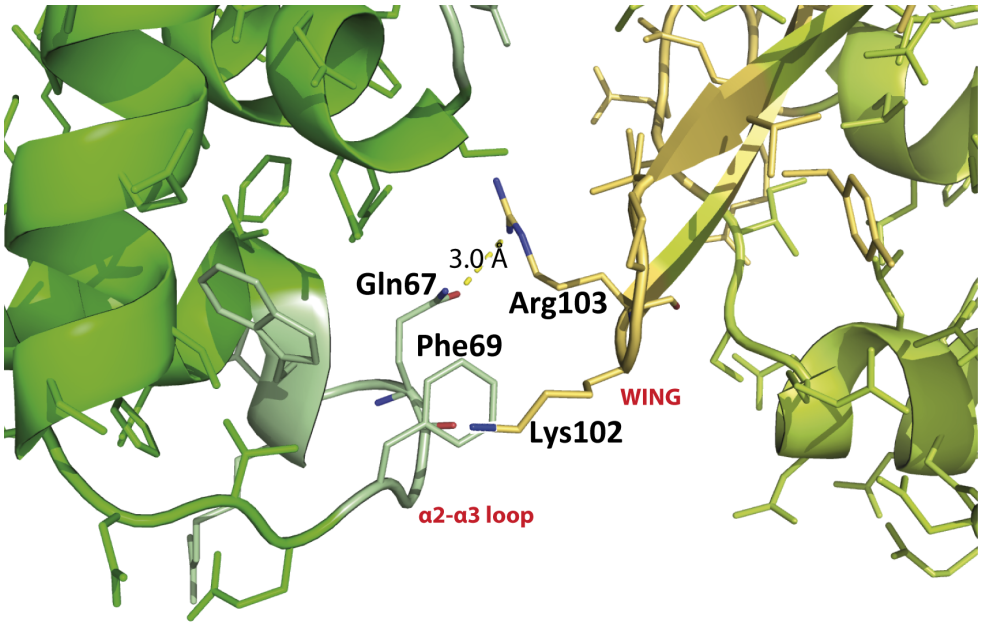

Figure 63. H-bond between Gln67 of monomer C and Arg103 of monomer D and electrostatic interaction between Phe69 of monomer C and Lys102 of monomer D.

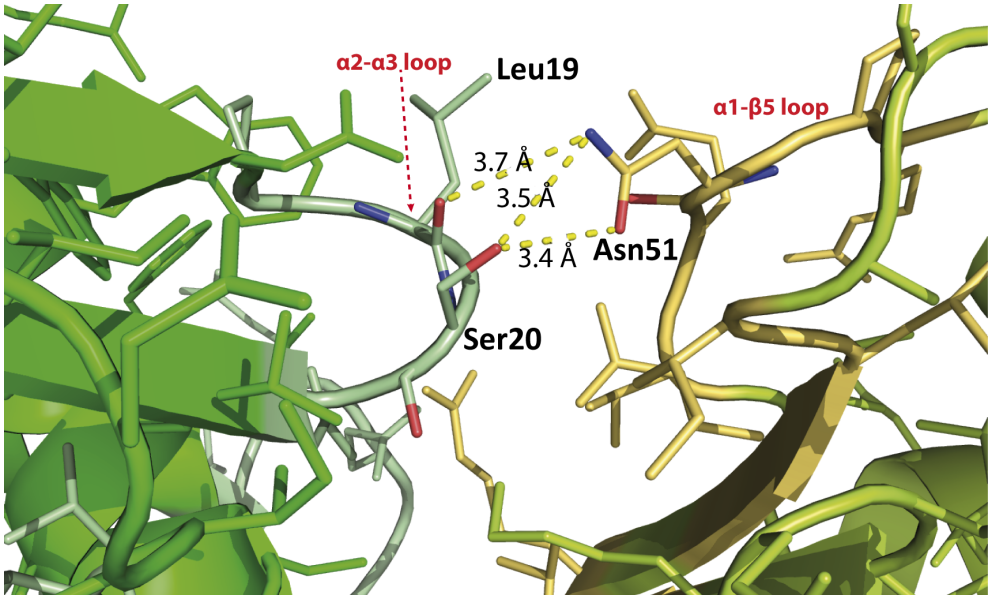

Figure 64. H-bonds between Leu19 and Ser20 of monomer C and Asn51 of monomer D.

4.4.7 DNA conformation

The DNA presents a curvature of 11.3°. As analyzed in the ToxRDNA20 complex, the A-T rich sequence region from A-92 to A-89 is producing a high compression of the minor groove, which forms a suitable cavity where the wing of monomer C can enter. The minor groove narrows up to 8.8 Å in this region, reaching the minimal width value of the 40-bp oligonucleotide. In fact, the rest of the minor groove, at the regions where monomers D and E are binding, narrow up only to 9.4 Å and 9.0 Å, respectively. There is a slight compression of the major groove in the CATA and CATG box regions: 17.8 Å in the monomer C binding site, 17.7 Å in monomer D binding site and finally 17.5 Å in monomers E binding site (Figure 65). ToxR binding may contribute to stabilize the compression of the grooves.

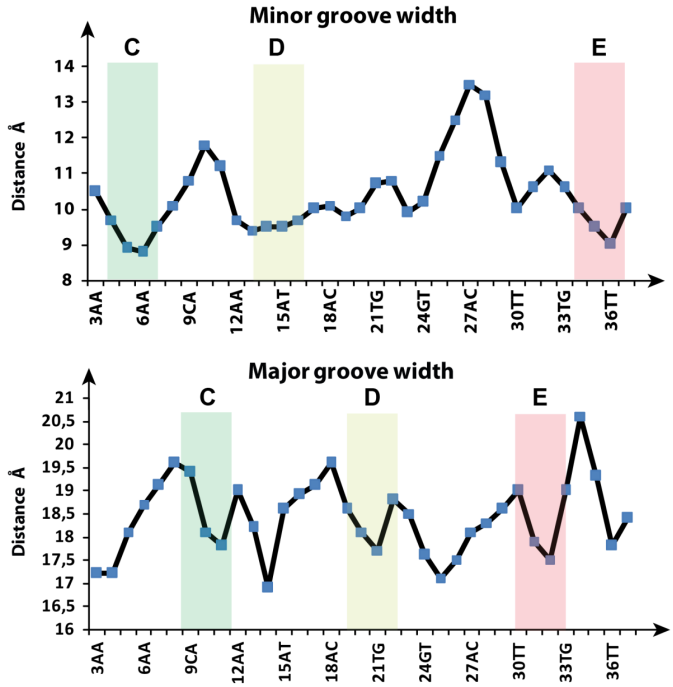

Figure 65. Minor and major groove width plots for the ToxRDNA40 structure. Major and minor groove widths are measured calculating the distance between phosphates. In the x-axis we indicate the number of the phosphate of the chain A and the corresponding DNA bases that are located before and after the phosphate. Binding sites for monomers E and F are shadowed following the usual color code: in the upper panel these DNA regions correspond to the binding site of the wing, in the lower panel they correspond to the binding site of the recognition helix.

All the geometrical parameters of the DNA are shown in Tables 14 and 15. As seen in ToxRDNA20, the A-tracts show a high value of propeller-twist. In particular, the DNA has the highest propeller value for adenine 14 of chain A (-27.34°), which is able to form a bifurcated H-bond with thymine 26 and thymine 27 of chain B. This is a common feature of the A-tracts that display a high compression of the minor groove, as described by Coll and his collaborators (Coll *et al.* 1987)

bp	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
1 C-G	-1.01	0.77	-0.29	17.43	-29.51	-1.64
2 C-G	-0.09	0.41	-0.19	9.85	-18.00	13.01
3 A-T	0.59	-0.48	-0.71	1.46	-24.33	-2.69
4 A-T	0.31	-0.01	-0.04	15.07	-24.28	6.30
5 A-T	0.59	-0.06	0.06	8.52	-26.47	2.33
6 A-T	-0.31	-0.02	0.16	6.20	-24.58	-2.39
7 A-T	0.04	0.01	0.20	5.48	-23.61	1.91
8 A-T	0.20	-0.12	0.75	6.16	-12.17	-4.78
9 C-G	0.17	0.08	-0.12	12.90	-20.41	2.56
10 A-T	-0.17	-0.09	0.23	-7.04	-3.23	-3.74
11 T-A	-0.21	0.10	-0.31	1.25	-20.23	0.62
12 A-T	0.03	-0.13	-0.26	1.46	-14.51	1.44
13 A-T	0.15	-0.01	0.13	-1.36	-21.94	4.24
14 A-T	-0.25	-0.04	-0.29	-2.94	-27.34	3.38
15 A-T	0.00	0.06	0.12	2.51	-23.18	1.81
16 T-A	0.02	-0.20	0.46	3.58	-14.99	-0.95
17 A-T	0.05	-0.01	0.18	-1.84	-13.01	-2.02
18 A-T	-0.01	-0.07	0.36	9.04	-11.94	-1.92
19 C-G	0.74	-0.09	-0.29	16.24	-22.06	-2.90
20 A-T	-0.43	-0.08	0.53	-5.18	-6.01	3.15
21 T-A	0.17	0.04	0.44	0.63	-12.84	1.15
22 G-C	-0.13	-0.04	-0.08	-0.93	-12.24	1.36
23 A-T	-0.47	-0.25	0.28	-0.95	-22.44	3.46
24 G-C	-0.29	-0.04	-0.36	-8.69	-19.14	-0.13
25 T-A	-0.33	-0.02	-0.46	-3.41	-19.12	9.47
26 T-A	0.72	-0.05	0.15	-4.23	-15.43	-0.22

27 A-T	-0.80	0.46	0.12	1.58	-19.20	9.96
28 C-G	0.07	-0.13	0.29	6.75	-13.45	2.37
29 T-A	0.22	0.11	-0.18	0.50	-12.73	1.20
30 T-A	0.36	0.12	-0.70	4.75	-18.88	1.18
31 T-A	-0.66	-0.16	-0.47	1.74	-24.45	-7.68
32 A-T	-0.11	0.11	0.14	3.43	-18.29	2.28
33 T-A	0.25	0.33	0.29	9.07	-1.81	4.15
34 G-C	-0.05	0.53	-0.07	-6.38	-12.55	1.32
35 T-A	-0.25	-0.09	0.47	-14.91	-20.48	-4.11
36 T-A	-0.31	-0.02	0.57	-7.80	-18.30	2.72
37 T-A	-0.17	-0.23	0.13	-9.07	-19.30	-6.36
38 T-A	0.36	0.18	-0.09	-10.87	-26.11	-5.88
39 T-A	-1.61	0.39	-0.26	-1.49	-18.75	-10.19
40 C-G	-0.15	0.13	-0.53	1.86	-6.08	-3.36
Average	-0.07	0.03	0.01	1.51	-17.83	0.51
S.D.	-0.45	0.23	0.35	7.42	6.57	4.64

Table 14. Local base-pair parameters for the ToxRDNA40 structure.

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
1 CC/GG	0.66	0.48	3.65	1.23	10.11	37.81
2 CA/TG	-0.56	2.20	3.76	5.39	2.82	45.63
3 AA/TT	0.42	0.32	2.77	-3.11	1.93	32.75
4 AA/TT	-0.18	0.07	3.58	-2.28	-2.11	37.27
5 AA/TT	-0.08	-0.24	3.20	2.08	-3.46	34.85
6 AA/TT	0.58	-0.37	3.23	3.07	-5.28	38.03
7 AA/TT	-0.00	-0.70	3.09	-3.83	-4.84	36.47
8 AC/GT	0.75	-0.90	3.22	8.41	-1.05	34.05
9 CA/TG	-0.44	-0.91	3.43	-5.45	10.02	37.78
10 AT/AT	-0.46	-0.50	3.13	1.85	-0.79	31.36
11 TA/TA	-0.08	0.40	3.34	-2.57	4.51	37.18
12 AA/TT	0.31	-0.08	3.20	-3.72	-5.22	38.88
13 AA/TT	-0.08	-0.30	3.26	1.44	0.85	32.59
14 AA/TT	-0.28	-0.19	3.02	-3.49	1.56	38.49
15 AT/AT	0.19	-0.71	3.23	-1.80	-4.53	31.98
16 TA/TA	0.41	0.37	3.36	4.17	-7.04	41.77
17 AA/TT	0.03	-0.43	2.99	-1.16	-3.31	33.52
18 AC/GT	0.48	-0.69	3.22	6.35	-0.92	38.26
19 CA/TG	0.14	-0.60	3.44	-5.06	4.60	36.75
20 AT/AT	-0.49	-0.50	3.23	0.99	-4.21	33.60
21 TG/CA	-0.28	0.09	3.29	0.71	4.27	38.73
22 GA/TC	0.27	-0.24	3.14	-2.42	-0.43	32.27
23 AG/CT	-0.42	-0.49	3.58	0.90	3.27	35.50
24 GT/AC	-0.15	-0.75	3.10	1.75	-1.40	34.87
25 TT/AA	0.26	0.12	3.28	-2.55	2.66	41.28
26 TA/TA	0.57	0.52	3.06	1.04	9.50	25.49

27 AC/GT	-0.24	0.07	3.19	-3.57	4.34	36.79
28 CT/AG	-0.32	-0.60	3.37	5.15	5.85	36.23
29 TT/AA	0.45	-0.31	3.02	3.04	3.40	34.72
30 TT/AA	-0.56	-0.23	3.40	-2.04	-4.01	33.41
31 TA/TA	0.68	0.33	3.36	-2.01	1.69	40.97
32 AT/AT	0.47	-0.54	2.96	-1.99	0.89	34.58
33 TG/CA	0.21	-1.17	3.41	4.65	6.35	35.57
34 GT/AC	-0.58	-1.28	3.56	-3.57	-2.12	32.25
35 TT/AA	0.11	-0.86	2.90	-0.51	-1.34	33.92
36 TT/AA	-0.93	-0.60	3.19	-0.53	-2.91	36.42
37 TT/AA	-0.18	-0.34	3.14	-0.67	-1.83	41.90
38 TT/AA	-0.25	-0.53	2.84	0.99	1.51	24.66
39 TC/GA	-0.17	0.13	3.32	-1.00	0.03	37.94
Average	0.01	-0.26	3.24	3.24	0.60	35.81
S.D.	0.42	0.61	0.22	0.22	4.34	4.03

Table 15. Local base-pair step parameters for the ToxRDNA40 structure.

4.5 ToxR in complex with a 25-bp oligonucleotide

The determination of the ToxRDNA40 structure revealed a third ToxR binding site in the *toxT* promoter sequence, which was due to a 5'-CATA-3' box from C-63 to A-66 of chain B. Since we observed a second 5'-CATA-3' box located 11-bp downstream of the first one in the chain B, we tried to obtain a complex of ToxR with a 25-bp oligonucleotide, named DNA25, containing both 5'-CATA-3' boxes (Figure 66). We called the new complex ToxRDNA25.

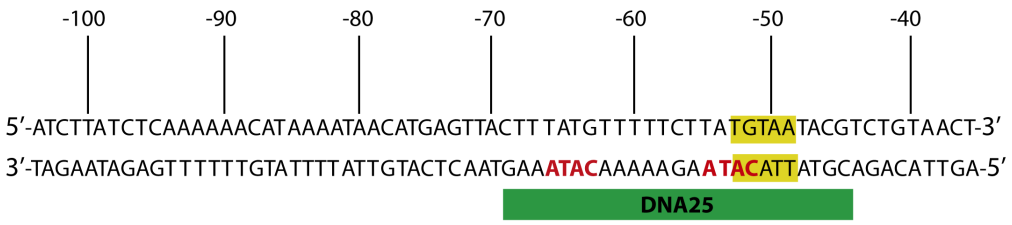

Figure 66. DNA25 sequence comprises the sequence of the *toxT* promoter from C-69 to G-45. 3'-CATA-5' boxes are highlighted and the second one overlaps with the TcpP binding site (TGTAA).

4.5.1 ToxRDNA25 complex formation

We incubated the protein and the DNA with a [1:1.2] ratio, and then we loaded the reaction product into a Superdex 200 size-exclusion chromatography column. The ToxRDNA25 complex eluted showing a major protein-DNA peak, confirming the formation of the complex. We verified by SDS-PAGE that all the fractions of the peak corresponded to the ToxRDNA25 complex and not to DNA alone (Figure 67).

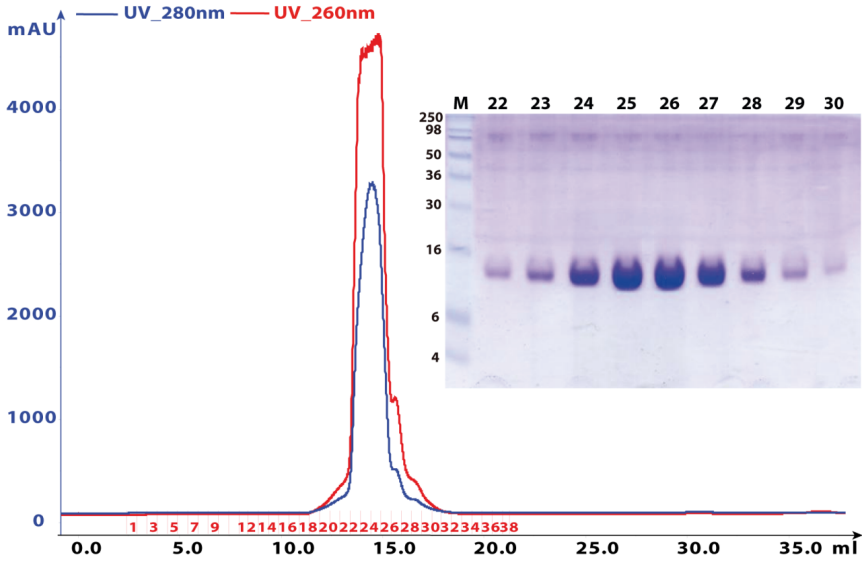

Figure 67. Peak of ToxRDNA25 complex after size-exclusion chromatorgaphy using a Superdex 200 (10-300) column and the SDS-PAGE analysis of the purified complex.

4.5.2 Crystallization of the ToxRDNA25 complex

We obtained crystal hits from different crystallization screening conditions in 100 nl robot nanodrops and we successfully scaled up to microdrops the condition PAC1-C7 (see Table 2 in section 3.4.1). We were able to get large bar- and hexagonal-shaped crystals after optimization of the crystallization condition (25% PEG4000, 0.3 M ammonium sulfate) (Figure 68). Crystals were cryo-cooled in liquid nitrogen using as cryobuffer the crystallization condition solution with 20% glycerol.

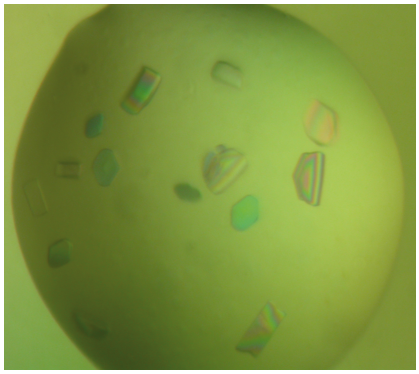

Figure 68. ToxRDNA25 crystals after optimization.

4.5.3 Data collection and structural determination of ToxRDNA25

In total, we tested twenty ToxRDNA25 crystals at Alba Synchrotron BL13-XALOC beamline and collected six complete data sets. The best crystal, measuring 0.36 x 0.16 x 0.14 mm^3, diffracted up to 3.2 Å resolution (diffraction pattern shown in Figure 69).

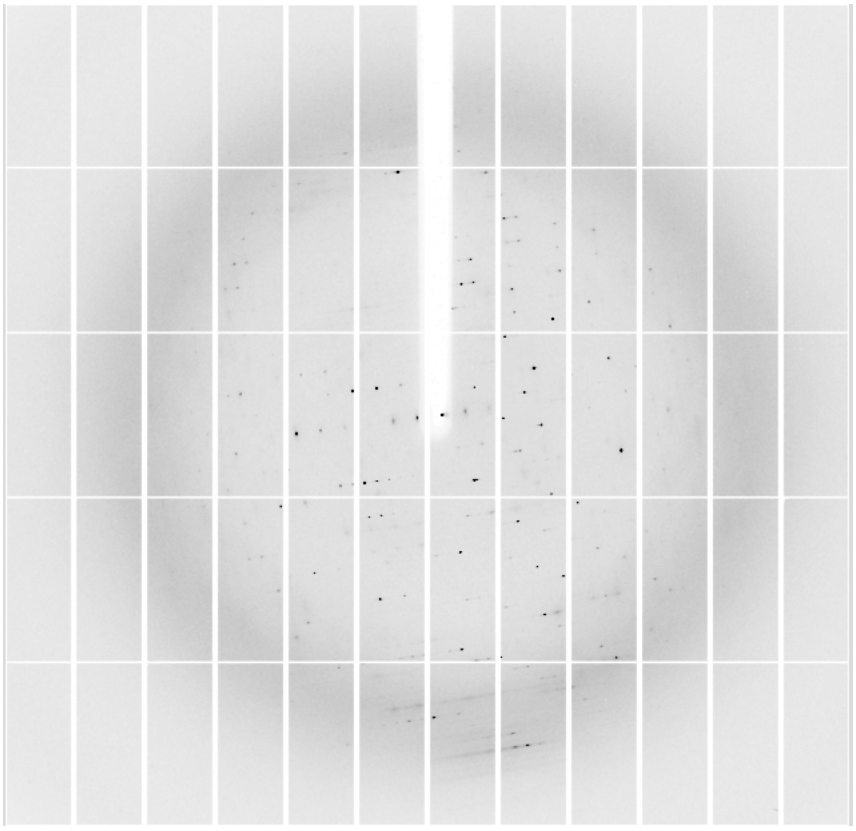

Figure 69. Diffraction pattern of a ToxRDNA25 crystal that diffracted up to 3.2 Å.

Diffraction data were processed using XDS. Data collection and processing statistics for this crystal are detailed in Tables 16 and 17. All the analyzed crystals belonged to the same space group and had similar cell dimensions.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
5.82	6568	3004	3134	95.9%	2.3%	2.4%	6013	26.30	3.0%	99.9*	-2	0.773	843
4.12	12201	5503	5672	97.0%	5.5%	5.3%	11208	14.12	7.1%	99.7*	1	0.805	1672
3.37	15583	7110	7314	97.2%	14.8%	14.5%	14203	5.18	19.0%	99.4*	5	0.758	2095
2.92	18635	8490	8692	97.7%	15.2%	14.9%	16934	1.64	19.4%	99.8*	5	0.620	2497
2.61	20559	9563	9791	97.7%	358.2%	378.0%	18440	0.27	465.1%	40.3*	3	0.553	2675
2.38	14612	9291	10916	85.1%	4295.4%	4608.1%	9886	-0.00	5870.4%	4.6	3	0.507	1127
2.21	8423	6483	11815	54.9%	5043.1%	5511.6%	3844	-0.01	7106.7%	2.3	2	0.494	315
2.06	4303	3690	12686	29.1%	2540.3%	2687.9%	1226	-0.01	3592.6%	-0.6	-5	0.509	80
1.95	1106	1059	13534	7.8%	-99.9%	-99.9%	94	-99.00	-99.9%	15.5	41	0.544	7
total	101990	54193	83554	64.9%	8.6%	8.7%	81848	3.87	11.2%	99.8*	3	0.653	11311

Table 16. Data processing outputs for the best dataset of ToxRDNA25.

Parameters	**Value**
Beamline	BL13-XALOC, Alba Synchrotron
Wavelength (Å)	0.979
Space group	$C222_1$
Cell dimensions	83.70 Å 162.37 Å 83.63 Å 90.00° 90.00° 90.00°
Number of unique reflections	9572
Resolution range (Å)- (overall/last shell)	81.18-3.25/3.36-3.25
R_{merge} (%) (overall/last shell)	6.68/53.28
Multiplicity (overall/last shell)	4.0/4.2
Completeness (%) (overall/last shell)	99.17/99.05
<I>/σ<I> (overall/last shell)	13.02/2.71
Wilson B-factor ($Å^2$)	114.98

Table 17. Data collection and processing statistics for ToxRDNA25. R_{merge} is defined in section 4.3.3.

We determined the structure of ToxRDNA25 by the molecular replacement method using Phaser (Mccoy *et al*., 2007). As searching models, we used the structure of ToxR

taken from the ToxRDNA20 complex and one molecule of idealized double strand B-DNA of 14 bp. Calculations gave one solution with a high TFZ score (8.4), confirming $C222_1$ as the correct space group. We grew the DNA strands manually, and refined the model by successive cycles of manual building of the model using Coot and restrained refinement using Refmac. No solvent molecules were added to the model because of the lack of clear electron density corresponding to the solvent. Rfactor and Rfree converged to 20% and 27%, respectively. ToxRDNA25 geometry was validated using Molprobity, which assigned to the structure an overall score of 2.72. This value was considered good enough taking into account the resolution of the structure (3.2 Å). Table 18 summarizes the refinement statistics.

Parameter	**Value**
Resolution range (Å)	81.18-3.25
Number of protein/DNA atoms	2798
Number of solvent molecules	0
RMSD for bonded angles $_{(1)}$ (°)	1.60
RMSD for bond lengths $_{(1)}$ (Å)	0.010
Rfactor/Rfree	0.1974/0.2723
Average B-factor (Å^2)	125.00
Ramachandran favored (%)	82.0
Ramachandran ouliers	2.3
Molprobity score	2.72

Table 18. Refinement statistics for ToxRDNA25. Rfactor= $\Sigma|Fobs - Fcalc| / \Sigma|Fobs|$. $_{(1)}$ RMSD from target values.

The final model showed a continuous and quite well defined electron density for all the structure (Figure 70).

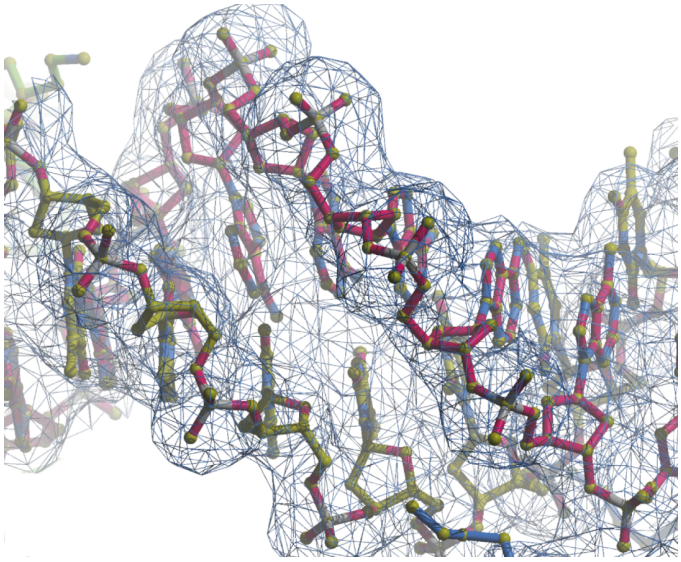

Figure 70. 2 F_o-F_c electron density map (contoured at 1.0 σ) with the final ToxRDNA25 model fitted in the electron density.

All the residues were located in allowed regions of Ramachandran plot with the exception of three residues (Figure 71).

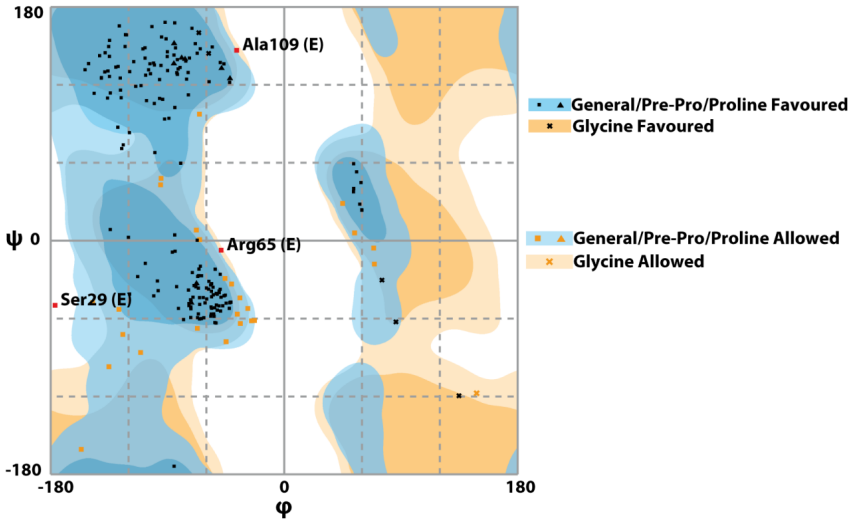

Figure 71. Ramachandran plot for the two ToxR molecules (monomers E and F) in the ToxRDNA25 structure.

4.5.4 Global description of the protein-DNA complex structure

In ToxRDNA25, two molecules of ToxR-DBD bind the DNA through the same elements analyzed in ToxRDNA20 structure: the recognition helix, the wing and the secondary wing (Figures 72 and 73). However, in this case, the two monomers are oriented in the opposite direction (note the arrows on Figure 72). The two proteins are far apart, with no interactions between them.

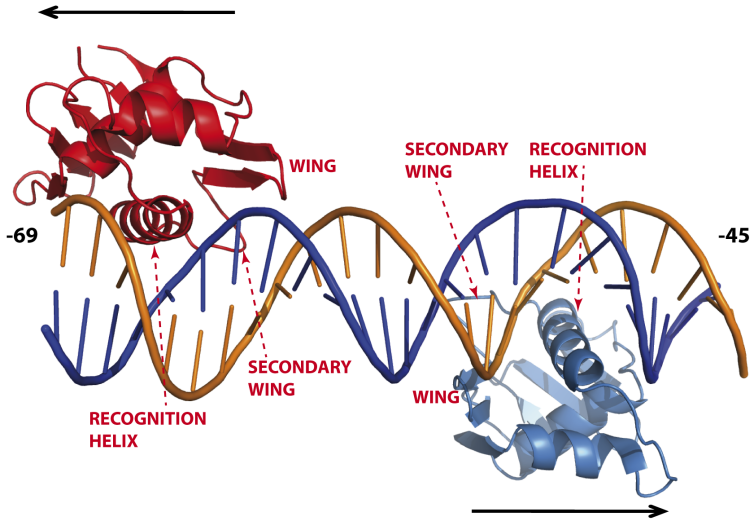

Figure 72. Overview of the structure of ToxRDNA25 complex.

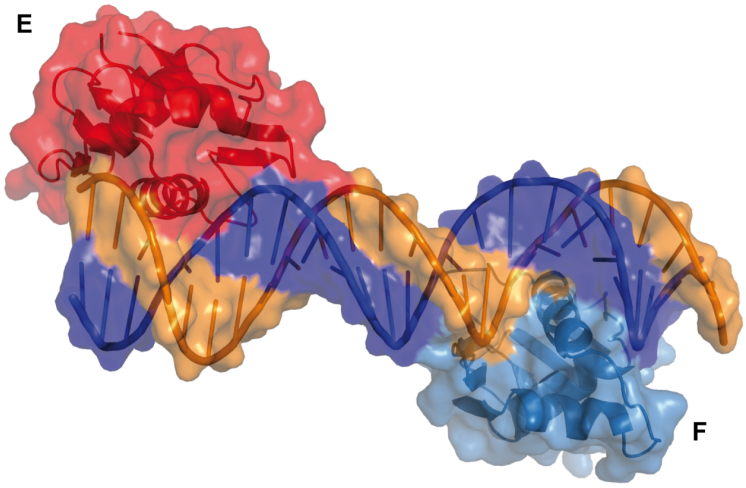

Figure 73. Surface representation of the ToxRDNA25 complex.

4.5.5 DNA-protein interaction

A plot of the DNA-protein interactions, which was made by using Nucplot (Luscombe *et al.*, 1997), is shown in Figure 76. Monomer E displays the same DNA-protein interactions that are described in the ToxRDNA40 complex. On the other hand, monomer F is surprisingly binding a TGTA sequence, which overlaps with the site specific for TcpP (Goss *et al.* 2010). In monomer F, the wing displays important interactions with the minor groove through residues Thr99 (H-bond between γ-OH of Thr99 and the phosphate of G-52 of chain A), Pro101 (Van der Waals' interaction between the side chain of Pro101 and the ribose of T-53, chain B) and Tyr105 (H-bond between η-OH of Tyr105 and the phosphate of G-52 of DNA, chain A). Lys102 forms an H-bond, very weak (3.73 Å), between the N of the side chain and the phosphate of A-58 (chain B) and a second H-bond between the N of the main chain and the phosphate of T-53 (DNA, chain A). The majority of the interactions are between the recognition helix and four DNA bases, from T-53 to A-50, a region that we named the TGTA box. The recognition helix interacts with the major groove through five residues: Thr77, Gln78, Ser81, Arg84 and Lys85.

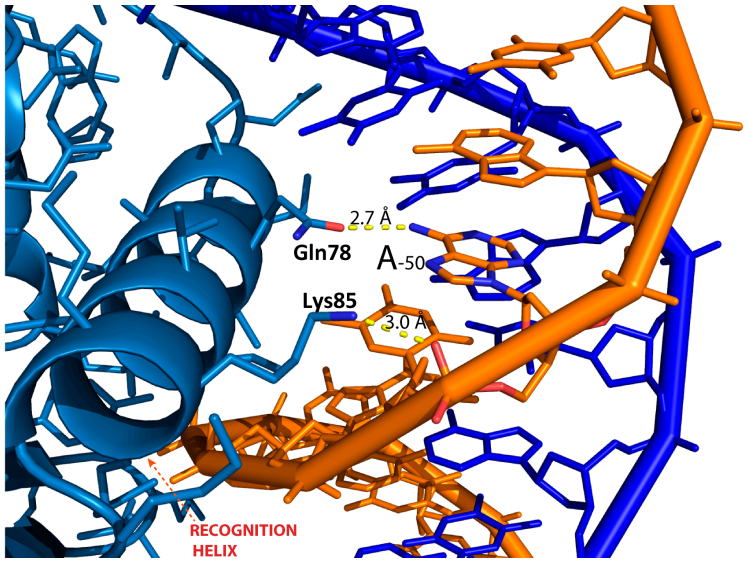

Figure 74. Molecular details of the contacts between two residues of the recognition helix of monomer F, Gln78 and Lys85, and DNA. Gln78 displays a direct contact with a DNA base (A-50) and Lys85 forms an H-bond with the phosphate of the same base.

As in ToxRDNA20 and ToxRDNA40 complexes, Gln78 seems to play an essential role in the recognition of this specific DNA sequence since it is able to interact directly with a DNA base. It forms an H-bond between the δ-C=O and the N6 of adenine A-50 (chain A) (Figures 74 and 76). In the recognition helix another key-residue is Thr77, which is able to perform direct interactions with DNA, displaying Van der Waals' interactions between the γ-CH3 and three DNA bases: thymine T-51 (chain A), guanine G-52 (chain A) and thymine T-53 (chain A). The secondary wing is also interacting with the major groove through the contacts between the γ-OH of Thr91 and the phosphate of T-51, chain A (H-bond) and between the N of the side chain of Lys92 and the phosphate of A-56 (H-bond) (Figures 75 and 76).

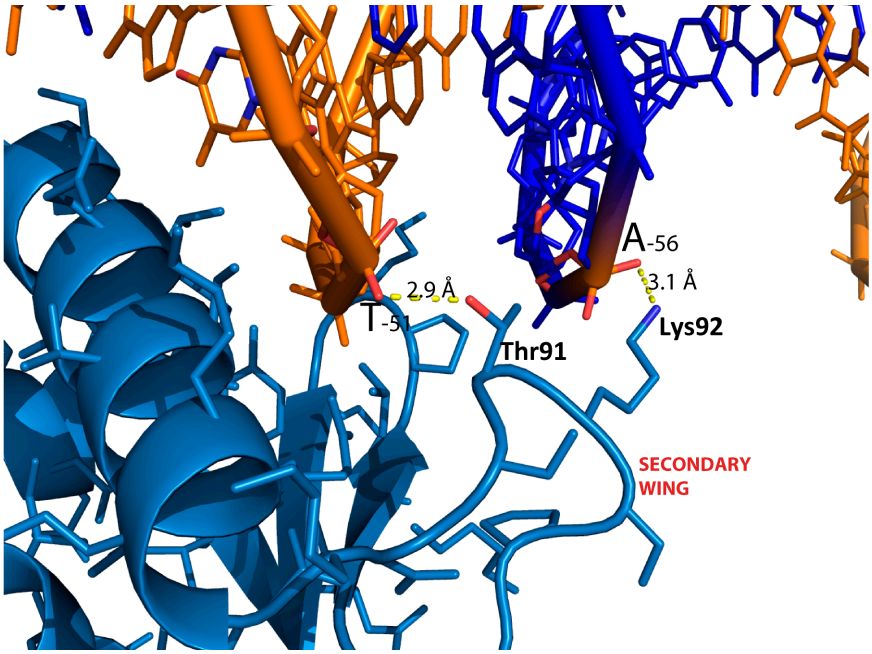

Figure 75. Molecular details of the contacts between Thr91 and Lys92 of monomer F and DNA.

Analysis of DNA-protein interaction allowed us to define the ToxR binding site for the two monomers: the binding site for monomer E goes from C-69 to T-58 (as seen in ToxRDNA40), while the one for monomer F goes from G-57 to G-46.

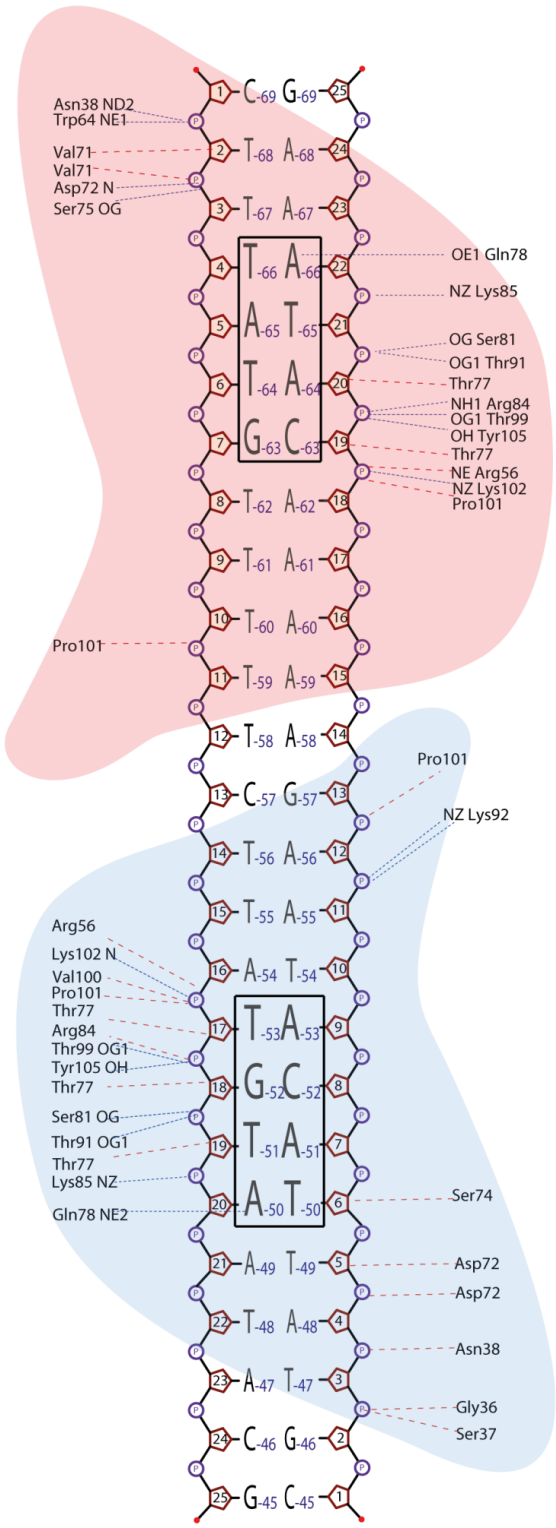

Figure 76. Interaction scheme of ToxR with DNA in the ToxRDNA25 structure. The CATA and TGTA boxes and the ToxR-DBD binding sites are highlighted. Monomers E and F are represented following the usual color code.

4.5.6 DNA conformation

The DNA presents a curvature of 9.6°. As analyzed in the ToxRDNA20 and ToxRDNA40 complexes, the A-T rich sequence regions show a high compression of the minor groove, which form a suitable conformation for the wing to bind both in monomers E and F. In particular, the minor groove, at the regions where monomers E and F are binding, narrow up to 9.1 Å and 9.3 Å, respectively. A slight compression of the major groove is observed in the CATA and TGTA box regions: 17.8 Å both in monomer E and monomer F binding sites (Figure 77). All the geometrical parameters of the DNA are presented in Tables 19 and 20. As seen in the ToxRDNA20 and ToxRDNA40 complexes, we observed high values of propeller-twist in the A-tracts of the DNA.

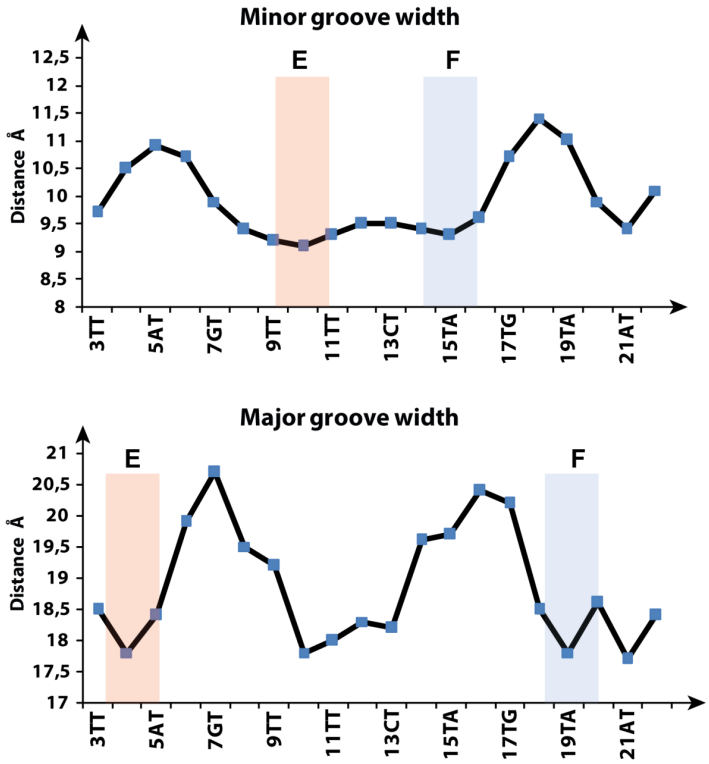

Figure 77. Minor and major groove width plots for the ToxRDNA25 structure. Minor and major groove widths are measured calculating the distance between phosphates. In the x-axis we indicate the number of the phosphate of the chain A and the corresponding DNA bases that are located before and after the phosphate. Binding sites for monomers E and F are shadowed following the usual color code: in the upper panel these DNA regions correspond to the binding site of the wing, in the lower panel they correspond to the binding site of the recognition helix.

bp	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
1 C-G	-0.73	0.21	0.99	-20.17	-7.94	8.15
2 T-A	-0.48	-0.17	0.52	-11.95	-17.79	-8.88
3 T-A	-0.04	-0.02	-0.19	-2.63	-15.27	4.82
4 T-A	0.00	-0.09	-0.21	0.07	-23.88	-5.92
5 A-T	0.08	-0.12	0.02	-13.81	-12.52	-0.94
6 T-A	0.37	0.05	-0.16	-0.44	0.36	-4.80
7 G-C	-0.42	-0.19	0.43	1.10	-4.46	-3.65
8 T-A	0.49	-0.10	0.37	3.45	-5.39	0.17
9 T-A	-0.24	-0.29	0.25	-2.53	-12.04	-7.08
10 T-A	-0.33	0.14	0.34	-3.37	-15.95	1.51
11 T-A	0.07	-0.24	0.02	2.20	-14.21	0.18
12 T-A	-0.12	-0.39	-0.11	6.12	-19.24	-6.62
13 C-G	0.12	-0.27	0.37	5.68	-22.17	2.60
14 T-A	-0.14	-0.21	0.00	3.24	-9.55	0.07
15 T-A	0.06	-0.16	-0.13	-0.49	-7.74	-2.46
16 A-T	0.31	-0.06	0.69	-5.52	-4.58	4.81
17 T-A	0.38	-0.23	0.20	0.09	-10.56	0.68
18 G-C	-0.29	-0.08	-0.38	-8.44	0.56	3.36
19 T-A	-0.05	-0.48	0.72	2.78	-8.32	-5.24
20 A-T	-0.39	-0.11	-0.42	2.85	-28.41	1.70
21 A-T	-0.52	0.17	-0.28	-9.10	-18.57	-0.39
22 T-A	0.88	0.08	-0.12	1.13	-17.88	-1.48
23 A-T	0.63	-0.23	0.67	-1.00	-9.38	1.56
24 C-G	0.76	-0.08	-0.19	9.16	-13.96	6.37
25 G-C	0.05	0.23	-0.06	-11.11	-17.44	10.66
Average	0.02	-0.11	0.13	-2.11	-12.65	-0.03
S.D.	0.41	1.18	0.38	6.96	7.25	4.89

Table 19. Local base-pair parameters for the ToxRDNA25 structure.

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
1 CT/AG	-0.75	-0.60	3.01	0.92	-1.03	37.10
2 TT/AA	0.86	-0.54	2.98	4.36	-3.25	35.03
3 TT/AA	-0.44	-0.58	3.15	0.75	1.14	36.95
4 TA/TA	0.48	0.24	3.45	-1.13	3.08	40.15
5 AT/AT	-0.35	-0.42	3.01	-1.03	-0.58	33.02
6 TG/CA	0.38	-1.41	2.93	-3.44	1.77	32.29
7 GT/AC	-0.23	-0.88	3.26	0.46	-3.67	37.73
8 TT/AA	-0.54	-0.99	3.42	0.17	-5.40	30.71
9 TT/AA	0.22	-0.81	3.26	-1.57	-5.15	35.89
10 TT/AA	-0.07	-0.01	3.00	0.58	-3.92	41.35
11 TT/AA	-0.46	-0.26	3.06	-1.93	-3.91	35.54
12 TC/GA	0.58	-0.41	3.25	0.02	-1.57	37.20
13 CT/AG	-0.25	-0.63	3.31	0.93	-1.65	33.25
14 TT/AA	-0.14	-0.51	3.44	0.20	-3.43	34.93
15 TA/TA	0.14	-0.58	3.34	-5.67	-4.44	41.35
16 AT/AT	-0.14	-0.65	3.17	2.88	-5.11	31.79
17 TG/CA	-0.39	-0.97	3.29	-2.31	4.41	37.29
18 GT/AC	-0.61	-0.64	3.19	-6.43	-6.22	28.05
19 TA/TA	-0.19	0.38	3.23	10.94	1.30	35.66
20 AA/TT	0.27	-0.42	3.51	-3.56	4.55	37.01
21 AT/AT	-0.38	-0.09	2.99	-2.58	-1.56	41.46
22 TA/TA	0.15	-0.50	3.29	-2.60	-5.24	35.51
23 AC/GT	0.76	-0.58	3.18	9.92	-1.41	36.15
24 CG/CG	0.18	1.37	3.63	1.46	15.34	37.22
Average	-0.04	-0.44	3.22	0.06	-1.08	35.94
S.D.	0.44	0.54	0.19	4.04	4.70	3.34

Table 20. Local base-pair step parameters for the ToxRDNA25 structure.

Chapter 5: Discussion

We have solved the 2 Å structure of a DNA-protein complex of ToxR-DBD with a 20-bp oligonucleotide, containing the sequence of the *toxT* promoter from C-97 to C-78. We chose this specific oligonucleotide based on the experiments described in literature (Goss *et al.* 2013). According to these experiments, this sequence encompasses two putative ToxR binding sites of the *toxT* promoter and thus, we expected two molecules of ToxR-DBD binding this DNA segment. Unexpectedly, in ToxRDNA20 only one molecule of ToxR-DBD binds the DNA. ToxR-DBD shows the characteristic fold of the winged helix-turn-helix family of transcription factors (Martínez-Hackert and Stock 1997b). However, when we compared ToxR-DBD with the structures of other proteins of this family, we found out that it has a unique turn that we called "secondary wing" (Figures 38 and 39). This secondary wing plays the role in enabling ToxR to bind the DNA in a more efficient way since a number of additional protein-DNA interaction appear. Two characteristic elements of ToxR are the wing and the recognition helix, which together with the secondary wing are responsible for the DNA recognition and binding (Figure 42). Another peculiarity of ToxR-DBD fold seems to be the N-term β-sheet, which is wider than in PhoB: an extra parallel β-strand (β8), located in the C-term, forms this β-sheet with the four antiparallel strands located in the N-term. This β-sheet might have a role in the transmission of the signal from the periplasmic domain to the DNA binding domain since β8 is preceding the linker that connects ToxR-DBD with the periplasmic domain. We analyzed all the DNA-protein interactions present in ToxRDNA20 and elaborated an interaction scheme, which allowed us to redefine the ToxR binding pattern (Figure 78). According to the ToxRDNA20 structure, the ToxR binding site is: 5'-AAAAA**CATA**AAA-3' from A-93 to A-82. The five adenines upstream of the CATA box are essential to narrow the minor groove and create a suitable cavity where the wing can enter and establish contacts with the DNA. Thus, the peculiar local DNA at this A-T rich region sequence is critical for the DNA sequence specificity. In fact, the minor groove narrows up to 8.9 Å and allows the interaction with Pro101 and Lys102 of the wing (Figures 44, 46 and 72). The secondary wing is also interacting with the minor groove through residue Thr91.

Regarding the CATA box, it is important to highlight the direct contacts of ToxR with DNA bases: an H-bond between Gln78 and A-85 and Van Der Waals' interactions between Thr77 and A-87 and T-86 bases. These interactions are the only direct and specific contacts with DNA bases present in the ToxRDNA20 complex.

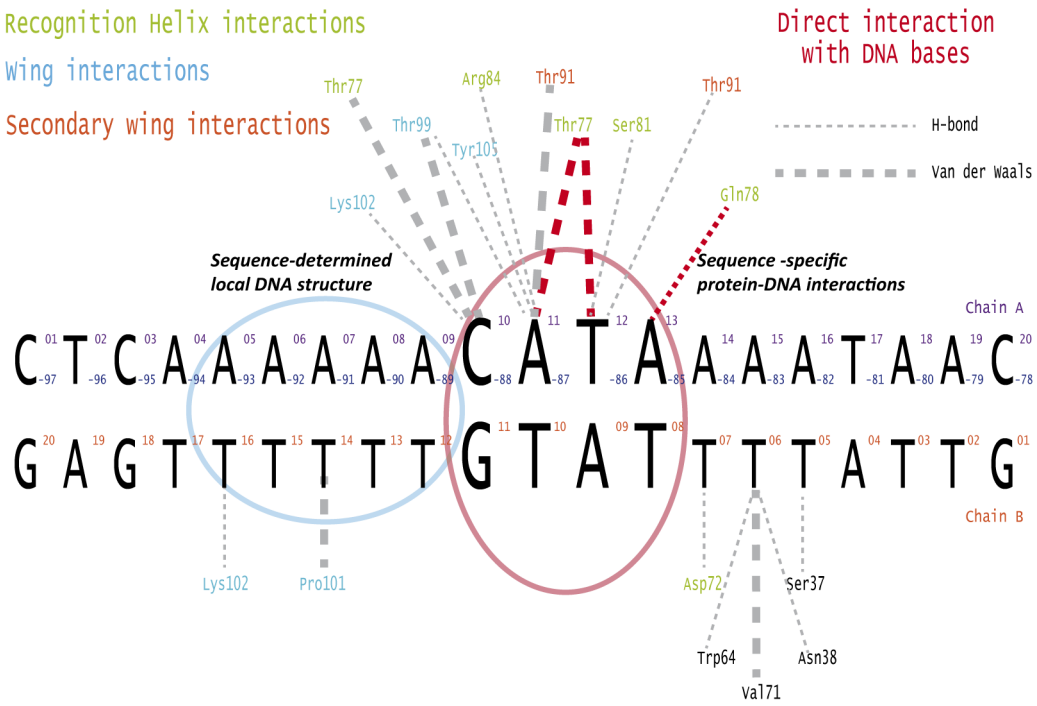

Figure 78. ToxRDNA20 interaction scheme.

The analysis of ToxR-DNA interactions allowed us to explain the data published on the ToxR mutations that abolish *toxT* activation (Morgan *et al*. 2011) (Table 21).

ToxR mutation	Effect on *toxT* promoter activation	Explanation by our structures (lost interaction)
W64R	Defective	H-bond
V71A	Defective	Van der Waals' interaction

D73A	Defective	H-bond
S74A	Defective	H-bond
S75A	Defective	H-bond
Q78R	Defective	H-bond
T99R	Defective	H-bond
P101L	Defective	Van der Waals' interaction

Table 21. ToxR mutations that are defective in the activation of *toxT* promoter (Morgan *et al*. 2011).

After solving the ToxRDNA20 complex, we aimed at getting a bigger complex of ToxR with a 40-bp oligonucleotide, taking into account the result of the footprinting experiments (see section 1.4.2). Hence, we designed a longer oligonucleotide, starting from T-96 (this thymine is replaced by a cytosine to enhance stability of the dsDNA) to C-57 (Figure 48). This 40-bp oligonucleotide (DNA40) contains the ToxR binding site found in ToxRDNA20 and an additional putative binding site, which had been described as a degenerated toxR binding site (from C-69 to T-56) (Goss T.J *et al.* 2013). ToxR seems to have a low affinity for this degenerate binding site and would not have a biological relevance in the *toxT* promoter activation. In spite of this, our working hypothesis consisted in getting a dimer of ToxR in complex with the DNA40. We solved the structure of ToxRDNA40 complex at 2.6 Å. Surprisingly, the structure showed three monomers attached to the DNA. Two monomers (C and D) are binding in tandem in a head-to-tail conformation, while a third monomer (monomer E) binds the DNA downstream, in the opposite direction (Figure 56). Monomer C binds to the same binding site of ToxRDNA20 complex, monomer D binding site is 5'-AATAA**CATG**AGT-3' from A-83 to T-72, and monomer E binding site is 3'-AAAAAA**CATA**AAG-5' from C-69 to T-58 (Figure 79). It is particularly interesting that the guanine G-75 of the CATG box is able to display a direct interaction with Gln78. This key-amino acid can interact indiscriminately with an adenine or a guanine thanks to its plasticity: by rotating its side chain χ3 angle by 180° it can interact with the adenine of a CATA box by its nitrogen or with a guanine of a CATG box by its oxygen (Figure 58).

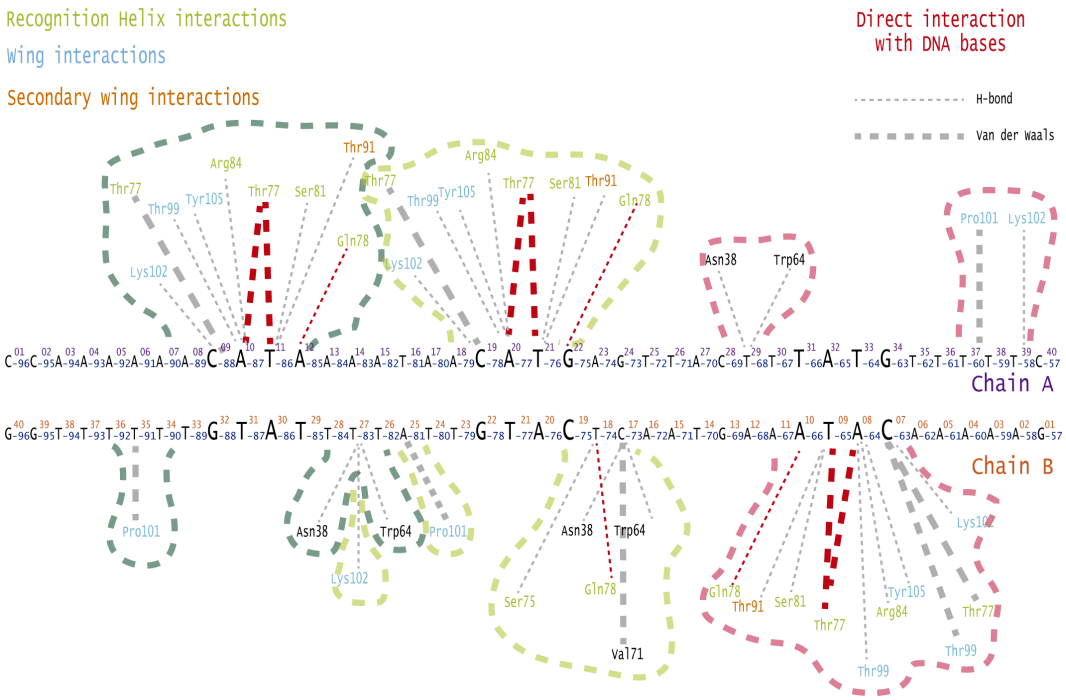

Figure 79. ToxRDNA40 interaction scheme.

When we analyzed the tandem interactions between monomer C and monomer D (see section 4.4.6) it seems that they are not significant enough to stabilize a dimer in the absence of DNA. So, we suggest two possible hypotheses: i) the dimer observed in the ToxRDNA40 structure is the same dimer of ToxR in solution in the absence of DNA, but it is a weak complex interaction ii) the dimer observed in the ToxRDNA40 structure is different from the ToxR dimer in solution and these conformational changes would occur in the presence of the DNA. Monomer E is binding a 5'-CATA-3' box in the opposite direction of monomers C and D at -63 (DNA, chain B) and this unexpected result is very interesting because it indicates that ToxR is able to bind its binding sites in both directions. This finding indicated that the "degenerate binding site" has indeed affinity for ToxR and that the protein is able to bind both in tandem to direct repeat and in opposite direction to an inverted repeat. Once the A-tract-CATA box was determined as the binding region of ToxR-DBD, we observed another putative binding site further downstream, at -52 of chain B. Hence, we tried to get a third ToxR-DNA complex with an oligonucleotide that includes the monomer E binding site plus the 5'-(-52)CATA(-55)-3'

box (chain B), downstream with respect to the first one. We picked a 25-bp oligonucleotide ranging from C-69 to G-45 (DNA25) instead of a very long oligonucleotide of 52-bp, increasing our chance of success (Figure 66). Our working hypothesis was that two molecules of ToxR could bind the DNA in tandem as monomer C and D in ToxRDNA40, but in the opposite direction (Figure 80).

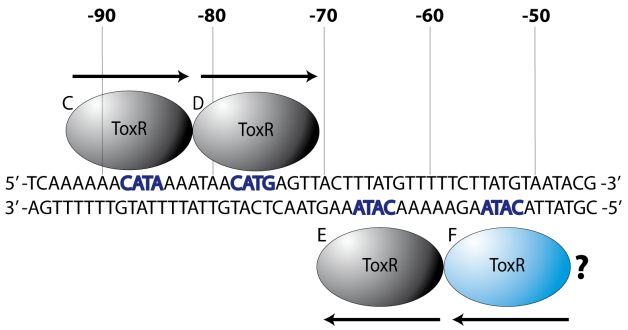

Figure 80. Working hypothesis for ToxRDNA25.

We solved the ToxRDNA25 structure at 3.2 Å. One ToxR monomer is binding at the same binding site of monomer E in ToxRDNA40, while a second monomer (monomer F) does not bind to the 5'-(-52)CATA(-55)-3' (chain B), but binds a further downstream sequence, TGTA, which overlaps with one of the TcpP binding sites (Figure 81).

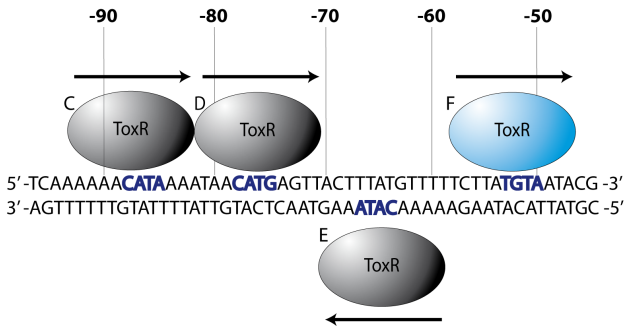

Figure 81. ToxR-DBD monomers binding the *toxT* promoter. In the ToxRDNA25 structure, monomer F binds a TGTA sequence instead of the 5'-(-52)CATA(-55)-3', contradicting our working hypothesis.

We got no ToxR tandem in the opposite direction to monomers C and D, and this might be due to the different distance between the two 5'-CATA-3' boxes of chain B (seven DNA bases) and the two 5'-CATA-3' boxes of chain A (six bases). Instead, the F ToxR protomer is binding inverted to the E monomer. This surprising result raises questions about the biological meaning of the interaction between monomer F and the *toxT* promoter. ToxR and TcpP seem to compete for the same binding site and this suggests two different scenarios: (i) ToxR binds temporarily the TcpP binding site to relieve the H-NS protein and then it is replaced by TcpP, (ii) ToxR binds this binding site and this event precedes TcpP recruitment (interaction with protomer ToxR-F) and the subsequent activation of the *toxT* promoter. The binding site for monomer F is 5'-TCTTA**TGTA**ATA-3' (Figure 82).

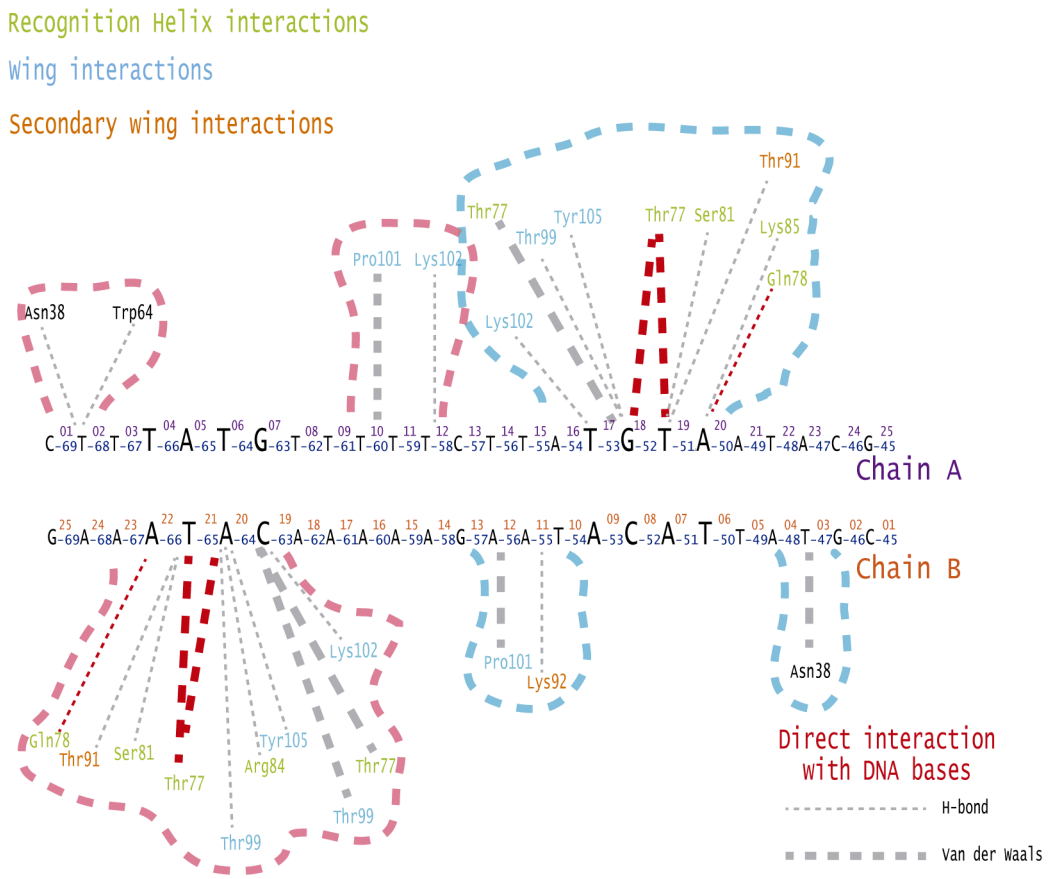

Figure 82. ToxRDNA25 interaction scheme.

In ToxRDNA25 structure, monomer F interacts with the minor groove of the DNA through two residues of the secondary wing, Thr91 and Lys92 (Figures 75 and 82) instead of only one residue as in ToxRDNA20 and ToxRDNA40 for monomers C, D and E. We superposed the structures of the four monomers with the bound DNA (eleven bases corresponding to the ToxR binding site) to detect any conformational change, but they present only minor differences (RMSD < 1 Å). Gathering the three structures, we are able to get the consensus ToxR binding site sequence, which is represented by using a sequence logo in Figure 83 (Crooks *et al*. 2004).

Figure 83. Sequence logo of the consensus ToxR binding site sequence.

We can deduce that an A-T rich sequence region is essential for *toxT* recognition from ToxR and that last two bases **T A/G** of a CATA/CATG/TGTA (Py-Pu-T-Pu) box are also needed since they displays direct contacts with the protein. In Figure 84, we present a summary of the structures solved. According to the three structures solved, ToxR covers a *toxT* promoter region from -97 to -45 and four ToxR molecules bind this *toxT* promoter region. This is not fully in keeping with the footprinting experiments reported in the literature (Krukonis *et al.* 2000) and opens for discussion about the role of ToxR, as transcription factor of *toxT* promoter.

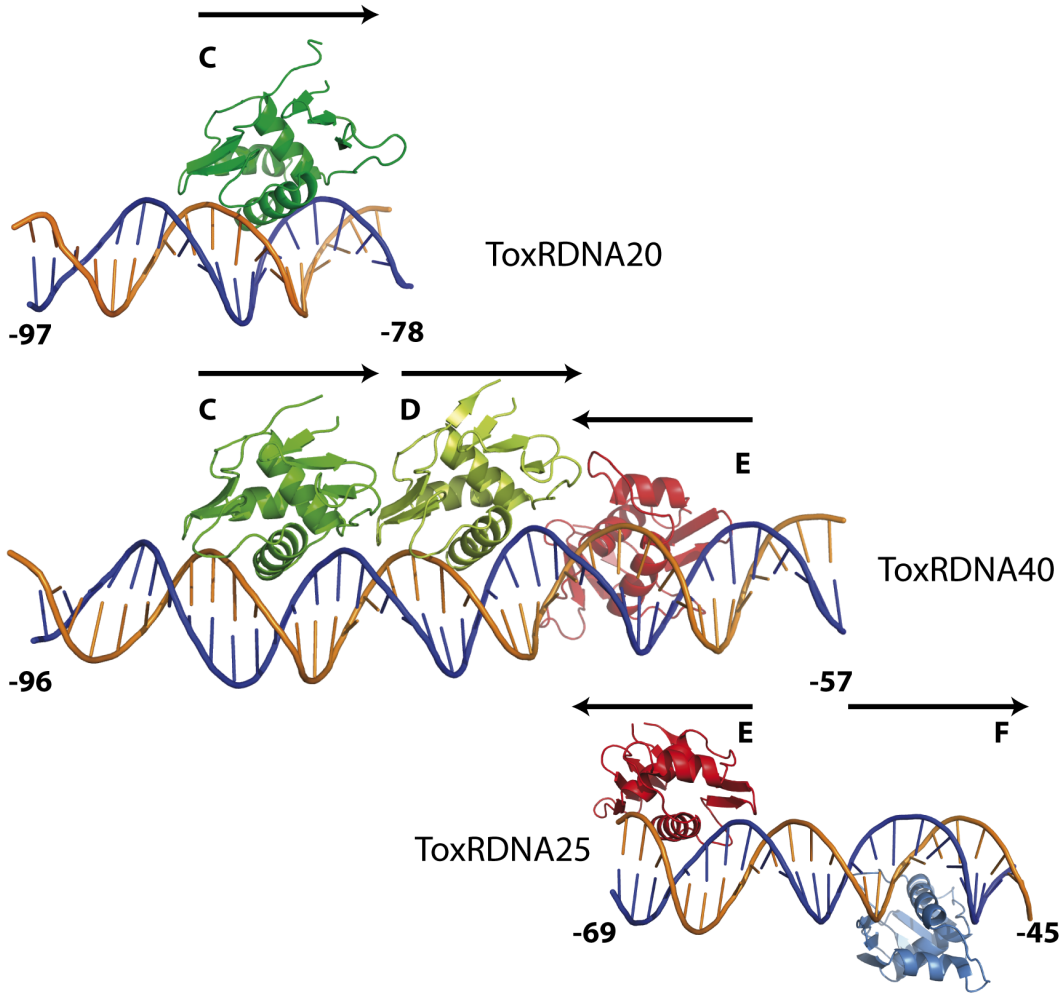

Figure 84. Summary of the ToxR structures solved. The arrows indicate the orientation and binding to either direct repeat or inverted repeat.

Moreover, we created a model of a ToxRDNA52 merging the ToxRDNA40 and the ToxRDNA25 structures, comprising the *toxT* promoter region that goes from -96 to -45 and the four ToxR monomers that are binding DNA (Figure 85). To the best of our knowledge, this is the first time that a transcription factor is described to be able to bind the same promoter in tandem (head-to-tail) and inverted (head-to-head).

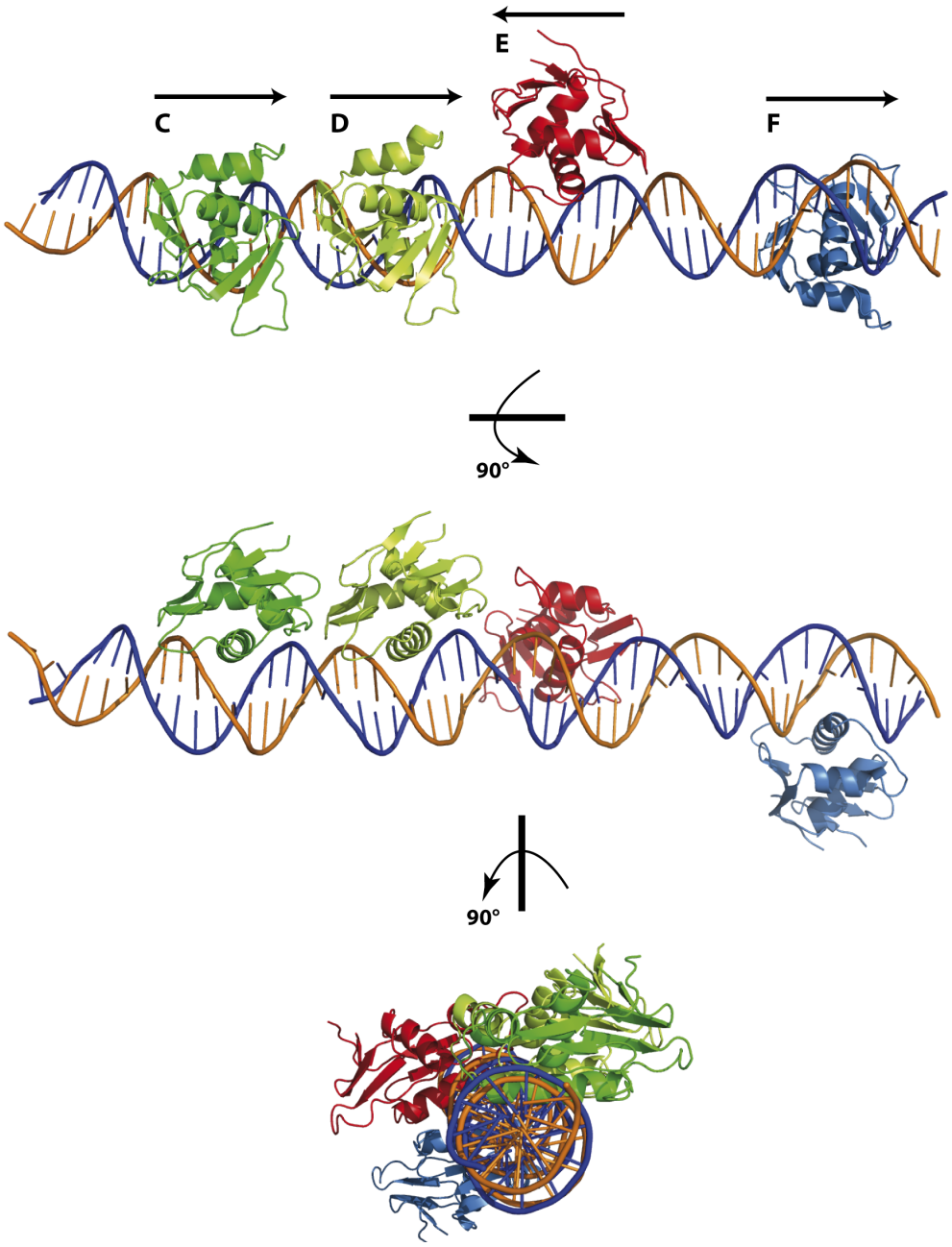

Figure 85. Top, front and side views of the ToxRDNA52 model.

In Figure 86, we put forward a model of the possible membrane architecture of the transcriptional complexes we characterized. Monomers C, D and E have the N-terminal oriented in the same direction, but monomer F does not. However, the cytoplasmatic

domain of ToxR is endowed with a flexible linker that makes monomer F orientation reasonable.

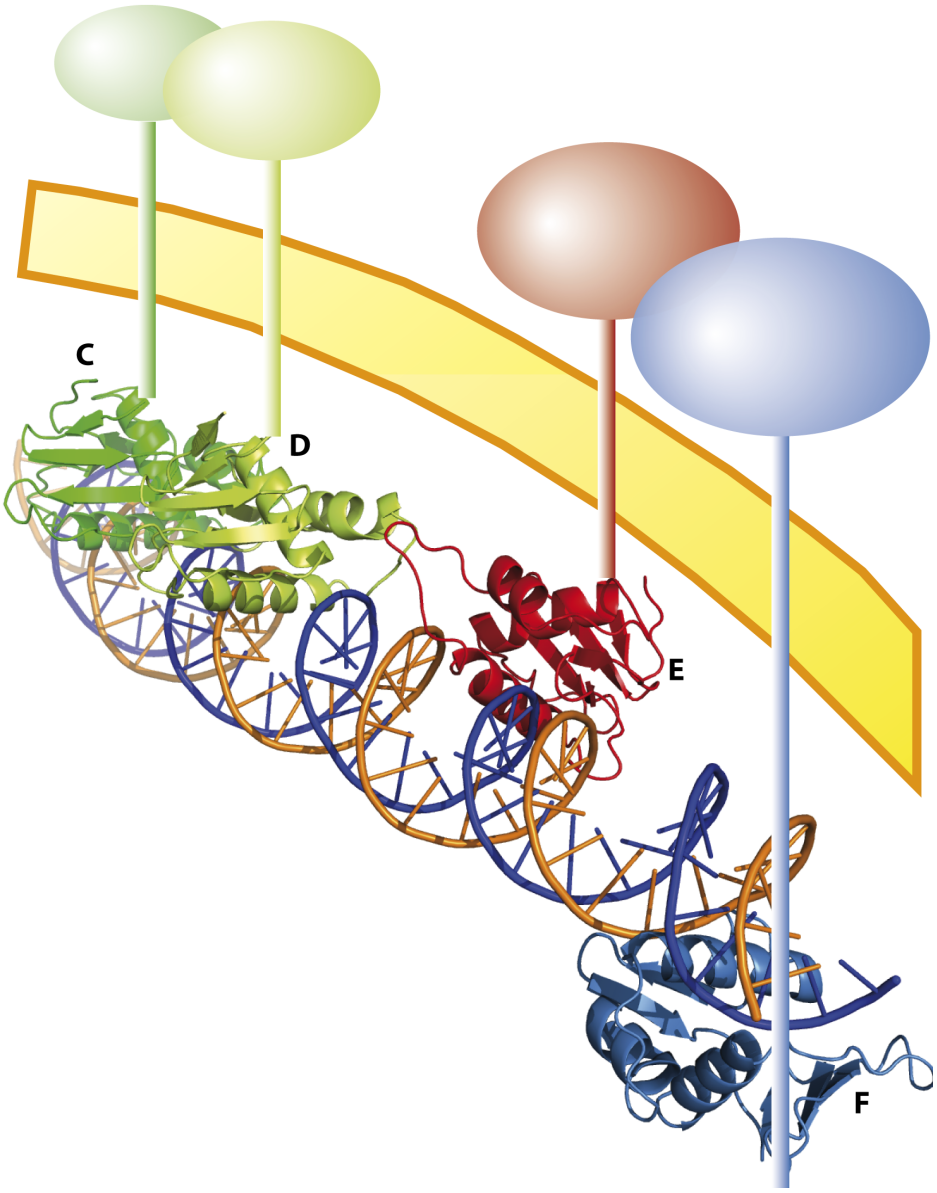

Figure 86. Diagram of ToxR architecture in the membrane.

We suggest an integrated model of the action mode of ToxR in *toxT* activation, taking into account our three ToxR-DNA structures and the models proposed in literature. The *toxT* promoter is normally repressed by "histone-like nucleoid structuring protein H" (H-NS) (Figure 87A). When ToxR is expressed, it can recruit the DNA to the membrane according to the "membrane recruitment" model (see Figure 19). The DNA is fished by ToxR and leaded to the membrane. The "membrane recruitment" model fits with our results, in fact DNA needs to be placed near the membrane before transcription can start, and this would be one of the functions of ToxR-DBD. Indeed, ToxR acts in the first steps of *toxT* activation since it seems to have a fundamental role in removing H-NS protein from *toxT* promoter (Figure 87B). According our structures, ToxR molecules bind the *toxT* promoter in the region from -97 to -45, and probably this is the region normally occupied by H-NS in the repressed state of *toxT* transcription. Afterwards, ToxR-DBD recruits TcpP: this interaction would be essential to stabilize the TcpP complex with the DNA (Figures 87C and 87D). This interaction would also occur between their periplasmatic domain, their transmembranal helix and maybe between their cytoplasmic linker. The "hand-holding" model would in fact fit with our results because a ToxR monomer binds the 5'-TGTA-3' box from T-53 to A-50 and it could form a heterodimer with a TcpP molecule binding the 5'-TGTA-3' box from T-42 to A-39. Our results are in keeping with the hand-holding model, and do not support the "catch and release" model because according to our structures, ToxR binding site is very close to TcpP binding site. The role of ToxR would be to stabilize TcpP facilitating its interaction with the *toxT* promoter. On the other hand, ToxR does not seem to alterate the promoter architecture since the DNA in our structures is not particularly curved, but it only presents a compression of the minor groove and more slightly of the major groove. Hence, we could discard also the "promoter alteration" model. So, the action of ToxR as a co-activator of the *toxT* promoter seems to be accomplished at different levels and its presence is essential to prepare the ground for the second player that is TcpP. Once TcpP binds the DNA, it recruits and activates the RNA polymerase and the transcription of *toxT* can start (Figure 87E).

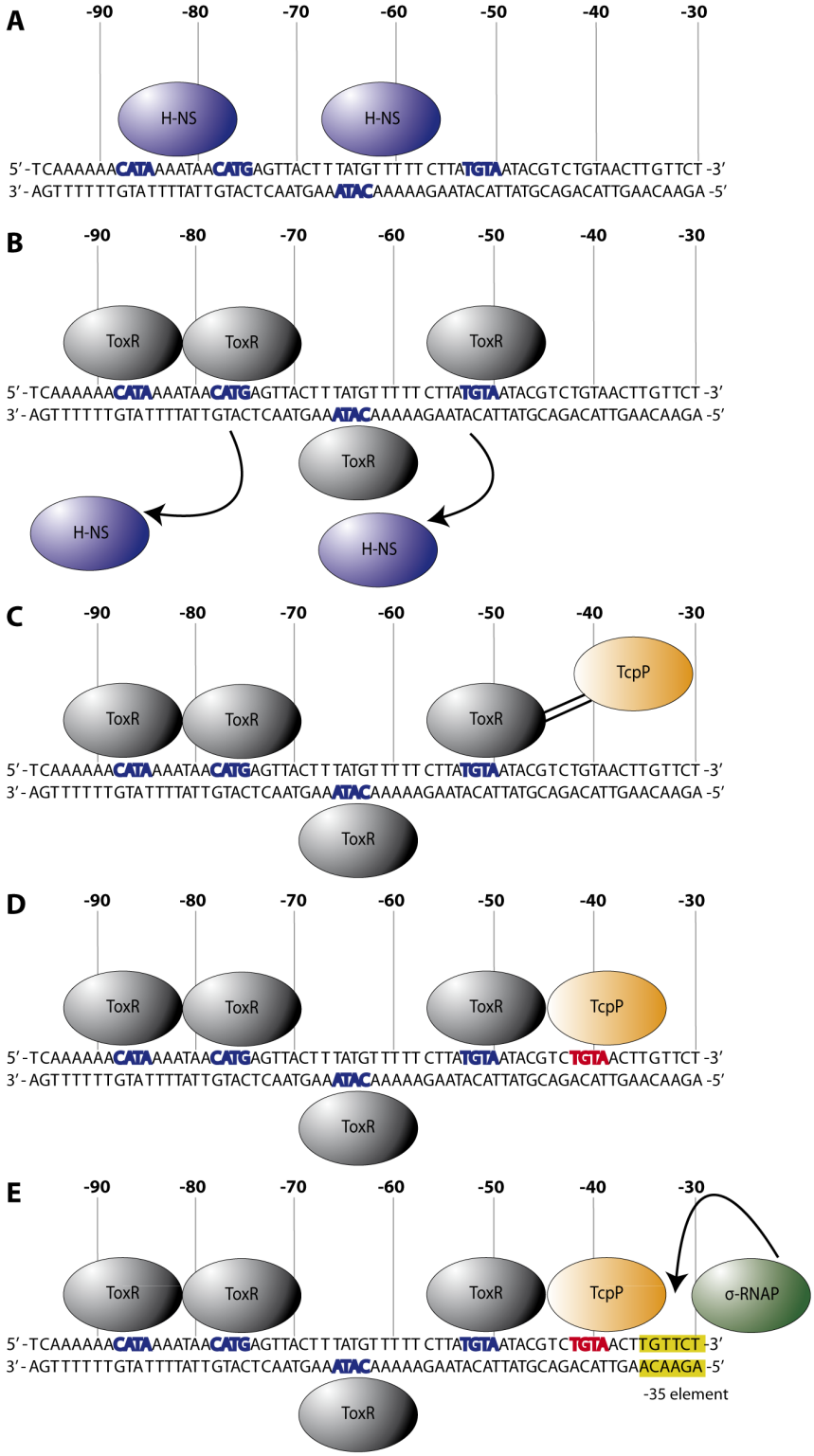

Figure 87. Model of ToxR, a co-activator of the *toxT* transcription.

Finally, we propose two possible variants of our model. The first variant encompasses a different mechanism of recruitment and activation of TcpP (Figure 88). In this case, TcpP would act as dimer and its recruitment could occur through the interaction between the periplasmic domain of both proteins and/or by their transmembranal portion. One monomer of TcpP would compete for the same binding site of ToxR (monomer F) and it would be able to replace ToxR (monomer F), probably for its better affinity for the binding site (-53TGTA-50) (Figure 88D). Hence, TcpP dimer can bind in tandem the *toxT* promoter and activate transcription (Figure 88E). In this variant of the model, ToxR would play a regulatory role and the interaction between the DNA binding domain of the two transcription factors would not be necessary.

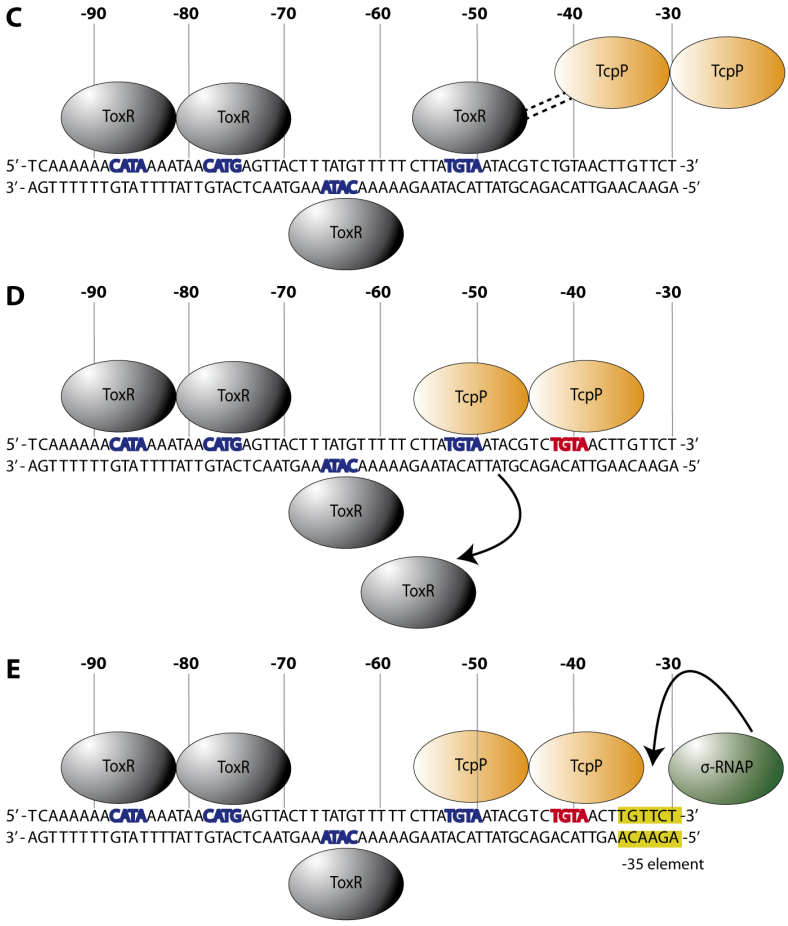

Figure 88. First variant of our model of ToxR as co-activator of the *toxT* transcription.

In the second variant of our model, two TcpP molecules activate the *toxT* promoter, but no one would compete with ToxR for the same binding site (Figure 89). After TcpP recruitment, ToxR-DBD (monomer F) would be able to stabilize the two molecules of TcpP on the *toxT* promoter: one TcpP monomer would bind the -42TGTA-39 box while the second TcpP monomer would bind another binding site between the two TGTA boxes (Figure 89D). The second TcpP monomer would form a transcriptional ternary subcomplex with ToxR-DBD and the *toxT* promoter similar to the one described by Blanco and his collaborators for PhoB (see section 1.3.1). ToxR-DBD structure was superposed to PhoB^{E} in complex with the σ_4 subunit of RNA polymerase and the DNA. This ternary complex provides some hints of how ToxR-DBD could interact with TcpP (Figure 90). Moreover, it could be a suitable model to study the activation of transcription in other promoters as the *ompU* promoter, which requires the presence of ToxR alone.

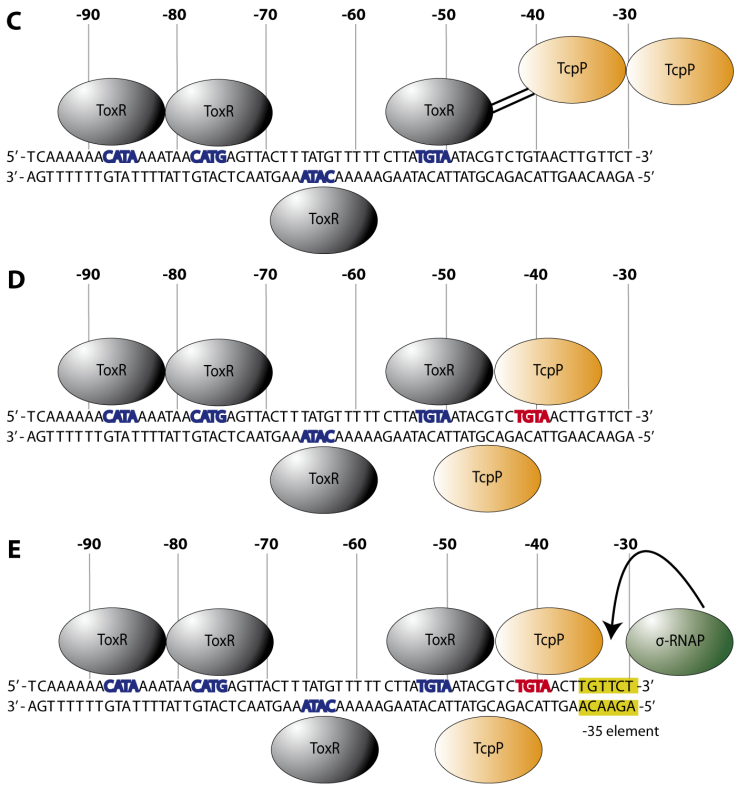

Figure 89. Second variant of our model of ToxR as co-activator of the *toxT* transcription.

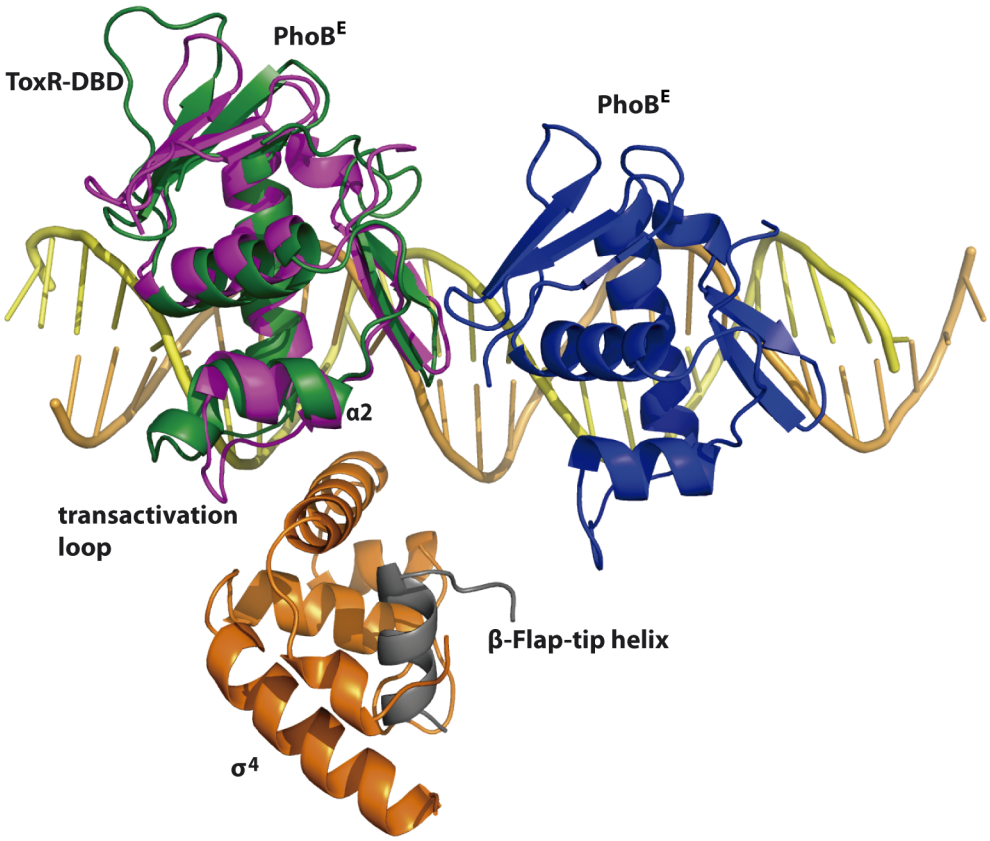

Figure 90. Superposition of ToxR-DBD (green) and PhoB^E (magenta) in a ternary complex with DNA and the σ_4 subunit of RNA polymerase (orange).

More experiments are needed to clarify the interplay between ToxR and TcpP in the activation of the *toxT* promoter. In particular, we aim to determine the structure of the ternary complex of ToxR-DBD with TcpP and the *toxT* promoter.

Chapter 6: Conclusions

1 The ToxR DNA binding domain (ToxR-DBD) from *Vibrio cholerae* shows a winged-helix fold, sharing the most important structural and functional elements with the PhoB DNA binding effector domain ($PhoB^E$).

2 With respect to the $PhoB^E$, ToxR-DBD shows an additional unique element connecting the α3 recognition helix and the β6 strand. This loop plays an active role in DNA binding, enabling additional interactions with the minor groove. We have called this unprecedented element "secondary wing".

3 ToxR-DBD shows a five-stranded, mixed β-sheet that is wider than in PhoB due to the addition of a final β8 strand, parallel to the β1 strand. The fact that the β8 strand is preceding the linker that connects the ToxR-DBD with the periplasmic domain suggests a role of this β-sheet in the transmission of the signal received at the other side of the inner membrane.

4 The ToxR binding to DNA is based on the interactions of the ToxR-DBD recognition helix with the major groove, and those of the wing and the secondary wing with the minor groove. The only direct interactions with DNA bases shown in our three structures are those of the residues Gln78 and Thr77.

5 The DNA in the ToxR-DNA complexes shows only a slight curvature. However, it shows a strong compression of the minor groove, high propeller-twist values and at least one bifurcated H-bond in the A-tract region. These special features seem to be determinant for ToxR binding.

6 The ToxR binding site, referred to as "A-tract-CATA" element, 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.

7 Four ToxR-DBD monomers bind the DNA along the region of the *toxT* promoter extending from position -97 to position -45. The two upstream monomers are bound in

tandem (head to tail) with weak protein-protein interface interactions, while the other two monomers are bound in an isolated manner (with no contact with other neighboring ToxR-DBD monomers).

8 As a unique feature in transcription regulation, the ToxR-DBD monomers bind the *toxT* promoter in two different orientations: three of the monomers bind with the wing element facing the upstream end of the promoter, and the other promoter binds with the wing facing the transcription start site.

9 The furthest downstream ToxR-DBD monomer bound to the *toxT* promoter is overlapping the predicted TcpP binding site. This suggests either a regulatory competition between ToxR and TcpP for the same binding site, or a very close interaction of both proteins binding to contiguous sites.

10 ToxR may act as a co-activator of the *toxT* promoter by several mechanism: (i) capturing the DNA in the cytoplasm and recruiting it to the membrane, (ii) relieving the H-NS proteins from the promoter to allow the TcpP binding and (iii) recruiting TcpP by direct interaction with ToxR-DBD and stabilizing the interaction of TcpP with the *toxT* promoter.

References

Ali M., Lopez A.L., You Y.A., Kim Y.E., Sah B., Maskery B. & Clemens J. (2012) "The global burden of cholera.", *Bulletin World Health Organization* **90**: 209–218.

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**: 9-13.

Aravind L., Anantharaman V., Balaji S., Babu M.M., Iyer L.M. (2005) "The many faces of the helix-turn-helix domain: transcription regulation and beyond." *FEMS Microbiol Rev.* **29**: 231-262.

Arribas-Bosacoma R., Kim S.K., Ferrer-Orta C., Blanco A.G., Pereira P.J., Gomis-Rüth F.X., Wanner B.L., Coll M., Solà M (2007) "The X-ray crystal structures of two constitutively active mutants of the *Escherichia coli* PhoB receiver domain give insights into activation." J Mol Biol. **366**: 626-641.

Beck N.A., Krukonis E.S., DiRita V.J. TcpH (2004) "Influences virulence gene expression in *Vibrio cholerae* by inhibiting degradation of the transcription activator TcpP." *J Bacteriol*. **186**: 8309-8316.

Blanco A.G., Sola M., Gomis-Rüth F.X., Coll M. (2002) "Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator." *Structure*, **10**: 701-713.

Blanco A.G., Canals A., Bernués J., Solà M., Coll M. (2011) "The structure of a transcription activation subcomplex reveals how σ(70) is recruited to PhoB promoters." *EMBO J.* **30**: 3776-3785.

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**: 2548-2561.

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**: 131-143.

Canals A., Blanco A.G., Coll M. (2012) "σ70 and PhoB activator: getting a better grip." *Transcription*. **3**: 160-164.

Chao D.L., Halloran M.E., Longini I.M. Jr. (2011) "Vaccination strategies for epidemic cholera in Haiti with implications for the developing world." *Proc Natl Acad Sci USA* **108**: 7081-7085.

Chen J., Xie J. (2011) "Role and regulation of bacterial LuxR-like regulators." *J. Cell Biochem.* **112**: 2694–2702.

Chen V.B., Arendall W.B. 3rd, Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., Richardson D.C. W (2010) **"**MolProbity: all-atom structure validation for macromolecular crystallography.**"** *Acta Crystallogr D Biol Crystallogr*. **66**: 12-21.

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**: 1330-1340.

Coll M., Frederick C.A., Wang A.H., Rich A. (1987) "A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin." *Proc Natl Acad Sci U S A*. **84**: 8385-8389.

Colwell R.R. (1996) "Global climate and infectious disease: the cholera paradigm." *Science*, **274**: 2025-2031.

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:** 1459-1473.

Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. (2004) "WebLogo: A sequence logo generator."*Genome Research* **14**: 1188-1190.

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**: 136-137.

Cuthbertson L., Nodwell J.R. (2013) "The TetR family of regulators." *Microbiol Mol Biol Rev*. **77**: 440-475

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**: 29-37.

DiRita V.J., Neely M., Taylor R.K., Bruss P.M. (1996) "Differential expression of the ToxR regulon in classical and El Tor biotypes of *Vibrio cholerae* is due to biotype-specific control over *toxT* expression." *Proc Natl Acad Sci USA*. **93**: 7991-7995.

Drew H.R., Dickerson R.E. (1981) "Structure of a B-DNA dodecamer III. Geometry of hydration." *J. Mol. Biol.* **151**: 535-556.

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

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

Gill D.M. (1975) "Involvement of nicotinamide adenine dinucleotide in the action of cholera toxin in vitro." *Proc Natl Acad Sci USA*, **72:** 2064-68.

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 USA*, **75:** 3050-3054.

Girardi A.C, Di Sole F. (2012) " Deciphering the mechanisms of the Na+/H+ exchanger-3 regulation in organ dysfunction." *Am J Physiol Cell Physiol.* **302**: 1569-1587

Goss T.J., Seaborn C.P., Gray M.D., Krukonis E.S. (2010) "Identification of the TcpP-binding site in the *toxT* promoter of *Vibrio cholerae* and the role of ToxR in TcpP-mediated activation." *Infect Immun*. **78**: 4122-3133

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**: 884-895.

Greenough W.B. III. (2004) "The human, societal, and scientific legacy of cholera." *J Clin Invest* **113**: 334-339.

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**: 12116-12149.

Haas B.L., Matson J.S., DiRita V.J., Biteen J.S. (2015) " Single-molecule tracking in live Vibrio cholerae reveals that ToxR recruits the membrane-bound virulence regulator TcpP to the *toxT* promoter. *Mol Microbiol.* **96**: 4-13.

Harris J.B., LaRocque R.C., Qadri F., Ryan ET & Calderwood SB (2012) "Cholera." *Lancet* **379**: 2466-2476.

Häse C.C., Mekalanos J.J. (1998) "TcpP protein is a positive regulator of virulence gene expression in *Vibrio cholerae*." *Proc Natl Acad Sci USA* **95:** 730–734.

Hobman J.L., Wilkie J., Brown N.L. (2005). "A design for life: prokaryotic metal-binding MerR family regulators." *Biometals* **18**: 429–436.

Hoch J.A., Varughese K.I. (2001) " Keeping signals straight in phosphorelay signal transduction." *J Bacteriol*. **183**: 4941-4949.

Huillet E., Velge P., Vallaeys T., Pardon P. (2006). "LadR, a new PadRrelated transcriptional regulator from *Listeria monocytogenes*, negatively regulates the expression of the multidrug efflux pump MdrL." *FEMS Microbiol. Lett*. **254**: 87–94.

Johnson S. (2006). "The ghost map: The Story of London's Most Terrifying Epidemic – and How it Changed Science, Cities and the Modern World."

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

Karaolis D.K.R., Johnson J.A., Bailey C.C., Boedeker E.C., Kaper J.B., Reeves P.R. (1998) "A Vibrio cholerae pathogenicity island associated with epidemic and pandemic strains." *Proc Natl Acad Sci USA* **95**: 3134-3139.

Kim S.K., Kimura S., Shinagawa H., Nakata A., Lee K.S., Wanner B.L., Makino K. (2000) "Dual transcriptional regulation of the *Escherichia coli* phosphate-starvation- inducible psiE gene of the phosphate regulon by PhoB and the cyclic AMP (cAMP)-cAMP receptor protein complex." *J Bacteriol*. **182**: 5596-5599.

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**: 559-592.

Kovacikova G., Skorupski K. (1999) "A Vibrio cholerae LysR homolog, AphB, cooperates with AphA at the tcpPH promoter to activate expression of the ToxR virulence cascade." *JBacteriol.* **181**: 4250-4256.

Kovacikova G., Skorupski K. (2002) "Regulation of virulence gene expression in Vibrio cholera by quorum sensing: HapR functions at the aphA promoter." *Mol Microbiol*. **46**: 1135-1147.

Krissinel E. (2007). "On the relationship between sequence and structure similarities in proteomics." *Bioinformatics* **23**: 717-723

Krissinel E. and Henrick K. (2007). 'Inference of macromolecular assemblies from crystalline state." *J. Mol. Biol*. **372**: 774-797.

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:** 67–84.

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**:157-165.

Laskowski R.A., MacArthur M.W., Moss D.S. Thornton J.M. (1993) "PROCHECK: a program to check the stereochemical quality of protein structures." *J. Appl. Cryst.* **26**: 283-291.

Laub M.T., Goulian M. (2007) "Specificity in two-component signal transduction pathways." *Annu. Rev. Genet*. **41**: 121-145.

Lu X. J. and Olson W.K. (2003). "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures." *Nucleic Acids Res.* **31**: 5108-5121.

Luscombe N.M., Laskowski R.A., Thornton J.M. (1997) "NUCPLOT: a program to generate schematic diagrams of protein-DNA interactions". *Nucleic Acids Res.* **25**: 4940-4945.

Maddocks S.E., Oyston P.C. (2008) "Structure and function of the LysRtype transcriptional regulator (LTTR) family proteins. *Microbiology* **154**: 3609-3623.

Makinoshima H., Glickman M.S. (2006) "Site-2 proteases in prokaryotes: regulated intramembrane proteolysis expands to microbial pathogenesis. *Microbes Infect*. **8**: 1882-1888.

Manning P.A. (1997) "The tcp gene cluster of Vibrio cholerae." *Gene* **192**: 63-70.

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

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

Martínez-Hackert E., Stock A.M. (1997b) " Structural relationships in the OmpR family of winged-helix transcription factors." *J Mol Biol.* **269**: 301-312.

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 USA* **102**: 16403-16408.

Matson J.S., Withey J.H., DiRita V.J. (2007) "Regulatory networks controlling *Vibrio cholerae* virulence gene expression." *Infect Immun*. **75**: 5542-5549.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C. and Read R.J (2007) "Phaser crystallographic software." *J. Appl. Cryst.*, **40**: 658-674.

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**: 5831–5834.

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**: 518-528.

Miller V.L., Taylor R.K., Mekalanos J.J. (1987) "Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein." *Cell* **48**: 271-279.

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**: 157-186.

Morgan S.J., Felek S., Gadwal S., Koropatkin N.M., Perry J.W., Bryson A.B., Krukonis E.S. (2011) "The two faces of ToxR: activator of ompU, co-regulator of *toxT* in Vibrio cholerae." *Mol. Microbiol*. **81**: 113-128.

Morris J. G. Jr. (2003) "Cholera and other types of vibriosis: a story of human pandemics and oysters on the half shell." *Clin Infect Dis* **37**: 272-280.

Murshudov G.N., Skubák P, Lebedev A.A., Pannu N.S., Steiner R.A., Nicholls R.A., Winn M.D., Long F., Vagin A.A. (2011) *"REFMAC*5 for the refinement of macromolecular crystal structures." *Acta Crystallogr D Biol Crystallogr*. **67**: 355-367.

New York: United Nations Office for the Coordination of Human Affairs (2009) "Zimbabwe: OCHA Cholera Update Situation Report" No. 22.

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**: 4295- 4303.

Pennella M.A., Giedroc D.P. (2005). "Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. *Biometals* **18**: 413–428.

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." *JBacteriol*. **180**: 4724-4733.

Pratt J.T., Ismail A.M., Camilli A. (2010) "PhoB regulates both environmental and virulence gene expression in *Vibrio cholerae*." *Mol Microbiol.* **77**: 1595-1605.

Ramos J.L., Martínez-Bueno M., Molina-Henares A.J., Terán W., Watanabe K., Zhang X., Gallegos M.T., Brennan R., Tobes R. (2005). "The TetR family of transcriptional repressors." *Microbiol. Mol. Biol. Rev*. **69**: 326-356.

Reidl J., Klose K.E. (2002) "*Vibrio cholerae* and cholera: out of the water and into the host." *FEMS Microbiol Rev.* **26**: 125-139.

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**: 12507-12515.

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**: 1311-1323.

Robert X, Gouet P (2014) "Deciphering key features in protein structures with the new ENDscript server." *Nucleic Acids Res*. **42**: 320-324.

Rutherford S.T, Van Kessel J.C., Shao Y., Bassler B.L. (2011) "AphA and LuxR/HapR reciprocally control quorum sensing in vibrios." *Genes Dev.* **25**: 397-408.

Sack D.A., Sack R.B., Nair G.B., Siddique A.K. (2004) "Cholera." *Lancet* **363:** 223–233.

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

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**: 41-55.

Skorupski K., Taylor R.K. (1999) "A new level in the Vibrio cholerae virulence cascade: AphA is required for transcriptional activation of the tcpPH operon". *MolMicro* **31**: 763-771.

Solà M., Gomis-Rüth F.X., Serrano L., González A., Coll M. (1999) "Three-dimensional crystal structure of the transcription factor PhoB receiver domain." *J Mol Biol*. **285**:675-687.

Srivastava D., Waters C. M. (2012) "A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP." *J Bacteriol*. **194**: 4485-4493.

Stock A.M., Robinson V.L, Goudreau P.N. (2000) " Two-component signal transduction." *Annu Rev Biochem*. **69**: 183-215

Swint-Kruse L., Matthews K.S. (2009). "Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol*. **12**: 129-137.

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 USA* **84**: 2833-2837.

Trucksis M., Michalski J., Deng Y.K., Kaper J.B. (1998) "The Vibrio cholerae genome contains two unique circular chromosomes." *Proc Natl Acad Sci U S A* **95**: 14464-14469.

Ulrich L.E., Koonin E.V., Zhulin I.B. (2005). "One-component systems dominate signal transduction in prokaryotes." *Trends Microbiol*. **13**: 52-56.

Varughese K.I. (2002) " Molecular recognition of bacterial phosphorelay proteins" *Curr Opin Microbiol.* **5**: 142-148

Wanner BL. (1996) "Phosphorous assimilation and control of the phosphate regulon." In: Neidhardt F.C., Curtiss R.I., Ingraham J.L., Lin E.C., Low K.B.J., Magasanik B., *et al*., "Eds. *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology." *Washington DC: American Society for Microbiology* 1357-81.

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**: 51–62.

World Health Organization. (2008) "Outbreak news – severe acute watery diarrhoea with cases positive for Vibrio cholerae, Viet Nam." *Wkly Epidemiol Rec* **83**: 157-158.

World Health Organization. (2010) "Cholera vaccines: WHO position paper." *Wkly Epidemiol Rec* **85**: 117-128.

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.

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

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**: 7080-7089.

Acknowledgements

Quiero agradecer a los que han hecho possible la realización de este trabajo de tesis y los que me han dado su apoyo durante estos duros años de doctorado.

Quiero dar las gracias al Prof. Miquel Coll por haberme dado la posibilidad de trabajar en su grupo de investigación y de haber confiado en mis capacidades durante mi camino predoctoral.

Un gracias especial va al Dr. Albert Canals por haber sido mi supervisor, un excelente mental coach y un maestro del arte de la cristalografia. Gracias por haberme dedicado mucho tiempo y energía y por haber entrenado mis capacidades.

I want to acknowledge Prof. Eric Krukonis from the University of Detroit Mercy to provide us the purified ToxR-DBD protein and for the fruitful discussions.

Quiero agradecer a la Prof. Josefa Badia por haber solucionado rapidamente cualquier tipo de problema burocratico con la Universidad.

Queiro agradecer La Caixa y el programa “Obra Social La Caixa” por haber financiado mi investigacion.

Quiero agradecer los miembros del “Tribunal Advisory Committee”: Prof. Ignasi Fita, Prof. Ernest Giralt y Prof. Manuel Palacin.

Quiero agradecer los miembros del IRB administration, en particular a Patricia Nadal, Clara Caminal, Alba Olivares y Leyre Caracuel.

Quiero agradecer los miembros de la plataforma de purificación y de cristalización de proteinas (Isabel, Joan, Sonia, Sandra).

Quiero agradecer todos los miembros del grupo “Structural biology of protein & nucleic acid complexes and molecular machines” , los CR1! Quiero empezar por Rosa, que ha ofrecido su experiencia para este trabajo de tesis y que que ha sido importante en los momentos mas duros para darme su apoyo. Sigo dando las gracias al nuestro super tecnico de laboratorio, Esther Ferrando, que me ha solucionado muchos problemas y facilitado mi investgación ademas de luchar cada dia contra el entropia del laboratorio. Gracias tambien a Montse, que ha sido una compañera fantastica durante este camino donde hay riesgo de perderse la joventud! Gracias a Roeland por las mejores pausas del dia, las mas divertidas y fundamentales. Gracias a Salvatore y Radek por haber sido buenos amigos y por haber dado la “vidilla” necesaria al laboratorio para poder trabajar a gusto. Gracias a Cristina por sus consejos y a Zuz por haberme dado un gran ejemplo de perseverancia. Un gracias a los antiguos miembros (Dani, Nayibe, Fabio, Juliana) y a los nuevos doctorandos, los jovenes que han aportado nueva energia positiva al lab: Jorge, Daniel y Mireia.

Un gracias especial va a Mai, Miquel, Jessi y Freixo por haber sido mi familia en estos años, con los martes como si furean domingos y por haber compartido conmigo experiencias muy chulas fuera del IRB.

Voglio poi ringraziare i miei genitori: mio padre Doriano e mia madre Giuliana. Mi avete insegnato che nella vita ci sono da fare dei sacrifici per ottenere le cose più belle ed arrivare ai traguardi piú difficili. Mi avete aiutato e supportato (anzi sopportato) costantemente, e anche se a distanza siete riusciti ad essere sempre presenti durante questa esperienza. Senza di voi non ce l’avrei mai fatta! Allo stesso modo voglio ringraziare mio fratello Luca. Sei stato veramente fondamentale, mi hai incoraggiato costantemente e hai saputo ascoltarmi e consigliarmi come un bravo dottore e soprattutto come un fratello eccezionale.

Dulcis in fundo, ringrazio te, Raffaella Iurlaro, perchè sei stata la mia chaperonina, perchè sei riuscita a darmi la forza di andare avanti anche quando avevo perso la fiducia in me stesso e per l'amore che mi dai ogni giorno. Abbiamo condiviso tanto in questi anni a Barcellona che ci hanno fatto crescere e maturare come persone e come scienziati; ora mi sento pronto per la prossima avventura, sempre insieme, da dottori!