Structural studies on inhibition mechanisms, oligomerization and DNA binding of the transcription regulator Fur : from in silico simulations to in vitro biological assays

Serge Nader

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 25 mai 2016

Présentée par

« Serge NADER »

Thèse dirigée par **Serge CROUZY,** Directeur de Recherche, CEA de Grenoble, et coencadrée par **Julien PERARD,** Ingénieur de Recherche, Institut de Biosciences et Biotechnologies de Grenoble

Préparée au sein du **Laboratoire de Chimie et Biologie des Métaux** dans **l'École Doctorale Chimie et Sciences du Vivant**

Études structurales des mécanismes d'inhibition, d'oligomérisation et de liaison à l'ADN du régulateur de transcription Fur : des simulations *in silico* aux tests biologiques *in vitro*

Thèse soutenue publiquement le **« 23 novembre 2018 »**, devant le jury composé de :

Mme Isabelle SCHALK **Rapporteur**
Directrice de Recherche, CNRS, Ecole Supérieur de Biotechnologie de Strasbourg

M. Thomas SIMONSON **Rapporteur**
Professeur, Ecole Polytechnique

Mme Joanna TIMMINS **Examinateur**
Chargée de Recherche, CNRS, Institut de Biologie Structurale – Grenoble

M. Renaud DUMAS **Président du jury**
Directeur de Recherche, CRNS, Institut de Biosciences et Biotechnologies de Grenoble

M. Serge CROUZY **Directeur de thèse**
Directeur de Recherche, CEA, Institut de Biosciences et Biotechnologies de Grenoble

M. Julien PERARD **Invité**
Ingénieur de Recherche, CNRS, Institut de Biosciences et Biotechnologies de Grenoble

Mme Isabelle MICHAUD-SORET **Invité**
Directrice de Recherche, CNRS, Institut de Biosciences et Biotechnologies de Grenoble

Communauté
UNIVERSITÉ Grenoble Alpes

THESIS

To obtain the degree of

DOCTOR OF PHILOSOPHY AT LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Specialty : **Structural Biology and Nanobiology**

Ministerial decree : 25th of May 2016

Presented by

« Serge NADER »

Thesis supervised by **Serge CROUZY,** Research director, CEA de Grenoble, and **Julien PERARD,** Research Engineer, Biosciences and Biotechnologies Institute of Grenoble

Prepared within the **Chemistry and Biology of Metals Laboratory** in the **Chemistry and Life Sciences Doctoral School**

Structural studies on inhibition mechanisms, oligomerization and DNA binding of the transcription regulator Fur : from *in silico* simulations to *in vitro* biological assays

Thesis publicly defended on the **« 23rd of November 2018 »**, in front of the jury composed by :

Mrs. Isabelle SCHALK **External Examiner**
Senior Researcher, CNRS, Ecole Supérieur de Biotechnologie de Strasbourg

Mr. Thomas SIMONSON **External Examiner**
Professor, Ecole Polytechnique

Mrs. Joanna TIMMINS **Internal Examiner**
Researcher, CNRS, Institut de Biologie Structurale – Grenoble

Mr. Renaud DUMAS **Jury president**
Senior Researcher, CNRS, Biosciences and Biotechnologies Institute of Grenoble

Mr. Serge CROUZY **Thesis Supervisor**
Senior Researcher, CEA, Biosciences and Biotechnologies Institute of Grenoble

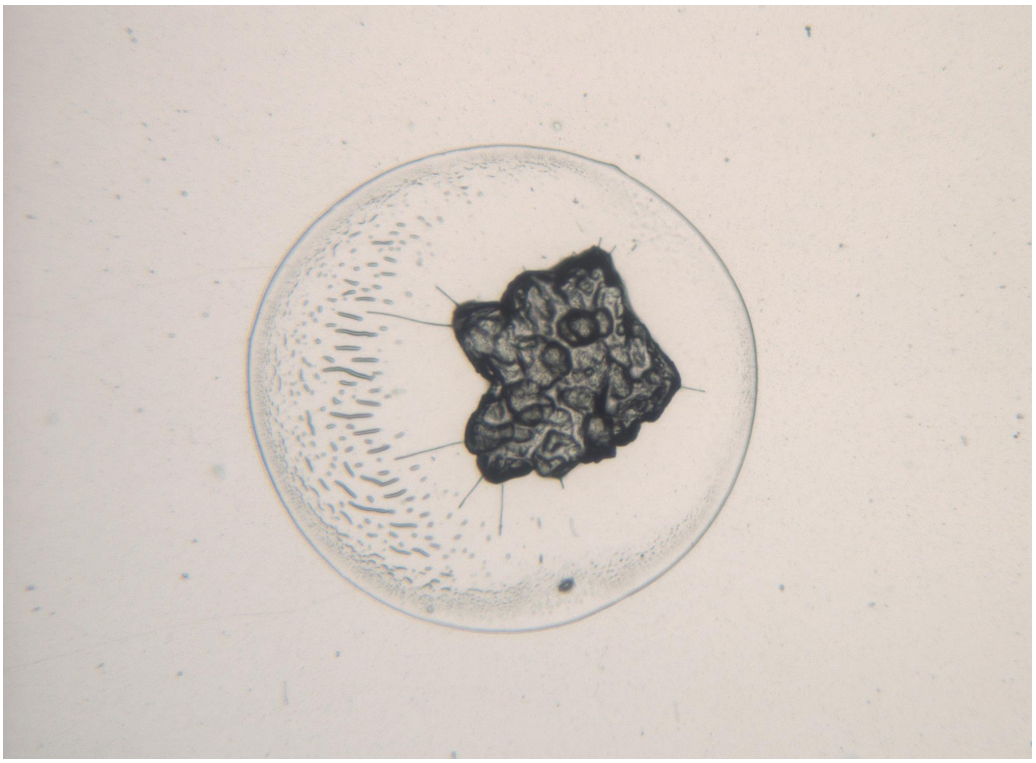
Mr. Julien PERARD **Invited guest**
Research Engineer, CNRS, Biosciences and Biotechnologies Institute of Grenoble

Mrs. Isabelle MICHAUD-SORET **Invited guest**
Senior Researcher, CNRS, Biosciences and Biotechnologies Institute of Grenoble

"Nothing in Biology Makes Sense

Except in the Light of Evolution"

Theodosius Dobzhansky (Dobzhansky, 1973)



Preface

This manuscript is submitted for the degree of Doctor of Philosophy at the University of Grenoble Alpes. The research described herein was conducted under the supervision of Dr. Serge Crouzy and Dr. Julien Pérard in the Chemistry and Biology of Metals Laboratory, at the French Alternative Energies and Atomic Energy Commission (CEA), between October 2015 and November 2018. Financial support was granted by the "Region Auvergne-Rhone Alpes" academic research community: ARC 1 "Santé".

To better introduce, and emphasise the significance of the results obtained from my studies, the introduction will start by describing antimicrobial resistance, highlighting its mechanisms and emergence in nature, in addition to world wide efforts put into surveillance and drug development. The following introduction chapter shows how in the search for new therapeutic targets, one can consider interfering with iron regulation in bacteria. This second chapter, presents how iron became a key element in living organisms and at the same time an exploitable weakness in our battle against infection. The introduction ends with details on the protein studied in this work, the Ferric Uptake regulator, by describing how it works, its structure and my team's previously obtained data.

This work presents original results, except when references are made to previous research. Part of this work has been presented in the following publication:

Julien Pérard, Serge Nader *et al.* "Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in Francisella tularensis virulence". In: Communications Biology 1.1 (July 2018), p. 93. issn: 2399-3642. url: https://doi.org/10.1038/s42003-018-0095-6

Contents

List of acronyms

AMR	Antimicrobial Resistance
BIF	Banded Iron Formation
CHARMM	Chemistry at HARvard Macromolecular Mechanics
COM	Center Of Mass
DNA	Deoxyribonucleic Acid
DMSO	Dimethyl Sulfoxide
DTxR	Difteria Toxin Receptor
EDTA	Ethylenediaminetetraacetic Acid
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic Mobility Shift Assay
ESRF	European Synchrotron Radiation Facility
EXAFS	Extended X-Ray Absorption Fine Structure
Fur	Ferric Uptake Regulator
GOE	Great Oxygenation Event
GROMACS	GROningen MAchine for Chemical Simulations
HTG	Horizontal Gene Transfer
IACT	Integrated Average Correlation Times
ICP-AES	Inductively Coupled Plasma Atomic Emission Spectroscopy
IC50	The half maximal inhibitory concentration
LECA	Last Eukaryotic common Ancestor
LUCA	Last Universal Common Ancestor
MALLS	Multiangle Laser Light Scattering
MD	Molecular Dynamics
MDR	Multidrug-Resistant
MRSA	Meticillin-Resistant *S. aureus*
NOE	Nuclear Overhauser Effect
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PMF	Potential of Mean Force
QS	Quorum Sensing
RI	Refractive Index
RNA	Ribonucleic Acid
SCOP	Structural Classification Of Proteins
SEC	Size-Exclusion Chromatography
SAXS	Small-Angle X-ray Scattering
TCEP	tris(2-carboxyethyl)phosphine
WT	Wild Type
WHO	World Health Organisation
XANES	X-ray Absorption Near Edge Structure
XAS	X-ray Absorption Spectroscopy
Zur	Zinc Uptake Regulator

Fur proteins abbreviations

CjFur	Fur from *C. jejuni*
EcFur	Fur from *E. coli*
FtFur	Fur from *F. tularensis*
HpFur	Fur from *H. pylori*
MgFur	Fur from *M. gryphiswaldense*
PaFur	Fur from *P. aeruginosa*
VcFur	Fur from *V. cholerae*
YpFur	Fur from *Y. pestis*

Chapter 1

Antimicrobial resistance

The discovery of antibiotics has dramatically changed human and veterinary medicine, preventing and curing infections and saving millions of lives. A bacterium eventually becomes resistant to antimicrobial treatment through the natural process of adaptative evolution. However, the misuse of antimicrobial agents greatly accelerates the rate at which resistance emerges. Nowadays, microbial resistance to antibiotics is considered to be a major public health threat as currently available antimicrobial agents lose their effectiveness, and very few new drugs are being developed, many types of infection are becoming life-threatening again especially in poor and overpopulated countries (The Regional Office for Europe of the World Health Organization, 2017; O'Neill, 2016). Even in developed nations, the excessive financial cost of such public health episodes reaches billions of euros. The World Health Organisation (WHO) predicts that by 2050 every 3 seconds one death will be linked to antimicrobial resistance (O'Neill, 2016). It is interesting to note that this is faster than the average person blinking rate of once every 5 seconds.

Antimicrobial drugs are medicines that are effective against infections caused by bacteria, viruses, fungi and parasites. Antimicrobial resistance (AMR) arises when the microorganism survives exposure to a medicine that would under normal conditions kill it or stop its growth. This allows the surviving strains to spread and grow due to a lack of competition, leading to the emergence of "superbugs". According to the Central Asian and Eastern European Surveillance of Antimicrobial Resistance last annual report *"the world is heading towards a post-antibiotic era in which common infections could once again kill"* (The Regional Office for Europe of the World Health Organization, 2017). Resistance to antimicrobial molecules is a natural process observed with a lot of interest since the discovery of antibiotics. Resistant strains to streptomycin were isolated about one year after the discovery of the antibiotic (Price et al., 1947). Even if genes conferring resistance to antibiotics are ancient, AMR has become a problem in recent times mainly due to the misuse of actual drugs, increasing the development and spread of resistance. This is making us face a growing enemy with a largely depleted armoury, risking to lose the ground we gained in the last century.

Previously, resistant strains were associated with hospitals and controlled laboratory settings, but in the last decades the number of resistant infections in communities is increasing (O'Neill, 2016). Another example of increasing resistance comes from fluoroquinolone resistance mechanisms. After their introduction in 1987, it was thought that a resistance to this type of gyrase inhibitors required two independent mutations and was therefore unlikely. However, mutants of the genes involved were later characterised and the fluoroquinolone resistance have increasingly been encountered since (Davies et al., 2010). Whenever resistance, or any other biological process, is biochemically possible it will occur if its evolutionary trajectory and selective pressure are provided.

1.1 Mechanisms of antibiotic resistance

Bacteria have evolved sophisticated mechanisms of drug resistance in order to survive. These resistance qualities are acquired from a pool of resistance genes from other bacterial species including antibiotic-producing organisms. The gene sequences involved in resistance were integrated by recombination, typically via integrons (Davies, 1994), that are DNA sequences structured in what is known as cassettes on which integrase enables genetic material to be integrated into the DNA. Resistance to antibiotics can be achieved through multiple biochemical pathways. In what will follow, a list of the major mechanisms is given in Figure 1.1 (Munita et al., 2016).

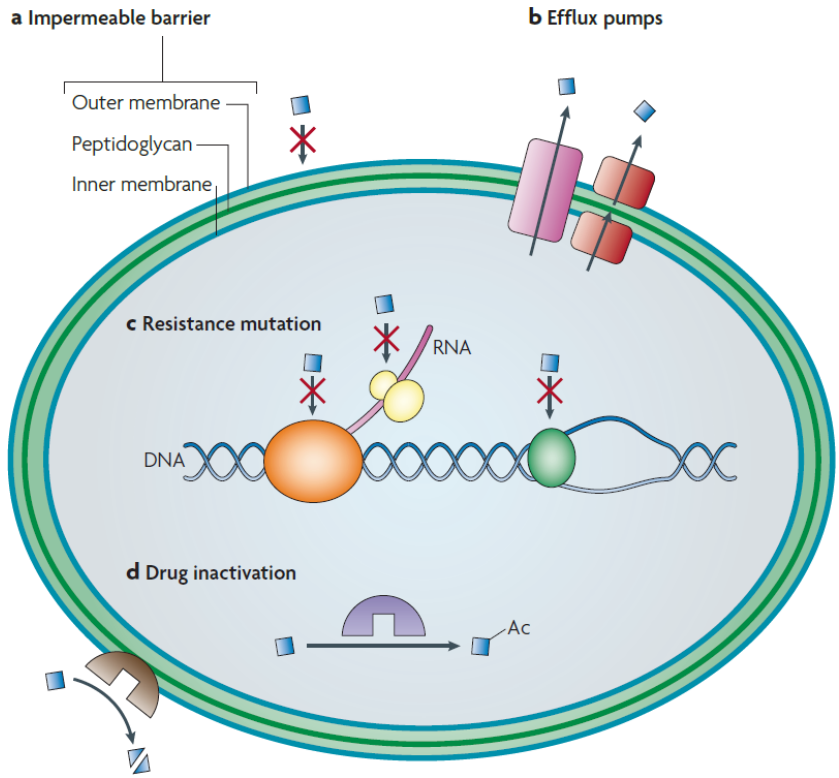


Figure 1.1: Mechanisms of antibiotic resistance in Gram-negative bacteria (Allen et al., 2010)

1.1.1 Decreased membrane permeability and use of efflux pumps

Decreased membrane permeability is frequent in gram-negative bacteria where molecules such as tetracyclines and some fluoroquinolones are affected by changes in permeability since they use porins to cross the outer membrane, Figure 1.1: a. Porin-mediated antibiotic resistance can be achieved through a change in the expression levels and types of porins, sometimes the impairment of porin function is observed. For example, in antibiotic resistant *P. aeruginosa*, mutations of the *oprD* gene, used for basic amino acids acquisition, leads to decreased uptake of antibiotics. Another way of resisting antibiotics is the production of complex machineries capable of extruding the toxic compound, Figure 1.1: b. These efflux pumps can act on a wide range of antimicrobial classes and are divided into families that differ in structural conformation, energy source, extruded substrates and bacterial host. For example, Tet efflux pumps extrude tetracyclines using proton exchange and are considered a classic example of efflux-mediated resistance (Munita et al., 2016).

1.1.2 Resistance mutations

Bacteria have evolved tactics such as protection and modifications of the target site to achieve resistance, Figure 1.1: c. An example of target protection is the tetracycline resistance determinants Tet(M) and Tet(O) acting as elongation factors they interact with

the ribosome and remove tetracycline from its binding site. In addition, this interaction alters the ribosomal conformation preventing rebinding of the antibiotic. Modification of target sites is the most common mechanism of antibiotic resistance. It can be achieved by point mutations, enzymatic alterations of the binding site and replacement of the original target. Point mutations occur in rifampin resistance where mutations of RNA polymerase decrease affinity of the drug for its target. Enzymatic alterations of the binding site has been characterized in macrolide resistance where methylation of the ribosome occurs through an enzyme encoded by the *erm* genes. Replacement or bypass of the target sites is encountered in bacteria capable of evolving new macro-molecules accomplishing similar functions as the original target but not inhibited by the antimicrobial molecule. Methicillin resistance in *S. aureus* (MRSA) uses this mechanism after acquisition of *mecA* genes encoding a penicillin binding protein PBP2a that has low affinity for all β-lactams (Munita et al., 2016), allowing transpeptidase activity in the presence of these molecules.

1.1.3 Drug inactivation

The most successful bacterial strategy is to produce enzymes that inactivate the drug. This can be done by adding specific chemical groups to the antibiotic or by destroying the molecule itself, which makes the antibiotic unable to interact with its target, Figure 1.1: d. Chemical alteration can be achieved by acetylation, phosphorylation, or adenylation creating a steric hindrance that decreases the avidity of the drug for its target. An example is the modification of chloramphenicol by acetytransferases known as CATs, *cat* genes have been described in both Gram-positive and Gram-negative bacteria. In the case of destruction of the antibiotic molecule, β-lactam resistance is a good example since it relies on β-lactamases that destroy the β-lactam ring.

1.1.4 Coupling of resistance mechanisms

In several bacterial species, resistance occurs through a combination of mechanisms. Fluoroquinolone resistance can occur using different biochemical pathways : over-expression of efflux pumps, protection of topoisomerases binding sites by the Qnr protein, and mutations in genes encoding DNA gyrase and topoisomerase IV modifiying the target sites. In other cases, evolution promotes some mechanisms of resistance over others. The main mechanism of resistance to β-lactams in Gram-negative bacteria is the production of β-lactamases, while resistance to these compounds in Gram-positive bacteria is accomplished by modification of their target : the penicillin-binding proteins (Munita et al., 2016).

Nevertheless, familiar mechanisms of antibiotic resistance, like the ones cited above, do not seem to be responsible for the protection of bacteria in biofilms. Several hypothesis are proposed, the first is the slow or partial penetration of antibiotics into the biofilm.

The second is the altered chemical microenvironment within the biofilm such as anaerobic niches, local accumulation of acids and variable osmotic pressure that can antagonise the action of antibiotics or affect bacterial permeability. The third hypothesis is the presence of highly protected subpopulations in spore-like states providing a powerful defence against antibiotics (Stewart et al., 2001).

1.2 Antibiotic resistance is ancient

Trying to understand the origins of genes conferring resistance to antibiotics, scientists debate on the presence of such genes in the pre-antibiotic era and the selection pressure placed on them as a result of human activity. For example, *oxa* genes encode β-lactamases and confer resistance to β-lactam antibiotics, these genes are found both on chromosomes and on plasmids. Phylogeny of plasmid-borne *oxa* genes show that they have existed since the Cambrian Explosion, for over 500 million years, and were mobilized from chromosomes to plasmids a least two times independently at 116 and 42 million years ago. Plasmid-encoded beta-lactamase are suggested to have been originally penicillin-binding proteins involved in the synthesis of peptidoglycans (Martinez, 2009).

This observation contradicts the common impression that mobilization of resistance genes is a strict result of modern use of antibiotics (Barlow et al., 2002; Allen et al., 2010). Additional research on other β-lactamase groups such as the serine β-lactamase and metallo-β-lactamase groups from remote areas showed that both are very ancient to the point that homology between them is lost. Using structure based phylogeny, it was shown that they both originated on bacterial chromosomes more than two billion years ago (Davies, 1994). Moreover, genetic elements encoding resistance to β-lactam and tetracycline have been isolated from 30000 year old permafrost sediments (D'Costa et al., 2011).

Another origin of current resistance genes comes from housekeeping genes such as the sugar kinases and acyltransferases that may have evolved to modify aminoglycoside antibiotics (Davies, 1994). A phylogenetic analysis of antibiotic resistance genes encoding ribosomal protection proteins (RPP) that confer resistance to tetracyclines, shows a branching and diversification of clusters before the modern antibiotic era. The same observation was obtained for the *erm* gene family encoding enzymes that protect ribosomes from macrolide antibiotics, the quionolone resistance genes (*qnr*) and of the *vanHAX* cluster confering high-level vancomycin resistance, proving their existence way before the use of modern antibiotics (Aminov et al., 2007; Davies et al., 2010).

A study of the Murray collection comprising several bacterial strains, mainely *Enterobacteriaceae*, collected in the pre-antibiotic era between 1917 and 1954 showed that numerous old plasmids belong to the same group as today's resistance plasmids indicating that the lineage of the latter is from plasmid resident in enterobacteria before synthetic

antibiotics were used (Datta et al., 1983). More recent work showed that genetic determinants of mobilisable AMR are indeed present in the collection (Baker et al., 2015). Gene encoding for antibiotic resistance are often associated with the mobilome, the ensemble of mobile genetic elements, and can be transferred between distantly related species. Transposable elements with almost all combinations of antibiotic resistance genes have been identified (Davies, 1994).

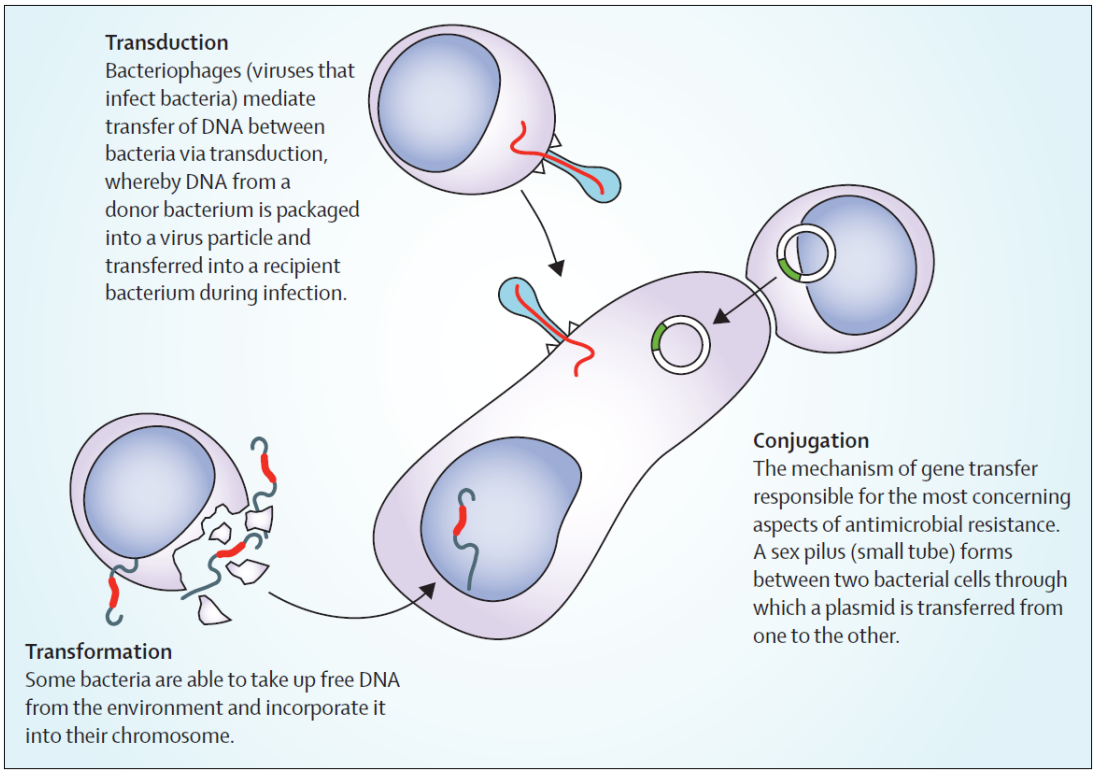


Figure 1.2: Horizontal gene transfer between organisms (Holmes et al., 2016).

The most common mechanism of horizontal gene transfer (HTG) is plasmid-mediated transmission and is shown in Figure 1.2. The prevalent mechanism of genetic exchange being conjugation, when genetic material is transferred through direct contact between two bacteria. Transduction is the process by which DNA is transferred from one bacterium to another by a virus. Bacteriophages carrying antibiotic resistance genes are rarely identified in hospital isolates of resistant strains but are common in the case of *S. aureus*. And transformation that involves taking up DNA molecules form the external environment to be incorporated into the genome of the recipient cell. Other processes may exist, for example cell to cell fusion such as those found in biofilms (Davies et al., 2010).

In several cases, the presence of antibiotic producing species in the same environment as other bacteria provokes the transfer of their own resistance genes. In addition some antibiotics have been shown to promote plasmid transfer between different bacterial species and by doing so being labelled as bacterial pheromones (Davies, 1994) playing major roles in bacterial evolution and speciation (Mazodier et al., 1991).Given the appropriate transfer machinery, resistance genes can in theory be acquired from any source. Luckily,

gene flow in nature is structured by ecology (Wellington et al., 2013).

HTG played a key part in evolution, but what happened during the evolution of bacteria over billions of years cannot be compared to the phenomenon of antibiotic resistance that took place over the last century. What we are witnessing in our lifetime is an evolutionary process intensified by anthropogenic influence rather than the slower course of natural evolution (Davies et al., 2010).

1.3 Alternative roles of antibiotic resistance in nature

Antibiotic producing strains carry genes encoding resistance to the antibiotics that they produce. In addition to providing self protection, resistance genes, found in the same gene cluster as the antibiotic biosynthesis pathway, could be involved in the regulation of antibiotic production (Allen et al., 2010).

Toxic compounds can be found in non-clinical settings, some have biotic origins such as antibiotics and antimicrobial agents produced by plants and fungi, others have abiotic origins such as heavy metals derived from the earth crust and it's erosion. The presence of offensive compounds or environments might select for antibiotic resistance mechanisms. Bacteria cultured form the marine air-water interface were shown to be more resistant to antibiotics than bacteria cultured directly from water. The same was observed for environments with radiation or pollution. So some genes conferring antibiotic resistance are likely to have other primary roles.

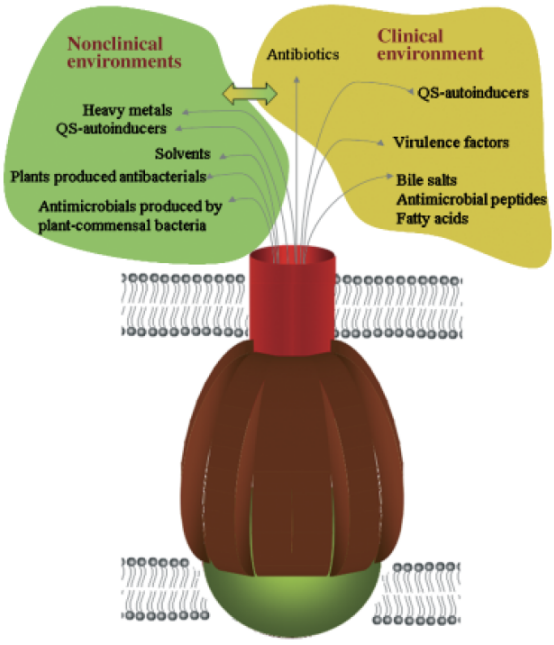


Figure 1.3: Role of multidrug resistant (MDR) pumps in nonclinical and clinical environments (Martinez et al., 2009). QS : Quorum Sensing.

Genes encoding multidrug efflux pumps are present in chromosomes across the three domains of life and are generally conserved, showing that they are not exclusive to antibiotic producing species and suggesting that their major roles are not resisting the antibiotics used in therapy. The potential roles proposed are relevant to the behaviour of bacteria in their natural ecosystems. It has been demonstrated that efflux pumps are necessary for detoxification of intracellular metabolites, bacterial virulence, cell homeostasis and intercellular signal trafficking (Martinez et al., 2009). Certain classes of efflux pumps are known to evacuate various toxins and heavy metals, offering a general mechanism of resistance not only to antibiotics Allen et al., 2010. Figure 1.3 summarizes these mechanisms in

the case of an RND pump from Gram-negative bacteria.

The first step of bacterial infection is colonization. This can be inhibited by host-produced compounds such as bile salts, long-chained fatty acids and antimicrobial peptides. To avoid these compounds and colonize a tissue bacteria use MDR efflux pumps. For example, the AcrAB system in *E. coli* confers resistance to bile salts *in vitro*. In the case of *Salmonella typhimurium* and *Francisella tularensis* inactivation of the AcrAB system impairs colonization, showing that MDR efflux pumps are required for full virulence in certain species (Martinez et al., 2009).

In quorum sensing bacteria produce small molecules, sometimes called autoinducers, to regulate its population. Quorum sensing includes biosynthesis of antimicrobial peptides, polysaccharide synthesis and virulence factor such as elastases and proteases. An interesting aspect of quorum sensing is how the autoinducers accumulate in the environment as the population density increases. Some diffuse freely but others require transporters and this is where multidrug resistant efflux pumps play an important role. In *P. aeruginosa* a mutation of the MexAB-OprM efflux pump affects release of N-3-oxododecanoyl homoserine lactone (3-Oxo-C_{12}-HSL) decreasing the production of quorum sensing controlled virulence factors, the same was observed for other efflux pumps mutants. Loss of multidrug resistant efflux pumps in *Vibrio cholerae* decreased the expression of the virulence gene regulator *tcpP*, possibly by affecting cell-to-cell signalling (Martinez et al., 2009). In some carbapenems producing saprophytic bacteria, living on decaying organic matter, the genes implicated in the synthesis of antibiotics are shown to have a role in the quorum-sensing apparatus and formation of biofilms (Holmes et al., 2016).

These efflux pumps are important for some species that thrive in extreme environments. *Cupriavidus metallidurans* grows in biotopes rich in heavy metals, volcanic biotopes for example, and uses like all micro-organisms trace amounts of heavy metals as cofactors of several proteins. Therefore, to maintain a finely tuned metal homeostasis, bacteria utilize efflux pumps to regulate their intracellular heavy metal concentrations. In addition, organic solvents present in non-anthropogenic petroleum in soil and water, as a product of biosphere activity, can be toxic compounds to bacteria, their tolerance is mediated by efflux pumps. This mechanism is also used in species that are not specialized in growth on high concentrations of heavy metals, since genes coding for efflux pumps are shared with great degree of conservation (Martinez et al., 2009).

Some bacterial species can grow on antibiotics as sole carbon and nitrogen sources. *Pseudomonas fluorescens* grows on streptomycin, *Burkholderia cepacia* grows on penicillin, *Streptomyces venezuelae* grows on chloramphenicol. Without their resistance genes these organisms would not be able to survive. Some of these isolates exhibit resistotypes not seen before, emphasising the necessity of their study for a better prediction of the future on resistance (Allen et al., 2010).

1.4 Alternative roles of antibiotics in nature

Microbial populations are capable of producing a wide spectrum of small bioactive molecules, this diverse mix of molecules is termed the parvome (Davies, 2013). Usually used for quorum-sensing reactions and affecting pathogenesis, these molecules can activate biochemical pathways in target organisms at low concentrations. Interestingly at high concentrations, some molecules of the parvome have antibiotic activity Barlow et al., 2002; Allen et al., 2010. The different effects of the same molecules can be interpreted as signalling responses that adapt the bacterial metabolism in mixed microbial communities. In addition, antibiotics have an important role in some cases of symbiosis. For example, some ant species rely on fungus growing as food source. They carry on their cuticle an antibiotic producing actinomycete, *Pseudonocardia sp.*, which is used to biologically control their fungal garden. In other cases of parasitism, bacteria evolved to recognise antibiotics secreted by plants as signalling molecules to induce adaptation to the plant defence system allowing colonisation of the rhizosphere (Allen et al., 2010).

1.5 Antibiotic pollution and resistance in the biosphere

In contrast to intrinsic resistance, developed by bacterial population over the past millions of years, acquired antibiotic resistance is a recent event in the evolution of human pathogens with the main selective force being the human use of antibiotics (Martinez, 2009). The major role played by human activities in the generation of environmental reservoirs of antibiotic resistance cannot be disputed. Starting from the 1940s, increasing amounts of antibiotics have been produced and widely disseminated in the environment, thus providing constant selection pressure for resistant strains in the biosphere. For example, genomic and microbial studies of waste-water treatment plants have pointed out that they are now rich reservoirs for resistant genes and organisms (Davies et al., 2010). Antibiotics released into the sewage system after use in humans will be degraded, associated with sewage sludge or released into rivers. Sludge-associated drugs will pollute agricultural systems with the use of sludge as fertiliser. Some antibiotics are not easily biodegradable and some persist in soils for long periods of time. Fluoroquinolones can persist in the environment for month or even years (Wellington et al., 2013).

To take into account the extended reach of this antibiotic pollution, recent studies have shown the presence of antibiotic resistance genes in the gut flora of people who live in isolated areas and not exposed to modern civilization antibiotic therapy. Researchers were able to demonstrate that the acquired resistance genes were of the same type as those found in antibiotic-exposed settings. These data suggest that the resistance in such remote communities is likely to be the consequence of clonal expansion and horizontal transfer of genetic elements harbouring resistance genes from antibiotic-exposed settings

and not the result of an independent *in situ* selection (Pallecchi et al., 2007).

But how can antibiotic resistant bacteria travel great distances ? Physical forces like wind and watershed can spread antibiotic resistance genes. Even bacteria from environments that are thought to be stationary such as soil, can be moved, one example is the intercontinental transport of bacteria on desert dust. Proximity to human activities can also influence antibiotic resistance profiles in wild animals. Bacterial isolates from captured mice in rural England are resistant to β-lactam antibiotics. Likewise, isolates from wild animals in Mexico and Australia have a hight frequency of antibiotic resistance. In Africa, apes that are in contact with humans harbour more antibiotic resistant strains than those that inhabit more remote areas of the continent (Allen et al., 2010).

Moreover, a key reservoir for long distance dissemination of antibiotic resistance is wild migratory birds with their long distance travel and the wide variety of environments they inhabit. Their impact can be seen on Arctic birds that are showing pattern of antibiotic resistance similar to clinical isolates, this can be explained by the fact that many migratory birds breed in the Arctic and migrate to other continents. The environmental reservoirs of resistance are poorly understood demonstrating the need for more investigations on this subject (Allen et al., 2010).

Antibiotics are needed in agriculture and aquaculture to maintain animal health and welfare as well as food security. This wide scale use of antibiotics encourage the development of resistance. Studies suggest that 75-90% of used antibiotics are excreted from animals un-metabolised and enter water courses and sewage systems. The same problem is observed in human waste from hospitals and in the environment of active pharmaceutical ingredients manufacturers (O'Neill, 2016).

More antibiotics are used in food production than human medicine with an estimation of 400 mg of antibiotics per kg of meat produced in some European countries. In the United Kingdom alone, around 400 tonnes of antibiotics are used each year for food producing animals. In the European Union the use of antibiotics to promote growth in livestock was banned in 2006. It is interesting to note that such practices are still present worldwide like in the United States of America for example. The use of copper, or other metals, as bactericide in fields or its natural occurrence in certain areas raises concerns since metals can co-select for resistance to antibiotics and heavy metals. Alarmingly, evidence of β-lactamase genes transmission from livestock to humans have been reported. Educating farmers to reduce antibiotic use in addition to legislative enforcement can help improve the current situation (Wellington et al., 2013; Holmes et al., 2016).

Large scale manufacturing of antibiotics is also problematic in countries where regulations tend to be vaguely defined. In India and China this problem arises as studies show extremely high antibiotic concentrations, few milligrams per litre instead of the accepted nanograms per litre, in river water downstream of pharmaceutical industries (Wellington et al., 2013). In addition to agriculture and pharmaceutical industries, often neglected

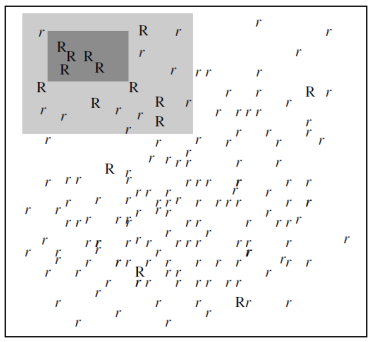


Figure 1.4: Ecological landscapes of antibiotic resistance (Martinez, 2009). R: resistance determinants present in antibiotic producers; r : metabolic or signalling determinant that can be used for resistance at high antibiotic load.

social issues can drive antibiotic resistance through prescription of unnecessary antibiotics to retain patients loyalty. Moreover, the increased internet access opened the way to even more unregulated purchasing of antibiotics, automedication and unregulated disposal in the environment (Wellington et al., 2013).

Most works studying the impact of human activity on the biosphere, such as the over use of antibiotics, are founded on the study of higher organisms. However, the majority of life on earth is microbial and the consequences of environmental changes on the microbiosphere are largely ignored. What should be taken into account is the existence of three different landscapes important in the evolution of resistance of human pathogens (Martinez, 2009), Figure 1.4.

The first level, all white panel, is the natural biosphere where there is a large number of genes capable of conferring antibiotic resistance to human pathogens through HGT (R genes). other genes, (r genes) have also been selected for detoxification, communication or other functions different from antibiotic resistance, unless they are expressed at high concentrations. The second level, grey square in Figure 1.4, consists of habitats where contact of human associated bacteria is frequent with the environmental microbiota. These places, like waste waters, are considered to be hotspots for antibiotic resistance genes acquisition by pathogenic bacteria. Furthermore, since waste waters usually contains antibiotics the selection for resistant strains increases. The third level, dark grey square in Figure 1.4, is the clinical environment or even the patient where selection will favour already acquired resistance genes and mutation-driven diversification genes whose function is only resistance (Martinez, 2009).

Humans can be exposed to antibiotic resistance genes or antibiotic resistant bacteria through several ways : crops exposed to contaminated sludge or manure, livestock treated with veterinary drugs and containing resistant flora, fish exposed to drugs soluble in water, contaminated water sources and groundwater used for drinking, contaminated coastal waters used for bathing or shellfish production (Wellington et al., 2013).

1.6 Problems facing new drug development

It is clear that investment in the development of new antimicrobial drugs and strategies is urgently needed. Lack of such investments shows the fears that resistance will eventually develop, limiting returns on investments due to restrictions in use of the no longer efficient newly developed drugs. This makes development of new antibiotics seem less attractive than business investment in medicines for chronic diseases, pushing major pharmaceutical companies to stop research in this area. And instead, investing in scientifically challenging but commercially lucrative disease area like cancer (World Health Organization, 2015a). Moreover sales of patented antibiotics make up only 10%, around 4 billion USD, of a relatively large total market of antibiotics worth 40 billion USD of sales a year. Adding to that the unattractive commercial return on R&D investment until widespread resistance has emerged against previous generations of drugs. So it is not surprising that firms are not investing in antibiotics despite the very high medical needs, Figure 1.5.

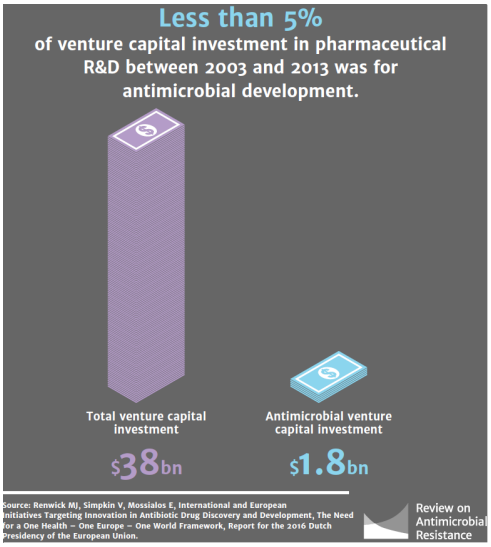


Figure 1.5: Distribution of investments in pharmaceutical R&D (O'Neill, 2016).

The same scheme is true in the allocation of public research funds by governments, the United States of America National Institutes of Health gave 1.2% of its grants to fund AMR related research compared to 18.6% to cancer research between 2009 and 2014 (O'Neill, 2016). What is needed is new processes that facilitate investments in research and development of new antibiotics and ensure a governed public health framework for the use of new products to conserve their effectiveness and longevity (World Health Organization, 2015a; O'Neill, 2016).

1.7 Worldwide surveillance and control centers

Worldwide, several control centers are on the lookout for every new case of infection by a resistant micro-organism. All these centres share the same objective of preventing any possible high scale spread of the antibiotic resistance infections. The main surveillance networks are the following :

- The European Antimicrobial Resistance Surveillance Network (EARS-Net) coordinated by the European Centre for Disease Prevention and Control. The Central Asian and Eastern European Surveillance of Antimicrobial Resistance (CAESAR) network, an initiative that aims to support all countries of the European Region that are not part of the EARS-Net.
- The Global Antimicrobial Resistance Surveillance System (GLASS) promotes a worldwide understanding and engagement to support the global effort to control AMR by collecting an unprecedented set of information related to AMR at global level through a network of surveillance sites linked to national reference laboratories and national coordination centers. This World Health Organisation (WHO) supported system aims to analyse and share data in order to optimize decision-making and drive local and regional action.
- The Interagency Coordination Group on Antimicrobial Resistance (IACG). The United Nations Secretary-General has established IACG to improve coordination between international organizations and to ensure effective global action.

All these organisations share the same goals to improve awareness and understanding of AMR, strengthen surveillance and research, reduce the incidence of infection, optimize the use of antimicrobial medicines, develop the economically sustainable investment in new medicines (The Regional Office for Europe of the World Health Organization, 2016; World Health Organization, 2015a; World Health Organization, 2015b).

In addition to all international networks, joint public-private partnerships like the Global Antibiotic Research and Development Partnership (GARDP) are encouraging research and development in the aim of delivering new treatments. Annual events are also used to raise awareness like the World Antibiotic Awareness Week an WHO initiative held every November since 2015 to increase awareness of global antibiotic resistance and to encourage correct practices among the general public.

1.8 Species under high surveillance

With the alarming rate at which antibiotic resistance infections are rising, especially in the case of community acquired resistant infections, some bacterial species draw attention more than others. This is the case for example of *E. coli* and *S. aureus* infections with one in four infections being resistant to classical, or first-line, antibiotic treatment (The Regional Office for Europe of the World Health Organization, 2017). To face this, the WHO placed the species listed below under GLASS surveillance (World Health Organization, 2015b). Figure 1.6 summarizes the different classes of antibiotics and their modes of actions.

Acinetobacter spp.

Acinetobacter spp., especially species belonging to the *A. baumannii* group, are intrinsically resistant to many antimicrobial agents due to their ability to exclude various molecules from penetrating their outer membrane. Acquiring a multidrugresistant strain is usually caused by prolonged mechanical ventilation and intensive care unit hospitalisation. Colistin is the only effective antibiotic left but emerging resistance is being detected.

Escherichia coli

Despite the fact that it is part of the normal flora in the intestine in humans and animals, *E. coli* is the cause of most community and hospital-acquired urinary tract and bloodstream infections. in the last few years, *E. coli* became the leading cause of foodborne infections worldwide. Resistance in *E. coli* develops through mutations and acquisition of mobile genetic elements. Usually carbapenems, members of the beta lactam class of antibiotics, are the only available treatment for severe infections, although carbapenem resistance mediated by carbapenemases is emerging. Alternatively, colistin use is increasing with rare cases of resistance first described in China.

Klebsiella pneumoniae

The majority of human infections caused by *K. pneumoniae* are health-care associated. Resistance in *K. pneumoniae* develops through mutations and acquisition of mobile genetic elements. With high proportions of cephalosporin resistance the treatment of severe *K. pneumoniae* rely on carbapenems. However, *K. pneumoniae* is today the main cause of carbapenem-resistant bacterial infections worldwide rendering almost all available treatment options ineffective. Which leads to the use of last resort drugs like tigecycline, a tetracycline derivative, and colistin, a polypeptide antibiotic.

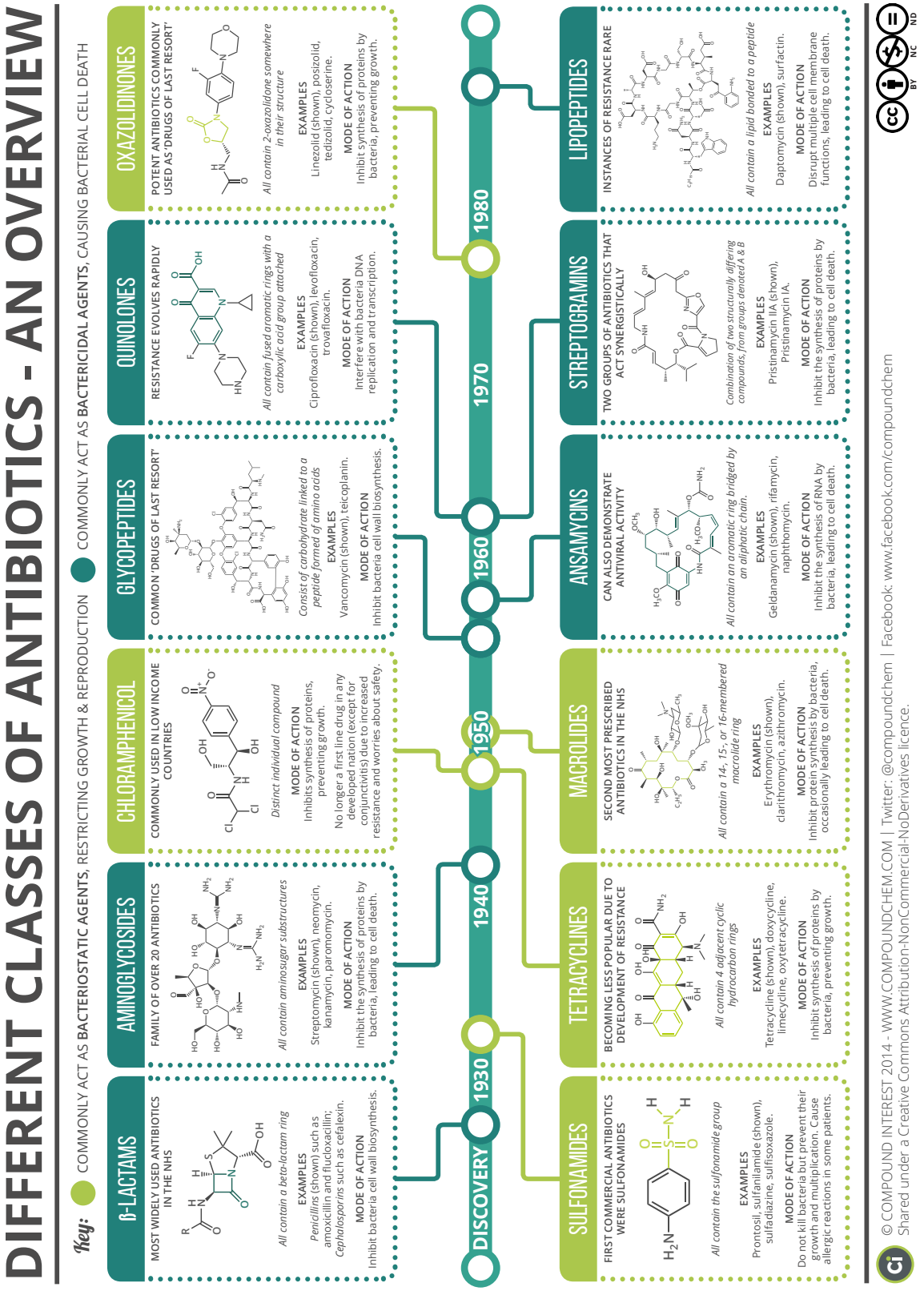


Figure 1.6: Timeline showing the discovery of major classes of antibiotics and their modes of action (*Different classes of antibiotics - An overview*).

Neisseria gonorrhoeae

N. gonorrhoeae is the causal agent of gonorrhoea, the acute sexually transmitted infection of the reproductive tract in humans. Actually the last remaining option for treatment are third-generation cephalosporins. With few new treatment options in the drug development pipeline and several countries reporting treatment failures the WHO fears that gonorrhoea will emerge as a silent epidemic in the foreseeable future.

Salmonella spp.

Salmonella spp. can be found in the intestines of food-producing animals and in human or animal faeces. Infections are usually acquired by consumption of contaminated food which can lead to foodborne outbreaks with severe cases of enteric fever caused by *Salmonella typhi*. First detected in 1989, multidrug-resistant strains of *Salmonella spp.* spread worldwide. Fluoroquinolones are suggested for the treatment of enteric fever caused by species resistant to first-line antibiotics. Reduced susceptibility to oral drugs and reports of treatment failures are of concern.

Shigella spp.

Shigella spp. are the major cause of dysentery and diarrhoea throughout the world. They are usually transmitted by ingestion of contaminated food or water and through human contact in poor sanitation conditions. After resistance to first-line antibiotics like cotrimoxazole and ciprofloxacin, *Shigella spp.* strains are now resistant to second-line antimicrobial drugs, such as the third-generation cephalosporin ceftriaxone and the macrolide azithromycin.

Staphylococcus aureus

S. aureus can be part of the normal human microbiota but remains one of the most important human pathogens, causing a variety of infections, most notably skin, soft tissue, bone and bloodstream infections. *S. aureus* developed resistance to penicillin, a few years after it's introduction in the 1940s, by the production of a beta-lactamase. Therefore beta-lactamase inhibitors were developed to be administered with the antibacterial drugs. However, strains of *S. aureus* have become resistant to these penicillinase-stable antibacterial drugs and termed meticillin-resistant *S. aureus* (MRSA). Second-line drugs used to treat MRSA infections are expensive and treatments of last resort using glycopeptides such as vancomycin and teicoplanin require special care to avoid adverse side-effects.

Streptococcus pneumoniae

S. pneumoniae is the leading cause of community acquired pneumonia worldwide causing also high mortality invasive disease such as meningitis. Resistance to antibacterial drugs occurs by acquiring mutations in the genes coding for the penicillin-binding proteins. In addition, *S. pneumoniae* has a variety of virulence factors that facilitate the host's infection. Resistance has been linked to worse clinical outcomes in patients with pneumococcal meningitis, most likely leaving survivors with permanent residual symptoms.

Based on the rising drug resistance in pathogens under surveillance, estimations show that unless action is taken, death toll from AMR will reach 10 million each year by 2050. This is a staggering one person every three seconds (O'Neill, 2016). Which explain why the listed species are on the Priority Pathogens List of antibiotic-resistant bacteria as critical priorities for research and development of new and effective antibiotic treatments (World Health Organization, 2015b).

1.9 Sustaining the therapeutic life of current antibiotics

Several approaches have been proposed to to prolong the useful therapeutic life of present antibiotics. Drug cycling can be used to maintain heterogeneity of antibiotic agents and to avoid acute selective pressure. It involves the periodic replacement of first-line antibiotics with alternative structural classes. The weakness in this drug cycling, is that it does not provide a long-term solution since resistant strains will be reselected when related antibiotics are used at each cycle. In addition, in large hospitals it may be difficult to decontaminate infected intensive care units while at the same time cycling between different antibiotics.

A related tactic is a combination of compounds that have different modes of action in the hope that resistance will not have time to emerge since carefully studied combinations will quickly eliminate the strain. This combinatorial approach has been applied with success in the treatment of infections in cancer and HIV patients (Holmes et al., 2016).

Non-antibiotic approaches exist for the treatment of bacterial infection and involve the stimulation and recruitment of the innate immune system and the human gut microbiome. Another proposition, that is the work base of this manuscript, is the inhibition of bacterial virulence to stop the infection process without the need for antibiotics. The main advantage of this strategy is that selection for resistance might not occur because the growth of the infecting organism would not be impaired. Nevertheless, it is noteworthy to mention that in an ideal world, the effective use of vaccines against all infectious diseases would reduce drastically the use of antibiotics that will be limited to surgical procedures under strict controls (Davies et al., 2010).

In addition to innovative techniques, effort should be made on improvement of di-

agnosis and prescription practices, perusing studies on withdrawn or underused antimicrobial and encouraging de-novo drug discovery. Areas of particular interest for de-novo drug discovery include molecules capable of attenuating bacterial virulence and disrupting biofilm formation, bacteriophage therapy, eco-biological approaches (Nood et al., 2013) and enhancement of host immune responses. In addition allowing guaranteed access to medicines and ensuring the use of adequate serum drug concentrations in low-income countries through the prosecution of falsified, substandard and generally low quality drugs (Holmes et al., 2016).

Moving forward and reducing demand for antibiotics requires a series of key strategic objectives. Starting with a massive global public awareness campaign, followed by improvement in hygiene standards and the prevention of infection spread. Besides unnecessary use of antimicrobials in agriculture and their dissemination into the environment should be reduced. Improvements in global surveillance of drug resistance and antimicrobial consumption are to be undertaken. Rapid diagnostics to cut unnecessary use of antibiotics need to be promoted as well as the use and development of vaccines and alternatives. Improvements in the numbers, pay and recognition of people working on infectious diseases are to be made. Establishing a global innovation fund for early-stage and non-commercial research as well as better incentives to promote investments in new drugs and their development are to be carried out. All this global coalition for action is to be supported by the G20 and the UN (O'Neill, 2016).

1.10 Chapter conclusion

The most commonly prescribed drugs in human medicine are antibiotics but up to 50% of the prescriptions are considered unnecessary (Holmes et al., 2016). The period after World War II saw a "golden era" of antibiotic discovery with new products reaching the market from 1940s to the 1970s. Since then the rate of discovery has dramatically fallen (O'Neill, 2016). In addition, drug resistant bacteria are omnipresent in the biosphere, but their consequences are aggravated in situations such as civil unrest, famine and natural disasters (Davies et al., 2010).

For the pharmaceutical industry, medicines that are no longer effective lose their value. Hence the need for industry leaders to support a responsible use of medicines in order to prolong their effectiveness. Through research and development the industry can slow down the emergence of resistance. New concepts are needed to promote innovation and cooperation among policy makers, academia and the pharmaceutical industry, to establish new globally available technologies to prevent and treat resistant infections (World Health Organization, 2015a).

The main efforts should be concentrated on understanding how resistance spreads within and between humans and animals through food, water and the environment. More-

over, the ability to rapidly characterize emerging resistance and elucidate the underlying mechanisms should be the main objective. This is crucial for keeping surveillance and diagnostic tool up to date. In addition, effort should be made to endorse the use and development of existing vaccines that can prevent infectious diseases whose treatment would require antimicrobial drugs. In fact, vaccination reduces the prevalence of viral infections which are often inappropriately treated with antibiotics, resulting in secondary infections that require antibiotic treatment (World Health Organization, 2015a).

With these warnings and recommendations in mind, the development of new strategies to fight pathogens should be a priority. The new ideal therapeutic targets should exert weak evolutionary pressure, disarm or weaken the pathogen and be unique to the micro-organism.

One way to do so is by interfering with the iron regulation and its homeostasis within bacteria. The next chapter shows how iron became important for life and at the same time, an exploitable weakness in our battle against infection.

Chapter 2

Iron: a key element in living organisms

The evolutionary path that early life took was not a product of chance, it was constrained by the changing thermodynamic equilibrium of the environment which forced a one way progression (Williams et al., 2003). The majority of organisms require the first row transition metals that include manganese, iron, cobalt, nickel, copper and zinc. The ability of these metals to cycle between oxidation states contribute to their catalytic importance and toxicity (Palmer et al., 2016). The regulation of such important elements resides in several sensing mechanisms able to regulate the import and export of metals. With such processes in place, a competition between organisms gave rise to what is known as the battle for metals between hosts and pathogens. We can make use of this battle and its insights to explore new opportunities for anti-virulence therapeutic potential.

The availability of elements and their oxidation states were extremely strong constraints placed on primitive cells and their energy systems. Chemical equilibrium depended on the solubility, redox potentials and stability of inorganic complexes (Williams et al., 2003). Transition metals played an extremely important role in the evolution of the biosphere. Their bioavailability varied over geological times, driven by the emergence of oxygenic photosynthesis that gave rise to the Great Oxygenation Event (GOE) around 2.45 billion years ago Boyd et al., 2014. Many biochemical pathways require metalloenzymes, proteins containing ions of inorganic elements such as transition metals, forming around one third of all proteins (Dupont et al., 2010).

Many of the essential elements for early life, needed in their reduced state in the cytoplasm, became oxidized by the environment and had to be scavenged then reduced. This is the case of iron (Williams et al., 2003).

2.1 Evolution was chemically constrained

Iron is an abundant element in the universe due to its nuclear stability and its freezing of nuclear kinetics of element formation in evolving stars. Once iron is reached, fusion is halted and the core temperature drops leading to stellar death. On Earth, iron is close in abundance to oxygen though most of it is in the central metallic core of our planet (Williams, 2012). Nevertheless, it is the fourth most abundant element in the earth's crust, following oxygen, silicon and aluminium. At first glance it would seem unlikely that iron can limit biochemical processes, however the bioavailability of iron is dependant of its redox state and has forged the evolution of our biosphere (Falkowski et al., 1998).

2.1.1 Iron and early life

Hydrothermal vents are still today, as they were in the Archean ocean, a major source of dissolved elements in oceans. They are continuously providing high concentrations of precursors compounds with biological importance by circulating fluids through temperature gradients and various reaction surfaces. Less than 500 million years after the formation of the ocean, they provided multiple pathways for abiotic synthesis of chemical compounds later used as building blocks for early life (Baross et al., 1985). The predominant source of Fe^{2+} in deep Archean ocean basins was hydrothermal water circulating through basalt ridges (Boyd et al., 2014).

Fe^{2+} often forms complexes with sulfide near hydrothermal discharges giving rise to an iron-sulphur mineral phase, in particular pyrrhotite shown to catalyse the reduction of N_2 and CO_2, and the production of H_2 under high temperature and pressure (Boyd et al., 2014). The same chemical reactions occur in living organisms with Fe-S cluster containing enzymes.The strong similarity between the reactivity of Fe-S minerals and Fe-S contain-

ing enzymes in addition to the presence of these enzymes in early evolving lineages are key arguments for the "Fe-S World" theory for the origin of life (Boyd et al., 2014). This theory proposes the exergonic formation of pyrite from hydrogen sulfide H_2S and ferrous iron Fe^{2+} as energy source for an autotrophic origin of life (Wächtershäuser, 1988):

$$Fe^{2+} + 2\,H_2S \longrightarrow FeS_2\downarrow + 4\,H^+ + 2\,e^-$$

Four billion years ago, semipermeable membranes of iron sulfide aggregates called botryoids, shown in Figure 2.1, containing highly reduced hydrothermal solution could have been the precursors for life. Fe-S clusters catalysing reactions thereby modifying membrane properties of pyrite FeS_2 spheres 0.1-1 mm across are a good example . Indeed, chronologically the "Fe-S world" preceded the "RNA world" (Joyce, 1989). More recent research, shows that prebiotically plausible iron–sulphur peptide, called the primordial tetrapeptide and composed of four ECG peptides, can experimentally generate a pH gradient across model membranes of late protocells, which are vesicles with NADH as electron donor and hydrogen peroxide as electron acceptor (Bonfio et al., 2018).

Iron-sulphur metalloproteins are ubiquitous in both aerobic and anaerobic Archaea, Bacteria and Eukarya, suggesting that they were integrated into an essential metabolic pathways early in the evolution of life (Boyd et al., 2014). In plants, algae and other plastid-bearing organisms the sulphur utilization factors (Suf) system is the oldest identified Fe-S cluster biogenesis system (Pérard et al., 2018). SufB, a scaffold protein, may represent the most ancient system for the intake of abiological Fe-S compounds from the environment into proteins and may have been present in the last eukaryotic common ancestor (LECA) (Tsaousis et al., 2014; Boyd et al., 2014).

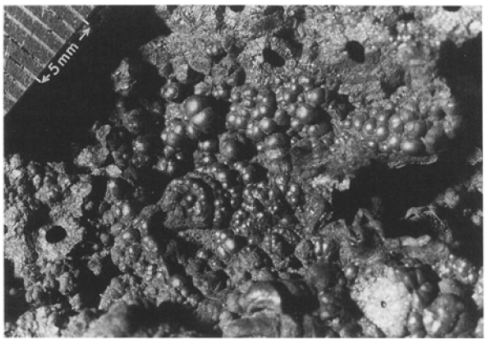


Figure 2.1: Pyrite botryoids, Ireland (Russell et al., 1994).

2.1.2 The impact of early ocean chemistry

In order to understand the impact that iron had on the evolution of life and its emergence, we should understand the chemistry of early oceans. Since we cannot sample seawater from the distant past, researchers use rocks from former sea floor sediments to infer the chemical characteristics of ancient environments. "Banded Iron Formation" or BIFs are deposits of iron minerals older than 1.8 billion years, they show that iron rich oceans in the early earth history gave way to iron scarcity later on (Anbar, 2008).

Given the different behaviour of metals in their oxidized and reduced states, their bioavailability was determined by the oxidation state of oceans. Reduced iron (ferrous

Fe^{2+}) is highly soluble in water whereas oxidized iron (ferric Fe^{3+}) is insoluble (Falkowski et al., 1998; Boyd et al., 2014).

Most studies show that the redox state of the environment evolved through three stages with two oxygenation events. The first, known as the Great Oxygenation Event (GOE) occurring 2.4 billion to 1.8 billion years ago, where mainly coastal surface waters contained dissolved O_2, iron was still abundant in the form of dissolved Fe^{2+} complexes in deep waters. The second oxygenation event took place 0.8 billion to 0.5 billion years ago increasing the concentration of Cu, Zn and Mo while decreasing Fe, Mn and Co concentrations and is known as the Neoproterozoic Oxygenation Event (NOE) (Anbar, 2008; Dupont et al., 2010) (Figure 2.2).

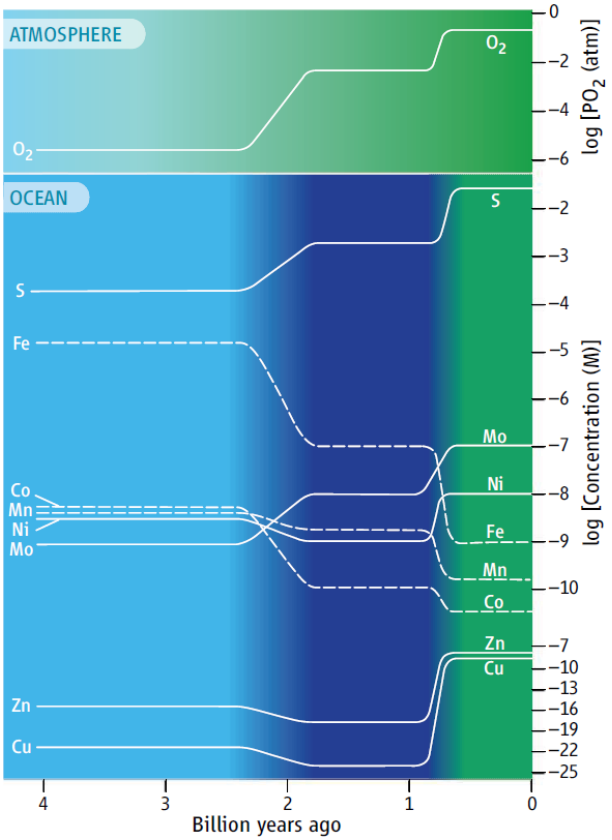


Figure 2.2: Fluctuation in element abundance through time (Anbar, 2008). Colour gradients show transitions from anoxic, S-poor oceans (light blue) to H_2S-rich oceans (dark blue) before complete ocean oxygenation (green).

Before oxygenic photosynthesis, anoxic oceans had very high concentration of iron (Falkowski et al., 1998). Iron minerals are usually deep in Earth's mantle, but olivine ($MgFeSiO_4$) was on the surface of the seabed resulting, with hydrothermal flows, to the initial high availability of iron in the sea. Later on, photosynthesis led to the oxidation of Fe^{2+} into Fe^{3+} and its precipitation as Fe_2O_3 in BIFs (Williams, 2012). In actual oceans iron concentration does not exceed a few nanomolars (Falkowski et al., 1998; Anbar, 2008).

Since thermodynamic equilibrium was important for early life, the early ocean shared many features with early cells (Williams et al., 2003). Experiments show that Fe^{2+} concentration in cytoplasms is between 10^{-6} and 10^{-7}M. Interestingly, Fe^{2+} is still held in the cytoplasm at the same concentration as it was in the sea more than 3 billion years ago (Williams, 2012).

In fact, with the precipitation of iron, its concentration in the sea changed drastically from 10^{-6}M of Fe^{2+} to 10^{-10}M of Fe^{3+} in mineral complexes and 10^{-17}M of free Fe^{3+} leading to a very low iron availability. Overcoming this difficulty required scavenging for iron. Unicellular organisms devised organic molecules called siderophores to chelate and uptake Fe^{3+} specifically (Schalk, 2013). Around 3.5 billion years ago, shortly after the emergence of life, porphyrin complexes of

Fe, Co, Ni and Mg appeared. The complex synthesis of theses rings is thought to have started form cyanide. On the other hand, for multicellular organisms, like plants and animals, the uptake and distribution of iron to the whole organism increased in difficulty (Williams, 2012).

The changes in the environment were a penalty at first, leading to an increased pay-off later on in evolution. The possible evolution pathways were constrained by the need to adapt to the new chemistry. This was done by using extended cellular compartments like vesicles and later on multicellular strata to avoid conflicting chemistry within the organism. This led in term to the need for communication and feedback fulfilled by RNA and DNA (Williams et al., 2003).

Incorporating certain elements became more difficult requiring increasing complexity but the new chemistry increased potential energy sources. At that time, the reducing buffer capacity of the ocean was high, slowing the rate at which oxygenation was imposing important chemical changes, giving life enough time to evolve without going to extinction (Williams et al., 2003). It is incredible how we owe the reducing primitive ocean the ability to write and read these words !

2.2 Emergence of metal binding proteins

Early forms of life lacked metal-binding proteins and intracellular metal concentration control. The emergence of metal-specific structures and the abundant metal in the Archean ocean, lead to the development of metal homeostasis regulation, potentially promoting the diversification of Archaea and Bacteria lineages (Dupont et al., 2010). The control of iron concentration is important for cellular biochemistry, hanges in iron concentration can control key reactions such as Krebs cycle pathways, switching facultative bacteria from anaerobic to aerobic metabolism (Williams, 2012).

Phylogenetic analysis of proteomes from all domains of life show the influence of trace-metals chemistry on the evolution of proteins (Dupont et al., 2010), and that selection for metalloproteins reflects the availability of metals through geological times (Boyd et al., 2014). The high concentration of Fe^{2+} in ancient oceans lead to a good representation of protein folds specifically binding Fe in early evolving lineages. In fact, their presence can be mapped back to the Last Universal Common Ancestor (LUCA). In contrast, protein folds specific for Cu and Zn are well represented in organisms that appeared after the GOE (Dupont et al., 2010; Boyd et al., 2014) when oxidative weathering of continental Cu or Zn sulfide minerals increased the concentration of these metals in oceans within a few hundred kilometres of shorelines (Holland et al., 1986).

For each domain of life, the number of protein fold-families specific to each transition metal reflects the hypothetical chemical environment in which it evolved, providing evi-

dence for the influence of trace metal geochemistry on biological evolution. Studies show that low Zn, Mo and Cu concentrations might have prevented the widespread evolution of the eukaryotic domain until the planet-wide shift in redox state took place (Dupont et al., 2010). The majority of Bacteria and Archeae lack recently evolved protein architectures. Archeae share with Eukarya late-evolving Zn-binding architecture mainly involved in transcription and translation. In contrast, eukaryotes possess late-evolving metal-binding structures, particularly those that bind Ca, Zn and Fe (Dupont et al., 2010).

Enzymes responsible for key steps in molecular biology (replication, transcription and translation) are all dependant of divalent ions for catalytic or structural purposes. Which makes us think about the environment they evolved in and the cations that were present during that time. In modern day organisms, Mg^{2+} is acting as a cofactor for these enzymes. Fe^{2+} and Mg^{2+} have similar coordination chemistry and intracellular concentrations (Dlouhy et al., 2013) they could be interchangeable and play the same roles (Athavale et al., 2012). Experiments show that Fe^{2+} like Mg^{2+} can fold the 23s rRNA and support DNA replication, transcription and translation (Okafor et al., 2017). These results shed light on the role of Fe^{2+} in early life chemistry being abundant at that time. Life emerged using Fe^{2+} in key molecular biology enzymes, then switched to Mg^{2+} as iron precipitated after the GOE.

With an increasingly oxygenated biosphere, a rapid diversification of bacterial lineages and genetic innovation during the Archaean eon is observed. However, one striking contradiction is that phylogenomic patterns show an increase in iron-using genes over time, even with the decreased iron bioavailability. This may reflect the success of organisms that evolved iron-acquisition proteins, rather than replacing their existing set of iron-binding proteins (David et al., 2011).

2.3 Metal transport in bacteria

Despite being primordial for life, a strict equilibrium needs to be maintained to prevent metal toxicity but at the same time fight against its low bioavailability problems. Metal homeostasis control has been optimized during evolution, where organisms adapted to this iron regulation problem through several ways including solubilization of extracellular iron, reduction, chelation, internalization via specific transporters and the use of ferritin molecules as iron stock (McHugh et al., 2003). The transcription of genes coding for metal transporters is under the regulation of metal sensors (Waldron et al., 2009a).

The intracellular accumulation of metals is dictated by the balance between importers and exporters. Organisms, such as yeast, can increase iron bioavailability and facilitate its uptake by acidifying their environment. Ferric iron (Fe^{3+}) solubility shifts from 10^{-18}M at pH 7 to 10^{-3}M at pH 2 (Dlouhy et al., 2013).

In Gram-negative bacteria metals diffuse through the external membrane through porins such as ompF. The cytoplasmic membrane harbours metal transporters such as ATP-binding cassette (ABC), resistance and nodulation proteins (RND) and cation diffusion facilitator (CDF) proteins amongst others. ABC-type ATPases import manganese, zinc, nickel or iron across the cytoplasmic membranes of bacteria and archaea. CDF proteins get rid of the excess metal from the cytoplasm. RND proteins are involved in the externalisation of cobalt, zinc, nickel or copper, across the membranes of Gram-negative bacteria. In addition energy-coupled importers are used to acquire large complexes that can not be internalized through porines such as iron bound to siderophores (Waldron et al., 2009a). Figure 2.3 shows metal transporters in *E. coli*.

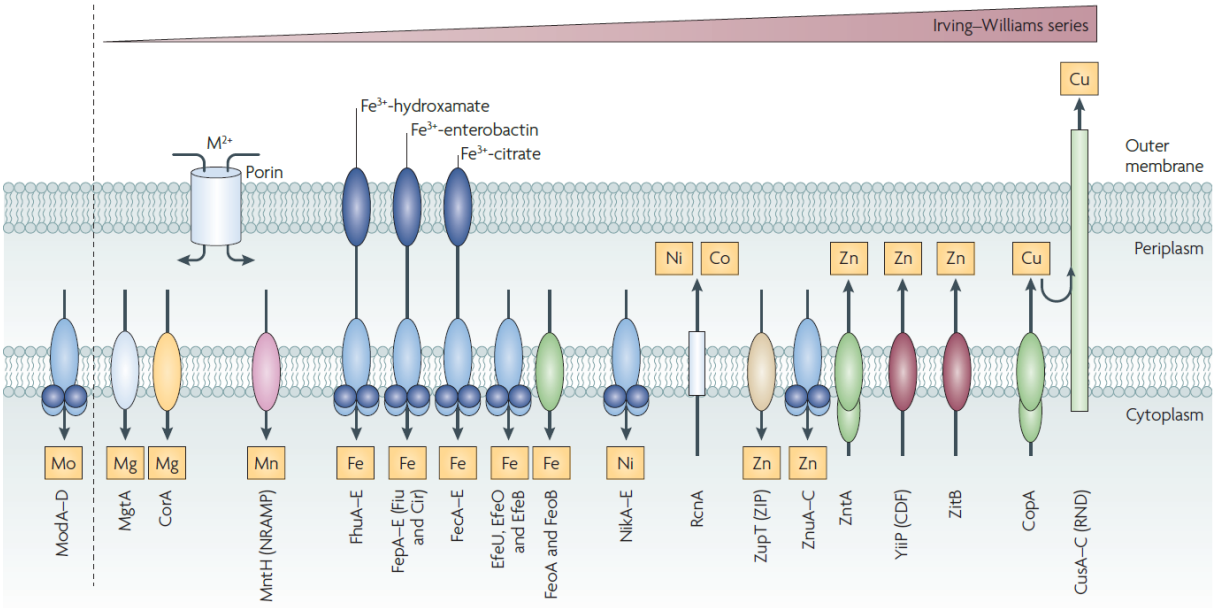


Figure 2.3: Metal transporters in *Escherichia coli* (Waldron et al., 2009a) ModA, molybdate-binding protein; MgtA, magnesium-transporting ATPase; CorA, cobalt/magnesium transport protein; NRAMP, natural resistance associated with macrophage protein; FhuA-E, ferrichrome outer membrane transporter; Fep, ferric enterobactin-transporting protein; Fiu, ferric iron uptake; Cir, colicin I receptor; FecA-E, Fe^{3+} dicitrate transport protein; EfeU, elemental ferrous iron uptake; Feo, ferrous iron-uptake system; RcnA nickel/cobalt efflux system; ZupT, zinc-uptake transporter; Znu, zinc uptake; ZntA zinc-transporting P-type ATPase; CDF, cation diffusion facilitator; ZitB, zinc transporter; CopA, Copper-exporting P-type ATPase; CusA–D, copper-sensitive operon.

Microorganims can also synthesise and secrete siderophores. These low molecular weight compounds, with more than 500 chemical structure, form high affinity complexes with ferric iron, around 10^{-33}M, making it available for uptake by transporters on the cell surface. In *P. aeruginosa*, iron acquisition difficulty can be overcomed by the use of its two main siderophores: pyoverdine and pyochelin (Schalk et al., 2013; Gasser et al., 2015). *S. cerevisiae* has an extensive system of transporters to uptake iron and Fe^{3+}-siderophores complexes from the environment (Dlouhy et al., 2013). *E. coli* has six siderophore receptors (Cir, FecA, FepA, FhuA, FhuE, Fiu) in addition to its anaerobic mechanism of iron uptake via FeoB (McHugh et al., 2003).

2.4 The need for metal sensors

Almost half of all enzymes have metal related functions. In eukaryotes, a study of 1371 enzymes with known structure showed that almost 50% required metals, 41% of them being in catalytic centers. Iron in particular, constitutes a significant portion of the prokaryote and eukaryote metallome (Waldron et al., 2009b).

Metalloproteins binding iron are widely used in oxygen binding, energy metabolism, DNA repair and antioxidant functions. Since iron is needed for survival, organisms optimized pathways to uptake and store sufficient iron concentrations (Dlouhy et al., 2013).

However, iron is toxic at high concentration catalysing the formation of reactive oxygen species (ROS), namely the hydroxyl radical $HO^{\bullet}$, which damages proteins, membranes and DNA. Toxicity of transition metals comes from Fenton chemistry where Fe^{2+} reacts with hydrogen peroxide to form Fe^{3+}, a hydroxyl radical, and a hydroxide (Palmer et al., 2016):

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + HO^- + HO^{\bullet}$$

A delicate control and coordination of the equilibrium between iron deficiency and iron overload ensures optimal iron trafficking (Dlouhy et al., 2013). Metal sensors are key to this metal homeostasis.

In addition to homeostasis, the challenge in biology of metals resides in correctly populating each metalloprotein (Waldron et al., 2009a). The difference of affinity between metals and organic molecules reflects the stability of such complexes and is described by the Irving–Williams series (Irving et al., 1948) with copper and zinc forming the tightest complexes :

$$Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$$

Since metal sensors exchange their metals with the cytosolic pool, their affinity becomes the threshold for homeostasis, thus altering the production of proteins to acquire, expel or sequester metal. The affinity of sensors to their metals increases as the sensors detect metals farther up the Irving–Williams series (Waldron et al., 2009b). Figure 2.4 shows metal sensors in *E. coli*.

In the cytoplasm, metal availability is regulated through regulation of metal importers, exporters and cellular stocks. It is very important to regulate the supply of competitive metals such as copper.

Theoretically, proteins taking their metals from the cytosolic pool, where all the others metals are present, would all bind copper because of the Irving-Williams series. A system must be used to keep metals at the top of the stability series away from the binding sites of metals that happen to be less stable. In fact, cells restrict the bioavailability of metal atoms in the cytoplasm in a way that makes proteins compete for different metals,

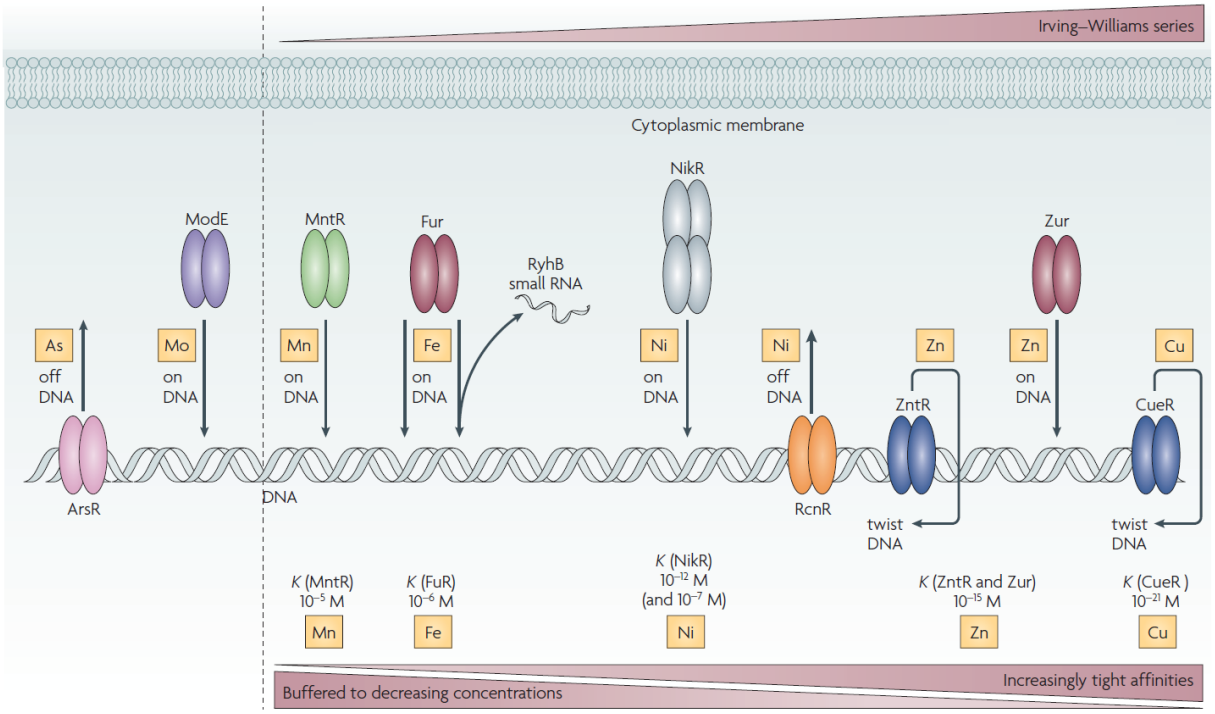


Figure 2.4: Metal sensors in *Escherichia coli*, adapted from Waldron et al., 2009a. Arrows indicate the mode of action of sensors. Metal derepression : arrow away from DNA; Metal corepression : arrow towards DNA; Metal-dependent activation by DNA under-winding : arrow around DNA. Colors indicate sensor families: arsenic resistance transcriptional regulator ArsR (pink), molybdenum dependant transcriptional regulator ModE (purple), manganese dependant transcription regulator MntR (green), ferric uptake regulator family (red) showing ferric uptake regulator Fur and zinc uptake regulator Zur, nickel-responsive regulator NikR (grey), nickel efflux system RcnA (orange), MerR family (blue) showing zinc sensor ZntR and copper sensor CueR. K represents the affinity for the sensed metal.

instead of metal atoms competing with each other for binding sites (Waldron et al., 2009a; Waldron et al., 2009b).

It is important to note that metal sensing in multicellular eukaryotes enables them to perform complex tasks. Their cells can respond to both their own metal status and the metal status of the whole organism. In the case of iron, these responses take place by adjusting messenger-RNA stability through iron-responsive proteins IRP1 and IRP2 or by adjusting intracellular protein trafficking and degradation. Another example is hepcidin, when secreted from the liver after an increase of iron concentration in the circulation, it inhibits dietary iron absorption by inducing internalization and degradation of ferroportin (Waldron et al., 2009b).

2.5 Iron pools in organisms

Iron in living organisms is predominantly bound to three types of metallorproteins. First, heme-containing proteins are used in oxygen transport, energy metabolism and transcriptional regulation. Heme consists of iron bound to a porphyrin ring, Figure 2.5,

and is found in proteins such as hemoglobin and cytochrome P450.

Second, iron-sulphur containing proteins are formed by iron (Fe^{2+} or Fe^{3+}) and sulfide S^{2-} that can be arranged in different ways the most common being [2Fe-2S] and [4Fe-4S] clusters, see Figure 2.6. Two systems control the biogenesis of these clusters in eukaryotes: the mitochondrial iron-sulphur cluster (ISC) and the cytosolic Fe-S protein assembly (CIA) machinery. Archaea and bacteria compared to eukaryotes have more Fe-S proteins and fewer haem proteins (Waldron et al., 2009b).

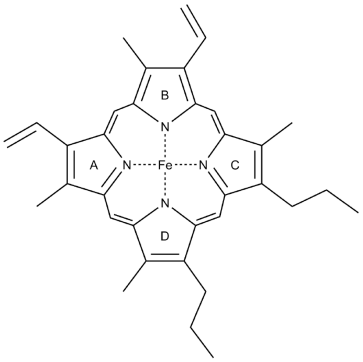


Figure 2.5: Structure of protoporphyrin IX with ferrous iron (Dlouhy et al., 2013).

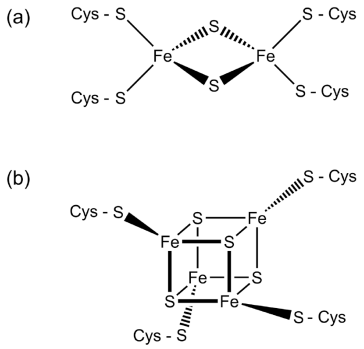


Figure 2.6: Common Fe-S clusters (Dlouhy et al., 2013): (a) [2Fe-2S] and (b) [4Fe-4S].

Third, non-heme proteins where iron bounds directly to the protein with different possible amino acid ligands enabling the catalysis of a wide range of reactions (Dlouhy et al., 2013).

Cytoplasmic iron concentration vary with species and cell type. In *S. cerevisiae* iron concentration range form 250 µM to 600 µM in the cytoplasm and peaks at 800 µM in mitochondria and vacuoles where excess iron is stored. In higher eukaryotes, excess iron is stored in cytosolic and mitochondrial ferritin. In humans, erythroid cells have 400 µM of intracellular iron whereas in rat hepatocytes this concentration is close to 1 mM. The majority of this iron is present as Fe-S clusters in proteins and heme centers as cofactors or stored in ferritin, vacuoles and lysosomes. The remaining iron in the cell is known as the labile iron pool (LIP) and constitutes 0.2-3% of cellular iron. In mitochondria, LIP constitutes 0.4% of mitochondrial iron and is measured between 1 and 16 µM (Dlouhy et al., 2013).

2.6 The battle for iron

In response to bacterial infections, the host iron homeostasis undergores important changes that include the poduction of iron sequestering proteins, in a process termed nutritional immunity, in addition to local increase of toxic metal levels in specific cases. In response to these changes, bacteria evolved systems to overthrow iron sequestration

and toxicity. This coevolution between hosts and bacteria was given the name of "battle for iron".

In 1973 the evolutionary biologist Leigh Van Valen proposed the Red Queen hypothesis (Valen, 1974) as a metaphor for an evolutionary arms race that is still taking place : with each development in host defense, pathogens evolve a new offense.

The aerobic environment and neutral pH of serum ensures that extracellular iron is insoluble and hence difficult to access by invading pathogens (Skaar, 2010). To get their required metals, bacteria evolved three main metal acquisition systems: elemental metal import, metal capture by siderophores, and metal acquisition from host proteins termed metal piracy.

Two-thirds of iron in humans is in erythrocytes, bound to hemoglobin. In the serum, the physiological pH and oxic conditions render free Fe^{3+} insoluble, it is mainly bound to transferrin (Palmer et al., 2016). Transferrin can bind two Fe^{3+} ions with high affinity. It provides iron for several human cell types through the cell surface receptor TfR1 that internalizes the complex by endocytosis (Dlouhy et al., 2013). Moreover, Fe^{3+} is bound to lactoferrin in milk, saliva, tears and mucus. Inside cells, Fe^{3+} is bound by ferritin (Palmer et al., 2016), see Figure 2.7.

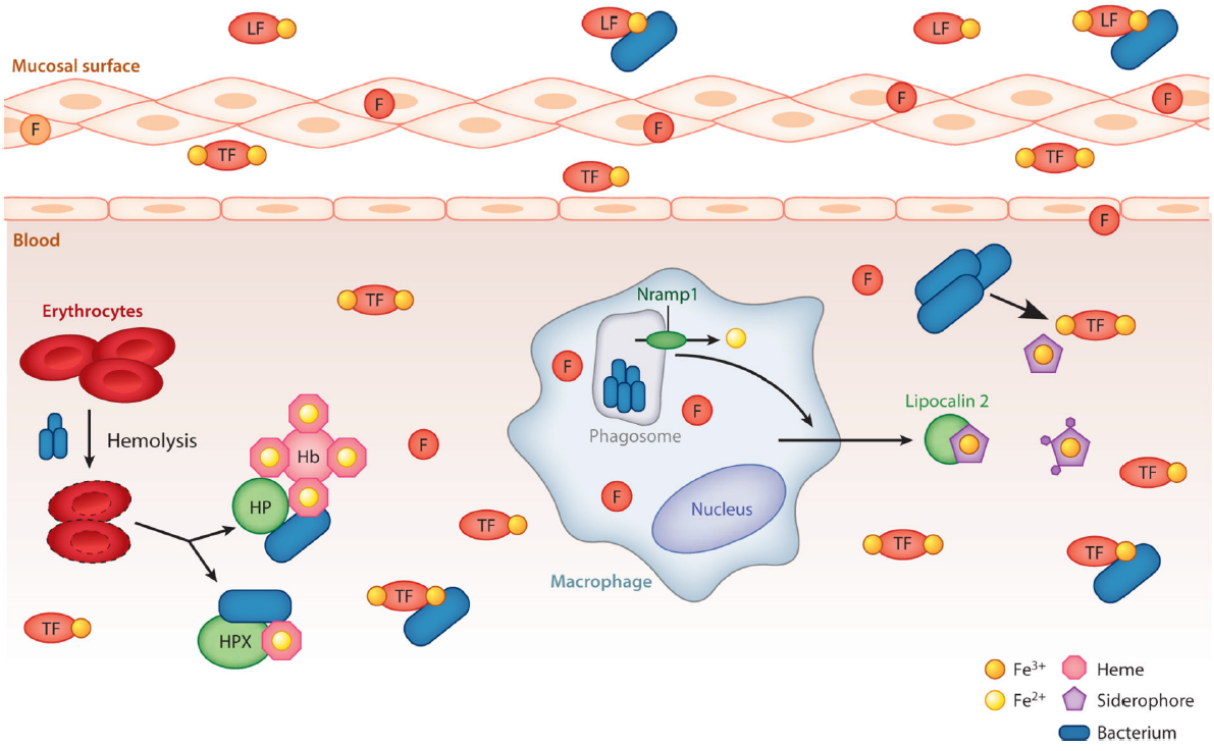


Figure 2.7: Iron at the host-pathogen interface (Palmer et al., 2016). Lactoferrin (LF) at the mucosal surface, transferrin (TF) in blood and tissue, and ferritin (F) in the cellular cytoplasm. Hemoprotein complexes liberated after hemolyse by pathogens include : hemoglobin (Hb), hemoglobin-haptoglobin (HP), and hemopexin (HPX). Nramp1-mediated iron efflux helps in the iron starvation of intracellular pathogens.

Since the majority of iron in the human body is found in erythrocytes, many pathogens evolved to access heme iron. They lyse erythrocytes, bind hemoglobin and extract heme to free iron. In addition, other bacteria, such as *P. aeruginosa*, release hemophores with 10^{-11}M affinity to heme that is brought back to the hemophore receptor HasR on the bacterial membrane, see Figure 2.8.

Most commonly, bacteria acquire iron from heme, but some pathogens can obtain iron directly from nutritional immunity proteins through iron piracy. For example, transferrin-binding proteins TbpA/TbpB bind transferrin and extract its Fe^{3+}, other pathogens express lactoferrin receptors (Palmer et al., 2016). It was shown that transferrin, which binds iron with a high affinity (Skaar, 2010), was and still is engaged in an evolutionary tug of war with TbpA. A High frequency of point mutations in certain domains of transferrin prevent TbpA binding, providing a counter attack to bacterial iron piracy among great apes (Barber et al., 2014).

To overcome iron limitations, bacteria use siderophores to scavenge extracellular iron. Siderophores are of low molecular weight and have a high affinity for Fe^{3+}. This affinity can be sufficiently high to steal iron from host proteins. Enterobactin, a siderophore secreted by *Enterobacteriaceae*, binds Fe^{3+} with an estimated Kd equal to 10^{-50}M out-competing transferrin that has a Kd of 10^{-22}M for Fe^{3+} (Palmer et al., 2016).

In Gram-negative bacteria TonB-dependent system and ATP-binding cassette (ABC) transporters are used to transport and internalize Fe^{3+}-siderophore complexes. In Gram-positive bacteria import is also performed by ABC transporters in addition to lipoprotein substrate-binding proteins (Palmer et al., 2016).

Once inside bacteria, iron is released from siderophores by one of two mechanisms. For siderophores with high Fe^{3+} binding affinity, like enterobactin or bacillibactin, they are hydrolyzed by specific esterases then Fe^{3+} is reduced to Fe^{2+}. For other siderophores with a relatively lower affinity to iron, ferric reductases reduce Fe^{3+} to Fe^{2+} that will be released by competitive sequestration with other intracellular proteins (Palmer et al., 2016).

To fight against siderophores, vertebrate hosts produce lipocalin, see Figure 2.8, also known as siderocalin, to sequester Fe^{3+}-bound siderophores during the innate immune response to infection. In response, bacteria evolved stealth siderophores, not recognised by lipocalin such as salmochelin and petrobactin (Palmer et al., 2016). Studies also show that hosts can secrete siderophore-like compounds mainly catechol and catechol-like structures. In addition to competing with pathogen siderophores, the host siderophore-like compounds are used for iron trafficking (Dlouhy et al., 2013).

Interestingly, instead of entering in long and fastidious battle for iron, some species chose different strategies. The causative agent of Lyme disease *Borrelia burgdorferi* solved the problem by substituting iron with manganese in its iron-requiring enzymes (Skaar, 2010).

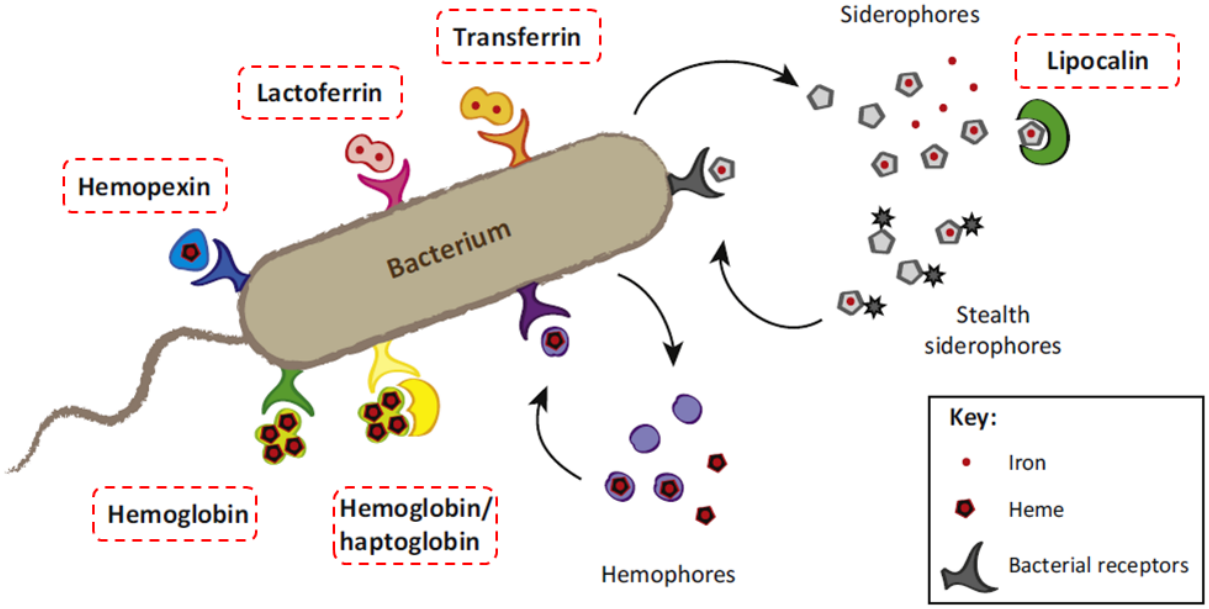


Figure 2.8: Interplay between nutritional immunity and iron piracy (Barber et al., 2015). Dashed boxes indicate host nutritional immunity proteins to which pathogens evolved receptors.

In the case of opportunistic pathogens that have not evolved metal acquisition systems, host metal status and increased metal availability can affect susceptibility to infections. In humans, β-thalassemia, caused by hereditary mutations in the hemoglobin β-chain, triggers iron overload and is associated with increased rates of infection. An unbalanced metal homeostasis can also develop from environmental exposure. Factory workers and miners exposed to high levels of iron dust have increased rates of respiratory tract infection (Palmer et al., 2016).

Evolution led vertebrates, by the means of their nutritional immunity, to be devoid of easily available free iron. This ensures that upon infection, bacteria will encounter an iron starvation period that will drastically limit their development. Evolving with such constraints, bacterial pathogens sense iron depletion as a sign of vertebrate tissue and modify their behaviour through virulence to trigger iron piracy. This sensing typically involves transcription control mediated by iron sensors in particular the Ferric Uptake Regulator that is shown to contribute in virulence of animal and plant pathogens (Abed et al., 2007; Skaar, 2010; Pérard et al., 2018).

2.7 Chapter conclusion

The bioavailability of metals strongly influenced early biological evolution and the metabolic strategies that sustained life during that time (Boyd et al., 2014). Evidence shows that diversification of eukaryotes coincided with rising redox potential of ancient oceans (Anbar, 2008).

Transition metals are necessary for all forms of life. Their unique inorganic and redox properties made them essential cofactors for enzymes. For pathogenic bacteria, a vertebrate host is a rich source of metals so it evolved diverse metal acquisition strategies (Palmer et al., 2016).

Research on transition metals in biology at the host-pathogen interface presents opportunities for therapeutic potential since it is a cross-disciplinary work combining biochemistry, microbiology, evolutionary biology, human genetics, and environmental sciences. Other areas of interest are the study of evolution of bacterial pathogenesis, bacterial social behaviour, and the effect of host metal on bacterial virulence (Palmer et al., 2016).

An important defence against bacterial infection is the withholding of nutrients to hinder bacterial growth in a process called nutritional immunity. The most significant form of nutritional immunity is the sequestration of nutrient iron. Most pathogens fight against iron starvation through high-affinity iron uptake mechanisms regulated by iron sensors essential for bacterial virulence (Skaar, 2010). The next chapter will introduce the Ferric Uptake Regulator protein, an interesting anti-virulence target for which our group developed inhibitors.

Chapter 3

The Ferric Uptake Regulator

As described in the previous two chapters, the scientific community needs to tackle the problem of microbial resistance to antibiotics urgently. Given the importance of iron to all forms of life and the battle for iron that takes place between hosts and pathogens, every organism evolved to optimize the uptake iron in its environment. In the case of bacteria all transcriptional modifications are dependant of a metal sensor : the Ferric Uptake Regulator protein (Fur), making it an ideal target for drug development. This chapter descries the Fur protein and previous research carried out in our laboratory that identified small peptides capable of inhibiting Fur.

3.1 Fur discovery

In 1956 Garibaldi shows that iron down regulates the expression of iron acquisition systems (Garibaldi et al., 1956). In 1978 Ernst isolates a *Salmonella typhimurium* mutants constitutively expressing siderophores, suggesting a gene mutation involved in iron acquisition regulation. The gene was designated *fur* for the ferric uptake regulator (Ernst et al., 1978). Later on, Hantke describes an *E. coli* mutant expressing constitutively iron regulation functions, his hypothesis was that of a mutation in the *fur* gene (Hantke, 1981). A few years later, the gene in question was cloned (Hantke, 1984) and sequenced (Schäffer et al., 1985). *fur* is located at min 15.5 of the *E. coli* genetic map (Hantke, 1984).

Fur homologues were identified in pathogens such as *Yersinia pestis* (Staggs et al., 1991), *Pseudomonas aeruginosa* (Prince et al., 1993), *Helicobacter pylori* (Bereswill et al., 1998), *Staphylococcus aureus* (Kuroda et al., 2001), *Bacillus subtilis* (Bsat et al., 1998) and many others species. The majority of these homologues can rescue an *E. coli* Fur deficient strain, a sequence alignment is presented in Figure 3.1. Fur is present in all Gram-negative and some Gram-positive bacteria, but not in eukaryotes.

3.1.1 Fur-like proteins

The Fur superfamilly has five proteins other than Fur : zinc uptake regulator (Zur), nickel uptake regulator (Nur), manganese uptake regulator (Mur), peroxyde stress regulator (PerR) and the heme availability regulator (Irr). In *H. pylori*, members of the Fur metalloregulator family like Zur, Nur, PerR are absent and only Fur is present. Thus it is considered as a global regulator playing a major role in adaptation to different stress conditions (Pich et al., 2012).

Other metal regulators are also expressed in low iron conditions. The diphtheria toxin repressor (DtxR) constitutes another family of iron regulators, its members regulates manganese transport in organisms with Fur-dependant iron regulation. DtxR-like proteins named iron dependant regulator (IdeR) exist in *streptomycetes* and *mycobacteria*. In Gram-positive bacteria with high GC content they regulate genes similar to those regulated by Fur in many Gram-negative bacteria and Gram-positive bacteria with low GC content (Hantke, 2001).

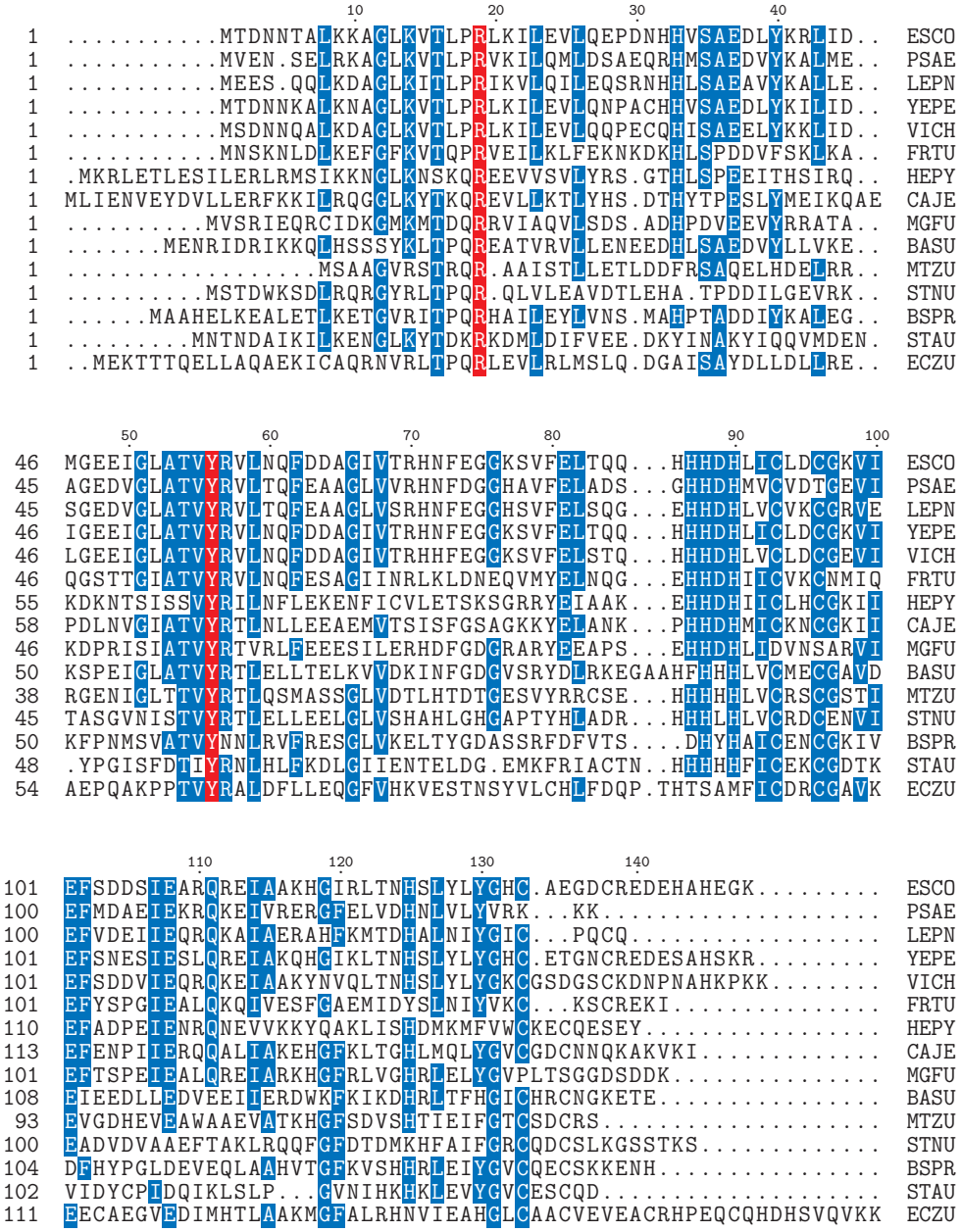


Figure 3.1: Sequence alignment of Fur and Fur like proteinsof various species. Red residues are found amongst all the presented species, blue residues are highy conserved. From top to bottom: Fur from *Escherichia coli* (ESCO), *Pseudomonas aeruginosa* (PSAE), *Legionella pneumophila* (LEPN), *Yersinia pestis* (YEPE), *Vibrio cholerae* (VICH), *Francisella tularensis* (FRTU), *Helicobacter pylori* (HEPY), *Campylobacter jejuni* (CAJE), *Magnetospirillum gryphiswaldense* (MGFU), *Bacillus subtilis* (BASU), the Zinc uptake regulator from *Mycobacterium tuberculosis* (MTZU), the Nickel uptake regulator from *Streptomyces sp.* (STNU), the Peroxide operon regulator from *Bacillus subtilis* (BSPR), Fur from *Staphylococcus aureus* (STAU) and the Zinc uptake regulator from *Escherichia coli* (ECZU).

3.2 Fur is a global regulator

Fur is considered as a key case study of metalloregulatory proteins that convert metal ion concentration into metabolic changes (Lorenzo et al., 1988). Because of its role as a global regulator, the literature sometimes refers to it as the *fur* modulon (Andrews et al., 2003). The *fur* gene was first identified as a negative regulator of genes involved in iron control (Hantke, 1984; Bagg et al., 1987) but later on it was discovered that Fur is involved in other cellular processes. In his work, Hantke showed that Fur regulates 100 genes in *E. coli*, from which 60 are known to be involved in iron acquisition. The others are associated with respiration (*cyoA*), mobility (*FlbB*), glycolysis (*gmpA*), protection against oxydative stress (*sodA*) and virulence (*hly*) (Hantke, 2001).

In addition, *fur* inactivation in *E. coli* leads to inability of growth on media with non fermentable carbon sources like succinate. Indicating that *fur*, in addition to regulating iron uptake, influences succinate uptake and metabolism probably because of a defect in the respiratory chain iron (Hantke, 1987). In addition, these mutants are sensitive towards iron induced oxidizing stress due to an increase in free cytosolic iron concentration, demonstrating once again the central role played by Fur (Touati et al., 1995). Further studies showed that *fur* mutation in *Neisseria meningitidis* induces heat shock response, indicating the existence of an intertwined circuit between heat shock response and Fur regulation (Delany et al., 2006).

The 100 genes identified to be regulated by Fur are either repressed or induced (McHugh et al., 2003). In *Neisseria meningitidis* the transcription of 83 genes was shown to be controlled by Fur (44 positively regulated and 39 negatively regulated) (Delany et al., 2006). Later studies showed that Fur regulates 196 genes in *E. coli* (Abed et al., 2007). With this large span of Fur targets, all organisms with Fur mediated gene regulation use several operating modes to achieve the desired gene expression.

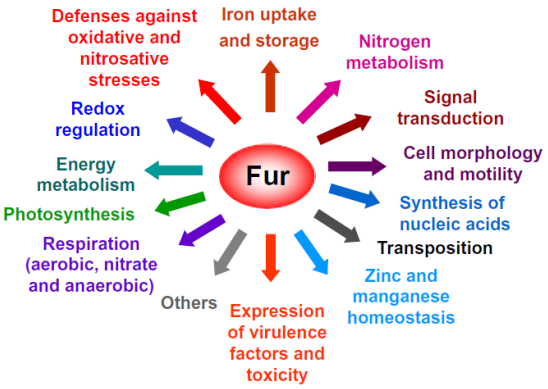


Figure 3.2: Schematic representation of the Fur modulon (Fillat, 2014).

3.2.1 Auto-regulation of *fur*

The expression of *fur* gene is controlled by Fur itself, like other transcription factors or regulators. A second control mechanism takes place through the CAMP-CAP system, establishing a correlation between iron absorption and metabolic status of bacterial cells (De Lorenzo et al., 1988). Later studies showed that in *E. coli*, *fur* is dually regulated by the generation of oxygen radicals and iron (Hantke, 2001).

Even if some species have proteins that regulate *fur* expression (CAP in *E. coli* (De Lorenzo et al., 1988), RpoS in *Vibrio vulnificus* (Lee et al., 2003), NikR in *Helicobacter pylori* (Delany et al., 2005), PerR in *Bacillus subtilis* (Fuangthong et al., 2002), the most widely used regulation mechanism is regulation by Fur itself. In iron deficiency conditions Fur represses its own expression, this was shown in *E. coli* (De Lorenzo et al., 1988; Lorenzo et al., 1988), *Helicobacter pylori* (Delany et al., 2003) and *Edwardsiella tarda* (Wang et al., 2008).

In normal conditions, Fur is constitutively expressed. Its intracellular abundance has been measured in *Vibrio cholerae* and *E. coli* and is estimated at 2500–5000 copies per cell (Watnick et al., 1997).

3.2.2 Negative regulation

Repression by Fur is done through its binding to a consensus sequence located near the Pribnow box, an essential sequence for transcription, of promoters (McHugh et al., 2003). In the presence of sufficient cytosolic iron concentration, Fur is activated and binds its specific DNA sequence. By doing so, it inhibits the binding of the RNA polymerase to the DNA sequence thus negatively regulating gene expression.

Negatively regulated genes are highly expressed in *fur* mutants. *Salmonella typhimurium* and *Escherichia coli fur* mutants overproduce siderophores and their outer membranes receptors as if they are growing in iron limited conditions. This is true since under iron-rich conditions, siderphore synthesis and siderophore transport are repressed (Hantke, 2001). Moreover, the regulation of iron acquisition and storage is mediated by Fur using Fe^{2+} as cofactor (McHugh et al., 2003). *E. coli* has 7 iron-acquisition systems controlled by 35 iron-repressed genes and mediated by Fur (McHugh et al., 2003). In *H. pylori*, Fur represses the expression of a wide array of genes like *fecA* an outer membrane receptor protein in the Fe^{3+}dicitrate transport system, *exbB2* important for maintaining outer membrane integrity, *frpB1* an haemoglobin receptor, *nikR* the nickel-responsive regulator and *feoB* a Fe^{2+} uptake system (Pich et al., 2012).

In some cases, Fur can repress gene expression without being activated by iron. Apo-Fur repression in *Helicobacter pylori* affects 16 genes like *pfr* encoding ferritin and *sodB* encoding the iron superoxide dismutase (Ernst et al., 2005). Interestingly, similar regu-

lation was mentioned for Fur from *Campylobacter jejuni* Holmes et al., 2005 and *Desulfovibrio vulgaris* (Bender et al., 2007). No transcriptional regulation by apo-Fur has been described in *E. coli*.

3.2.3 Direct positive regulation

Despite being often described as a transcription repressor, Fur can play the role of an activator (McHugh et al., 2003). In *E. coli*, a metal activated Fur positively regulates the *sodB* gene Massé et al., 2002 and in *P. aeruginosa* it is the case of *BfrB* coding an iron storage protein (Wilderman et al., 2004).

Proteomic analysis on *Vibrio cholerae fur* mutants showed that the expression of some proteins required the presence of holo-Fur (Litwin et al., 1994). This regulation system was also observed for specific genes in other organisms: *nifS* in *Helicobacter pylori* that encode Fe-S cluster synthesis (Alamuri et al., 2006). In addition to genes involved in iron homeostasis and oxidative stress resistance like *bfrAB* and *sodB* in *Neisseria meningitidis* (Delany et al., 2006) and *Yersinia pestis* (Gao et al., 2008).

Analysis of promoter sequence of positively regulated genes show that Fur boxes are located further upstream than in negatively regulated genes. In *Neisseria gonorrhoeae*, genes exhibiting Fur boxes more than 70 nucleotides upstream from transcription start site are up-regulated by Fur and iron. They usually encode complexes involved in anaerobic and aerobic respiration (Pich et al., 2012).

3.2.4 Indirect positive regulation

Fur fonctions mainly as a repressor, but can also activate transcription. Fur regulates some genes indirectly by modulating the expression of a small RNA molecule designated RyhB (Waldron et al., 2009a).

In 2002, Massé and Gottesman showed that RyhB, a small RNA (sRNA) can negatively regulate the expression of iron-storage and iron-using proteins in *E. coli* when iron is scarce (Massé et al., 2002). In fact, the transcription of *ryhB* is repressed by Fur (Wilderman et al., 2004). When expressed under low iron conditions, RyhB pairs with specific mRNA sequences and trigger their degradation by the RNA degradosome (Prévost et al., 2007). A model of Fur and RyhB iron regulation is shown in Figure 3.3.

RyhB regulates around 18 operons, the *isc* operon being one of RyhB's direct targets, indicating that along with Fur, RyhB plays an important role in iron metabolism (Massé et al., 2005). In addition, RyhB RNA levels are inversely correlated with messenger RNA of the *sdhCDAB* opreon, encoding succinate dehydrogenase, *acnA* and *fumA*, involved in the tricarboxylic acid cycle, *ftnA* and *bfr*, two ferritin genes, *entA-F* enterobactin biosynthesis genes, and *sodB*, a gene for superoxide dismutase. All these genes were shown

to be positively regulated by Fur by an unknown mechanism. The discovery of RyhB and its negative regulation by Fur helped uncover the mechanism of positive regulation of gene expression by Fur through the repression of another repressor (Massé et al., 2002; McHugh et al., 2003; Wilderman et al., 2004).

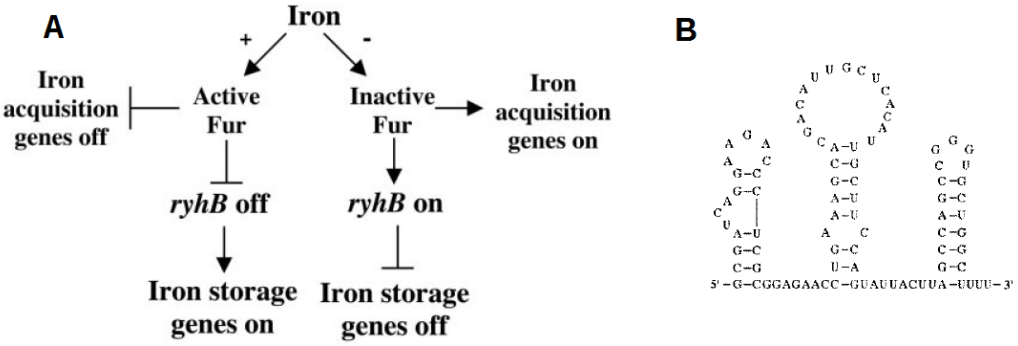


Figure 3.3: A: Model of Fur and RyhB iron regulation. B: Predicted secondary structure of *ryhB* (Massé et al., 2002).

Since RyhB has been identified in *E. coli*, it has been shown that its sequence, including its promoter and operator, are well conserved amongst several bacterial species like *Salmonella*, *Klebsiella*, *Shigella* and *Photorhabdus luminescens* (Wilderman et al., 2004). In some cases, RyhB can also positively regulate gene expression, as in the case of *shiA* involved in siderophore biosynthesis. This positive regulation is achieved by increasing the stability of *shiA* mRNA (Prévost et al., 2007).

The advantage of such two-tiered regulation system in which Fur regulates iron uptake and RyhB regulates iron-storage and iron-using proteins, is that the use of sRNA is one of the most economical and fast ways to globally repress genes. In the case of a sudden decrease in iron availability the cell can rapidly stop the synthesis of target proteins helping reorient free iron to more crucial functions. In addition, sRNA use may be the result of the need for a simple yet effective mechanistic way to convert the negative Fur regulator into an indirectly positive regulator (Massé et al., 2002).

However, in *P. aeruginosa* no RyhB sequence homologs were found. Yet two sRNA sequences were identified by bioinformatics as functional homologs of Ryhb. They encode iron regulated transcripts, and are preceded by a Fur box and their expression is induced under iron starvation. These sRNAs are PrrF1 and PrrF2, for Pseudomonas regulatory RNA involving iron (Fe), sharing 95% sequence identity with each other. In *P. aeruginosa* genes highly expressed under iron-rich conditions include *sdh*, encoding succinate dehydrogenase, showing a similar pattern to genes regulated by RyhB in *E. coli* (Wilderman et al., 2004).

3.3 The Fur box

Over the last twenty years, several models for the binding of Fur on DNA were proposed. A consensus was found from promoters of different genes on a 19 bp palindromic, AT rich, sequence known at first as the iron box, then as Fur box : **GATAATGATAATCATTATC** (Lorenzo et al., 1987; Escolar et al., 1998; Hantke, 2001). The different binding models include binding on the palindromic motif with each Fur subunit interacting with 9 bp with one spacer in the middle, w in Figure 3.4A and B (Rohs et al., 2010), or a binding on the hexameric GATAAT motif as shown in Figure 3.4C (Escolar et al., 1998; Escolar et al., 1999; Escolar et al., 2000). Later work, that was confirmed by crystal structures, showed that Fur actually binds DNA in a 7-1-7 or 6-1-6 motif (Figure 3.4D and E) with two dimers binding one iron box from opposite sides (Baichoo et al., 2002; Lavrrar et al., 2002; Pich et al., 2012).

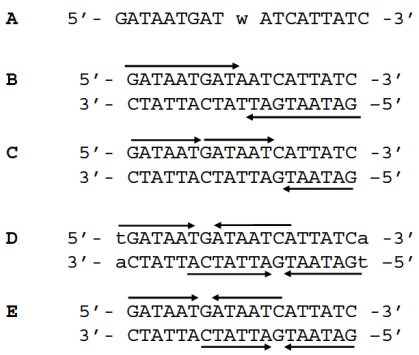


Figure 3.4: Consensus Fur box sequence (A) on which are represented the proposed Fur interaction models. Arrows indicate one Fur subunit. A: w corresponds to A or T nucleotides. B: binding on the palindromic motif. C: binding on hexameric motifs. D: binding to the 7-1-7 motif. E: binding to the 6-1-6 motif (Vitale, 2009).

3.4 Structural description of Fur proteins

In 1995, Stojiljkovitch and Hantke showed that in *E. coli*, translation of *fur* gives a 16795 Da protein with 148 aminoacids. They also identified two functional domains: the C-terminal part (74-147) involved in dimerization and the N-terminal part (1-81) involved in DNA binding (Stojiljkovic et al., 1995). Later, in order to find what oligomeric state of Fur is active in *E. coli*, a wild type strain was complemented with an inactive mutant of *fur* and a reporter gene (*lacZ*) placed under Fur regulation. Results show that the reporter gene was derepressed indicating that, *in vivo*, Fur acts as an oligomer (at least as a dimer) in *E. coli* (Braun et al., 1990).

Over the last twenty year, sufficient structural information on Fur proteins was gathered and showed that they all share a common fold with an average length of 120 amino acids. Their N-terminus composes the DNA binding domain with a winged-helix motif and

their C-terminus constitutes the dimerization domain and harbours metal binding sites. Members of this protein family, and superfamily, have a histidine-rich `HHHXHX2CX2C` motif in their dimerization domain (Fillat, 2014). Figure 3.5 shows the structural classification of Fur proteins in the SCOP2 database.

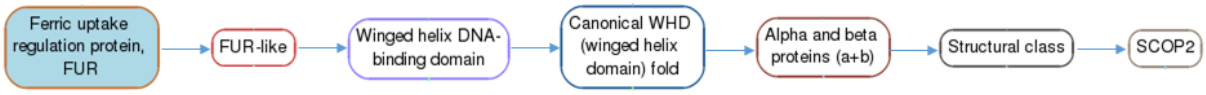


Figure 3.5: Structural classification of the Fur protein (SCOP) in the SCOP2 database (*SCOP2*).

Ten Fur structures and one DNA binding domain are available in the PDB they present the structure of Fur from six different species. Table 3.1 provides a list of all the available structures of Fur proteins in the PDB.

Organism	PDB ID	Resolution (Å)	Reference
P. aeruginosa	1MBZ	1.8	Pohl et al., 2003
E. coli (N-ter)	2FU4*	1.8	Pecqueur et al., 2006
V. cholerae	2W57	2.6	Sheikh et al., 2009
H. pylori	2XIG*	1.8	Dian et al., 2011
C. jejuni	4ETS	2.1	Butcher et al., 2012
M. gryphiswaldense	4RAZ	1.9	Deng et al., 2015
M. gryphiswaldense	4RB0 & 4RAY	1.55	Deng et al., 2015
M. gryphiswaldense	4RB1-3	2.6 & 2.75	Deng et al., 2015
C. jejuni	6D57	1.8	Sarvan et al., 2018
F. tularensis	5NHK*	1.8	Pérard et al., 2018
F. tularensis	5NBC*	1.7	Pérard et al., 2018
E. coli	-	2.3	This work
P. aeruginosa	-	2.6	This work

Table 3.1: Available Fur structures in the PDB. (*) indicate structures obtained by our research group. PDB ID 2FU4 is the structure of the N-terminal domain of EcFur.

3.5 Metal binding in Fur proteins

As expected for an iron sensor, Fur proteins contain iron binding sites. However, purification of Fur from *E. coli* and *V. anguillarum* showed the presence of a structural zinc atom per monomer required to stabilize a Fur dimer (Althaus et al., 1999; Zheleznova et al., 2000). This site will be designated site one or S1, and can be seen in the structure of Fur from *H. pylori* in Figure 3.6 and 3.7 where the other metal sites are also described. In the case of Fur from *H. pylori*, the S1 site is composed of cysteines 102, 105, 142 and 145. S1 is close to the C-terminus and is usually coordinated by four cysteines arranged

in two CXXC motifs. In addition to the zinc binding site, Fur structures contain one or two metal binding sites whose roles are still controversial; they are needed for protein activation and activity regulation.

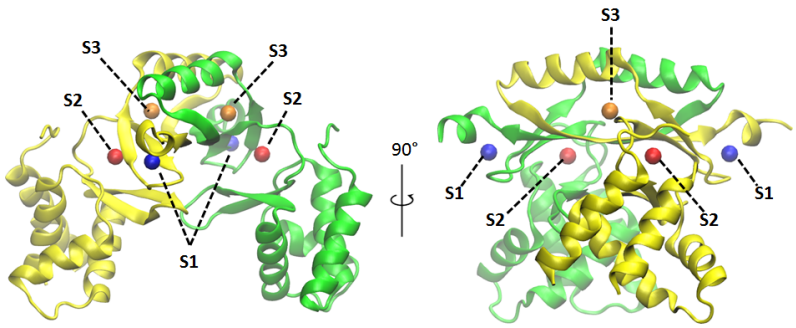


Figure 3.6: Cartoon representation of Fur from *H. pylori* (PDB ID: 2XIG), visualized in VMD (Humphrey et al., 1996), showing the three metal sites: S1 in blue, S2 in red and S3 in orange (Dian et al., 2011).

In all Fur proteins, a metal site, called site two or S2, is present in the hinge region between the DNA binding domain and the dimerization domain and corresponds to a high affinity metal sensing site that involves amino acids from both domains (Figure 3.7A). In the case of Fur from *H. pylori*, the S2 site is composed of His42, Glu90, His97, His99 and Glu110. The S3 site is proposed to be an accessory site, able to tune the affinity of the protein towards DNA. In the case of Fur from *H. pylori*, the S3 site is composed of His96, Asp98, Glu117 and His134.

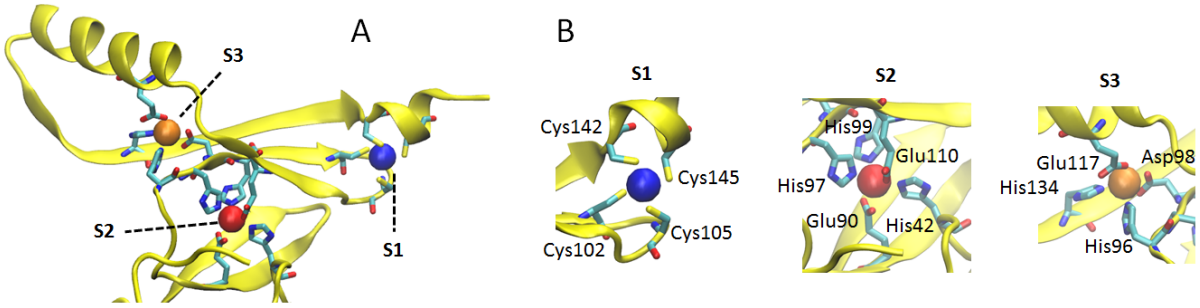


Figure 3.7: Detailed view of metal sites in Fur from *H. pylori*.A: Global positioning of the three metal sites, note how S2 is composed of residues from both dimerization and DNA binding domains. B: Detailed view of each metal site showing the residues involved in metal coordination.

In this section, the case of Fur from *H. pylori* was detailed because it has all three metal sites found in Fur proteins and was thoroughly studied in our team (Vitale et al., 2009; Dian et al., 2011). Fur metal sites are usually equivalent between bacterial species, however, some differences are observed in specific cases. Details of metal sites in the structures obtained in this work will be detailed later.

Taken together the information about metal sites and DNA binding mechanism of Fur proteins, a schematic representation of the activation can be described in Figure 3.8,

where zinc binds its site and enables the dimerization of Fur that becomes active when it metal sites S2 and S3 coordinate a divalent metal element.

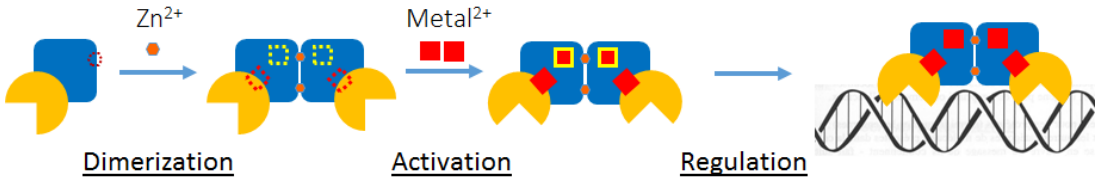


Figure 3.8: Different steps that take place in order to activate a Fur protein. Upon zinc binding to S1 the protein dimerizes and after binding of divalent metal elements in S2 and S3 the protein becomes active and bind to its specific DNA sequence.

3.6 The case of Fur from *E. coli*

When this work started, there was no full structure for Fur from *E. coli*. Previously, Saito proposed a model for the N-terminal domain following NMR studies (Saito et al., 1991) that was not validated by later secondary structure predictions (Holm et al., 1994; Peredo et al., 2001). Later work by our team enabled the resolution of the structure of the N-terminal DNA binding domains PDB ID : **2FU4** (Pecqueur et al., 2006). However no successful crystallization trial of the full protein was achieved after that, even though several experiments were carried out, until this work.

In vitro, Fur from *E. coli* can be activated by Fe^{2+}, Mn^{2+}, Cu^{2+}, Cd^{2+}, Ni^{2+}, Co^{2+} or Zn^{2+} (Bagg et al., 1987; Lorenzo et al., 1987). Unpublished results obtained by Sylvia Vitale during her PhD studies relieved the affinity of metal sites of Fur from *H. pylori* towards different elements. The affinity of site S1 to Zn^{2+} was estimated to be in the range of several nM, in comparison with S2 that has a Kd in the range of µM and binds $Co^{2+} > Fe^{2+} > Mn^{2+} > Ni^{2+} \gg Zn^{2+}$. The difference in affinity between S1 and the other sites is logical since S1 is structurally needed and without it the protein could not function, compared to the other sites that ensure the regulatory and iron sensing role of Fur and where an easily reversible binding is required. The determination of Kd for each site in Fur from *E. coli* is still in progress in the lab using Isothermal titration calorimetry (ITC) experiments, the task is somewhat complicated since for each metal site the others need to be mutated. Now that some of the mutants are available, after X-ray absorption experiments detailed later, characterisation can be continued. However, some barriers need to be overcomed, for example, in site S1 no direct mutation can be made since this site is responsible for the dimerization of the protein.

Fur from *E. coli* is of great importance to our project, since, as described later in this chapter, our first inhibitors were developed on this protein. Obtaining a full structure from was important to confirm acquired data on a previous model obtained by homology to Fur from *V. cholerae*, in addition to its use as a tool for drug discovery.

3.7 Developing Fur inhibitors

As detailed in the previous chapters, the urgent need for new strategies to fight antibiotic resistant bacteria, the importance of iron for all forms of life and the role Fur plays in regulating iron uptake makes this protein a key target for drug development. This section describes previous research from our laboratory that enabled the identification if Fur inhibitors.

3.7.1 Fur and virulence

As an iron responsive master regulator Fur is critical for bacterial growth and iron toxicity prevention. Several studies show that Fur plays an important role in bacterial virulence. It regulates for example, the Feo transporter that is utilized by bacteria to import Fe^{2+}. This transporter plays an important role in colonization and virulence of pathogens like *Helicobacter pylori*, *Salmonella enterica* and *Campylobacter jejuni* (Fillat, 2014; Palmer et al., 2016; Sarvan et al., 2018). The mutation or deletion of the *fur* gene attenuates the virulence of *S. aureus* (Torres et al., 2010; Horsburgh et al., 2001), *V. cholerae* (Mey et al., 2005) and *L. monocytogenes* (Rea et al., 2004) in murine model.

In fact, in *Vibrio cholerae* Fur represses genes located in the pathogenicity island encoding the toxin-coregulated pilus in addition to genes in the *V. cholerae* mega-integron. These *fur* mutants have a significant disadvantage in colonizing the small intestine in a mouse model of intestinal colonization (Mey et al., 2005). Likewise, *Helicobacter pylori fur* mutants have a 50 fold decrease rate in stomach tissue colonization compared to the wild type strain (Gancz et al., 2006). In *Neisseria meningitidis* virulence is also influenced by Fur through its repression of *tbp2* transferrinbinding receptor, a virulence-associated gene involved in iron uptake (Delany et al., 2004). Similarly, *fur* mutants reduce *Staphylococcus aureus* virulence in a murine skin abscess model (Horsburgh et al., 2001). In *Staphylococcus aureus* haemolytic and cytotoxic activities are under Fur regulation, if *fur* is lacking *S. aureus* will increase exoprotein production. At the same time becoming more susceptible to neutrophils decreasing the amount of viable bacteria in a mouse models (Torres et al., 2010). In *Salmonella enterica* Fur was also shown to be required for virulence in mice (Troxell et al., 2011). Our recent work shows the impact of *fur* deletion in *Francisella tularensis* strains using *in vitro* or *in vivo* infection models (Pérard et al., 2018) and will be detailed in the following chapters.

Fur also regulates the expression of virulence factors like the Shiga-like toxin in *E. coli* (Calderwood et al., 1987), cytolysins in *S. aureus* (Torres et al., 2010) and cytotoxin associated gene A (*cagA*) (Pich et al., 2012). The expression of shiga-like toxin (sltA), that inhibits protein synthesis in eukaryotic cells, is produced in low iron conditions. Deletion of *fur* results in a constitutive high expression of *sltA* regardless of iron status. These data indicate a negative regulation of *slt* operon by Fur. The coordinated expression

of siderophore iron uptake and shiga-like toxin shows the major role played by Fur in bacterial virulence. Global regulators like Fur are important for pathogenesis through their response to environmental signals indicating entry into host organisms. As seen in the previous chapter, low availability of free iron in mammalian tissues can trigger the expression of bacterial virulence determinants (Calderwood et al., 1987).

3.7.2 *In vitro* screening for Fur inhibitors

The screen for inhibitors was made using the yeast two-hybrid technique. Discovered by Fields and Song in 1989, the yeast two-hybrid system introduced a new way to study protein-protein interactions (Fields et al., 1989). Yeast two-hybrid assays are of interest in drug discovery (Hamdi et al., 2012). The classical yeast-two hybrid system is made from three main components. First a DNA binding domain linked to the protein of interest constituting the bait complex. Second a transcription activation domain linked to the tested protein, establishing the prey complex. And third a reporter gene to visualise the potential interaction. As shown in Figure 3.9, if the prey and the bait interact, the DNA binding domain and the activation domain are close together, reconstructing an active transcription factor and expressing the reporter gene. In other cases, if their is no interaction between bait and pray or if a small molecule inhibits this interaction, Figure 3.9 right panel, no transcription of the reporter gene is observed. In drug discovery pipelines, the disruption of interaction by small molecules is used to screen for potential inhibitory molecules.

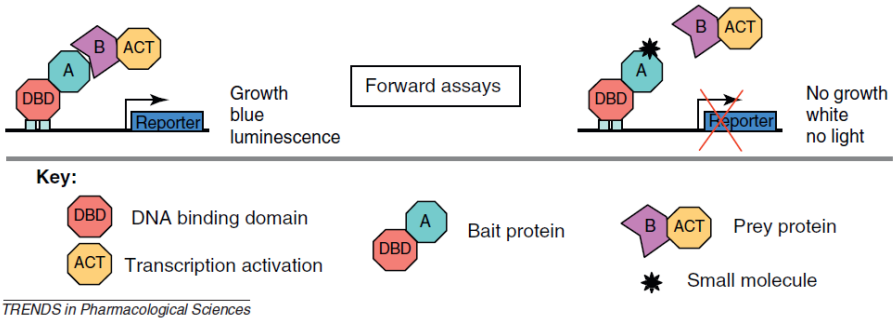


Figure 3.9: Yeast two-hybrid method for protein-protein and protein-small molecule interactions, adapted from Hamdi et al., 2012

The search was carried out using the peptide aptamer technology. Peptide aptamers were developed in the nineteen nineties as an equivalent of immunoglobulines in pharmaceutical research, since their production is faster and do not require expression in animals (Colas et al., 1996). Peptide aptamers are genetically engineered combinatorial proteins made from well characterized scaffold protein to which a variable peptide loop is added. This loop is exposed to the surface and will be able to interact and inhibit target proteins (Colas et al., 1996; Blum et al., 2000). The scaffold protein of the aptamer consisted of

thioredoxine A that is present in procaryotes and eukaryotes with similar three dimensional structures known as the thioredoxin fold consisting of a central core of five β strands surrounded by four α helices. Thioredoxines reduce disuldfide bridges with their exposed active site (CGPC) and are critical for thiol redox control (Arnér et al., 2000; Collet et al., 2010). Because of its stability, solubility and know structure the tioredoxine is used to study the properties of variable peptide loops in biological systems.

For the identification of potential Fur inhibitors, the yeast two hybrid system consisted of B112-aptamer fusions as prey constructs, where B112 is a short unstructured peptide encoded by *Escherichia coli* genomic DNA (Ruden et al., 1991). The bait constructs were LexA-Fur fusions, LexA is a bacterial protein that represses genes involved in the response to DNA damage like *recA* and *lexA* (Little et al., 1981). The peptide aptamer library developed by Aptanomics SA (Bickle et al., 2006) contained $2.7*10^7$ peptide aptamers with a variable loop of 13 amino acids.

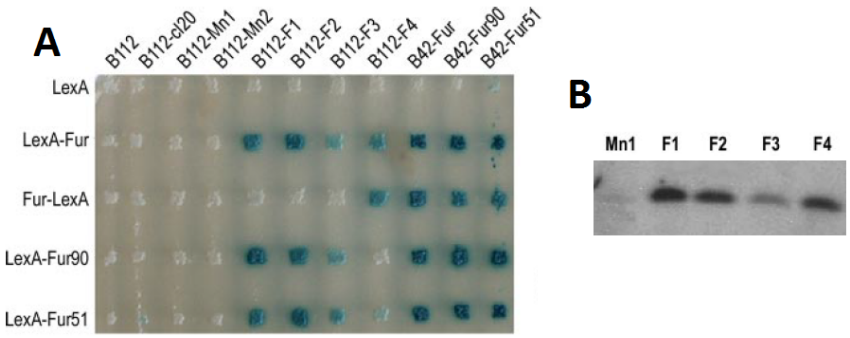


Figure 3.10: A: Yeast two-hybrid assay between LexA-Fur fusions and B112-aptamer or B42-Fur (a positive control) fusions. Mn1 is a negative control. B: Glutathione S-transferase (GST) pulldown of Fur by specific aptamers (Abed et al., 2007). Using a GST tag, aptamers are immobilized on a chromatographic column on which an *E. coli* protein extract is added. The elution volumes are analysed on a Western-blot using Fur specific rabbit polyclonal antibody.

4 clones presented an interaction between the peptide aptamer and Fur of E. coli, they were named F1 to F4, Figure 3.10A. Pulldown experiments were performed to confirm the interaction between these aptamers and Fur from *E. coli*, Figure 3.10B. No sequence homology was found between the variable regions of the aptamers, shown in Table 3.2 and *E. coli* proteins.

Aptamer	Variable region
F1	RLWCRYPHPPLTD
F2	RQCNICGASLYSY
F3	ETCKCGSQVWRHS
F4	CARCGARVNVYKY

Table 3.2: Sequences of peptide aptamer variable regions

Figure 3.11 also shows the interaction between Fur and Fur51 and Fur90, both are

inactive Fur mutants, showing that inactif homodimers can form heterodimers with the wild type protein. Interestingly, F4 is the only aptamer interacting with Fur when it is bound to LexA with its C-terminal domain indicating that F4 does not bind to the same region as the other aptamers. To get more details about this mechanism, another yeast two-hybrid screen was made using several truncated Fur constructs as baits.

Negative controls used in yeast two-hybrid assays are cl20, a randomly chosen aptamer that does not bind Fur, in addition to Mn1 an aptamer that confers resistance to high manganese concentration but without interacting with Fur.

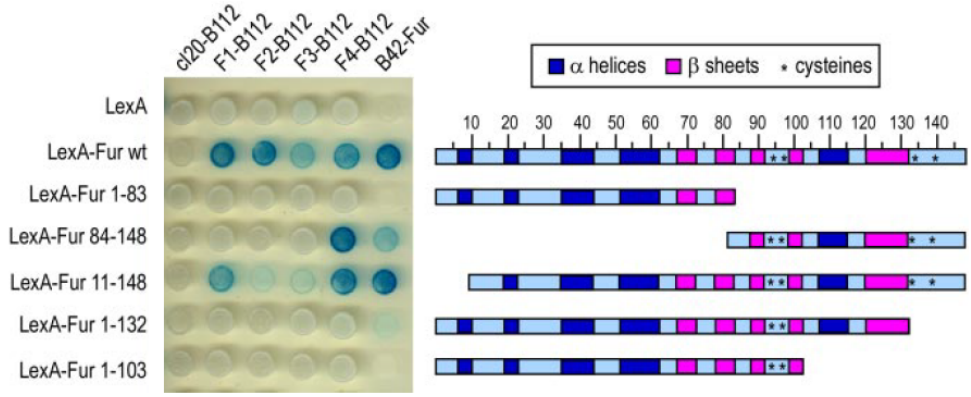


Figure 3.11: Identifying aptamer binding sites on Fur (Abed et al., 2007). A yeast two-hybrid assay between B112-aptamer fusions and different truncated Fur proteins. cl20 is a negative control.

This assay showed that none of the aptamers were able to interact with the N-terminal domain of Fur (1-83 construct) and that F4 was the only one to interact with the 84-148 construct. Indicating that F4 interacts with the C-terminal domain of Fur. In addition, the interaction of F1 was independent of the first 11 residues indicating that F1 and F4 interact with different areas of the protein, Figure 3.11.

In yeast two-hybrid assays, colour intensity can give a qualitative idea about the affinity of the complex, Figure 3.10A. The interaction between Fur and F3 aptamer is weaker than with the other aptamers, the same result is also visible in Figure 3.10B. In order to obtain quantitative data, the same study was repeated using a luciferase reporter gene *luc*. The results, shown in Figure 3.12, indicate that the apparent binding affinities of F1, -2, and -4 are similar to one another and higher than that of F3.

Inhibition of Fur repression by peptide aptamers

When expressed, the described peptide aptamers can inhibit Fur repression of gene expression. Using aptamers F1 to F4 in addition to a negative control cl20 and a positive control Fur90, the following test was established. The reporter gene *lacZ* is placed downstream of the *fiu* promoter containing four Fur boxes, therefore *lacZ* expression is highly repressed by Fur (Hantke, 1984). In addition, aptamer expression is dependant of arabinose. When the expression of aptamers is induced, their effect on Fur repression can

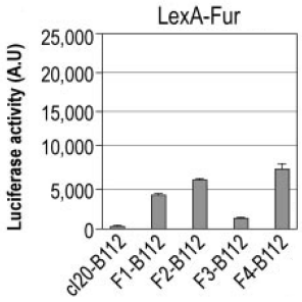


Figure 3.12: Quantitative yeast two-hybrid experiment using a luciferase reporter gene (Abed et al., 2007). cl20 is a negative control.

be seen through the colour of the colonies as seen in Figure 3.13. If they interact with Fur, the repression will be inhibited and the expression of lacZ gene, in the presence of X-gal in the culture medium, will give blue colonies. This is the case of F1 and F2 that show similar results to the dominant negative allele Fur90. F4 shows a slightly weaker inhibition of Fur. Moreover, F3 shows the weakest inhibition in agreement with the previous results.

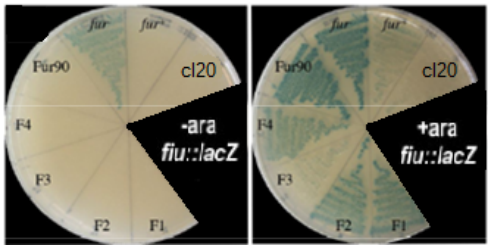


Figure 3.13: Peptide aptamer effect on transcirptional repression by Fur (Abed et al., 2007). Blue colonies indicate an inhibition of Fur.

Homodimerization inhibition

In order to understand how each aptamer is interacting with Fur, a yeast two-hybrid assay was used to study their effect on Fur homodimerization. To do so three different constructions are expressed in yeast : LexA-Fur, B42-Fur and the aptamer of interest. If Fur is able do dimerize the reporter gene will be expressed and releases a signal. If the aptamer interferes with Fur homodimerization the signal will be reduced. Figure 3.14 shows the results obtained. The expression of wild type Fur is considered as the positive control since wild type Fur will tend to dimerize with LexA-Fur and B42-Fur constructions thereby decreasing the expression of the reporter gene. F4 aptamer is the only one to significantly affect homodimerization with 55% inhibition, indicating that it interacts with the C-terminal domain of Fur responsible for its dimerization.

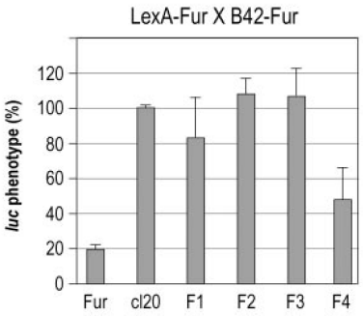


Figure 3.14: Yeast two-hybrid competition assay to study the effect of aptamers on Fur homodimerization (Abed et al., 2007). A luciferase reporter gene was used and results results are normalized with the value obtained for aptamer cl20 considered to be the negative control.

3.7.3 Effects of inhibiting Fur *in vivo*

To determine whether Fur-binding peptide aptamers affect *E .coli* virulence *in vivo*, they were tested *in vivo* in an immunodeficient *Drosophila* model. This TAK1 (transforming growth factor-β-activated kinase 1) *Drosophilia* mutant was infected with E. coli 1106 pathogenic strain expressing the peptide aptamers to be tested or the Fur90 protein. In this specific insect model, lethality occurs in 3 to 4 days after infection. When aptamer producing *E. coli* were used a slower mortality rate was observed, as shown in Figure 3.15. Indicating that the expression of Fur specific peptide aptamers decreased the virulence of pathogenic *E. coli* in *Drosophila* (Abed et al., 2007).

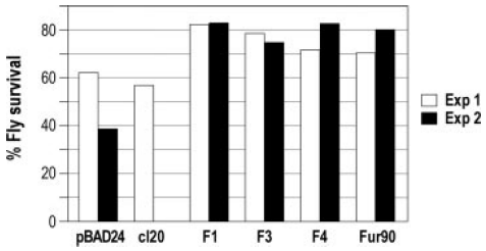


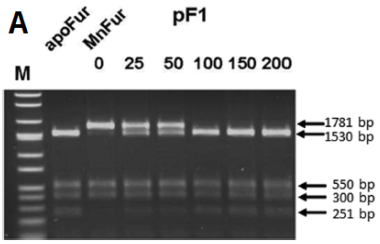
Figure 3.15: Virulence assay showing the percentage of surviving flies following septic injury (Abed et al., 2007), cl20 and pBAD24, the empty expression plasmid, are used as a negative controls. Fur90 is a positive control with an inactive Fur protein.

3.7.4 Coupled *in silico* & *in vitro* approach

As described previously, anti-Fur aptamers denoted F1 to F4 with thioredoxine A as scaffold and 13 amino acids variable loops were identified as Fur inhibitors. Using a mixed theoretical and experimental approach the pF1 peptide, corresponding to the variable loop of F1 aptamer with the following sequence: RLWCRYPHPPLTD was studied in detail in the work of Cheikna Cissé (Cissé et al., 2014).

pF1 was studied in vitro using a nuclease protection assay to determine its capability to interfere with the DNA binding activity of Fur. Following an enzymatic digestion, if a

1781 bp band is observed it indicates an active protein. Otherwise a 1530 bp and 251 bp bands indicate an inactive protein. More details on the nuclease protection assay can be found in the Methods chapter. Figure 3.16A shows that pF1 interaction with Fur has an IC50 of 53 µM, corresponding to the concentration of peptide necessary to inhibit 50% of the DNA binding activity of EcFur.



B

	sequence	IC50 (μM)
pF1(1–13)	RLWCRYPHPPLTD	53 ± 6 (12)
pF1(1–10)	RLWCRYPHPP	64 ± 10 (3)
pF1(1–9)	RLWCRYPHP	67.5 ± 5 (4)
pF1(1–8)	RLWCRYPH	30 ± 11 (4)
pF1(2–9)	LWCRYPHP	67 ± 10 (8)
pF1(3–9)	WCRYPHP	50 ± 15 (10)
pF1(4–9)	CRYPHP	inactive
pF1(3–8)	WCRYPH	26 ± 6 (8)
pF1(3–7)	WCRYP	inactive

Figure 3.16: A: Nuclease protection assay showing Fur inhibition by pF1. Lanes 1 and 2 show the activation of Fur through the addition of metal (Mn in this case). B: IC50 values for pF1 derivatives in interaction with Fur from *E. coli*. In the left column the numbers indicate the positions of the first and the last amino acids. In the right column number of assays is given in parentheses (Cissé et al., 2014).

In order to find the minimum active sequence of pF1 more than 20 pF1 derivatives were tested and their IC50 determined, results are shown in Figure 3.16B. The IC50 values are within the same range for the majority of pF1 derivatives. The peptide pF1(3-8) has the minimal active sequence with the lowest IC50 value. This experiment showed that the main residues involved in the interaction are in the middle of the peptide, in agreement with the fact that they are the most exposed to the solvent when constrained in an aptamer (Cissé et al., 2014).

In addition, docking results show high interaction energy between Fur and pF1 residues Y6, R5, and H8. Since at that time the structure of Fur from *E. coli* was not known, the model was constructed by homology to Fur from *V. cholerae* (PDBID : 2W57), details on the modeling conditions can be found in the Methods chapter. To confirm this result, these residues were mutated and the peptides were not active *in vitro*. The binding site of the peptide is located at the interface between the two subunits, with the arginine 5 residue inserted in a groove between the monomers Figure 3.17A. When pF1 derivates were docked on the EcFur model, the moiety that spans from W3 to P9 superimposes perfectly. In all the simulations, EcFur residue R70 is essential for peptide binding in addition to Y56, S126, Y128, Y130, and G75. To confirm this binding pocket, S126 and Y128 and were mutated into alanine *in vitro* and *in silico*. The docking results show a decreased affinity of pF1 toward this mutated protein. The experimental test consisted of a yeast two-hybrid assay where almost a complete loss of interaction was observed Figure3.17B.

An in-depth characterization, of these linear peptides derived from anti-Fur aptamers was carried out in the work of Sophie Mathieu (Mathieu et al., 2016)

Based on previous results (Abed et al., 2007), an optimized yeast two-hybrid assay was

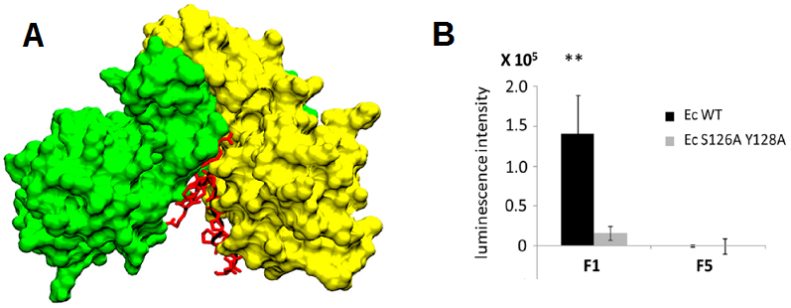


Figure 3.17: Licorice model of pF1(1-13) docked into EcFur, subunit A (green) and B (yellow). B: Comparison of pF1 (1-13) interaction with wild type EcFur and the S126A Y128A mutant by a yeast two-hybrid, F5 is a negative control (Cissé et al., 2014).

used to compare the interaction between anti-Fur aptamers and their peptide derivatives with EcFur. Figure 3.18 shows that when aptamers are used, strong interaction is observed compaired to the negative control. However, when peptides are used only pF2 shows a significant interaction compared to the negative control. In both assays, F2 and pF2 are the ones that interact the most with Fur from *E. coli* (Mathieu et al., 2016).

The difference in interaction strength between aptamer or peptides with EcFur can be explained by the constraints placed on the peptides from the scaffold of aptamers, probably limiting the flexibility and promoting a specific three dimensional organisation of the side chains. In contrast, linear peptides can adopt several conformation since no constraints are applied on them (Mathieu et al., 2016). In addition, the stability of the complexes and their expression levels due to the high solubility and easy expression of the thioredoxin used as a scaffold in the aptamers (Abed et al., 2007).

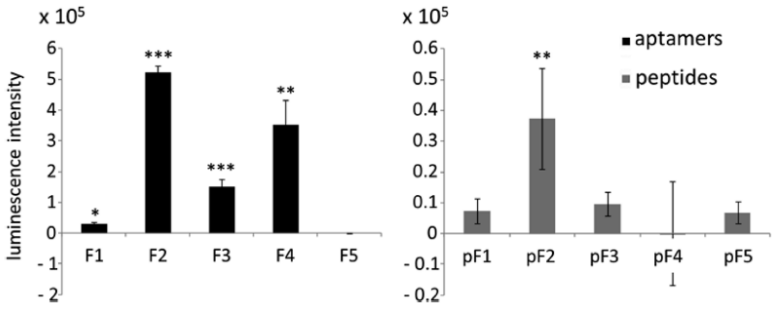


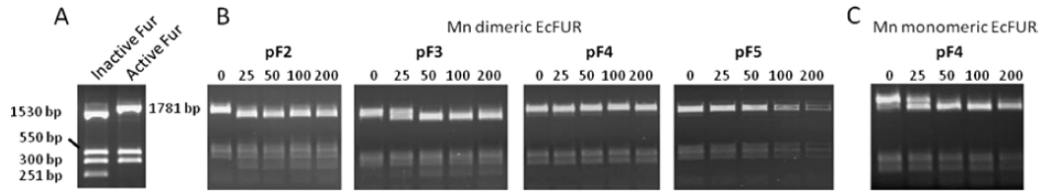
Figure 3.18: Interactions of EcFur with peptide aptamers (F1 to F5) and peptides (pF1 to pF5) by yeast two-hybrid assay, F5 and pF5 are negative controls (Mathieu et al., 2016).

As in the case of pF1, the IC50 of the other peptides were tested in a nuclease protection assay, see Figure 3.19. Peptides pF2 and pF3 show better inhibitory activity than pF1 3.19D. The profile of pF4 is similar to the negative control in the case of the interaction with a Fur dimer 3.19B. However its inhibitory action is present when interacting with a monomer 3.19C, validating previously obtained results about its interaction mechanism with the C-terminal domain (Abed et al., 2007).

It is important to note that the variation in the IC50 values shown in 3.19D is may

be related to the redox instability of the cysteine containing peptides. When oxidized, the peptides were not active on EcFur but their activity was restored after reductive treatment with TCEP (Mathieu et al., 2016).

Moreover, pF2 derivatives were tested and their inhibitory activity suggests that the N-terminal residues are essential for the inhibition. In addition, for some constructs where the N-terminal is conserved and the C-terminal lacks the last two residues (pF2(1-11) and pF2(1-10)), the inhibition is stronger than when pF2(1-13) is used.



D

	Sequence	EcFUR dimer / monomer
pF1(1-13)*	RLWCRYPHPPLTD	53 ± 6 (12) / ND
pF2(1-13)	RQCNICGASLYSY	12.5 ± 5 (11) / ND
pF3(1-13)	ETCKCGSQVWRHS	30 ± 17 (10) / ND
pF4(1-13)	CARCGARVNVYKY	NI (3) / 37.5 ± 12.5 (3)
pF5(1-13)	RSLRGRCLSQHQD	NI (5) / NI (3)

Figure 3.19: Interactions of EcFur with peptides pF2 to pF5 by nuclease protection assay (Mathieu et al., 2016). A: profile patterns for inactive and active protein. B and C: Activity tests on the dimeric and monomeric forms of the protein respectively. D: IC50 values, in µM, of the interaction between peptides and Fur from *E. coli*.

Docking of pF2 on the EcFur model shows two separate symmetrical binding sites, making the EcFur dimer able to bond two pF2 peptide at the same time, Figure 3.20. Similarly to pF1, residues such as R70, N72, G75, and Y128 seem crucial for the binding. A good binding free energy of -21.0 kcal/mol is obtained for this complex. Isothermal titration calorimetry also favours the binding of two pF2 peptide per EcFur dimer with a stoichiometry of 1 pF2 per Fur subunit (Mathieu et al., 2016).

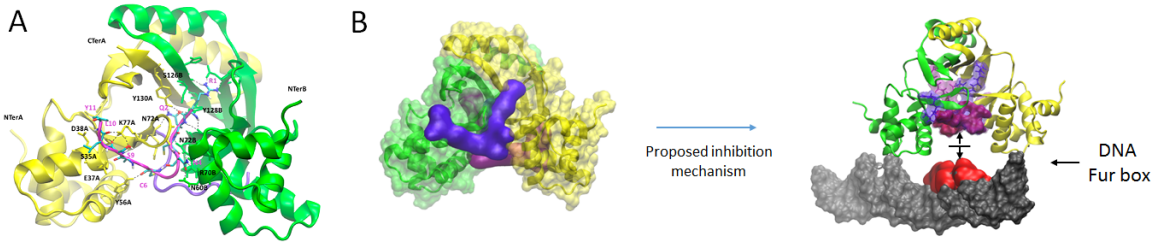


Figure 3.20: Docking of two pF2 peptides on the EcFur model. A: highlighting the main residues involved in the interaction. B: proposed inhibition mechanism (Mathieu et al., 2016).

3.8 Chapter conclusion

Since its discovery in the 1980s, the Fur protein has been thoroughly characterised as a key regulator and iron sensor within bacteria. Due to multiple regulation mechanisms,

Fur control a large set of genes, some of which are involved in the expression of virulence factors. In fact the mutation or inhibition of Fur was shown to decrease the virulence of pathogenic bacterial strains.

Fur has been characterised in different organisms and several PDB structures are available. However, when this work started, the structure of Fur from the model organism *E. coli* (EcFur) was still unavailable except for the N-terminal domain resolved in our laboratory (Pecqueur et al., 2006). Previously in our lab, a screening of a large peptide aptamer data base revealed four peptide aptamers capable of inhibiting EcFur (Abed et al., 2007). Following that work, small peptide inhibitors of EcFur were developed using a coupled experimental and theoretical approach (Abed et al., 2007; Cissé et al., 2014; Mathieu et al., 2016).

Thesis objectives

The work presented in this manuscript is divided in two main parts. The first is the study of Fur protein inhibition; this will be the subject of chapter 4. The second part is the study of Fur oligomeric states through *in silico* simulations; this will be discussed in chapter 5.

This division in two parts is the result of two main questions that were asked at the start of this project: how does the inhibition of Fur proteins work? and, why are there, depending on the bacterial species, Fur tetramers and Fur dimers ? These are relatively broad questions, and would probably require more than one thesis project to be answered. Nevertheless, they define the objectives that this work tends to accomplish using a combined experimental and theoretical approach.

In chapter 4, the optimization of previously discovered inhibitory peptides is discussed, as well as a study of inhibitors conceived *in silico*. The objective is to understand these inhibitions in each case, since Fur proteins are not all the same, in addition to investigating the role, if any, of the metal sites in the inhibition. To do so, docking simulations and X-ray absorption spectroscopy were used. In addition, extensive crystallization trials were done to obtain structures of Fur proteins in the presence of inhibitors. By doing so, two new structures of Fur proteins were obtained and characterized *in vitro*, and are described in this chapter.

In chapter 5, the goal is to understand the molecular mechanisms behind the interactions of Fur with DNA, and the dimer-dimer interactions in the case of Fur tetramers. This *in silico* study, based on crystal structures, has two main objectives. The first is to identify important residues in each type of interaction. This will enable a better understanding of Fur proteins mechanistics. The second is the search for a specific dimeric or tetrameric signature in the sequences of the studied proteins. This kind of signature will help in the understanding of Fur interactions in the absence of protein structures and hint towards the protein oligomeric state *in vivo*. In addition, by investigating the individual interactions at the molecular level, this study aims to understand why Fur does not bind a mutated DNA sequence, and sheds light on differences in interactions between Fur and two DNA sequences that it can bind *in vitro*.

The objectives of this work being described, the following chapter details the different methods used to investigate the mechanisms behind EcFur inhibition, as well as the inhibition of Fur from *P. aeruginosa* (PaFur), using a coupled experimental and theoretical approach.

Chapter 4

Studying Fur inhibitors

As discussed in the introduction, four peptide aptamers were shown to interact with EcFur and their variable parts (linear peptides) have been characterised. The properties of the interactions of EcFur with pF1, pF2, pF3 and pF4 have been characterized and inhibition pockets proposed on the basis of proteins and peptide mutations, in addition to docking simulations (Cissé et al., 2014; Mathieu et al., 2016).

The work detailed here, describes studies on the inhibition of Fur from two well studied species: *E.coli* (EcFur) and *P.aeruginosa* (PaFur) by inhibitors developed by our team. EcFur is well characterised, however, no X-ray diffraction structure is currently available, except for the N-terminal DNA binding domains (Pecqueur et al., 2006). The structure of PaFur was first described by Pohl as a dimer (Pohl et al., 2003) and later as a tetramer in solution by our team (Pérard et al., 2016).

This chapter focuses on three main inhibitors: the linear peptide pF2 presented in the introduction, previously unpublished pL1 and pL2 inhibitors conceived *in silico* and a small chemical molecule, that we will designate "molecule B", discovered after a chemical library screen. Their respective IC50 value, when available is presented in Table 4.1. In what will follow, dockings on EcFur and PaFur will be detailed, together with X-ray absorption spectroscopy experiments. At the same time, crystallization trials will be described leading to the crystal structures of PaFur and EcFur mutant proteins in addition to their biophysical characterisation.

Inhibitor	Sequence	IC50 EcFur (µM)	IC50 PaFur (µM)
pF1	RLWCRYPHPPLTD	53	-
pF2	RQCNICGASLYSY	6	-
pL1	RQCRYCHW	15	20
pL2	WCRYCHPR	15	20

Table 4.1: IC50 values of peptide inhibitors when used on EcFur and PaFur.

4.1 Docking of cyclic peptides on Fur from *E. coli*

Prior to the resolution of the EcFur-140 structure, described later in this manuscript, dockings of inhibitors were carried out on a model of EcFur obtained by homology to Fur from *V. cholerae* (VcFur) as described in (Cissé et al., 2014). A series of cyclic peptides, with various lengths of glycine linker residues, were docked on this model in an attempt to find new hits that would be later synthesised by our collaborator Melissa Degardin at the Département de Chimie Moléculaire (DCM) of the University of Grenoble Alpes.

Cyclic peptides were studied since they would be more stable in solution than linear ones, and more importantly, *in vivo*, they would be less prone to degradation by exopeptidases. Moreover, when the peptides were discovered, they were constrained by the aptamers within the loops and their interactions with EcFur was 10 times higher than the interaction of linear peptides and EcFur (See 3.18). So, in addition to stabilising and protecting from exopeptidases, the cyclic peptides were studied to mimic the restraints imposed by aptamer loops. A cyclic pF1 peptide was already synthesized and had a four folds increase in ints affinity towards EcFur (Cissé et al., 2014).

Before running new synthesis, dockings were carried out on the inhibitor pF2 that was cyclized using glycine residues added to its N-terminal and C-terminal ends. Figure 4.1A shows three different constructions that used 2, 4 or 6 glycines to create the cycle. When their docking position is compared to the one of the linear peptide (magenta), key interactions with the protein are missing especially in the case of pF2_2gly where three interactions are missing (red arrows). When the cycle is made bigger by adding more glycine residues, interactions are regained. When 6 glycines are used the interactions are similar to those of the linear peptides. This gain in interaction when the cycle size in increased is reflected by the binding free energy of each complex with ΔG = -16; -18; -22 kcal.mol^{-1} when 2, 4 or 6 glycines are used, respectively. Figure 4.1 compares pF2_2gly to the linear peptide which had a ΔG = -23 kcal.mol^{-1} showing why small cycles lost some of the interactions with the protein.

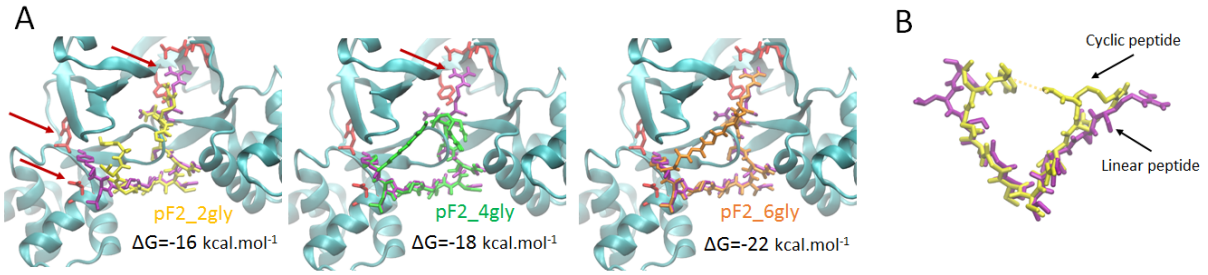


Figure 4.1: A: Docking of cyclic pF2 peptides on a homology model of EcFur. Three constructions of pF2 cyclized using two (yellow), four (green) or 6 (orange) glycine residues, are compared to the linear docking of pF2 (magenta). Red arrows indicate interactions present between the linear peptide and the protein but are missing in the case of a cyclic peptide. B: Spacial arrangement of a pF2 cyclized using two glycine residues (yellow) compared to the linear peptide (magenta).

This series of tests showed that small cyclic peptides based on pF2 were not better than the linear peptide. The best hit was when 6 glycines were used enabling the cycle to have the same interactions as the linear one. However, since this initial idea was to obtain small cycles, this cyclic peptide was not synthesized and other strategies to stabilize the peptides were chosen (D-retro-inverso-peptides, briefly detaild in the next section).

4.2 Inhibitory peptides : pL1 and pL2

Previously discovered peptide inhibitors were not able to inhibit Fur from *P. aeruginosa*. New peptides, pL1 and pL2, were designed *in silico* by adapting previously discovered peptide sequences in docking simulations by Lindsey Flanagan, a previous Master student from York, UK. Table 4.2 shows the sequence of each peptide while highlighting the resemblance between pL2 and pF1. Both pL1 and pL2 inhibit EcFur and PaFur, IC50 were found to be around 20 µM.

Inhibitor	Sequence
pL1	RQCRYCHW
pL2	WCRYCHPR
pF1	RLWCRYPHPPLTD

Table 4.2: Sequences of inhibitors pL1 and pL2. Orange residues highlight the similarity between pF1 and pL1.

pL1 and pL2 inhibit PaFur *in vitro*. In our simulations they were docked on a dimer of this protein, since at that time PaFur was thought to be dimeric in solution as the dogma was that all Fur proteins were dimers. Dockings were initiated by submitting the protein structure and the inhibitor sequence to on line docking servers (GalaxyPepDock (Lee et al., 2015), FlexPepDock (London et al., 2011) and CABS-dock (Kurcinski et al., 2015)) in order to have a consensus on the binding pocket before starting our study.

Once the starting point is determined, several loops of energy minimization with CHARMM (38b1), using the EEF1 potential (see section 7.3.2), and docking with Autodock (4.2) were done to determine the binding pocket and the binding free energy. In the case of pL1 the binding pocket was found between the two PaFur subunits as shown in Figure 4.2 with a ΔG = -18.84 kcal.mol^{-1} (mean of -16.54 kcal.mol^{-1}) with the major interactions involving the C-terminal part of pL1. Energy details of the interaction can be found in Table 4.3.

The same protocol was applied to pL2 where two binding pockets were predicted by on line servers. They were both studied and gave ΔG = -16.87 kcal.mol^{-1} (mean of -14.91 kcal.mol^{-1}) for conformation A and ΔG = -15.68 kcal.mol^{-1} (mean of -15.29 kcal.mol^{-1}) for conformation B (Figure 4.3). In both cases, the N-terminal of pL2 is mainly involved

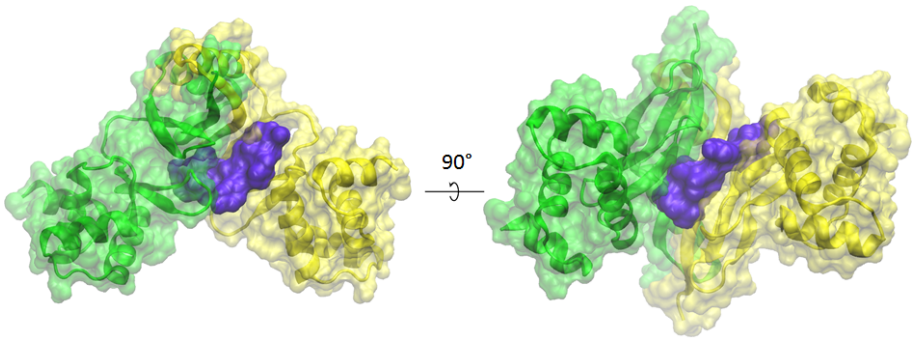


Figure 4.2: Docking of pL1, shown in blue, on a dimer of PaFur visualized in VMD (Humphrey et al., 1996). Yellow and green representations indicate each subunit of the dimer.

Protein residue	Energy	Peptide residue	Energy
Arg 69	-13.53	Trp 8	-16.97
Asn 125	-7.54	Tyr 5	-13.33
Asn 71	-6.23	Gln 2	-11.34
Asp 123	-5.74	Arg 4	-4.65

Table 4.3: Protein and peptide residues involved in the interaction of pL1 docked on PaFur. Energies are given in kcal.mol^{-1}.

in the interaction with the protein. In the case of conformation A, pL2 is located between the two chains and interacts with both. In comparaison, in conformation B, pL2 is closer to one chain than the other and interacts exclusively with it, see Figure 4.3. Table 4.4 shows the details of the interaction in each conformation.

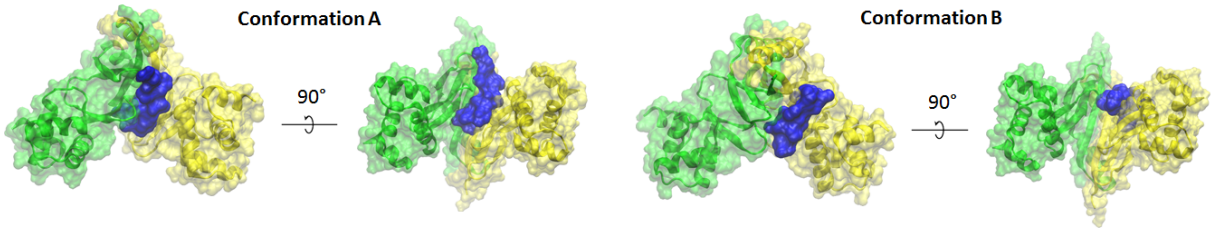


Figure 4.3: Docking of pL2 , shown in blue, on a dimer of PaFur visualized in VMD (Humphrey et al., 1996). Yellow and green representations indicate each subunit of the dimer.

Moreover, in a try to optimize these inhibitors and protect them from exonuclease *in vivo*, an attempt was made to create D enantiomers of each peptide where the sequence is in reverse order giving rise to D-retro-inverso-peptides. Preliminary docking results were not good with the binding free energy dropping to half its value when the D-peptides are used. These results are based on less than ten simulation cycles (see figure 4.32 shown in later sections), additional cycles should be carried on to confirm current results. In addition, the D-retro-inverso-peptide of pL2 (named pL2D) was synthesized by our collaborator and found to be less efficient than pL2.

Conformation A				Conformation B			
Protein	Energy	Peptide	Energy	Protein	Energy	Peptide	Energy
Arg 69	-12.59	Trp 1	-16.02	Arg 69	-5.18	Trp 1	-16.30
Tyr 129	-10.50	Arg 3	-12.95	His 124	-3.60	Arg 3	-4.85
Asn 71	-8.78	His 6	-5.31	Ala 82	-3.68	Tyr 4	-2.21
His 70	-4.95	Tyr 4	-3.85	Leu 128	-2.24	Cys 5	-1.87

Table 4.4: Protein and peptide residues involved in the interaction of pL2 docked on PaFur. This table shows details for both binding conformations of pL2. Energies are given in kcal.mol^{-1}.

4.3 Chemical library screening: the case of molecule B

In addition to the development of modified peptides, the search for new inhibitors of Fur proteins was ongoing. Since peptide aptamers were shown to interact and inhibit Fur (Abed et al., 2007), it was decided to use this interaction to screen for new small-molecule inhibitors. In fact, if a molecule binds Fur in the same inhibition pocket as the peptide aptamer does, it will disrupt the Fur-aptamer complex and it could be considered as a potential Fur inhibitor. Nevertheless, an interaction with the peptide aptamer instead of Fur cannot be excluded.

The Prestwick Chemical Library (PCL®) was used for a high throughput screening by Aynur Ahmadova, a previous postdoctoral researcher in our laboratory. Of the 1200 small molecules tested, the yeast two-hybrid assay revealed 84 hits. After eliminating those that can be toxic to yeast or generate false positives, and following a manual screen, only one molecule was left. Designated molecule B, this molecule was able to inhibit *in vivo* interaction of the PaFur/F4 complex and PaFur homodimerization, indicating a possible direct interaction with PaFur.

Molecule B is an already existant molecule, previously used as a drug for several years before being withdrawn from the therapeutics market. However, recent studies showed potential new targets for this molecule, opening the way for possible future therapeutical uses.

After the *in vivo* screening, the molecule was bought and tested *in vitro* in a nuclease protection assay. Surprisingly, no inhibition of PaFur was detected *in vitro* in the presence of molecule B. This could be linked to the tetrameric state of PaFur in solution, and since molecule B was shown to inhibit homodimerization, it might have some trouble interacting with an already formed and stable tetramer (the stability of this tetramer is described in Pérard et al., 2016).

To understand how molecule B works, and since the main information we are looking for is its binding site, docking simulations were carried out on Fur from *E.coli* with 15 chemical derivatives of molecule B. At the same time, to get more insights into the interac-

tion of molecule B with Fur proteins, crystallization trials were initiated for [Fur+molecule B] complexes.

4.3.1 Docking of molecule B on Fur from *E. coli*

Previously, unpublished data on dockings of molecule B on a Fur dimer from *P. aeruginosa* were carried out, showing binding pockets in the dimerization domain of the protein and no interaction with residues involved in DNA binding, validating its discovery as an inhibitor of PaFur dimerization. However, docking scores were relatively weak and in the absence of a structure of a subunit of PaFur alone, the docking simulation could not be continued.

In prior *in vivo* assays, molecule B was also shown to inhibit EcFur dimerization. As a complementary study during my thesis work, dockings of molecule B and 15 of its chemical derivatives, shown in supplementary information, were carried out with Audodock Vina (Trott et al., 2010) on a dimer model of Fur from *E. coli* obtained by homology modelling with Fur from *V. cholerae*. As in the previous dockings, scores obtained were low and varied between -4.2 and -6.5 depending on the molecule and the initial binding site.

Seeing that the results were similar to the previous ones, not concluant that is, this project was put on the side allowing crystallization trials to be the main focus for approximately six months.

4.4 X-ray absorption spectroscopy studies on EcFur

In the absence of crystallographic structures of Fur proteins bound to their inhibitors, X-ray absorption spectroscopy (XAS) techniques were used to acquire structural data on these protein-inhibitor complexes. XAS techniques can deliver information about the chemical environment of metal atoms, and since Fur proteins have metal sites they can be studied, details on XAS can be found in section 6.1.3. In addition to obtaining structural information about Fur metal sites, the XAS experiments described here were done to validate the hypothesis that peptides inhibiting Fur do not interact with metal sites. In other words, the inhibition is not due to the peptides chelating metals from Fur proteins.

To test this hypothesis, Fur from *E. coli*, *P. aeruginosa* and *F. tularensis* were produced as wild type proteins or metal sites mutants and purified in high concentrations (around 30 $mg.mL^{-1}$) before being mixed with inhibitors (Table 4.5 shows all prepared samples). Mutations affected the S2 and S3 sites, they were produced to reduce the affinity of one site and to populate the other in order to have only one metal absorber per protein. Since XAS could be considered as a non destructive technique on frozen liquids, our samples were also used in crystallization attempts as discussed earlier.

Sample	Composition	Sample	Composition	Sample	Composition
EcFur-WT	+Co ±pF2	FtFur-WT	+Co	PaFur-WT	+Co ±pL1
	+Mn ±pF2		+Mn		+ Mn
	+Co +Mn +pF2				+Co +Mn +pL1
EcFurΔS2	+Mn	FtFurΔS2	+Co	PaFurΔS3	+Mn
EcFurΔS3	+Mn ±pF2				+Co

Table 4.5: Samples prepared for XAS studies.

In the case of Fur from *E. coli* and its inhibitors pF1 and pF2, previous results showed that the inhibition pocket is located on the inner surface groove of the EcFur subunits and does not involve residues from metal sites (Cissé et al., 2014; Mathieu et al., 2016). In addition, in the molecular dynamics simulations, metal atoms and their sites were constrained mainly to avoid loosing the metal atoms during the simulations and to avoid interactions between the inhibitors and the metal atoms. The following paragraph shows the results obtained for EcFur-WT in the presence or not of the pF2 inhibitor. They were acquired on the FAME beamline at the ESRF with the help of Denis Testemale. Results were analysed by Giulia Veronesi and myself after following the "FAME+" training course.

Spectra of EcFur metal sites mutants were not acquired due to time limitation on the beamline, samples are still conserved for future use. Result analysis of Fur samples from *P. aeruginosa* and *F. tularensis* are still to be done and will not be discussed in this manuscript.

Proteins were loaded with Mn^{2+} and Co^{2+} since both are known to activate Fur. Fe was not used as a metal probe to avoid anaerobic production, purification and sample preparation in glove boxes. Mn^{2+} was used since it mimics Fe^{2+} *in vitro* without the need for glove boxes, all preparation steps were made under aerobic conditions. Co^{2+} was also used for its hexacoordinated chemical environment and ability to bind sulphur containing ligands such as cysteines found in our peptide inhibitors.

X-ray absorption spectra were acquired in EcFur WT loaded with Co^{2+} in different stoichiometry (1 eq. and 2 eq.) to check for potential differences in metal sites since spectra will indicate the average metal coordination environment for sites populated by Co in the sample. Using a 2 eq. Co solution makes sure that both sites are occupied and an ”average” metal site will be analysed. The 1 eq. solution will either fill 50% of the sites or depending on metal affinity, favour one site over the other.

The following paragraphs describe the results obtained for cobalt activated EcFur-WT, in the presence or not of pF2. X-ray Absorption Near Edge Structure (XANES) spectra (Figure 4.4A) show identical features for 1 eq. (black) or 2 eq. Co (green), suggesting that the metal binding environment is nearly the same in the S2 and S3 sites.

The experimental data corresponding to the sample laded with 1 eq. cobalt was fitted

with a model consisting of [4 histidines + 1 gluatamate] ("4 His 1 Glu_{bid}"), where the glutamate is free to rotate around $O_{\delta 1}$ enabling both mono and bidentate conformations. This model was based on information from the crystal structure detailed later in 4.6.4. The presence of several His residues is also suggested by the marked contributions from high frequency oscillations in the Extended X-Ray Absorption Fine Structure (EXAFS) spectra (Figure 4.4B, dashed lines) due to multiple scattering paths, as the ones generated by the imidazole ring structure (Strange et al., 1987; Binsted et al., 1992). When the relative model ("4His 1Glu bid") is built (Figure 4.4C(a)) and used to fit the experimental data, a good agreement is obtained (Figure 4.4B, solid lines) as confirmed by the low R factor (Annex Table 4.19).

In addition, the calculated angle formed by the Co-ligand-C bond $\beta(Glu_{bid}) = 90 \pm 2°$, that indicates a bidentate glutamate residue in the presence of cobalt. This is consistant with the fact that cobalt binding favours an octahedral environment. Considering that the XANES of 1 eq. and 2 eq. samples are nearly identical, Figure 4.4A, we can say that both metal sites have a similar coordination sphere, thus excluding that the coordination number in S3 is lower than 6. The coordination of site S3 was investigated since in the crystal structure detailed later, the site has four ligand in the presence of zinc, however the electron density map is not sufficiently resolved to propose a site that fits XAS data. The "4His 1Glu bid" model provides a good agreement also with the experimental data of EcFur WT loaded with 2 eq. Co, and equal best-fitting distances within the error as in the 1 eq. sample (Annex Table 4.19). This confirms that the two metal-binding sites S2 and S3 are nearly the same, by being octahedral, involving N/O donors only, and predominantly composed by His residues. Nevertheless, some differences in the nature of cobalt first neighbours cannot be excluded.

We therefore hypothesise that the aminoacids involved in cobalt binding are the ones suggested by the X-ray diffraction structure of site S2, arranged in a way to form an octahedral coordination sphere. Two possibilities can be tested for the S3 site for which the structure is not conclusive. One where the site is made of two histidines, a glutamate and an aspartate, with both carboxylates in bidentate configurations. A second possibility, is a site made from three histidines, a glutamate and an aspartate, with one carboxylate in a bidentate configuration and the other being monodentate, and assuming that the third histidine comes in and completes the coordination sphere (Figure 4.4C(b)). This additional histidine could come from a rearrangement of the metal sites with His86 interacting with site S2, rendering His88 free to interact with site S3. In EXAFS we see an average of the two sites, the models based on these two hypotheses and used to fit the experimental data are : "3 His 1.5 Glu_{bid}" and "3.5 His 1 Glu_{bid} 0.5 Glu_{mono}", see Table 4.19 in the annex.

The "3 His 1.5 Glu_{bid}" model provides a worse fit, R=8.8%, with respect to the "3.5 His 1 Glu_{bid}" model, R=7.3%. This suggests that when cobalt is bound in metal site S3,

a third histidine residue participates to the metal binding, in addition to the two revealed by the crystal structure, detailed later. This is consistent with the fact that XANES spectra are very similar for 1 eq. or 2 eq. of cobalt. Both sites would be octahedral, with identical residues except for the replacement of one histidine with one carboxylate monodentate.

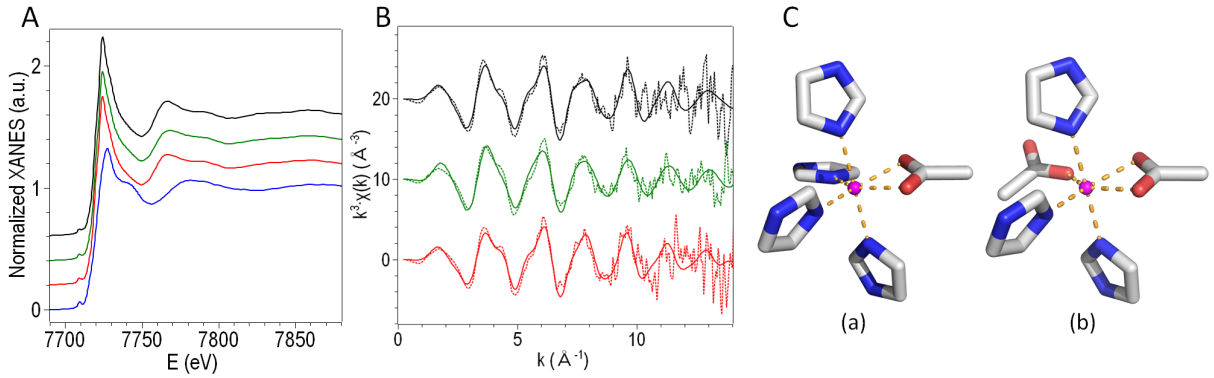


Figure 4.4: XAS results obtained on EcFur WT, curves were shifted for clarity. A: XANES spectra showing identical features for 1 eq. (black) or 2 eq. Co (green), suggesting that the metal binding environment is the same in both sites. XANES of EcFur WT + 2 eq. Co + pF2 (red) is indistinguishable from that of the protein in the absence of inhibitor (green). Spectral features differ between pF2 alone in the presence of Co (blue) and pF2 bound to the protein (red). B: EXAFS spectra of the same samples shown in A, dashed lines are experimental results, the relative best-fitting curves are presented as solid lines. C: 3D model of the best fitting sites (a) shows the site with 4 histidines and 1 glutamate bidentate whereas (b) shows 3 histidines 1 glutamate monodentate and one glutamate bidentate.

In order to understand whether the inhibition observed in presence of pF2 is due to an interaction between the peptide and the metal sites, the X-ray absorption spectra of EcFur WT loaded with 2 eq. Co (with both S2 and S3 occupied) and in presence of pF2 were measured. The XANES spectrum in this condition is indistinguishable from that of the metallated protein in the absence of the peptide (Figure 4.4A, red and green curves, respectively), suggesting that peptide anchoring to the protein does not modify the metal binding environment. The qualitative observation of the Co K-edge XANES spectra provides a proof of the lack of direct interaction between pF2 and the protein metal sites, considering that the spectral features differ dramatically between a pF2:Co 2:1 complex and in the case of pF2 anchoring to Co-EcFur (Figure 4.4, blue and red curves, respectively). The EXAFS spectrum of the metalloprotein-inhibitor complex can be fitted with the models used in absence of pF2, providing comparable fit quality and quantitative results (Figure 4.4B red line, Annex Table 4.19). We can therefore confirm that, regardless of the presence of pF2, 2 Co atoms in EcFur WT occupy two distinct sites S2 and S3, both octahedral, involving at least 3 His residues each, one Glu bidentate, and a fourth amino acid side chain. This proves that pF2 inhibition does not involve an interaction of this peptide with metal sites, validating previously obtained results.

4.5 Structure of Fur from *P. aeruginosa*

The structure of PaFur was the first Fur to be described in the literature by Phol (Pohl et al., 2003). Figure 4.5 A and B show the structure of the subunit of PaFur-WT in the PDB (PDB ID: 1MZB) and its description as a dimer. However, Pérard later showed that this protein is tetrameric in solution (Figure 4.5 C) (Pérard et al., 2016).

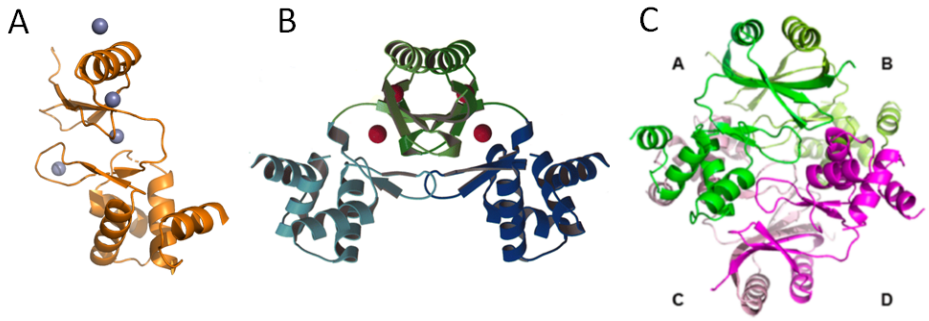


Figure 4.5: A: Structure of PaFur-WT (PDB ID: 1MZB (Pohl et al., 2003)) showing one Fur subunit and 4 Zn atoms. B: Structure of the dimer described by Pohl (Pohl et al., 2003) showing 2 Zn per subunit. C: Structure of the proposed tetramer from PISA analysis of Pohl's PDB data (Pérard et al., 2016).

The S3 metal site mutant of Fur from *P. aeruginosa*, H86A - H124A, named PaFurΔS3 was produced in large quantities for the X-ray absorption spectroscopy described earlier. After the experiment, the sample was recovered and used in crystallization trials. It is important to mention that this protein construction can be activated by the addition of metals. All crystallization trial were carried out through INEXT PID: 2217 - Inhibition of Fur from pathogenic bacteria: P. aeruginosa, F.tularensis and M. tuberculosis.

Crystals of Zn-PaFurΔS3 were obtained using the crystallization screens of Hampton Research Grid Screens™ and Qiagen protein crystallization suites at HTXLab high throughput robot screening facility (HTX Lab at EMBL-Grenoble). Using the harvester robot several crystals were tested in order to obtain the first diffraction patterns around 4 Å. After data analysis, the symmetry space group $P6_122$ was determined in an initial crystallization condition that was later manually optimized to obtain a final diffraction close to 2.2 Å (Figure 4.6).

Initial condition: Zn Acetate 0.2 M ; MES 0.1 M pH 6 ; PEG 8000 5%

Optimized condition: Na Acetate 0.2 M ; BTP 0.1 M pH 6 ; PEG 3350 17.5%

The Zn-PaFurΔS3 structure was solved by molecular replacement using Phenix 1.10.1-2155 with AutoSol and 95% of the model was built automatically. *Coot* was used for manual correction before a final refinement cycle in Phenix. The structure, shown in Figure 4.7B, in contrast to the the structure described by Pohl, consists of two Fur subunits belonging to different dimers.

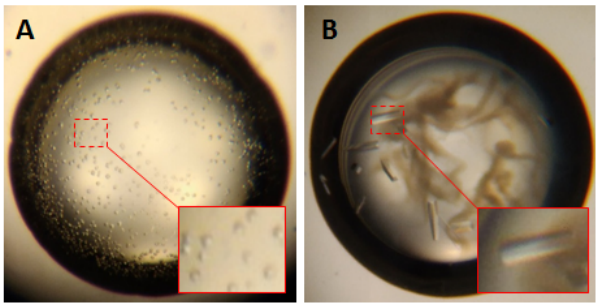


Figure 4.6: Crystals of PaFurΔS3 before (A) and after (B) optimization.

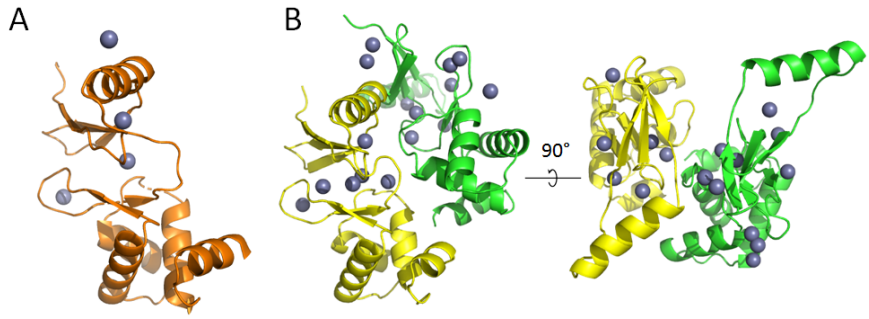


Figure 4.7: A: structure of PaFur-WT as obtained by Phol, showing one Fur subunit and 4 Zn atoms. B: structure of Zn-PaFurΔS3 obtained in this work, showing two Fur subunits, belonging to different dimers, and 16 Zn atoms.

Since the structure of PaFur-WT was already solved in the presence of Zn (PDB ID: 1MZB), trials to get the structural information in the presence of iron, its physiological cofactor, were initiated in anaerobic conditions. After several unsuccessful assays, iron was replaced with Mn^{2+} to conduct experiments in aerobic condition and at the same time conserve a chemical behaviour similar to iron. Following multiple tests, crystals of Mn-PaFurΔS3 were obtained in the crystallization condition of Mn-FtFur, described in Pérard et al., 2018, that contained: 50 mM MES pH 5.8, 20% w/v PEG 3350, 200 mM $MgCl_2(H_2O)_6$, 10 mM $MnCl_2$. This condition was then manually optimized and the resulting crystals diffracted at 2.6 Å. Before crystal collection, crystals were carefully washed to remove excess free Mn in a solution equivalent to their mother liquor but lacking the metal, before being cryoprotected using a solution obtained by adding 25% (v/v) glycerol to the mother liquor.

To validate the presence of metal in our crystals, X-ray fluorescence spectra were realized. For Zn-PaDS3fur crystals only a Zn signal was detected. In contrast, in the case of Mn-PaFurΔS3 crystals Zn and Mn signals were detected as expected. However, due to the physical properties of Zn and Mn their discrimination in electron density maps is not obvious. To validate this information, the metal content was measured in the purified proteins by ICP-AES for PaFur-WT and PaFurΔS3. Zn was only detected in PaFurΔS3 at 0.8 equivalent per subunit. The absence of Zn in PaFur-WT could be due to the purification protocol.

The structure of Mn-PaFurΔS3, like Zn-PaFurΔS3, was solved by molecular replacement using Phenix and manually corrected in *Coot*. To differentiate the localisation of each metal, data sets were collected at 0.99 Å to be able to detect both elements and at 1.29 Å to detect Mn solely (Figure 4.8). With this data collection method, an anomalous map was built where the discrimination between metal elements was performed. Mn atoms were placed in each peak superior to 5 σ in the anomalous map, this value was arbitrarily chosen and is data set dependent.

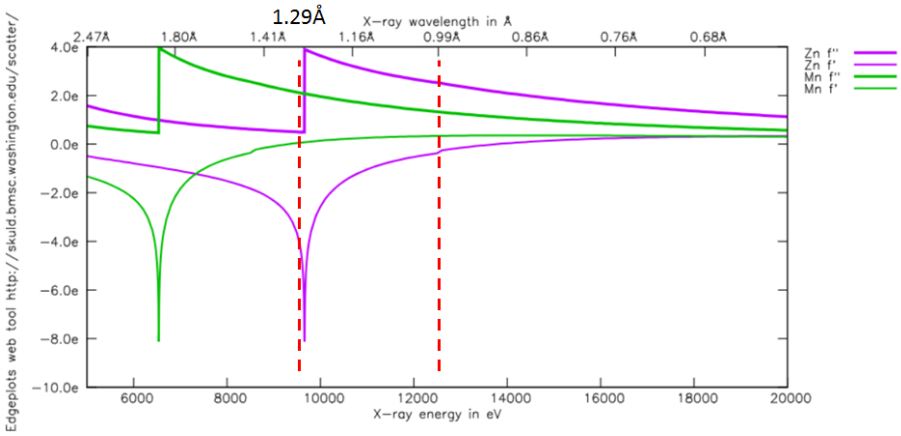


Figure 4.8: Mn (green) and Zn (magenta) edges difference. The red dashed lines show the wavelengths chosen for data collection (*X-ray Anomalous Scattering*).

Table 4.18, in the annex, shows data collection, phasing and refinement statistics generated by Julien Pérard using XDS/XSCALE and Phoenix for both Mn-PaFurΔS3 and Zn-PaFurΔS3 structures.

4.5.1 Conventional metal sites of Mn-PaFurΔS3 and Zn-PaFurΔS3

In PaFur-WT, the S2 site is formed by His32, Glu80, His89 and Glu100 (Figure 4.9A). This site is present, with the same ligands, in both of our structures Zn-PaFurΔS3 and Mn-PaFurΔS3 as shown in Figure 4.9B and C respectively. This site ligands are: His32, Glu80, His89 and Glu100.

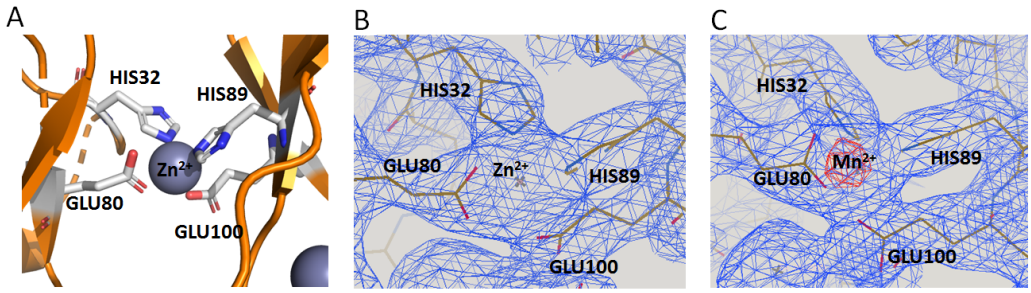


Figure 4.9: Comparing S2 sites chemical environment in three PaFur structures. A: S2 site of PaFur-WT (Pohl et al., 2003). B: electron density map of site S2 in Zn-PaFurΔS3. C: electron density map of site S2 in Mn-PaFurΔS3. All structures show four ligands in the S2 site.

The S3 site is formed by His86, Asp88, Glu107 and His124 in the wild type protein as shown in Figure 4.10A. However, since we crystallized a metal site mutant of PaFur using the PaFurΔS3(H86-A, H124A) construction, the residual S3 site in Zn-PaFurΔS3 and Mn-PaFurΔS3 can be seen in Figure 4.10B and C where only Asp88 and Glu107 can be detected since they were not mutated.

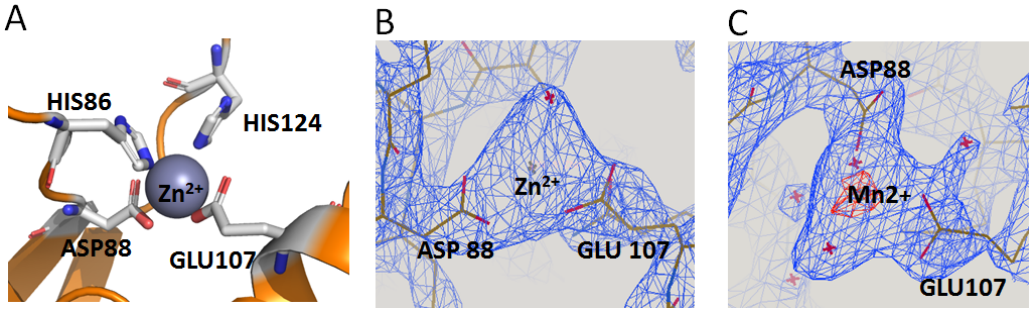


Figure 4.10: Comparing S3 site's chemical environment in three PaFur structures. A: PaFur-WT Pohl et al., 2003, showing a tetrahedral site with His86, Asp88, Glu107 and His124 as ligands. B: electron density map of residual site S3 in Zn-PaFurΔS3. C: electron density map of residual site S3 in Mn-PaFurΔS3. Due to the ΔS3 mutation, both the Zn or Mn loaded residual S3 sites only have Asp88 and Glu107 as ligand in addition to water molecules.

4.5.2 PaFurΔS3 tetramers and unconventional metal sites

Our biophysical characterisation of PaFur-WT and PaFurΔS3 indicates that they are tetramers in solution (Pérard et al., 2016), in order to study the physiological oligomeric state, the PISA web server was used to generate a PaFurΔS3 tetramer for both Zn and Mn structures since the tetramer was visible in the electron density map (Figure 4.11). PISA analyses the interfaces between macromolecules in their crystal environment allowing the prediction of their quaternary structure and calculations of the strength of contacts between them.

These structures show that some metal sites are also found in the wild type structure, and could possibly be crystallization artefacts linked to the protein packing, even after thouroughly washing the protein befor crystallization. However, in addition to the conventional S2 and S3 metal sites, the teramers show that some metal sites can form between dimers and are probably linked to tetramer stability. In fact, *in vitro* experiments show that even after 0.1M EDTA treatment, PaFur-WT is still tetrameric in solution (Pérard et al., 2016), indicating that some of these metal sites might not be crystallization artefacts and could be involved in tetramer stabilization. Results are still being analysed to understand the physiological importance of this high number of metal atoms present in PaFur structures.

Moreover, in the Mn-PaFurΔS3 structure, Mn^{2+} replaces some of the Zn^{2+} atoms found in the Zn-PaFurΔS3 structure, in addition to populating several other sites. This

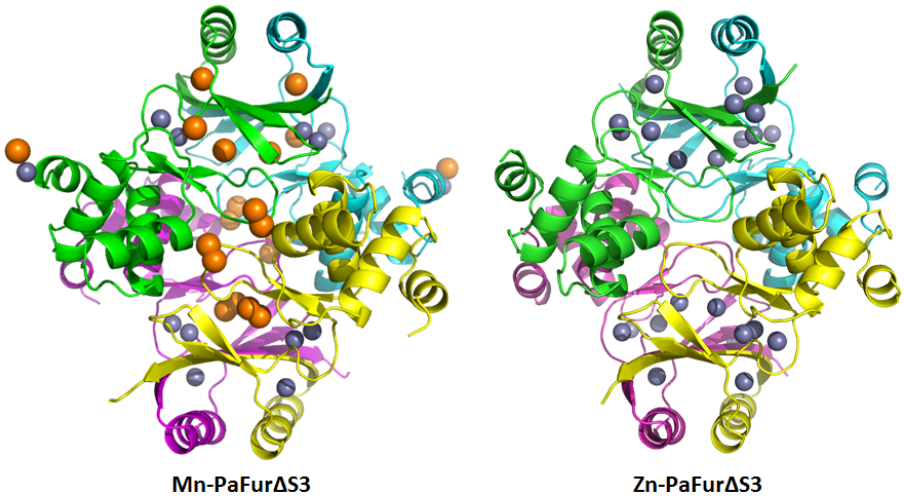


Figure 4.11: PaFurΔS3 tetramers generated by PISA, Zn is represented in grey and Mn in orange.

indicates that probably, a similar behaviour could be expected for Fe^{2+} when bound to PaFur *in vivo*.

Taken together, these observation about unconventional metal sites may indicate a previous role, during their evolutionary path, of Fur protein ancestors as iron storage proteins buffering the cytosol metal content.

4.5.3 Crystallization trials with molecule B

Since we resolved the structure of PaFurΔS3 in the following condition: Zn Acetate 0.2 M ; MES 0.1 M pH 6 ; PEG 8000 5%, crystallization trials were initiated in order to obtain a structure of PaFurΔS3 in the presence of molecule B, details on crystallization protocols can be found in 6.2.

The initial strategy was to use already obtained PaFurΔS3 crystals in soaking experiments (Figure 4.12). The first trials were not successful with the majority of crystals dissolving in their mother liquor when molecule B was added. The few crystals that survived the initial soaking step did not diffract when exposed to X-ray radiation.

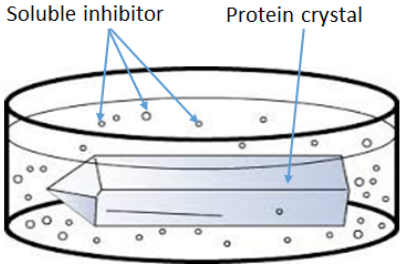


Figure 4.12: Crystal soaking experiment where a single protein crystal is placed in its mother liquor to which the inhibitor was added.

Following this first series of soaking attempts, a new batch of crystals was produced

through a manual optimization of the initial crystallization condition by varying the concentration of Zn acetate (between 0.1 M and 0.3 M) in 3 different conditions of PEG 8000 (5±2.5%) and fixing the pH at 6 or by varying the PEG 8000 concentration (between 0 and 10%) in three pH conditions (pH 6±0.5) and fixing the Zn acetate concentration at 0.2 M. The resulting crystals were used for a new series of soaking experiments that gave similar results to the first one. In addition, identical plates were incubated at 4°C for two month but showed no sign of crystal growth.

As a last attempt, a small amount of molecule B powder was sprinkled on several fresh crystallization drops containing PaFurΔS3 crystals and left to soak overnight. Remarkably, the crystal survived this treatment, they were fished and carefully washed then cryoprotected before freezing every half an hour for the first three hours of soaking and on the next morning. However, the diffraction quality were not good enough.

It is important to note that DMSO, in which the molecule B is solubilized, does not dissolve the crystals in control drops. In addition, molecule B soaked in the mother liquor of PaFurΔS3 does not crystallize on its own.

After the failure of various soaking protocols in yielding good crystals for structural analysis, co-crystallization was used to obtain crystals of PaFurΔS3 in complex with molecule B. Co-crystallization takes place in two steps, first a soluble solution of the protein is mixed with a soluble solution of molecule B, then the crystallisation solution is added to the soluble mix.

By providing a PaFurΔS3 protein sample mixed with molecule B to the HTXLab platform (EMBL), standard crystallization screens were performed and 11 hits were detected from 576 tested conditions. Crystals obtained in these conditions had poor diffraction quality mainly because of their small size (Figure 4.13A).

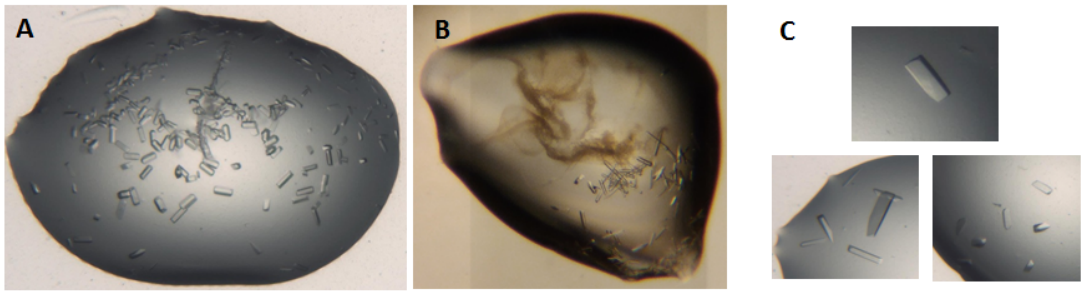


Figure 4.13: Co-crystallization of PaFurΔS3 with molecule B (F06 well of the PACT_MD screen). A: Crystals after the standard crystallization screens. B: Crystals after manual optimization. C: Crystals after the automated additive screen with 5.6 Å diffraction. A and C are 0.2 µL drops, whereas B is a 2 µL drop.

Of these 11 conditions, 3 were already known to crystallize the protein alone. The remaining 8 conditions shown in the annex Table 4.16 were chosen to be used in a manual optimization procedure where plates with concentration gradients of each component helped determine an optimized crystallization condition. However, the diffraction quality

of crystals was still poor (Figure4.13B).

Qualitative analysis of crystals obtained after this manual optimization lead to the use of a specific condition, shown below, in all the following crystallization assays with little success in improving diffraction patterns.

Initial condition: Zn Acetate 0.2 M ; MES 0.1 M pH 6 ; PEG 8000 5%

Optimized condition: Na Acetate 0.2 M ; BTP 0.1 M pH 6 ; PEG 3350 17.5%

To resolve this problem, the HTXLab additive screen : Additive Screen™ HR2-428 from Hampton Research was used. This type of screening uses a variety of molecules from salts to dissociating agents and organic molecules, and can reveal conditions where the crystals are bigger and of better quality. In our case, results analysis revealed 20 new conditions, shown in the annex in Table 4.17, with crystals diffracting down to 5.6 Å(Figure 4.13C). After several trials of manual optimization with the 20 selected additives, crystal quality was enhanced and diffraction patterns reached 3.9 Å.

Following these co-crystallization trials, the main result that shows up is a change in crystal space group, from P6 with the typical PaFurΔS3 bipyramidal crystals of the protein alone (Figure 4.6A) to an elongated P2 space group in the presence of molecule B.

The question that remains to be answered is the contribution of molecule B to this matrix change, since the protein alone does crystallize in the optimized conditions (Figure 4.6B). Molecular replacement in the electron density map of PaFurΔS3 molecule B complex using parts of the already solved structure of PaFurΔS3 did not help resolve the collected diffraction data. Leaving all the crystal structures detailed here unsolved.

Protein complex	T_m (°C)
PaFurΔS3	66
PaFurΔS3 + pL1	66
PaFurΔS3 + pL2	66
PaFurΔS3 + molecule B	69

Table 4.6: Melting temperatures of PaFurΔS3 with or without inhibitors

Together with inducing a potential change in the crystal matrix, the addition of molecule B increased the thermal stability of the protein, as shown in Table 4.6. This effect is unlikely due to the presence of DMSO, used to solubilize molecule B, since there is no increase of the melting temperature in the cases of pL1 and pL2 both also dissolved in DMSO. So when molecule B is added, it is definitely interacting with the protein, its specificity remains to be determined, but the increase in melting temperature and the space group change indicate that the protein is not in its normal state in the presence of molecule B.

4.5.4 Co-crystallization trials with pL1 and pL2

In parallel to docking and molecular dynamics simulations, crystallization trials were carried out on PaFur in the presence of pL1 and pL2 to validate our models and to gain insight into their inhibition mechanisms. Experiments were performed at the same time as [PaFur + molecule B] crystallization trials, and used the same techniques described earlier (soaking and co-crystallization).

The first hits were obtained during an automated crystallization conditions screen at HTX Lab (Figure 4.14) where the protein crystallized in the presence of pL2. Manual trials in the same conditions did not yield any crystals even after optimization, puting this project on hold.

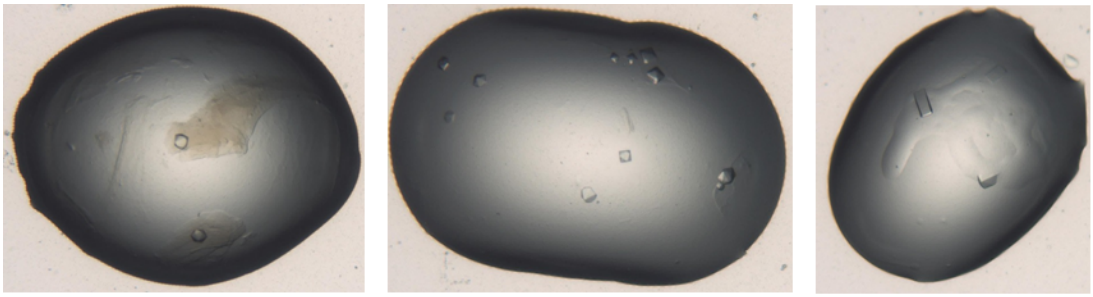


Figure 4.14: Co-crystallization of PaFurΔS3 in the presence of 1.5 mM of pL2. Crystal diffraction was not sufficient to resolve the structure. Conditions are A: Wizard_I+II_rigaku (A07): PEG 8000 10% w/v; MES 0.1 M; Zn acetate 0.2 M. B: Wizard_I+II_rigaku (F08): Reagent Alcohol 15 % v/v; MES 0.1 M; Zn acetate 0.2 M. C: PACT_MD (F07): Na acetate trihydrate 0.2 M; BTP 0.1 M pH 6.5; PEG 3350 20 % w/v (Annex Table 4.16).

4.6 Structure of Fur from *E.coli*

For the past 20 years, several researchers and PhD students from our team tried to crystallize Fur from *E.coli* without success. The last series of automated experiments dates back to 2013 and were carried out by Sandra Galop using EMBL's infrastructures and by Cheickna Cisse as part of the AURORA project in collaboration with "The Norwegian Structural Biology Centre - Universitetet i Tromsø".

Previous ICP-MS studies of metal activated Fur from *E. coli* showed 0.5-0.8 atoms of Zn per monomer. Indicating the presence of one Zn metal binding site in each subunit (Michaud-Soret et al., 1997). Later, Althaus showed the presence of two metal sites with different affinities towards Zn (Althaus et al., 1999).

Previously, in our laboratory structural studies were carried out using NMR and X-ray crystallography on the N-terminal domain of Fur from *E. coli* (Pecqueur et al., 2006) resolving it's structure between residues 1-82 (PDB ID: 2FU4), showing a well structured helix-turn-helix DNA binding domain. This indicated that EcFur should have the same

global structure of PaFur solved by Phol at that time. However, no information could be retrieved on the metal sites nor the docking models for inhibitors could be confirmed. To better understand how EcFur is structured and how our inhibitors work, crystallization trials on this protein were carried out.

4.6.1 EcFurΔS3 crystallization trials

In the absence of a structure of the wild type Fur from *E. coli*, and after the success with PaFurΔS3, the EcFurΔS3 (H87A, H125A) production used for XAS experiments was tested in crystallization assays in the same condition as PaFurΔS3 with no success, even after manual optimization of the condition.

Using a new production batch, an EcFurΔS3 sample was sent to HTXLab for a crystallization condition screen. From the 576 conditions, around 20 showed some possible positive results by having micro-crystals or sea urchins (Figure 4.15A). All these conditions had PEG 3350, sodium salts and a pH between 6.5 and 8.5. The same screen was carried on the EcFurΔS3 production used for XAS experiments as well as the first manual crystallization trials without any positive results, indicating that the observed results could be batch specific.

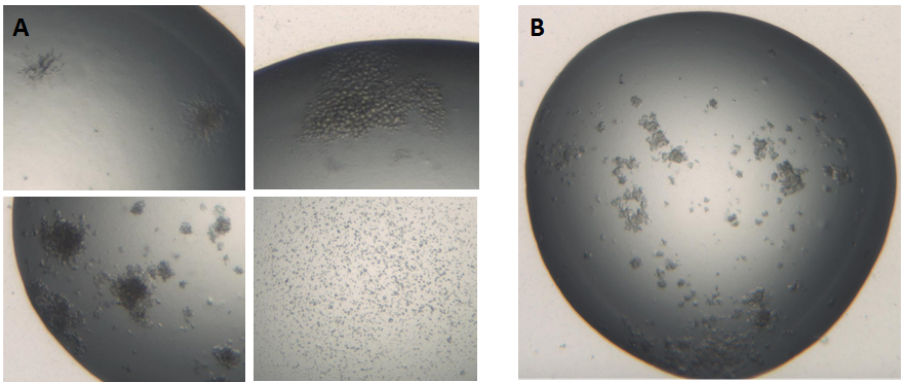


Figure 4.15: EcFurΔS3 crystallization drops after a crystallization condition screen. A: possible positive hits with sea urchins and micro-crystals. B: H07 condition from the PACT_MD plate considered as the main hit from this screen.

Following the first automated crystallization conditions screen, the H07 condition from the PACT_MD plate was chosen as the best hit (Figure 4.15B) with crystals appearing 19 days after the start of the screen. This condition was used for further manual optimization and in additive screening. It contains the following components:

sodium acetate trihydrate 0.2 M ; BTP 0.1 M pH 8.5 ; PEG 3350 20 % w/v

On the same additive screen assay, in addition to EcFurΔS3, a sample of EcFurΔS2 and EcFur wild type were sent to check whether they can crystallize in this condition. For all three samples, no hits were detected.

As a last attempt, a new production of EcFurΔS3 was done where the protein poly-histidine tag was not removed at the end of the purification, as was the case in the previous production batches. The tagged sample was sent for an automated crystallization condition screen where no hits were detected.

4.6.2 EcFur-140 crystallization

In order to facilitate the crystallization, the deletion of the floppy terminal part seemed a good strategy. In addition, former studies showed that C-terminal residues of EcFur were not essential for activity (D'Autreaux et al., 2002). After the EcFurΔS3 trials, it was decided to test the EcFurΔC-terminal(140-148) construction for crystallization. EcFurΔC-terminal(140-148) lacks the last eight amino acids of the wild type protein, to simplify the annotation, it will be designated EcFur-140.

In the first series of tests, it was interesting to see that EcFur-140 had a higher melting temperature than EcFurΔS3 or the EcFur monomer (Table 4.7) indicating a higher stability in solution probably due to the lack of the flexible C-terminal residues. Considering the time left before the end of my thesis, this information was crucial in shifting our focus from EcFurΔS3 to EcFur-140 in trying to obtain a structure of Fur from *E. coli*.

Protein complex	T_m (°C)
EcFur monomer	43
EcFurΔS3	59
EcFur-140	65

Table 4.7: Melting temperatures of EcFur constructions.

Previously, to investigate the role of the C-terminal region in EcFur, several C-terminal deletion constructions were produced, purified and characterised helping determine a sub-domain crucial for dimerization, located between residues 125-140. Due to their solubility and activity issues only the ones shown in Table 4.8 were tested for crystallization without success.

Other than highlighting the increase in melting temperature, the first automated crystallization screen of EcFur-140 did not reveal any positive hits. Knowing that in this particular test, the protein concentration was relatively low (4 $mg.mL^{-1}$), the experiment was repeated with a more concentrated sample (10 $mg.mL^{-1}$), at the same time the poly-histidine tag used for purification was kept. In the following text, EcFur-140 will always indicate a poly-histidine tagged protein (EcFurΔC-terminal(140-148)-polyhistidine tag) unless indicated otherwise. From the 576 conditions tested, only one drop contained a crystal (Figure 4.16A) in the following condition:

Mg nitrate 0.2 M ; PEG 3350 20% w/v

EcFur construction	C-terminal domain	Oligomeric state
EcFur wild type	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	dimer
EcFur-125	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	monomer
EcFur-131	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	?
EcFur-132	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	?
EcFur-133	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	?
EcFur-136	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	?
EcFur-140	...RLTNHSLYLYGHCAEGDCREDEHAHEGK	dimer

Table 4.8: Soluble constructions of EcFur used in crystallization assays, grey amino acids are not present in the construction. Oligomeric states were determined by SEC-MALLS-RI. However, some constructions were not stable enough in solution to allow a correct characterisation. All experiments were done after cleaving the polu-histidine tag.

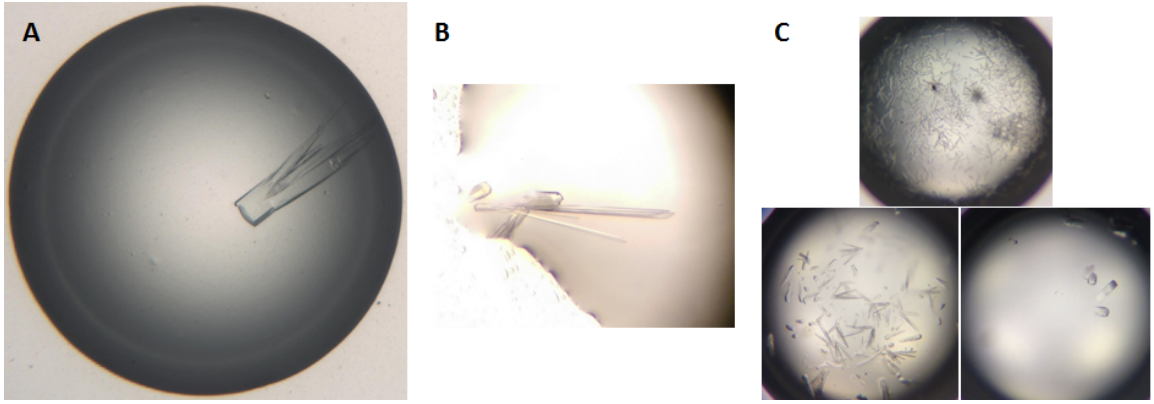


Figure 4.16: A: the first crystal of EcFur-140, obtained in a 0.2 µL drop at the HTXLab automated screen (F04 well of the PEGs-I_qiagen), details of this condition can be found in the text. B: successfully obtaining EcFur-140 crystals manually. C: seeding experiments using different seed concentrations, the lower the seed concentration, the bigger the crystals.

The crystal was fished manually and tested for diffraction to verify its nature. Due to its small size and polymorphic aspect, the diffraction pattern was bad but good enough to verify that it was a protein crystal and not a salt crystal. The next step was to reproduce manually this crystal in the laboratory.

For some reason, Mg nitrate was out of stock at all French chemical suppliers and the purchase of this individual reagents from Hampton Research or its stock solution from Qiagen was not possible since it was no longer produced. To be able to recreate the crystallization condition and start the optimization process the only choice was to produce our own Mg nitrate using the chemical reaction shown below:

$$2\,KNO_3 + MgCl_2 \longrightarrow Mg(NO_3)_2 + 2\,KCl$$

Crystallization screens were carried out using gradients of 5 different PEG polymer sizes (3350, 5000, 6000, 8000 and 10000 KDa) and 3 pH conditions (6.5, 7, 7.5) for 10

mg.mL^{-1} and 16.2 mg.mL^{-1} protein concentrations. These screens did not recreate the crystallization condition and no crystals were obtained. This is mainly due to the poor quality of the Mg nitrate and the excess KCl produced during the chemical reaction.

The same screen was repeated using the left over buffer, approximately 45 µL, from the F04 well in the PEGs-I_qiagen screening plate where the first crystal was fished. A white precipitate was observed after the addition of the buffer on the protein drops in all conditions. Remarkably, when using the 16.2 mg.mL^{-1} protein solution, crystals were obtained in only one drop with the following condition, and are shown in Figure 4.16B:

Mg nitrate 0.2 M ; PEG 3350 20% w/v ; protein:buffer ratio 1:1.5

With our recovered 45 µL of Mg nitrate running low, crystal production and optimization was carried out using crystal seeding. In this technique, crystals are broken down to microscopic size and then added to a fresh crystallization drop where they initiate nucleation and crystal formation. The crystals shown in Figure 4.16B were broken down and their seeds added to a fresh drop of protein. Depending on the seed concentration, crystals were obtained in relatively large numbers (Figure 4.16C) and the biggest crystals were obtained when using the lowest seed concentration. The resulting crystals were exposed to X-ray radiation and diffracted between 10 Å and 2.5 Å revealing a P6(3) space group.

Fluorescence spectroscopy showed the presence of Zn and Ni in the crystals (Figure 4.17). Zn atoms were expected since they are bound to the S1 metal site, however the presence of Ni could be linked to the purification protocol where Ni-NTA affinity columns were used.

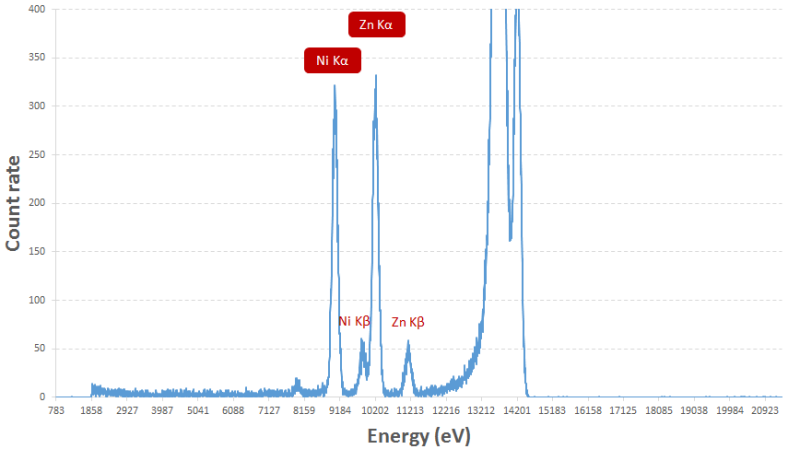


Figure 4.17: X-ray fluorescence spectrum of an EcFur-140 crystal showing the presence of Ni and Zn.

Metal content determinantion by ICP-AES

The metal composition of EcFur-WT and EcFur-140 were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES) (table 4.9). EcFur-WT with

the polyhistidine tag has 1 eq. of Zn per monomer and no Ni, however in the case of EcFur-140 with a polyhistidine tag, 0.6 eq. of Ni is detected and a similar amount of Zn indicating that this construction binds equal amounts of the two elements. These results are in accordance with X-ray fluorescence spectra collected for EcFur-140 crystals that showed the presence of Zn and Ni (Figure 4.17).

As a control, EcFur-WT with a polyhistidine tag was tested and revealed similar Zn results to EcFur-WT without a polyhistidine tag, in addition to similar Fe results to EcFur-140 with a polyhistidine tag that could be due to the presence of the tag. The presence of Fe and Zn is due to bacterial culture conditions, Ni however comes from Ni-NTA chromatography.

EcFur construction	Fe	Zn	Ni
EcFur-WT	0	1	0
EcFur-WT + polyhistidine tag	0.2	1.1	0
EcFur-140 + polyhistidine tag	0.2	0.6	0.6

Table 4.9: Metal composition per monomer of different EcFur construction obtained by ICP-AES.

Shortly after our first diffraction results, a crystallization screening kit with 10 mL of Mg nitrate solution was kindly provided by Christine Cavazza, our crystallographer collaborator. Crystallization optimization and additive screening were performed and identified 6 additives that enabled crystallization (Figure 4.18). Crystals obtained from the additive screen were fished and the best diffraction was observed for the crystal that used 0.1 M of phenol as additive.

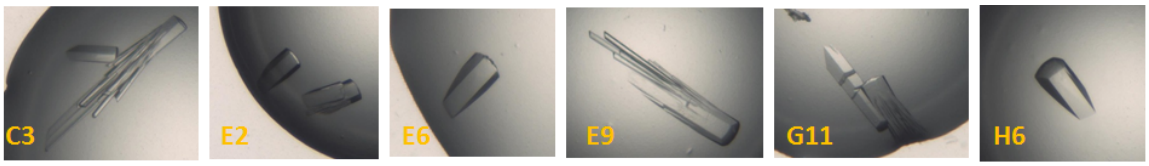


Figure 4.18: EcFur-140 crystals in the additive screen plate. The letters and numbers indicate the well and the additive used. C3: 0.1 M phenol. E2: 5% w/v polyvinylpyrrolidone K15. E6: 30% w/v D-(+)-glucose monohydrate. E9: 30% w/v D-sorbitol. G11 30% v/v 2-propanol. H6 40% v/v 1-propanol.

Subsequently, manual optimization of EcFur-140 crystals continued by using phenol as additive and crystal production was relatively easy. The optimization focused on obtaining homomorphic crystals since mosaic diffraction patterns were always observed for polymorphic crystals.

Actually, phenol was used in an attempt to minimize the appearance of polymorphic crystals. It was thought that the phenol solution would slow down the crystallization speed allowing homomorphic crystals to grow. In fact, the white precipitate that forms directly when adding the crystallization buffer to the protein drop was still present when

using phenol as an additive. This precipitate would on average disappear after five to ten minutes with the appearance of crystals. In an attempt to slow the crystallization process, several plates were placed at 4 °C but this halted the crystallization and no crystals were obtained by this method. Another attempt to slow down the crystallization was the use of mineral oil that would float on the crystallization buffer in the reservoir and decrease its surface slowing the rate of vapour diffusion between the crystallization drop and the buffer, giving crystals more time to grow. Different mineral oil volumes were used, none of them reduced the appearance of polymorphic crystals and all the drops showed the usual pattern of white precipitation followed by crystal formation.

With our current knowledge, homomorphic crystal formation is probably due to slight variations in reagent concentrations that are not under our control. The only solution is to prepare enough crystallization drops to be statistically sure that homomorphic crystal will be formed. Usually, a full crystallization plate (15 wells) is enough. As an example, Figure 4.19A shows how the same crystallization condition can give various crystal types. Indeed, the structure of Fur from *E. coli* was solved using a crystal from this production batch in the following crystallization buffer:

Mg nitrate 0.2 M ; PEG 3350 17-19% w/v ; phenol 0.1 M ; protein:buffer ratio 2:2

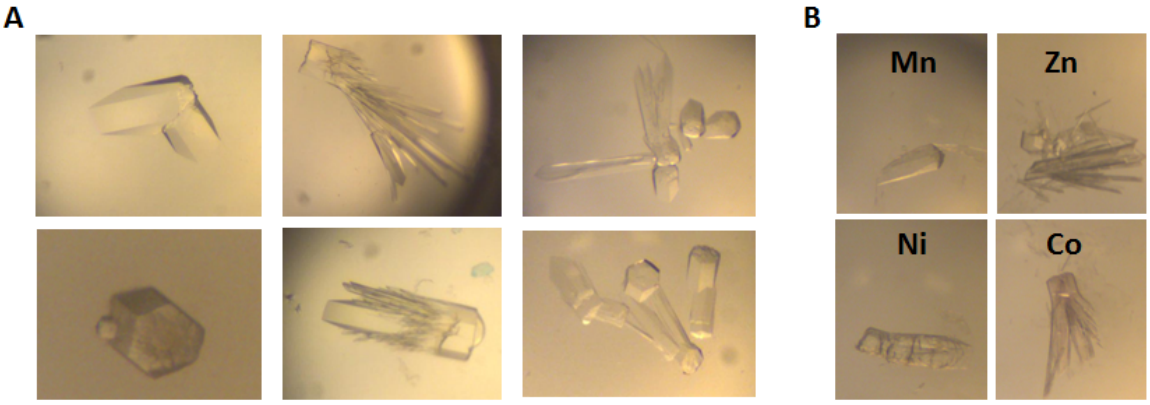


Figure 4.19: A: EcFur-140 crystals used in the final diffraction series, even when using fresh reagents, the polymorphic forms can not be avoided. The shown crystals were all obtained in different drops of the same crystallization condition. B: Soaking of EcFur-140 crystals in metal solution in preparation for SAD experiments. The crystals underwent major modification after soaking that decreased their diffraction quality. The Ni soaked crystal has visible cracks and in the case of cobalt the previously transparent crystal acquired a pink hue

Before continuing to other results related to the structure of EcFur-140, and to avoid any future investment in failed research, it is worth mentioning in this paragraph other experiments that did not yield crystals. The EcFur-140 construction that crystallized did so in its native form, without any metal treatment. We tried to crystallize a manganese loaded EcFur-140 without success in the same additive screen that gave 6 hits for the non metalated protein (Figure 4.18). In addition, several crystal soaking experiments

were carried out using different metals (Mn, Zn, Ni and Co) with varying incubation time and metal solution concentration. After this treatment, crystals were fragile with visible cracks appearing after soaking (see Ni example in Figure 4.19B). In other cases, crystals were intact but their diffraction quality decreased dramatically. Another crystallization experiment took place inside a glove box and involved loading the protein with iron. After one day, all the drops contained a dark precipitate. In addition, several assays of standard screening and additive screening were carried out on EcFur-140 in the presence of the pF2 inhibitor. Unfortunately, the addition of the inhibitor prevented crystallization in manual and automated screens. Moreover, crystallization trial of EcFur-140 in the presence of a DNA Fur box presented the same aspect of white precipitation when the crystallization buffer is added to the [protein + DNA] drop. However this precipitation never gave way to crystals, as in the case of EcFur-140 alone, even after several weeks of incubation at room temperature. Finally, an important aspect of the crystallization of EcFur-140 was its concentration and batch specificity. All the crystals presented here were obtained from the same production and all of them used a 16.2 $mg.mL^{-1}$ protein solution, except for the first crystal obtained in a 10 $mg.mL^{-1}$ solution (Figure 4.16). A higher or a lower concentration did not enable crystallization. None of the 4, 8, 10, 13, 15, 20 $mg.mL^{-1}$ solutions tested gave positive results. In addition, four independent EcFur-140 protein productions failed to crystallize even at 16.2 $mg.mL^{-1}$.

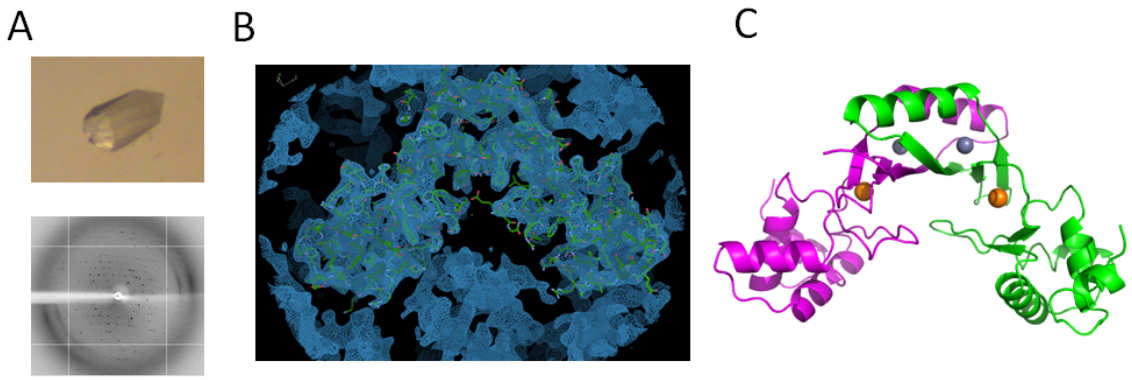


Figure 4.20: A: The EcFur-140 crystal that enabled protein structure resolution and its diffraction cliché. B: Electron density map visualized using *Coot* (Emsley et al., 2010). C: Cartoon representation of EcFur-140, Zn and Ni atoms are shown in grey and orange, respectively.

The structure of EcFur-140 was solved by Phenix Adams et al., 2010 with molecular replacement of VcFur (PDB ID: 2W57) structure and manual refinements (Figure 4.20). The frist hundred residues were built automatically and the reùainig part of the protein was done manually. It presents a dimer in the asymmetric unit (Chain A: residues 2-134 and Chain B: residues 2-132) with a resolution of 2.3 Å (Figure 4.20B and C). The global architecture of a Fur protein is conserved with a DNA binding domain (residues 2-82) and a dimerization domain (residues 83-134) however, the first amino acid and the last six terminal residues are not visible in the electron density map nor is the poly-histidine tag.

Previous NMR and the X-ray structure of the N-terminal domain of Fur from *E. coli* (PDB ID: 2FU4 Pecqueur et al., 2006) easily merged into our structure with an RMSD of 1 Å. In addition, PISA analysis of the protein interface shows 19 hydrogen bounds and three salt bridges. The dimer interface represents 2800 $Å^2$ in buried interface with 15.6 $kcal.mol^{-1}$ as a dissociation parameter. These data indicate the presence of a stable dimer, validated by further biochemical and biophysical characterisation described in the following paragraphs.

4.6.3 Metal site S1 in the EcFur-140 structure

Analysis of metal sites present in this structure showed that for the S1 site, usually involving Cys93, Cys96, Cys132 and Cys134, a disulphide bridge is observed between Cys93 and Cys96 in the electron density of monomer A (Figure 4.21). A similar case is observed in Fur from *V. cholerae* where the S1 site contains a disulphide bridge between Cys93 and Cys133 indicating the presence of alternative structural organisation of this protein domain.

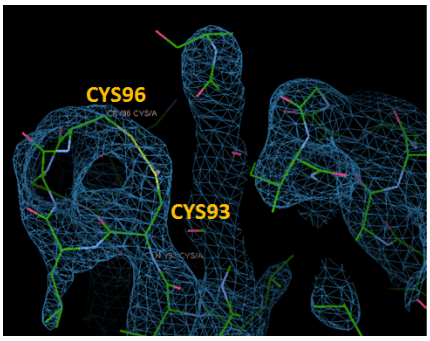


Figure 4.21: Electron density map of EcFur-140 chain A with a visible disulphide bridge between Cys96 and Cys93.

In our case, EcFur-140 structure does not show a metal coordination by Cys93 and Cys96, in addition Cys132 and Cys134 could not be located in the electron density map due to the poor resolution of this region. Previous studies showed that a structural S1 site can be formed by two disulphide bridges C92-C95 and C132-C137 in an oxidized monomeric protein (D'Autréaux et al., 2007).

EcFur-140 protein treatment with TCEP in the presence of Zn^{2+} completely inhibits crystallization of EcFur-140, preventing the study of both reduced and metal loaded protein samples. Moreover, as described earlier, soaking and co-crystallization experiments with various metals, under aerobic or anaerobic condition did not yield diffracting crystals. This indicates that only a native oxidized EcFur-140 is capable of forming stable crystals, the addition of Zn or TCEP destabilize the crystal packing and prevents further studies. However, in solution, the addition of a reducing agent like TCEP leads to the reduction of the cysteines that will coordinate Zn.

To validate this hypothesis, the number of free cysteines was measured by a DTNB assay for EcFur-WT and EcFur-140 samples (Table 4.10). This assay uses the Ellman's Reagent that reacts with available thiols in the protein giving a yellow TNB^{2-} quantified by measuring the absorbance at 412 nm.

Construction	Free cysteine per monomer
EcFur-WT-M	3.8
EcFur-140-M	3.9
EcFur-WT-D	0.7
EcFur-140-D	0.6

Table 4.10: DTNB assay results showing the relative number of free cysteines per monomer for samples of EcFur-WT and EcFur-140. M indicates monomeric species and D dimeric species obtained after adding Zn^{2+}.

EcFur-WT and EcFur-140 monomers have respectively 3.8 and 3.9 free cysteines per monomer after an overnight treatment with 1 mM TCEP at 4°C. These data are in good agreement with our results and confirm the presence of 4 cysteines in the reduced state of the protein. However in the case of dimers, 0.7 and 0.6 free cysteines per monomer are detected respectively for EcFur-WT and EcFur-140. This indicates that less than one cysteine is free with the three others coordinating Zn or bound in a disulphide bridge. Finally as a validation, when monomers are reduced with TCEP prior to dimerization through the addition of Zn similar free cycteines results are observed with 0.4 for EcFur-WT and 0.3 for EcFur-140.

The results showing the absence of a metallic ion in the S1 site of EcFur-140 structure together with thiol reactivity experiments suggest that this protein can be dimeric in a disulphide bridge is formed between His93 and His96 instead if having a Zn atom.

4.6.4 Unconventional S2 site in the EcFur-140 structure

The structure of EcFur-140 also contains the S2 regulatory metal site, described as essential and present in all active Fur structures (Pérard et al., 2016). A conventional S2 site is formed by 3 histidines and 2 glutamates. However, in our EcFur-140 structure, only one glutamate is bound to the metal and the site contains a fourth histidine. The described site binds Ni^{2+} and is composed of His33, His71, His88, Glu81 and His90 (Figure 4.22). Interestingly, the site adopts an octahedral geometry with 4 nitrogen atoms, from histidines, and 2 oxygen atoms in the coordination sphere, one from the glutamate and one from a water molecule. This S2 site still connects the DNA binding domain, through His33, His 71 and Glu81, to the dimerization domain, through His88 and His90.

Table 4.11 compares the S2 sites of Fur proteins from different species. The flexibility of measured distances can be due to difference in metal atoms since an hexa-coordinated

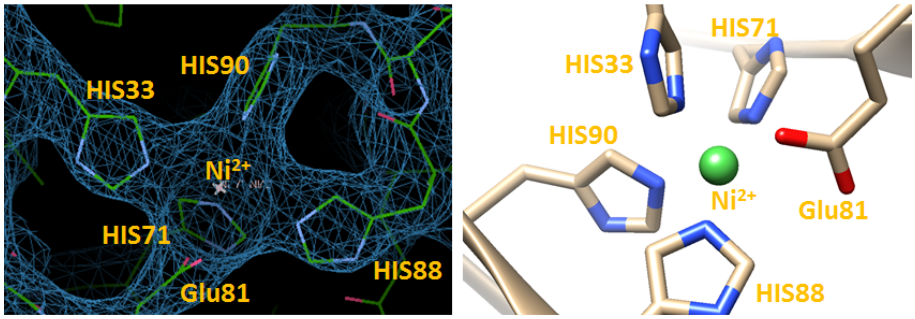


Figure 4.22: Electron density map metal site of S2 in EcFur-140 and its visual representation.

octahedral environment is favoured *in vivo* when Fur is bound to iron (Pérard et al., 2018). In the case of EcFur-140, Glu101 classically found in S2 sites of MgFur and FtFur is way too far and can not correctly coordinate the metal (way too far, insert correct distance). His71 could have replace Glu101 in the structure described here.

Amino acid	Distance from metal atom in Fur protein (Å)			
	EcFur-140	(Mn) MgFur	(Fe) FtFur	(Zn) PaFur*
His33	2.0	2.29	2.26	2.00
Glu81 (EO1)	2.20	2.22	2.19	2.04
Glu81 (EO2)	2.30 (H_2O)	2.35	2.19	-
His88	2.20	2.11	2.29	-
His90	2.05	2.08	2.23	2.04
Glu101	-	2.71	2.35	2.16
H71	2.20	-	-	-

Table 4.11: Distances between metal atoms and their ligands in Fur structures from different species. Metal atoms are not the same in all the structures, they are Ni for EcFur-140, Mn for MgFur, Fe or Mn for FtFur and Zn for PaFur. In the case of histidines, the distance shown is between N_ε atoms and the metal. * PaFur residues are shifted by **-1**, and for simplicity are shown here : His32, Glu81, His89 and Glu100.

It is interesting to note that His71 is present in other Fur proteins from *Y. pestis*, *V. cholerae* *P. aeruginosa* *M. gryphiswaldense* and *L. pneumophila* but not in Fur from *H. pylori* and *F. tularensis*. In PaFur for example, the histidine corresponding to His71 is located at 5 Å from the zinc ion in S2 and is clearly not involved in metal coordination in the structure but could be in the case of Fe^{2+} binding in *in vivo*. A similar case is found for VcFur with the histidine corresponding to His71 found at 4.1 Å from the metal. These results describe for EcFur-140 an S2 site with a core composed by His33, His88, His90 and Glu81. The metal binding flexibility of Fur proteins can explain the implication of His71 in metal site S2 in the structure of EcFur140. It could also be a way of metal selectivity for this regulatory site. In addition, due to its exposure to the solvent, His71 could potentially interact with metal ions and bring them closer to the S2 site triggering a conformational change in the hinge region.

This is the first description of such S2 site in Fur proteins, similar results could be expected for other species if the structures are resolved with the appropriate metal.

4.6.5 Classical accessory S3 site in the EcFur-140 structure

Finally, as predicted but never demonstrated, the structure of EcFur-140 also contains the S3 metal binding site, shown in Figure 4.23. The S3 site is described as an accessory site able to tune the affinity of the protein towards DNA. It is solved in most Fur structures bound to Zn^{2+} with four ligands His86, Asp88, Glu107 and His124 in the case of PaFur or H96, Asp98, Glu117 H134 in HpFur. In the case of our EcFur-140 data set, the electron density map is not sufficiently refined and work is still in progress. However, we can propose the following model where in the structure of EcFur-140, the S3 site binds Zn through the same corresponding residues: His87, Asp89, Glu108 and His125 in a tetrahedral conformation.

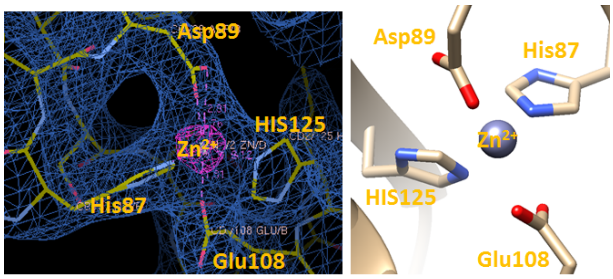


Figure 4.23: Electron density map of metal site S3 in EcFur-140 and its prompsed model, showing a tetrahedral geometry.

4.7 Biochemical characterization of EcFur-140

4.7.1 Activity tests

Nuclease protection assays were used to test the ability of the EcFur-140 protein to specifically bind the 19 bp Fur box DNA sequence, as described in a previous work (Mathieu et al., 2016). Figure 4.24 shows how EcFur-140, like EcFur-WT, is able to bind DNA in the presence of divalent metal elements like Mn, Co or Ni as previously described (Lorenzo et al., 1987; Bagg et al., 1987; Jacquamet et al., 1998; Mills et al., 2005). All proteins were prepared with 4 equivalent of metal, EcFur-WT fails to activate with Ni at this concentration, it needs to be incubated with 8 equivalents to become active.

The EcFur-WT polyhistidine tagged construction, was used to verify that the addition of the tag did not change the way the protein behave in solution, the results show that the tag has no effect on the proteins activity.

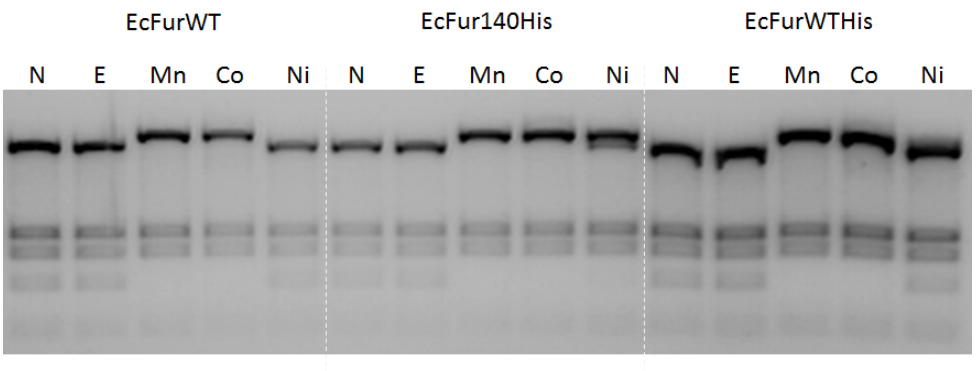


Figure 4.24: Nuclease protection assay on EcFur-WT, EcFur-140-histag and EcFur-WT-histag. N: native protein without any metal tratement. E: protein treated with EDTA. Proteins were incubated in the presence of 4 equivalent of Mn, Co or Ni.

These results show that the last eight amino acids, deleted in the EcFur-140 construction, are not essential for protein stability and do not affect its capacity to interact with DNA.

Interestingly, these last eight amino acids contain two histidines (1 in the case of VcFur and YpFur) and could be able to coordinate metal atom and participate in *fur* regulation. Some data suggests that a mutation of both terminal histidines (His142 and His144) induces a sensitive phenotype in the presence of Mn (Coy et al., 1994). However, EcFur-140 is still able to bind DNA *in vitro* and is stable in solution justifying our studies on it, instead of the wild type protein form *E. coli*.

4.7.2 Electrophoretic mobility shift assay (EMSA)

The following sequence `5'-GGGGATAATGATAATCATTATCGGG-3'` and its complementary strand were used in EMSA assays on EcFur-WT and EcFur-140. This technique demostrates the ability to bind DNA through a shift in the migration profile of free DNA when it is bound to a protein in a [DNA+protein] complex. Figure 4.25, left gels, shows the binding of EcFur-140 on a Fur box, the gradient used was not sufficient to determine the K_d exact value for this interaction, however, it is clear that it is lower than 25 nM. The right gels compares EcFur-140 and EcFur-WT showing a similar behaviour for both proteins. The literature indicates a K_d of 5 nM in the case of EcFur-WT (Lavrrar et al., 2002). More assays should be carried out to determine the K_d of EcFur-140 when interacting with a Fur box.

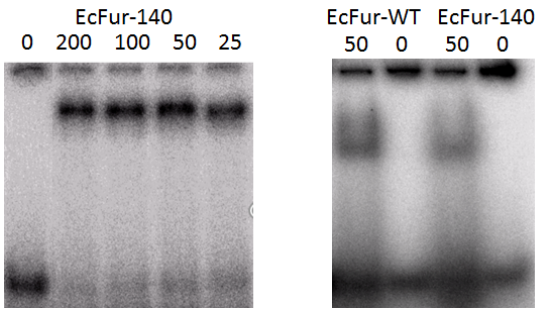


Figure 4.25: Electrophoretic mobility shift assay of EcFur-140 and EcFur-WT loaded with Mn in the presence of a Fur box. Values above the gels indicate Fur box concentrations in nM.

4.8 Biophysical characterisation of EcFur-140

4.8.1 SEC-MALLS

EcFur is mostly dimeric and stable in solution (EcFur-WT-D) and is activated by divalent metal ions. An overnight incubation with EDTA at 100 mM, to chelate metal ions, and TCEP at 1 mM, to reduce potential disulphide bridges, is able to convert the dimer into a monomer (EcFur-WT-M). The process is reversible with the production of a stable dimer upon incubation of monomer with 1 eq. Zn^{2+} per monomer overnight at 4°C (D'Autréaux et al., 2007; Pecqueur et al., 2006). Results obtained by SEC-MALLS-RI (Figure 4.26 light green and red curves) indicate the size of each species in solution: 32 kDa for the dimer and 16 KDa for the monomer, theoretical details on MALLS can be found in section 6.1.1. These results confirm the role of Zn^{2+} in stabilizing an EcFur dimer in solution as previously described (D'Autréaux et al., 2007).

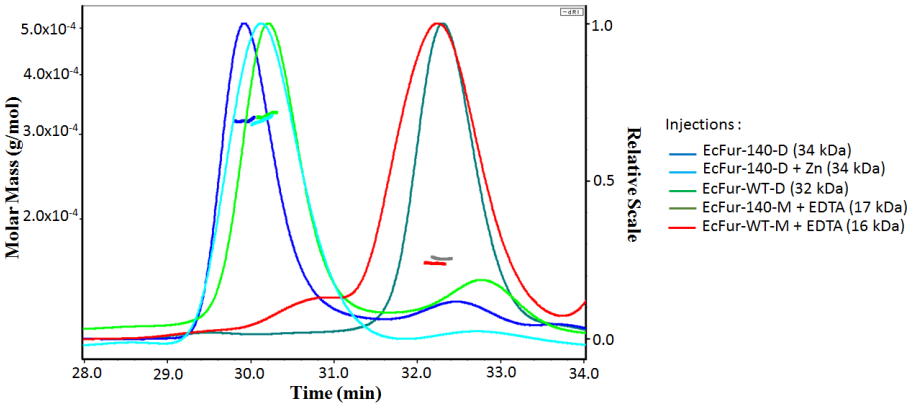


Figure 4.26: SEC-MALLS-RI analysis of EcFur-WT and EcFur-140 as dimers (D) and a monomers (M) showing the molecular weight difference between oligomeric species. The shif in elution time for equivalent oligomeric species is due to the use of different elution buffers and the presence of the polyhistidine tag in the EcFur-140 construction.

In addition, thermal shift experiments (TSA) indicate that the monomer has a T_m of 45 °C which is way lower than that of the dimer with a T_m of 65 °C. This indicates that

the dimer is more stable than the monomer and is probably the major species found *in vivo*.

Similarly to EcFur-WT, EcFur-140 can be monomerized after an EDTA and TCEP treatment overnight and shows similar oligomeric patterns on SEC-MALLS-RI (Figure 4.26 blue and dark green curves) the difference in the elution profile is mainly due to the presence of the poly-histidine tag in the EcFur-140 construction.

In addition, an EcFur-140/DNA complex was characterised. The theoretical MW of DNA is close to 16.5 kDa and the dimer is 34 KDa. A theoretical MW of the protein DNA complex should be equal to 84 kDa with two EcFur dimers binding one Fur box. In our experiment the MW observed was 78 kDa (Figure 4.27). The difference between the theoretical and the measured MWs could be due to the use of biophysical parameters adapted for proteins (dn/dc: 0.185) and not those adapted for DNA (dn/dc: 0.17) in the characterisation of the complex (Theisen, 2000).

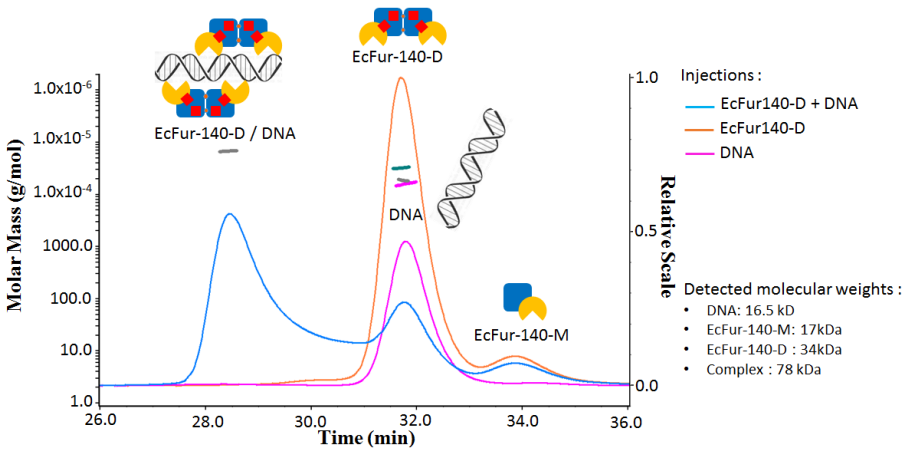


Figure 4.27: SEC-MALLS analysis of activated EcFur-140 and DNA and their complex showing the molecular weight difference between the protein, the DNA and the protein-DNA complex. Pictures inserted on the graph represent species eluted at each specific time.

However, using 4 wavelengths (214, 260, 280 and 340 nm) in our SEC-MALLS setup, the 260/280 ratio can be calculated and helps estimate the quantity of each species (DNA or protein) present in the complex in percentage Brescia et al., 2009. A 260/280 ratio equal to 1.55 was obtained for our complex and reflects a mix of 80% protein with 20% DNA (OD_{260nm}=0.107 and OD_{280nm}=0.069). Taking into account the theoretical mass of the complex, of 84 kDa, 80% of protein corresponds to 67.2 kDa and 20% of DNA corresponds to 16.8 kDa. These results are in agreement with our proposed model of two dimers binding the DNA box and confirm the presence of a homogeneous complex in solution. Similar experiments with other Fur proteins, like FtFur and MgFur, reveal the same complex types (Pérard et al., 2018).

4.8.2 Small-angle X-ray scattering experiments

SAXS is usually used as complementary technique to provide information about the molecular weight and oligomeric state of proteins. It is based on the small-angle scattering of light after passing through a sample solution and is described in section 6.1.2.

Our main objective is to understand how Fur proteins work and how the inhibitors affect their interaction with DNA. Key information can be obtained from structural characterisation of complexes in solution. SAXS experiments were carried out on Fur from *E. coli* and the results are shown in Table 4.12. Two Fur constructions were tested, the wild type protein EcFur-WT and the C-terminal deletion mutant EcFur-140 that crystallized.

Data was collected on the BM29 beamline at the ESRF by Julien Pérard, on of this work's supervisors, for samples with concentrations varing between 0.5 and 10 $mg.mL^{-1}$. The ATSAS 2.0 (Franke et al., 2017) program suit was used with PRIMUS and GNOM softwares for data analysis, and DAMMIF for *ab initio* modelling.

SAXS on EcFur-140

Comparison of data acquired from an Mn-loaded sample and a non-metalated sample shows that the addition of Mn does not drastically change the global shape of EcFur-140 in solution (Table 4.12 rows 2 and 3).

SAXS data of EcFur-140 and EcFur-140/DNA were acquired but will not be developed here since their SAXS scattering curves could not be correctly fitted with models of the X-ray structure of EcFur-140. This is due to the absence of the polyhistidine tag in the X-ray structure and its high flexibility in solution impacting SAXS scattering curves. However, these complexs were characterised in SEC-MALLS-RI experiments as shown previously.

SAXS on EcFur-WT

Similar to the case of EcFur-140, the presence of Mn^{2+} does not change the global shape of EcFur-WT in solution (Table 4.12 rows 4 and 5), indicating that the two proteins have similar behaviour in solution. Figure 4.28 compares the scattering curves of EcFur-WT in the presence (red) or abscence (black) of Mn^{2+}.

Fitting EcFur-WT SAXS envelope with the crystal structure of EcFur-140

Even if SAXS experiments did not shed light on the detailed inhibition mechanism by pF2, they nevertheless enabled a general characterisation of EcFur complexes. In the absence of a crystal structure of EcFur-WT, and due to the similarity of their structure in solution, the scattering curves of EcFur-WT were fitted with models of the crystall

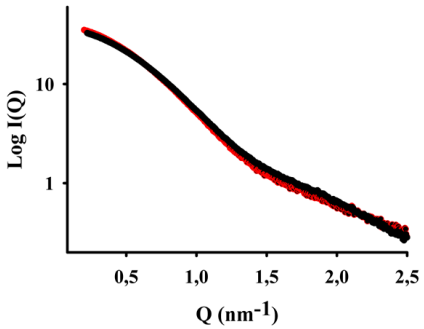


Figure 4.28: Scattering curves of EcFur-WT, with (red) or without (black) Mn^{2+}

Name	Rg (nm)	D_{max} (nm)	V_P (nm^3)	MW (KDa)
EcFur-140 + Mn	2.4	7.0	45	32
EcFur-140 (apo)	2.4	7.0	45	32
EcFur-WT + Mn	2.5	6.5	52	32
EcFur-WT (apo)	2.5	6.5	50	34
EcFur-140 + Mn + Fur box	3.0	9.5	90	74
EcFur-WT + Mn + Fur box	2.9	9.5	79	76

Table 4.12: SAXS biophysical parameters of EcFur-140 and EcFur-WT. Apo proteins are non-metallated proteins. Rg: radius of gyration. D_{max}: deduced from the pair distribution function. V_P: Porod's volume. MW: molecular weight.

structure of EcFur-140, and will be discussed below.

Figure 4.29 shows the fitting of the scattering curve of EcFur-WT with the resolved structure of EcFur-140. SAXS studies on EcFur-WT in solution show it to be more elongated than EcFur-140. To optimize the fitting on the experimental curve, the DNA binding domains in the crystal structure of EcFur-140 need to be opened up or elongated (this can also be seen in Figure 4.31A and C). The EcFur-140 structure shows a smaller distance between its DBDs, this distance was increased to better fit the experimental scattering curve of EcFur-WT, giving a χ^2 of 1.2, suggesting that EcFur proteins are elongated in solution.

In addition, the last 8 amino acids (141-148) deleted in the EcFur-140 mutant are highly flexible in the EcFur-WT construction and could be responsible for some of the difficulty in fitting the EcFur-WT scattering curve with the crystal structure of EcFur-140.

SAXS on EcFur-WT bound to DNA

To propose a model that fits the SAXS envelope of EcFur-WT bound to the Fur box (Figure 4.30) obtained by DAMMIF, the crystal structure of EcFur-140 was used. Figure 4.31A shows our first trials to fit the experimental scattering curve of EcFur-WT

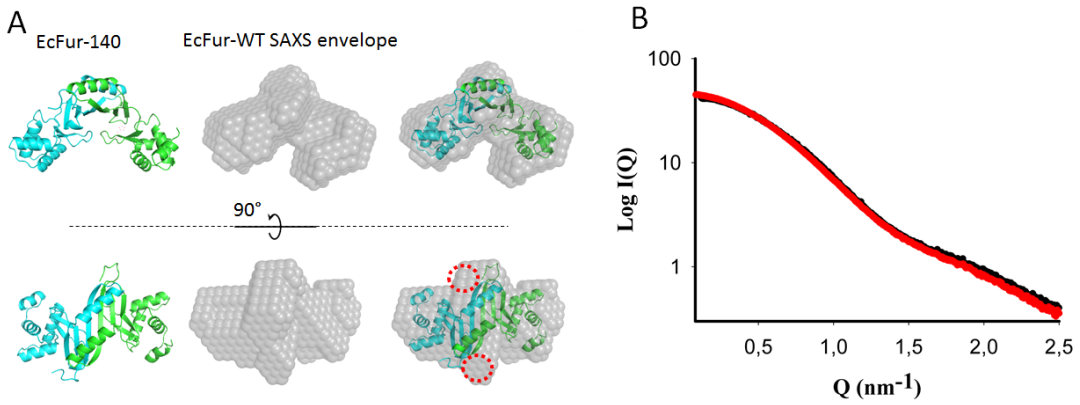


Figure 4.29: A: Structure of EcFur140 used to fit the SAXS envelope of EcFur-WT. The red circles indicate the C-terminal flexible region, missing in the structure of EcFur140 but present in the SAXS envelope of EcFur-WT. B: Scattering curve of EcFur-WT (black) fitted with the crystal structure of EcFur-140 (red) with a χ^2 of 1.2.

in complex with DNA with the resolved structure of EcFur-140. The fit gave curves that are not well superimposed with a χ^2 of 9. This could be due to the missing C-terminal part in the crystal structure of Ec-140 and the effects that the polyhistidine tag can have on it. Twisting or spreading the DBD outwards did not give a better fit, actually the quality decreased with a χ^2 of 23.

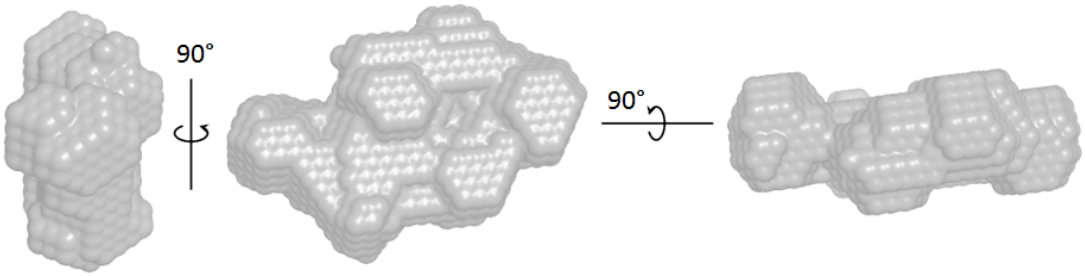


Figure 4.30: SAXS envelope of a EcFur-WT in complex with DNA, obtained by DAMMIF, viewed from different angles.

However, EcFur-140 in complex with DNA presents an elongated shape with a Porod volume of 90 nm^3 that is compatible with a model where two EcFur dimers are bound to a Fur box. The difference in V_p of EcFur-140 and EcFur-WT is probably due to the presence of the flexible polyhistidine tag in EcFur-140 that occupies more volume (Table 4.12 last two rows). This indicates that theoretically we should be able to fit the experimental curve of EcFur-WT + Fur box with two dimers of EcFur-140, we need to try different models.

To pursue the analysis, the experimental scattering curve of the EcFur-WT + Fur box complex was fitted with the solved structure of MgFur (PDB ID: 4RB1) where two MgFur dimers are bound tto he DNA sequence, the same sequence that was used in our SAXS experiments. This model gives a χ^2 of 4 indicating once again that the experimental data is compatible with a two dimers and one DNA sequence model (Figure 4.31B). By

replacing the MgFur dimers with EcFur-140 dimers the fit was improved with a χ^2 of 1.1 (Figure 4.31C and D) indicating with high confidence the possible structural arrangement of EcFur-WT bound to a Fur box. With this information, the X-ray structure obtained for EcFur-140 has a high probability to correspond to an activated conformation. All the results presented here are bering processed to be included in an article describing the structure of EcFur-140.

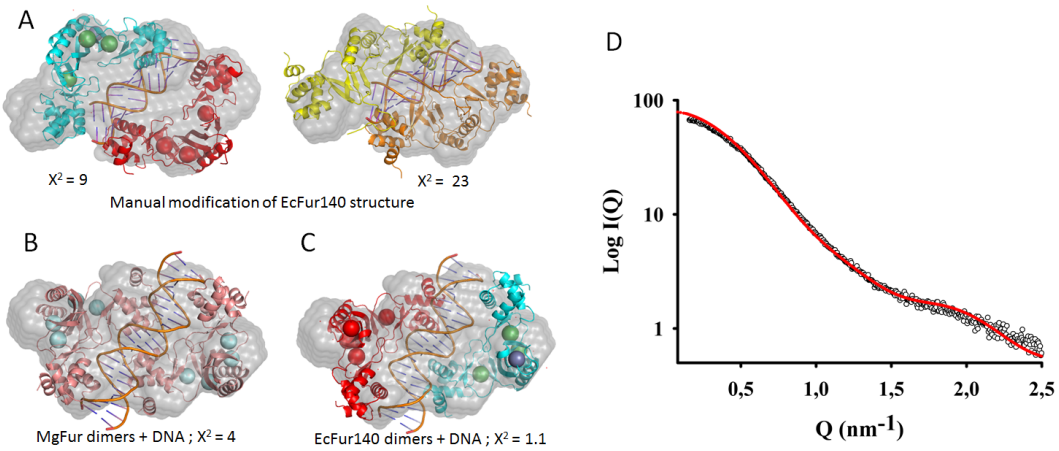


Figure 4.31: Generating models to fit the experimental SAXS scattering curve. A: manual modification of the resolved structure of EcFur140, mainly twisting and spreading the DBDs outwards. B: using the structure of MgFur (PDB ID: 4RB1) in complex with its DNA gives a better χ^2. C: EcFur140 dimers are placed in the same conformation as MgFur giving the best fit of the experimental curve. D: experimental scattering curve of EcFur-140 (black) fitted with the model (red) shown in C giving a χ^2 of 1.1.

SAXS on EcFur-WT in complex with pF2

Results obtained for EcFur-WT with or without pF2 (row and 4 in Table 4.13), show no impact on the structure of EcFur-WT, seen in the lack of variation in biophysical parameters that are relatively similar in both cases. This indicates that pF2 does not induce a global structural change when inhibiting EcFur.

Name	Rg (nm)	D_{max} (nm)	V_P (nm^3)	MW (KDa)
EcFur-140 + Mn	2.4	7.0	45	32
EcFur-WT + Mn	2.5	6.5	52	32
EcFur-WT + Mn + pF2	2.54	6.9	55	33

Table 4.13: Comparison of SAXS biophysical parameters of activated EcFur proteins with or without pF2. Rg: radius of gyration. D_{max}: deduced from the pair distribution function. V_P: Porod's volume. MW: molecular weight.

4.9 Docking of pF2 on the structure of EcFur-140

Having obtained the crystal structure of EcFur-140, the docking of pF2 was carried out. The results were compared to the previous docking, done on a model of EcFur built by homology to VcFur.

Several cycles of peptide docking to a fixed protein dimer, followed by molecular dynamics simulations were carried out. The AutoDock (version 4.2.6) (Morris et al., 2009) program was used for the docking of peptides on EcFur-140 (see section 7.4). For molecular dynamics, CHARMM was used (Brooks et al., 1983, see section 7.1.2). Water In CHARMM is treated implicitly with the EEF1 force field (7.3.2) and temperature is set to 300 K through Langevin dynamics simulations. Figure 4.32 shows the workflow diagram.

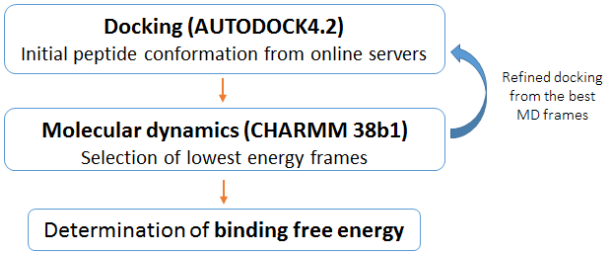


Figure 4.32: Workflow diagram showing how the docking of pF2 was done on EcFur-140.

The two steps of docking and MD simulation were repeated until the final docking free energy of the peptide and the CHARMM interaction energy between peptide and protein stopped decreasing.

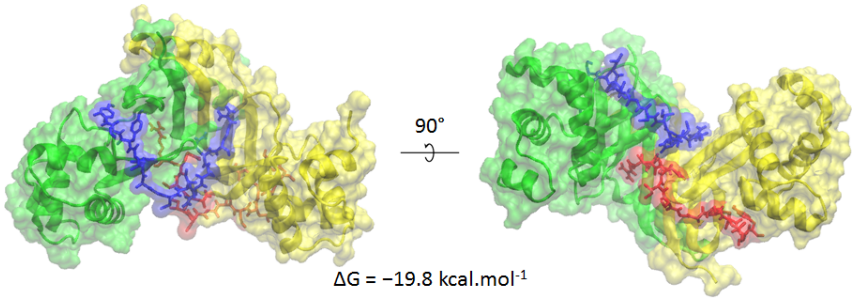


Figure 4.33: Docking of two pF2 peptides, pF2A in blue and pF2B in red, on EcFur-140 highlighting the symmetry of both inhibition pockets.

The difficulty of docking a 50 degree of freedom peptide to a protein was already emphasized in a previous paper published by our team (Cissé et al., 2014). Here, in the case of pF2, even more checking was introduced to try and produce valuable models which agree with experimental results. 5 ns MD simulations were run to allow the protein to move, then the frame with lowest peptide-protein interaction energy was elected for a symmetrisation of the structure. Assuming that pF2A (for instance) gave the lowest interaction energy, the matrix transformation converting monomer A into B was calculated

and applied to EcFur monomer A and peptide pF2A to build EcFur monomer B and pF2B, the best docking configuration for pF2A is shown in figure 4.33. Then the dockings were run on the symmetrized structure with a constant number of 15 free dihedrals. Further, two main checking lines were adopted: i) Dock one peptide (pF2A) and then its counterpart (pF2B) starting from the same protein structure ii) apply symmetry or not to reinitialize the position of one peptide after docking of the other.

Figure 4.34 shows the convergence of dockings after alternating CHARMM dynamics and minimizations and Autodock binding free energy calculations. A big improvement of the energies is seen when we start applying the symetrisations (indicated by the S letter). Green dots show the ΔG values, one for each pF2 peptide, when the best docking position of the previous EcFur model is used on the new structure of EcFur-140.

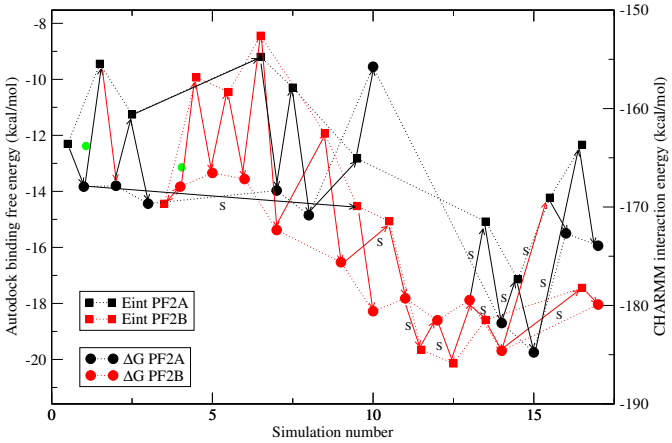


Figure 4.34: Graph showing the convergence of the docking of pF2A and pF2B on the EcFur-140 structure. The arrows depict the paths followed alternating CHARMM dynamics and minimizations (squares) and Autodock binding free energy calculations (circles). E_{int} is CHARMM interaction energy between the given peptide and the protein. The two green dots show the ΔG values, one for each pF2 peptide, when the best docking position of the previous EcFur model is used on the new structure of EcFur-140.

Table 4.14 details the final results, shown in blue, and compares them with previous data obtained on the model of EcFur built by homology to VcFur (in black) (EcFur-ΔCter15, abbreviated to EcFur-138). These simulations involve two pF2 peptides. The Autodock binding free energy, minimum energy cluster and number of poses in this cluster (out of 256 total) are indicated. Finally the total CHARMM potential energy of the system and average interaction energies between the selected peptide and the protein averaged over a 5 ns MD simulation are shown in columns 6 and 7.

Similarly, Table 4.15 summarizes the interactions between the protein and the peptide and compares the results with the ones obtained previously for the homolgy model.

Peptide	Protein	Free energy	Cluster avg. energy	Nbr. in cluster	CHARMM total energy	CHARMM interaction energy
pF2A	EcFur-138	-21.0 (+2.7)	-20.4	62	-6907 (42)	-121.2 (11.2) pF2B -120.0 (9.0)
pF2A	EcFur-140	-19.8 (-4.5)	-17.9	54	-6648 (43)	-127.8 (8.7) pF2B -136 (11.3)
pF2B	EcFur-140	-19.7 (-2.5)	-17.3	18	-6650 (43)	-136.9 (9.4) pF2A -126.0 (7.4)

Table 4.14: Docking free energies and interaction energies between pF2 and EcFur. All energies are given in kcal.mol^{-1} and all peptides were 1-13 full constructions. Published results are shown in black, where the docking of pF2 was carried out on a model of EcFur obtained by homology to VcFur. New results computed on the new X-ray structure of EcFur-140 are shown in blue. The free energy of the second cluster is shown in parentheses of colomn 3. The standard deviation of CHARMM energies are shown in parentheses in columns 6 and 7.

Peptide	Involved Fur residues	Invovled peptide residues
pF2	K77 N72 R70 Y130 E37 N60 S35 Y128 Y56 G75 H88 H125 S126 D38 H33	Q2 R1 S9 Y11 N4 W10 C6 C3 A8
pF2(A)	N72 K77 E74 K98 Y130 S126 Y128 V99 K41 H125 G76 H88 E37 S35 G75 R42 D38	R1 Q2 Y13 N4 I5 Y11 S9 A8 S12
pF2(B)	N72 K98 R70 S78 E74 K41 K77 E37 Y128 D38 H88 G97 D30 Y56 G76 V99 C96 L26	Y11 Q2 R1 Y13 S9 N4 S12 W10 G7 I5 C3 C6

Table 4.15: Interacting residues, with more than 5 kcal.mol^{-1} interaction energy and ranked in decreasing order, in the docking of the peptides to EcFur. Published results are shown in black, where the docking of pF2 was carried out on a model of EcFur obtained by homology to VcFur. New results computed on the new X-ray structure of EcFur-140 are shown in blue. Green residues have more than 10 kcal.mol^{-1} interaction energy.

The information presented in these tables is schematically represented in Figure 4.35 for pF2A and 4.36 for pF2B.

pF2B cysteine residues C3 and C6 both show a more than 5 kcal.mol^{-1} interaction energy (in Table 4.15) with the protein dimer in agreement with the fact that the pF2(1-10 C3S) mutant is inactive experimentally and that oxidized pF2 is inactive too. (pF2A cysteine residues with interaction energies of -4.6 and -0.9 kcal.mol^{-1}) with the protein do not appear in the table).

Models of pF2 docked to EcFUR-138 are in perfect agreement with the experimental results. Indeed each residue S126 or Y128 (two pairs of residues per dimer) presents more than 5 kcal.mol^{-1} interaction energies with pF2A or pF2B (Table S1) (The precise calculation gives a total interaction energy of S126 and Y128 with pF2A of -14.9 kcal.mol^{-1} and with pF2B of -20.2 in the best pF2A and pF2B dockings, respectively).

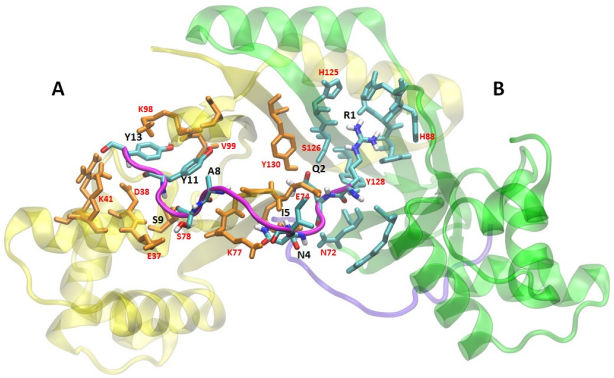


Figure 4.35: pF2A docked to EcFur-140, in the presence of pF2B, with a ΔG = -19.75 kcal.mol^{-1}. Fur monomer A is shown in yellow, monomer B in green, pF2A in magenta and pF2B in violet. Important residues of pF2A and the protein dimer are highlighted in black and red, respectively. This figure corresponds to Table 4.14 line 3 and Table 4.15 line 3.

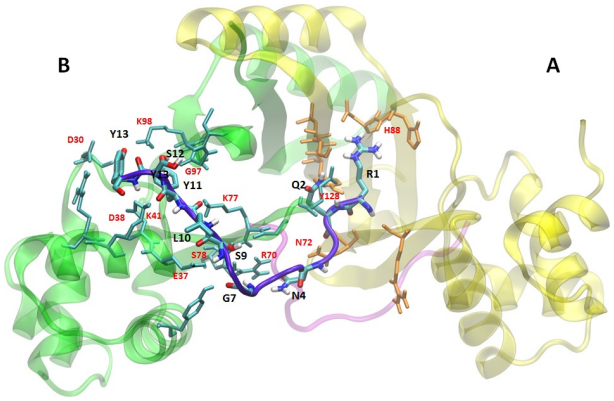


Figure 4.36: pF2B docked to EcFur-140, in the presence of pF2A, with a ΔG = -19.69 kcal.mol^{-1}. Fur monomer A is shown in yellow, monomer B in green, pF2A in magenta and pF2B in violet. Important residues of pF2B and the protein dimer are highlighted in black and red, respectively. This figure corresponds to Table 4.14 line 4 and Table 4.15 line 4.

4.10 Chapter conclusion

In a previous work, anti-Fur peptide aptamers, made of thioredoxin A from *E. coli* as a scaffold and a variable loop of 13 amino acids, and their corresponding linear peptides, pF1-pF4 were shown to specifically bind Fur from *E. coli* and inhibit its DNA binding activity. Their action was studied *in vitro* and *in vivo*, where a decrease in bacterial virulence was shown (Abed et al., 2007; Cissé et al., 2014; Mathieu et al., 2016). In addition, chemical library screening and *in silico* design allowed the identification of three new inhibitors that acted on Fur from *P. aeruginosa* (pL1, pL2 and molecule B). Following the same coupled theoretical and experimental approach, this work had as objective to investigate the mechanisms behind each inhibition.

To do so and in the absence of crystal structures of protein + inhibitor complexes, when this work started, docking simulations were carried out on a model of EcFur obtained by homology to Fur from *V. cholerae*. In an attempt to optimize and find new modified peptides, docking simulations were performed with cyclic peptides and showed that cycles obtained by addition of 6 glycines to the pF2 peptide had approximately the same binding free energy as the linear peptide. On its own, this information shows that cyclic peptides could potentially inhibit Fur as well as their linear equivalents. However, the idea was to obtain the smallest cyclic peptides possible, and by doing so, docking results showed a loss in free energy of binding. Once a consensus is found on the cycle size, the constructions will be transferred to our chemist collaborators to be synthesised and used in *in vitro* assays or fluorescence tracking experiments.

Docking simulations of pL1 and pL2 were also performed on the structure of PaFur obtained by Pohl (Pohl et al., 2003). Results show a binding pattern similar to what was observed for pF2, with both inhibitors binding to the region between the two subunits of a dimer. In order to obtain a more complete view of each inhibition mechanism and to be sure that the experimentally observed inhibition does not involve peptides binding to metal sites, or metal chelation form the protein, X-ray absorption spectroscopy was used. Fur protein samples from different species were loaded with Mn^{2+} or Co^{2+} and incubated in the presence of inhibitors. The results obtained for EcFur show that the inhibitor does not change the chemical environment of the metal, indicating that the inhibition is not due to interactions in this region of the protein. In addition, by fitting the experimental spectra, we were able to propose models where the metal sites are composed of 6 ligands, in accordance with the crystal structure resolved later and previous work on other Fur proteins.

As our objective is to study inhibition, one direct way of doing so was through crystallization assays. Following the XAS experiments, we were left with sufficient quantities of highly concentrated protein solutions. Manual and automated crystallization screens were carried out on a metal site mutant of Fur from *P. aeruginosa*, named PaFurΔS3,

with success. Interestingly, the tetrameric structures, resolved in the presence of zinc or manganese, revealed a considerable number of metal atoms bound to the protein that was thoroughly washed before crystallization. Compared to the structure of FtFur, resolved in our group (Pérard et al., 2018) these excess metal atoms raise question about their role in stabilizing the tetramer and are still being investigated. After the resolution of the structure of PaFurΔS3, soaking and co-crystallization trials were initiated to get the structure in the presence of one of the three inhibitors. Several months of optimization and many automated screens were used. We were able to obtain crystals but no diffraction data set was sufficient to resolve the structure. In the future, based on the work presented here, crystallization trials can be resumed in order to optimize the crystals that we were able to obtain.

At the same time, crystallization trials on EcFur were carried out. We first tried EcFurΔS3, the same metal site mutant that crystallized in the case of PaFurΔS3 without success. We decided then to remove the flexible C-terminal region (EcFur-140) of the protein and keep the polyhistidine tag on our construction during crystallization. This construction is stable in solution and has a similar DNA binding activity to the wild type protein. As for PaFurΔS3, we started with an automated screen that yielded one hit in this case. From that point, and after several weeks of manual optimizations, crystal diffraction quality was improved, enabling the structure of EcFur to be resolved for the first time after twenty years of failed trials.

Fluorescence spectroscopy showed the presence of nickel and zinc in the structure, a result validated by ICP-AES measurements, which show the binding of one zinc and one nickel atom on one EcFur-140 monomer. Nickel was not added to the protein and its presence can be linked to the purification protocol. When crystallization was tried with other metals (manganese, cobalt, zinc or even nickel), no diffraction was obtained. The analysis of EcFur-140 structure showed an organization similar to other dimeric Fur proteins with DNA binding domains between the resolved residues 2 to 82 and a dimerization domain formed by residues 83 to 134; the polyhistidine tag was not visible in the electron density map. Analysis of metal sites in the structure, revealed the absence of zinc in S1 and the presence of a disulphide bond, formed between two cysteins (93 and 96). A similar S-S bond is observed in VcFur (D'Autréaux et al., 2007; Sheikh et al., 2009), and could be due to an oxidized protein. The S2 site is an unusual metal site in EcFur-140, with 4 histidines (33, 71, 88, 90) and one glutamate (81) coordinating a nickel atom, with His71 being the novelty in this site. Usually, this site contains only three histidines and two glutamates. These results validate the previously discussed XAS data where a model of 4 histidines and one glutamate fits our experimental data. The EcFur-140 structure also allowed to validate the presence of an S3 site in Fur from *E. coli*. However, due to the resolution of the electron density map, only a model of this site can be proposed with two histidines (87 and 125), one aspartate (89) and a glutamate (108). With our current data set offering limited information on this specific metal site, a new

crystallization campaign should be initiated to acquire new data.

The inhibition of EcFur was also investigated in SAXS experiments, were data was acquired on the protein in complex with pF2 and showed no change in global structure in the presence of the inhibitor. This indicates that the inhibition by pF2 does not induce a global conformational change of the protein. At the same time, the structure of EcFur-140 helped us fit experimental SAXS data obtained on the wild type protein alone and in the presence of a Fur box. This allowed us to propose a model of EcFur binding DNA that can be used for further investigations.

Finally, previously obtained docking results on the model of EcFur, obtained by homology to VcFur, were validated by carrying out similar docking simulations on the new structure of EcFur-140. The two pF2 peptides still bind the same area of the protein confirming our published results. However, slight differences are visible mainly in the interacting protein/peptide residues. This shows the importance of the crystallization tool in the search for underlying inhibition mechanisms of Fur proteins. In addition to giving insights into an extensively studied model protein, the structure of EcFur-140 will serve as a docking platform for all the present, and future, inhibitory peptides developed in our laboratory.

Annex

Plate	Well	Composition
Wizard I (Rigaku)	A 07	PEG 8000 10% MES 0.1 M pH 6 Zn Acetate 0.2 M
Wizard I (Rigaku)	F 08	Isopropanol 15% MES 0.1 M pH 6 Zn Acetate 0.2 M
PACT premier™ MD1-29	A 03	PEG 2000 25% SPG 0.1 M pH 6
PACT premier™ MD1-29	F 01	NaF 0.2 M BTP 0.1 M pH 6.5 PEG 3350 20%
PACT premier™ MD1-29	F 02	NaBr 0.2 M BTP 0.1 M pH 6.5 PEG 3350 20%
PACT premier™ MD1-29	F 06	Na Formate 0.2 M BTP 0.1 M pH 6.5 PEG 3350 20%
PACT premier™ MD1-29	F 07	Na Acetate 0.2 M BTP 0.1 M pH 6.5 PEG 3350 20%
PEGs-I (Qiagen®)	E 01	NaF 0.2 M PEG 3350 20%

Table 4.16: Table showing the chosen conditions from standard crystallization screens of PaFurΔS3.

Tube number	Main precipitant	Drop Concentration
26	L-proline	0.01 M
29	sodium bromide	0.01 M
34	glycine	0.1 M
42	trimethylamine hydrochloride	0.01 M
50	polyvinylpyrrolidone K15	0.5 %(v/v)
52	pentaerythritol ethoxylate (3/4 EO/OH)	4 %(v/v)
53	polyethylene glycol 3350	1 %(v/v)
56	xylitol	3 %(v/v)
59	D-(+)-trehalose dihydrate	3 %(v/v)
61	ethylene glycol	3 %(v/v)
62	glycerol	3 %(v/v)
68	CYMAL-7	0.015 mM
73	trimethylamine N-oxide dihydrate	3 %(v/v)
79	(+/-)-1,3-butanediol	4 %(v/v)
83	2-propanol	3 %(v/v)
87	1,3-propanediol	4 %(v/v)
88	acetonitrile	4 %(v/v)
92	acetone	4 %(v/v)
95	2,2,2-trifluoroethanol	4 %(v/v)
96	1,1,1,3,3,3-hexafluoro-2-propanol	4 %(v/v)

Table 4.17: Compounds with which positive hits were detected in the additive screen.

Sample name	Mn-PaFurΔS3		Zn-PaFurΔS3
Data collection			
pdb deposition code	6H1C		n.a
Space group	$P6_122$		$P3_221$
Unit cell dimensions:			
a, b, c (Å)	85.57 85.57 179.62	85.05 85.05 177.46	84.584 84.584
α, β, γ(°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	96.626, 90, 90, 120
Wavelength (Å)	0.9797 (remote)	1.2925 (peak)	0.9797 (native)
Resolution range (Å)	68.5-2.34 (2.42-2.34)	42.52-2.61 (2.27-2.61)	40.33 - 2.75 (2.84-2.75)
R_{merge}	0.038 (0.17)	0.037 (0.17)	0.114 (0.763)
R_{meas}	0.05 (0.24)	0.06 (0.24)	0.080 (0.539)
Total reflections	33998 (3342)	20008(1923)	20332 (1956)
Unique reflections	17023 (1671)	15650 (1540)	10640 (1026)
I/σ(I)	14.12 (3.99)	10.2 (1.60)	7.92 (1.74)
Completeness (%)	99.22 (99.94)	99.72 (99.07)	98.68 (97.16)
$CC_{1/2}$	99.8 (47.6)	99.5 (57.8)	80.69 (53.97)
Redundancy	2.1 (2.0)	1.8 (1.7)	1.9 (1.9)
Refinement			
Resolution range (Å)	68.5 - 2.34		40.3 - 2.75
Reflections (refinement)	17023		10636
Reflections (R-free)	804		521
R_{work} (%)	21.07		23.00
R_{free} (%)	26.9		26.77
Wilson *B* factor	35.26		57.86
Number of atoms:			
total	2223		2137
macromolecules	2096		2072
ligands	22 (11Zn + 11Mn)		16 (16Zn)
water	106		507
Protein residues	267		49
RMS(bonds)	0.008		0.011
RMS(angles)	0.87		2.00
Ramachadran:			
favored (%)	98.2		93.82
allowed (%)	1.8		3.47
outliers (%)	1.15		2.7
Rotamer outliers (%)	1.27		1.83
B factors (Å^2):			
average	37.10		56.73
macromolecules	36.95		56.91
ligands	50.3		59.76
solvent	37.3		48.14

Table 4.18: Data collection, phasing and refinement statistics for the structures of Mn-PaFurΔS3 and Zn-PaFurΔS3. Values in parentheses are for highest-resolution shell. Integration statistics are from XDS/XSCALE (Kabsch, 2010), refinement statistics are from Phoenix (Adams et al., 2010). A refined version of this table is available in the article describing this structure.

Table 4.19: Co K-edge EXAFS fit results. The one standard deviation error relative to the last digit is reported in parenthesis. The parameter β is the angle defined by the Co-ligand-C bonds for each amino acid. In the model "3.5 His 1 Glu bid 0.5 Glu mon", a common Co-O distance is assigned to both Glu residues (Co-O_{Glu}). σ^2 is the the Debye-Waller factor. ΔE_0 is the energy shift. R_{factor} represents the relative error of the fit and data.

Sample	Model	Co-N_{His} (Å)	β_{His} (°)	Co-O_{Glu} (Å)	$\beta_{Glu\ bid}$ (°)	$\sigma^2{}_1$ (10^{-3}Å^2)	$\sigma^2{}_2$ (10^{-3}Å^2)	$\sigma^2{}_{MS}$ (10^{-3}Å^2)	ΔE_0 (eV)	R_{factor} (%)
EcFur WT 1 eq. Co	4 His 1 Glu_{bid}	2.13 (3)	132 (4)	2.07 (6)	93 (2)	4 (1)	2 (2)	7 (4)	-8.0 (9)	6.7
EcFur WT 2 eq. Co	4 His 1 Glu bid	2.13 (2)	131 (3)	2.05 (3)	93 (2)	4 (1)	11 (3)	17 (7)	-6.7 (8)	6.8
	3.5 His 1 Glu_{bid} 0.5 Glu_{mono}	2.16 (1)	131 (3)	2.04 (2)	96 (3)	5 (1)	10 (3)	6 (3)	-7.7 (8)	7.3
	3 His 1.5 Glu_{bid}	2.15 (2)	132 (3)	2.09 (3)	95 (2)	6 (1)	6 (4)	6 (4)	-5.7 (8)	8.8
EcFur WT 2 eq. Co +pF2	4 His 1 Glu_{bid}	2.14 (1)	131 (2)	2.05 (3)	96 (2)	5 (1)	9 (4)	7 (3)	-8.0 (9)	7.3
	3.5 His 1 Glu_{bid} 0.5 Glu_{mono}	2.14 (3)	132 (2)	2.08 (3)	95 (2)	5 (1)	5 (3)	4 (3)	-8.0 (9)	7.4

Chapter 5

Properties of Fur oligomeric states

As described previously, Fur proteins from different species adopt different oligomeric states *in vivo* (Pérard et al., 2016; Pérard et al., 2018), however, experiments and crystal structures indicate that they all bind DNA as dimers. Apart from experimental structural studies, no theoretical work has been done to investigate the reason behind this different behaviour. One may argue that depending on bacterial species Fur proteins can be involved in slightly different regulatory networks which can explain the need for different oligomeric states capable of interacting with different partners.

In the case of crystal structures of Fur from *P. aeruginosa*, they were resolved with 11 Zn atoms and 11 Mn atoms in the Mn-PaFurΔS3 structure and with 16 Zn atoms in the Zn-PaFurΔS3 structure (Table 4.18). This shows a possible role in metal storage in tetrameric Fur proteins. Currently, both structures are being analysed to investigate these unconventional metal sites (other than sites S2 and S3), since the protein was thoroughly washed before crystallization. This capacity to bind metal could possibly be related to previous functions that the evolutionary predecessors of Fur had in the past.

Another possible explanation for the difference in oligomeric states could be due to functional added values, this could mean two things. One, in species that express Fur dimers, natural selection favoured the evolutionary strategy where the cost of tetramers and what they can do more than dimers was less valuable than the production of dimers. Two, in species that express Fur tetramers, natural selection opted for a costly tetramer since its added value, in comparison with dimers, could not be replaced by other mechanisms. An example of an added value for tetramers, could be the protection of less stable dimers, or the regulation of DNA binding specificity through tetramer/DNA interaction before the dimer/DNA interaction.

In addition to the possible reasons described above, the difference in oligomeric states could be linked to different folding mechanisms of Fur and the chaperones used during its synthesis.

Taken together, all that is described above shows that the oligomeric division into two

groups could be due to several factors, known or not, playing a major role in bacterial metabolism. However what is certain is that the actual repartition of Fur dimers and tetramers is what suits the best each species in its actual environment.

Based on this fact our investigation started with a phylogenetic study to detect any possible evolutionary link between oligomeric state, Fur sequences and species kinship.

5.1 Phylogenetic analysis of Fur proteins

As the protein folding depends on the sequence, the first steps in our study consisted of grouping Fur proteins using their amino acid sequences. The bacterial species were chosen to represent different bacterial phyla and classes in addition to species where Fur proteins were structurally characterised, or extensively studied, and are later used in this work.

The right panel of Figure 5.1 shows the phylogenetic tree generated by the Maximum Likelihood method (Guindon et al., 2010) using the Seaview software (Gouy et al., 2010) for bacterial species depending on their Fur amino acid sequence, the method is described in section 7.5. Tetrameric Fur proteins are shown in green and are from *L. pneumophilia*, *P. aeruginosa* and *F. tularensis*. Dimeric Fur proteins are shown in orange. Interestingly, tetrameric Fur proteins are grouped together indicating that their oligomeric state is based on shared residues between the three Fur proteins.

To check if this resemblance in sequence reflects a common ancestry, the phylogenetic tree of the same species was built using the 16S ribosomal RNA (16S rRNA) sequences to check if Fur protein sequences can reflect bacterial phylogeny. The 16S rRNA is a component of the prokaryotic ribosome, 16S rRNA genes are used to construct phylogenies due to their slow rates of evolution. They were used in the pioneering work of Carl Woese and his colleagues where he proposed the division of life into three domaines or major evolutionary lines : the Eukarya, the Bacteria and the Archaea (Woese et al., 1990).

As previously described by Achenbach et al., 1997 Fur datasets reflect the phylogenetic signal reasonably well. In our analysis, Figure 5.1 left panel shows that the three species with tetrameric Fur proteins are grouped together indicating a possible link to a common tetrameric ancestor. Actually they are all members of the *Gammaproteobacteria* class, however, *V. cholerae* is aslo a *Gammaproteobacteria* but has a dimeric Fur as described by Pérard et al., 2016. Is VcFur an exception to the rule ? or is the common tetrameric ancestor closer to *L. pneumophilia*, *P. aeruginosa* and *F. tularensis* and further from their common ancestor with *Gammaproteobacteria* ? with our actual knowledge it is not possible to answer these questions. The difference could be due to the divergent speciation paths between species.

Based on these results, protein sequences of tetrameric Fur share some information in

common. In order to gather more information on these complexes, simulations were done to understand how tetramers dissociate into dimers that bind DNA.

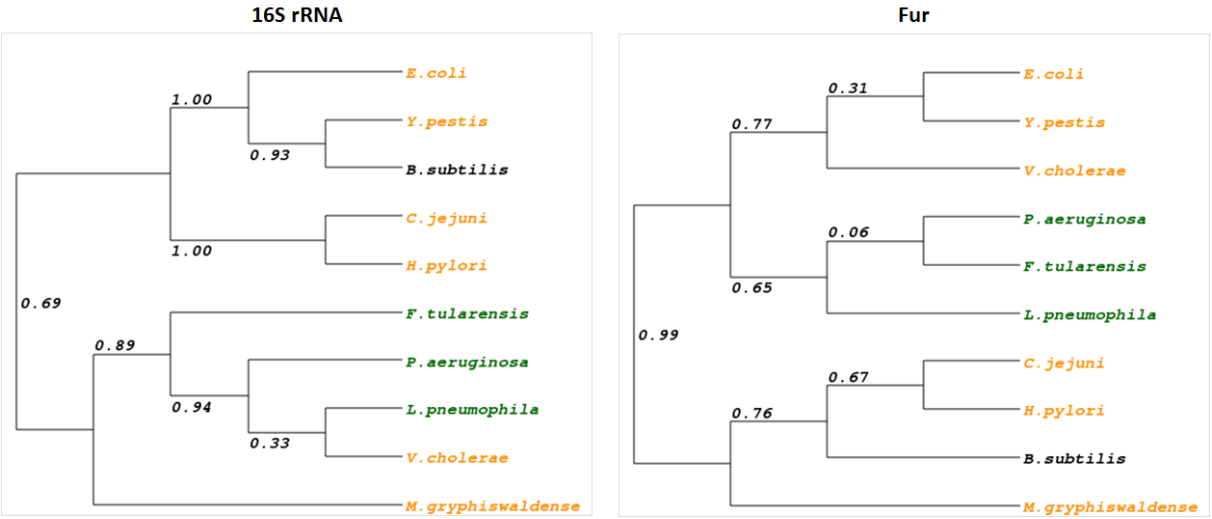


Figure 5.1: Phylogenetic trees of Fur proteins and 16S rRNA datasets on the same bacterial species. Tetrameric Fur proteins are shown in green, dimers are shown in orange. *B. subtilis* is the only Gram-positive in this study with no biophysical characterisation of its Fur and is shown black. Branch support values are indicated above the corresponding branch. Trees were constructed using the Maximum Likelihood method (Guindon et al., 2010) in Seaview (Gouy et al., 2010). Sequence alignemet used to generate each tree can be found in the Annex (Figures 5.60 and 5.61).

5.2 Studying Fur from *Francisella tularensis*

F. tularensis is the causative agent of tularaemia, a zoonotic disease, that can be transmitted to humans through many wild animal reservoirs (rabbits, foxes, wild boar,...). The pneumonic form of tularaemia in humans can be lethal if not treated. This intra cellular Gram-negative bacterium was first isolated in 1912 (McCoy et al., 1912) from squirrel carcasses in Tulare, California and was named after Edward Francis who studied tularaemia.

Due to its high pathogenicity, low infectious dose and ease of spread by aerosol *F. tularensis* is classified as a category A bio-terrorism agent by the Centers for Disease Control and Prevention (CDC) in the USA. Pathogens of this class are highly lethal and are treated as a high priority since their use can have a major impact on public health and disorganise economical and social structures of targeted communities.

In the absence of vaccines, studying Fur from *F. tularensis* helps understand and develop ways to control the virulence of this species.

To understand how tetramers dissociate and bind DNA as dimers, simulations of both [dimer+DNA] and tetramer complexes were carried out to find which complex is more

stable and what initiates their dissociation. This was done by identifying the free energy profiles and major interacting residues in each case.

In total, ten complexes, shown in Table 7.2 were studied, however only one will be fully detailed, the results of the other complexes will be described in the following sections. Here, Fur from *F. tularensis* will be used as a case study. The majority of the results were published this year in Nature - Communications Biology (Pérard et al., 2018), the full article can be found in the Annex at the end of this chapter.

System	Box size nm^3	Num. Metal M^{2+}	Num. Na^+	Num. Cl^-	Num. atoms	PDB ID
EcZur dimer + DNA	8.9*9.2*7.7	4	97	38	60119	4MTD
EcZur dimer + Chain C + DNA	11.6*8.9*8.1	4	112	51	80887	4MTD
MgFur-Furbox	9.4*9.2*5.9	4	76	30	51948	4RB1
MgFur-feoAB1	9.3*9.8*5.9	4	78	32	48073	4RB3
PaFur-Tetramer	10.2*8.7*6.9	8	44	36	57337	1MZB
PaFurΔS3-Tetramer	10.4*8.6*6.8	4	53	37	58232	this work
FtFur-Tetramer	10.5*8.0*6.5	8	37	33	52114	5NHK
FtFur-dimer and DNA (model)	9.3*8.9*6.4	4	77	32	50117	-
VcFur tetramer (model)	10.4*8.9*7.3	8	55	41	64287	-

Table 5.1: Size and ion composition of all systems studied in this work. Metal dications are represented as charged van der Waals spheres that do not discriminate between different atoms in our simple models of the metal binding sites. Number of counterions and total number of atoms are indicated.

5.2.1 Construction of the models

Construction of FtFur tetramer

Since the X-ray structure of FtFur (PDB ID: 5NHK) was solved by our team it was used as the initial model for the tetramer. The GROMACS program version 5.1.2 with the gromos54a7 united atom force field was used to perform long molecular dynamics simulations needed to compute free energy profiles. Fe^{2+} and Zn^{2+} were modelled as simple Lennard Jones hard spheres with charge +2, with Zn coordinated to charged deprotonated cysteines.

Construction of FtFur+ DNA models

In the absence of FtFur + DNA structure, the structure of *M. gryphiswaldense* (PDB ID: 4RB1) ref29 in the presence of DNA was used to model the wtDNA FurBox and correctly position the FtFur dimer on it by least-square fit matching of atom positions.

The 5'-GCCG**GATAATGATAATCATTATC**-3' fragment (consensus FurBox in bold) and its complementary 3'-5' sequence was used to model the double-stranded wtDNA.

Construction	Sequence
wtDNA	5'-GCCGGATAATGATAATCATTATC-3'
mutDNA	5'-GCCGGATACTGATAGTCCTGATC-3'

Table 5.2: DNA sequences used in the simulation of FtFur complexes.

The mutDNA sequence GCCGGATACTGATAGTCCTGATC contains four mutations, shown in red, with respect to the FurBox: A9 to C, A15 to G, A18 to C and T20 to G. They were constructed by simple matching of corresponding heavy atoms in the wtDNA model and building of missing hydrogens. Table 5.2 shows both sequences.

The three above vacuum systems were immersed in parallelepipedic SPC water boxes modelled with periodic boundary conditions after the addition of Na^+ and Cl^- counterions to ensure neutrality and a total ionic force of 0.1 $mol.L^{-1}$. The solvated systems were energy minimized and equilibrated under NPT (constant Number of particles, Pressure and Temperature) conditions at 310 K and 1 atm. More details on the construction of all complexes can be found in section 7.6.

5.2.2 Computing free energy profiles

Free energy profiles for the extraction (by translation along a fixed direction) of one FtFur dimer from the tetramer and of FtFur from DNA were computed. The simulations include a 'moving' subsystem (FtFur dimer, chains A and B) and a 'fixed' subsystem (FtFur dimer, chains C and D, wtDNA or mutDNA) as shown in Figure 5.2 in the case of the FtFur tetramer simulation.

The free energy profiles were built using the 'umbrella sampling' technique, detailed in section 7.1.8, and result from the overlapping of 26 computation windows that cover the reaction path of our systems. They are generated by translating the moving subsystem from the initial conformation along the X axis, in order to create snapshots of the different states the system will have during the dissociation. Their is one computation window for each translation distance. The meticulous translation protocol is shown in Figure 5.3.

Each window consisted of 100 ps NPT equilibration and 10 to 15 ns NPT production simulations. Position restraints on the 'fixed' subsystem and distance restraints on the whole protein, in the form of NOE-type restraints (Nuclear Overhauser Effect) between H-bonded H and O atoms to maintain its secondary structure, were applied. The 'moving' subsystem was subject to two harmonic biasing forces along the X direction only ('umbrella potential') applied between the centres of mass of the 2 Fur dimer subunits and the centre of mass of the 'fixed' subsystem.

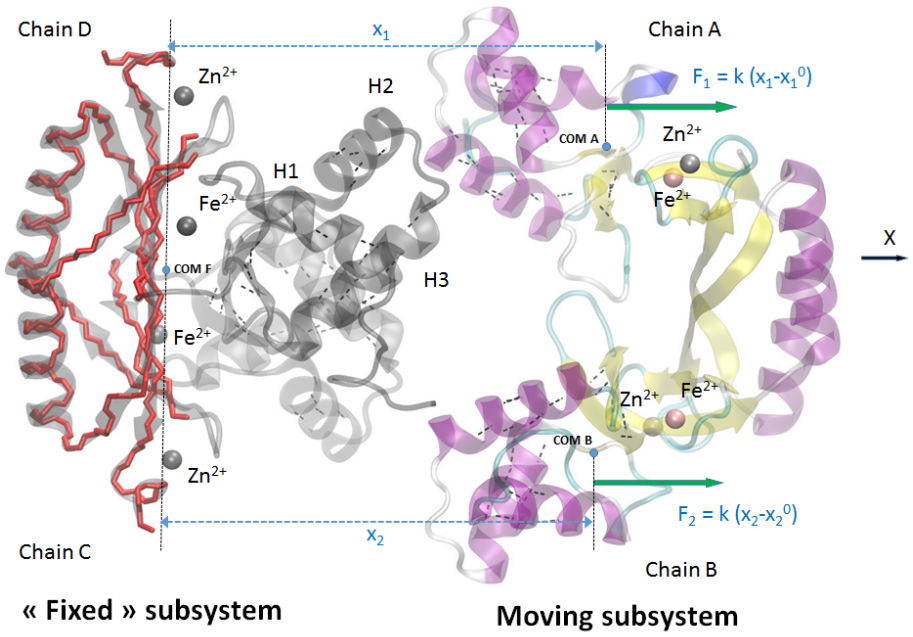


Figure 5.2: Structure of the FtFur tetramer corresponding to the last conformation in the pulling process. Chains A and B constitute the moving subsystem and have been translated (10*0.5+15*1.0) =20 Å from the initial tetramer along the X direction. Two equal forces, F_1 and F_2 are applied on their centers of mass (COM) to pull them away or maintain them at a fixed distance of the "fixed" subsystem. Part of the backbone highlighted in red is subject to position restraints during the pulling process. Black dashed lines show hydrogen bonds in the secondary structures which are maintained using NOE type distance restraints. The protein metal cations are shown as spheres, water and the counter ions are not shown for clarity.

After the dynamics runs, positions and forces were collected from the trajectories and the umbrella sampling harmonic potential was unbiased using the Wham algorithm implemented in the 'g_wham' program to yield the free energy profiles.

Potential of mean force calculations for the dissociation of the FtFur tetramer and FtFur dimer from DNA are shown in Figure 5.4. The calculated binding free energies were ΔG = 18.8, 10.5 and 8.8 kcal.mol^{-1} for dimer from Fur box, dimer from tetramer and dimer from mutated DNA, respectively.

The free energy profiles show that the FtFur dimer is less tightly bound to the other dimer in the tetramer than to the DNA Fur box, in agreement with the experimental deduction that the tetramer has to separate into two dimers in the close proximity of DNA. In addition, the simulation with the mutated DNA shows that this complex is the least stable, emphasising the *in vitro* observable fact that the tetramer only dissociates in the presence of its specific DNA sequence.

Three sigmoid functions were necessary to perfectly fit the profile of FtFur dimer binding to the DNA Fur box but the correspondence with the binding mechanism is not obvious especially for the first small sigmoid (2.8 kcal.mol^{-1} for a center of mass/center of mass distance of 2.29 nm or a pulling distance around 1 Å). This fitting will not be

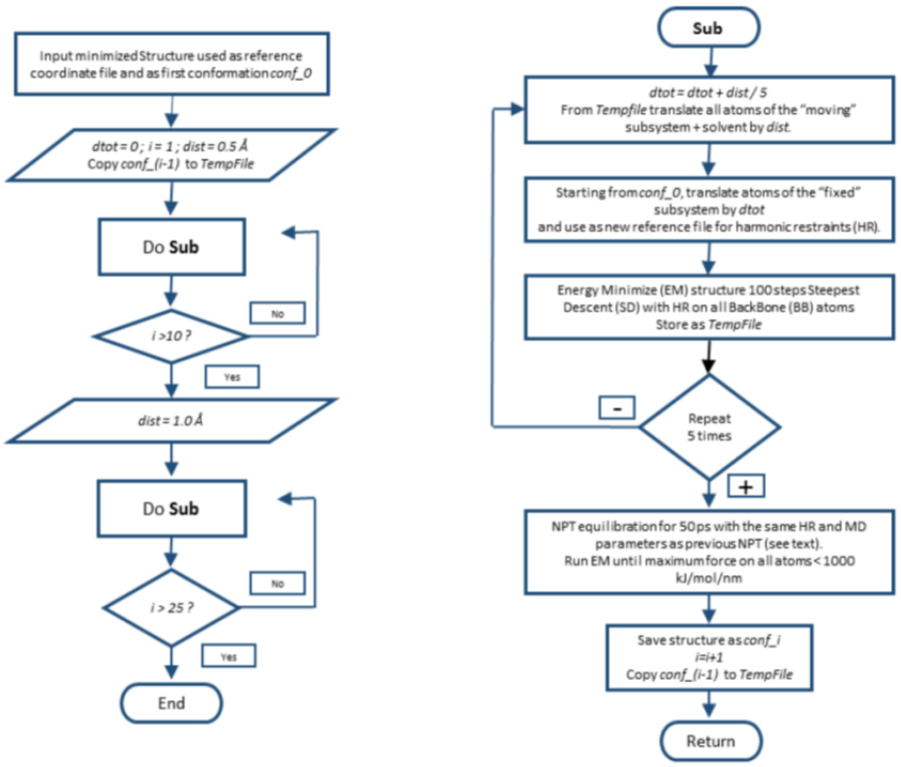


Figure 5.3: Algorithm for the preparation of initial coordinates for the computation of free energy profiles. The algorithm consists of a main subprogram doing the translations and pre-equilibrations, which is called twice: 1) 10 first windows corresponding to 0.5 Å translations followed by 2) 15 windows of 1 Å translations. Three variables are defined "dtot", the total translation distance for the "moving" subsystem, " i ", the conformation counter and "dist", the current translation distance in Å.

attempted in the profiles of the remaining systems presentend in this manuscript.

After the publication of our article, calculations were carried on for a better convergence and reliable error estimation in the case of the FtFur tetramer and FtFur DNA dissociations. The free energies involved have increased a lot as shown in Figure 5.5.

After 60 ns simulation time per window (compared to 15 ns in previous work), the new binding free energies are estimated to 23 and 17 kcal.mol^{-1} for dimer to Fur box and dimer to tetramer, respectively. However, the conclusions are unchanged, dissociating an FtFur dimer from a tetramer requires less energy than doing so on a dimer bound to DNA, indicating that the later is more stable. For both profiles, the error bars are shown in Figure 5.5, indicating that the differences between the two profiles is significant.

As previously mentioned (see section 7.2.2), four methods are proposed to estimate the error bars on the free energy profiles. The results are superimposed in Figure 5.7 and 5.6, for the tetramer and dimer DNA complexes, respectively.

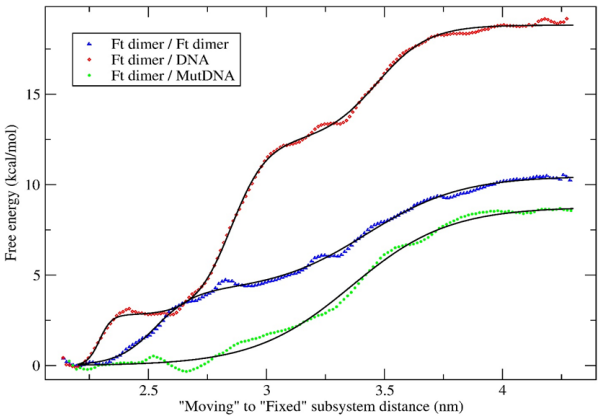


Figure 5.4: Potentials of mean force for the dissociation of FtFur dimer from the tetramer or from DNA. The x-axis, corresponds to the average centre of mass/centre of mass distance between the 'fixed' and the 'moving' subsystems. Wham output data were fitted with 1, 2 or 3 sigmoid functions with R.

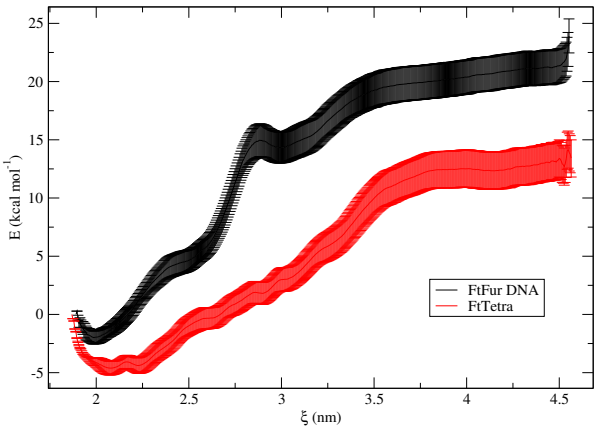


Figure 5.5: Potential of mean force (PMF) for the separation of FtFur tetramer into two dimers (FtTetra) and FtFur dimer from DNA (FtFur DNA). The error bars have been calculated from 100 histograms built with the bootstrap method and the b_{hist} option. ξ is the reaction coordinate.

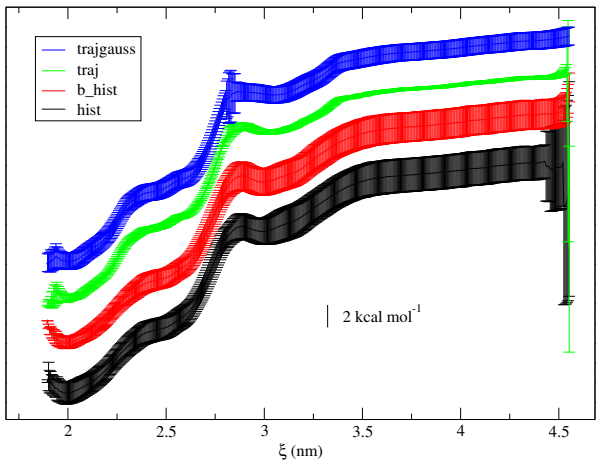


Figure 5.6: PMF for the separation of the FtFur dimer from DNA.

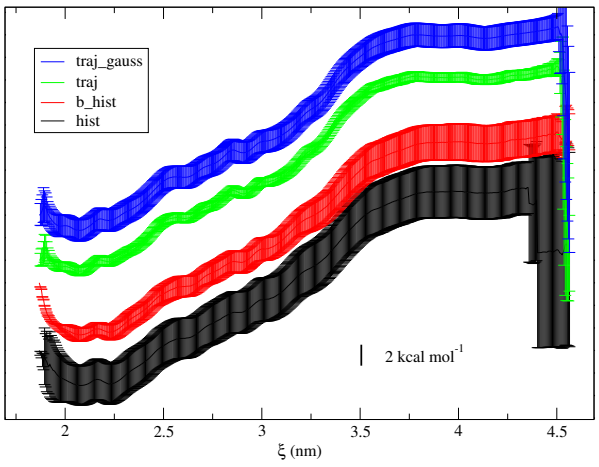


Figure 5.7: PMF for the separation of FtFur tetramer into two dimers.

Figure 5.6 and 5.7: Potential of mean force for the separation of each complex with error bars corresponding to the four methods proposed by g_{wham}. PMF have been translated in Y for better visualization.

From these graphs, it appears that the *traj* method using calculated autocorrelation times gives much lower error bars in free energy. These Integrated Average Correlation Times calculated for the 26 windows and the two reaction coordinates of the simulation of the FtFur tetramer dissociation are shown in Figure 5.8.

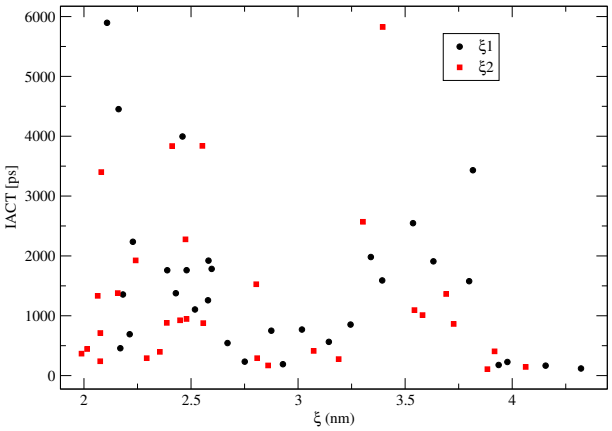


Figure 5.8: Integrated Average Correlation Times (IACT) calculated for 60 ns/window simulation of the separation of FtFur tetramer into two dimers. The two reaction coordinates ξ_1 and ξ_2 correspond to the pulling forces applied on chains A and B, respectively, to detach them from the other C and D chains in the tetramer.

These correlation times vary a lot, between 100 and 6000 ps, which could mean that the hypotheses of independence of the data points or equiprobability of picking the histograms (used in the other methods) may not be valid. Which method is the "best" between the recommended default b_{hist} and the *traj* methods remains questionable. The b_{hist} method has been used in the systems described in this manuscript to evaluate the maximum errors on the profiles.

5.2.3 Computation of average interaction energy profiles.

Once the free energy profiles were obtained, we focused on identifying key residues involved in the interaction within each complex. Interaction energy profiles, of dimer-dimer and dimer-DNA dissociation, were computed by extracting nonbonded interactions, electrostatic and Lennard Jones short range potential energies, from all the trajectories of the simulations for each amino acid, averaged for each window. Interaction energies were calculated between each residue in the Fur 'moving' dimer (chains A and B) and the 'fixed' subsystem (DNA or 'fixed' FtFur dimer) .Lennard-Jones (LJ-SR) and electrostatic interaction (Coul-SR) were extracted and summed for each residue from each window. For details on interaction energy calculation, see section 7.2.3.

When all energies are available, the conformation with the maximum interaction energy is determined and is used to normalize the visualization of the major interacting residues presented in the figures shown in the following sections. At the same time, this

maximum defines the 4% threshold below which residues are no longer considered as key residues contributing to the interaction. Once key residues are chosen, the visualisation graph is automatically generated, enabling results interpretation.

The scripts developed to perform all these steps helped us obtain a visualisation method of the evolution, throughout the whole dissociation process, of each interaction between residues from the moving subsystem and the fixed subsystem. This enabled us to get an idea about the molecular mechanisms associated with each protein complex.

Non bonded energy term choice

The gromos force field, detailed in section , includes four non bonded energy terms, only two of them (Coul-SR and LJ-SR) were used in our calculations. To make sure that we we could neglect the coulomb reciprocal and dispersion correction terms in our analysis, the four energy terms were plotted as a function of time for the whole system in the case of the starting conformations of the FtFur DNA and FtFur tetramer complexes. Results are presented in Figure 5.9 and show the relative low contribution of the coulomb reciprocal and dispersion correction terms to the total potential energy in comparison to the electrostatic and LJ potential. Taking this into consideration, in what will follow, every time we refer to the interaction energy, it will be the sum of the electrostatic and LJ potential only.

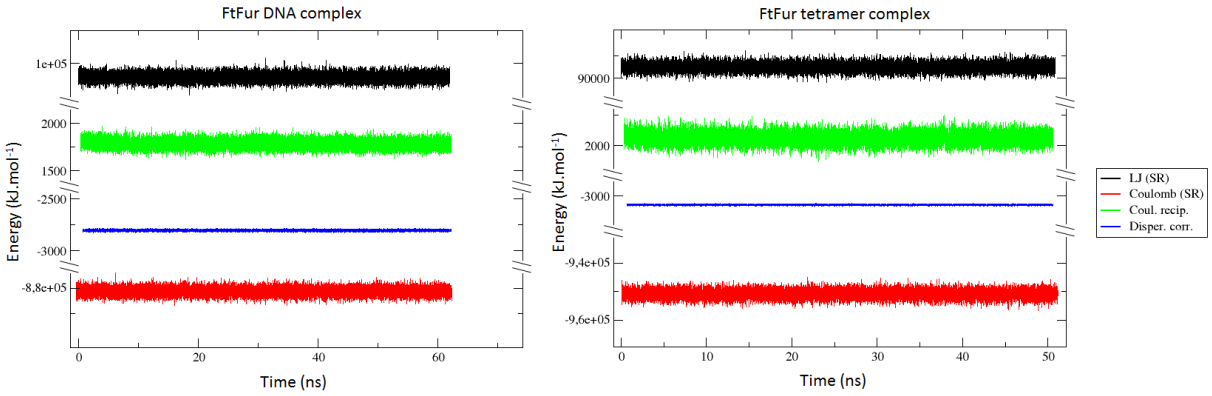


Figure 5.9: Gromos non bonded energy terms, in both FtFur simulations, are shown for the full simulation time at the first window (closest distance). LJ (SR) in black, Coulomb (SR) in red, Coul. recip. in green and Disp. corr. in blue.

5.2.4 Major interacting residues in the FtFur tetramer complex

The structure used to study the FtFur tetramer can be seen in Figure 5.10. Chains A (green) and B (yellow) were pulled away from the fixed dimer, chains C (red) and D (orange), using our simulation protocol described earlier. This visualisation method will be used to describe all the complexes studied in this manuscript. The same colours and chain names will also be respected to avoid any confusion. Chains A and B will always be the chains that were pulled away from the complex.

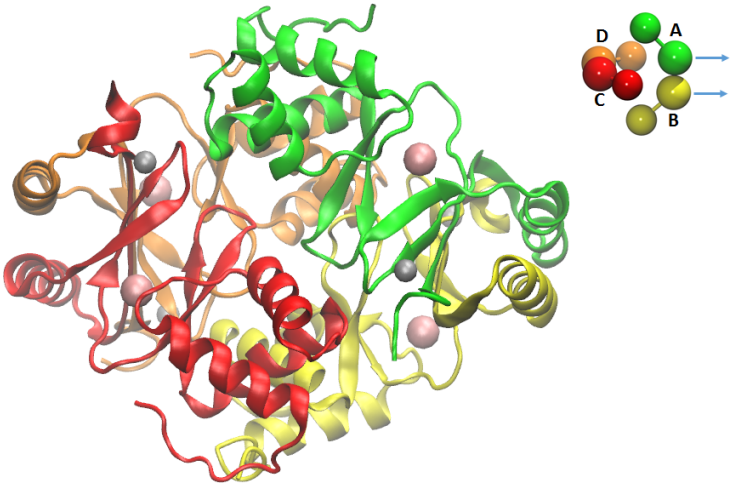


Figure 5.10: Calculated minimum energy structure of the FtFur tetramer complex. Chains A, B, C and D are represented in green, yellow, red and orange, respectively. In the upper right corner is a schematic representation of the model with coloured spheres representing each chain, and arrows indicating the dissociation direction. This representation will be used throughout this manuscript to help visualize the models.

As previously mentioned, in what will follow, the interaction energy refers to the sum of the electrostatic and LJ potential. By calculating the average individual interactions energies between each residue for a moving chain and the fixed dimer, the total interaction energy of that chain can be determined. When this total interaction energy is plotted against distance, Figure 5.11 can be generated where the behaviour of both moving chains during the dissociation can be compared.

In the case of the tetramer dissociation, both chains have a similar behaviour where their interaction with the fixed dimer decreases with distance, indicating that most probably, a symmetrical dissociation is observed, since both chains should have a similar set of interacting partners within the tetramer. This is validated by Figures 5.12 and 5.13, that show the individual interactions of residues in a chain throughout the whole simulation, where 15 common residues between chain A and B can be determined.

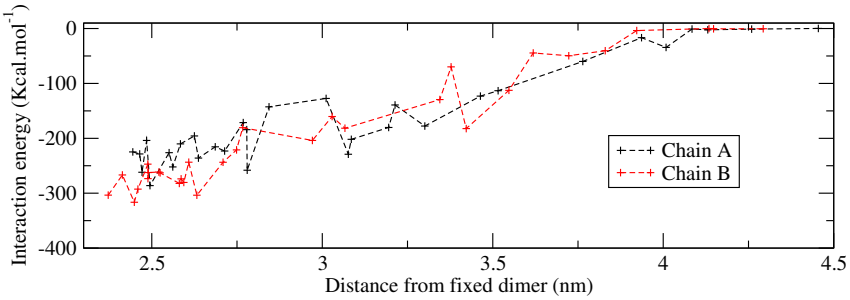


Figure 5.11: Comparison of average total interaction energies between Fur chains A and B and the fixed subsystem made by chains C and D.

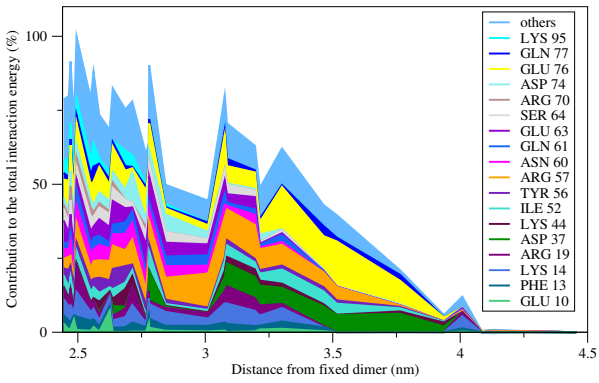


Figure 5.12: Major contributors of chain A.

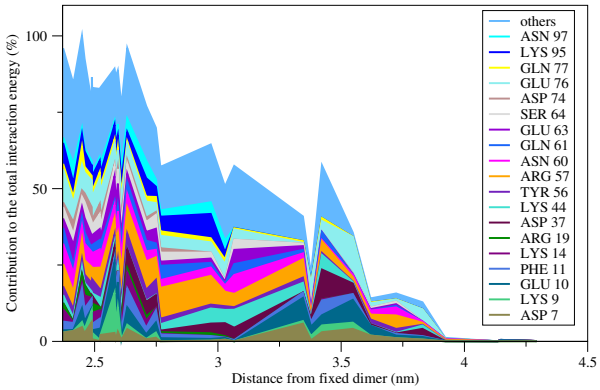


Figure 5.13: Major contributors of chain B.

Figure 5.12 and 5.13: Major contributors to the interaction energy between moving chains and the "fixed" dimer in the FtFur tetramer simulation.

In Figures 5.12 and 5.13, all residues shown have interaction energies with a value above 4% of the total interaction energy of their respective chain. This threshold was chosen to simplify the visualization and to avoid integrating various residues that had punctual interactions and could interfere with the analysis of major contributors. These graphs show contribution of individual residues, in percent, to the total interaction energy of their respective chain. For these selected residues, values are cumulated and displayed on top of each other to build, step by step, the total interaction energy profile of the chain studied. Since residue selection used a 4% threshold, and to achieve the 100% of interaction, the "others" category designates remaining interactions not displayed individually, but needed to achieve 100% interaction.

Other than displaying the contributions of major interacting residues in comparison with the total interaction energy, and comparing residues with each others, this type of visualisation gives a detailed idea about the evolution of individual interactions during the dissociation of the system. Since we are simulating a dissociation, by inverting the process we can gain insights into the complex formation. Three main types of residues can be distinguished:

- Residues present from the start to the end of the simulation, could be important for

recognition, initiation of oligomerization or DNA binding and involved in locking the complex in its final, bound, state.

- Residues present at the start but fading out before the end of the simulation, could be important for locking the final conformation in place. Theoretically this type of residues should be found on the inner surface of the interaction area.
- Residues absent at the start but present at the end of the simulation, could be important for the first series of recognition interactions, initiating complex formation.

Due to time and space limitation, only the top 5 interacting residues, of each chain, will be detailed in tables 5.3 and 5.4. They show their maximum contribution, in terms of energy or percent of the maximum interaction energy of the chain. The average distance between the center of mass of fixed and moving chains at the corresponding window is given in column 4.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Glu76	-43.1	15.0	3.51
Arg57	-39.4	13.7	3.07
Glu63	-25.3	8.8	2.76
Lys14	-23.9	8.3	2.49
Asn60	-17.1	5.9	2.49

Table 5.3: Top 5 interacting residues of chain A in the FtFur tetrameric complex, interaction energy corresponds to the average over all the conformation of the window, the contribution is given in percent of the maximum interaction energy of the chain. Average distance between the center of mass of chains in the window where this maximum is obtained is given in column 4.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg57	-34.8	11.0	2.58
Glu76	-37.5	11.8	3.54
Asn60	-21.9	6.9	2.59
Lys95	-25.1	7.9	2.97
Glu10	-31.4	9.9	2.58

Table 5.4: Top 5 interacting residues of chain B in the FtFur tetrameric complex, interaction energy corresponds to the average over all the conformation of the window, the contribution is given in percent of the maximum interaction energy of the chain. Average distance between the center of mass of chains in the window where this maximum is obtained is given in column 4.

Ideally, for each one of the top 5 residues the interacting partner(s) from the fixed subsystem can be determined by visualizing the window at which the maximum interaction energy is observed. This is done by extracting the mean conformation of that window

and minimising it before the visualization. However, this process is time consuming due to the number of residues to be checked in all chains, of all the conformations. This is why the complete list of partners predicted by the simulations will not be presented in this manuscript at this time. They will be probably presented during the PhD defence and included in the final version of this manuscript.

Meanwhile, as a temporary shortcut, partners of the top 5 residues, determined by our calculations, were identified by visualizing the conformation corresponding to the starting point of the simulation, and are shown in table 5.5. For some residues, the initial conformation did not deliver any information about the interacting partners (bad residue orientation). In these cases, the partners were found in the mean conformation of the window at which the involved residue had a maximum interaction energy, they are shown in blue in table 5.5.

Chain A		Chain B	
Residue	Partner	Residue	Partner
Glu76	Thr54(C)	Arg57	Glu63(C)
Arg57	Glu63(D)	Glu76	Arg70(D)
Glu63	Arg57(D)	Asn60	Ser64(C)
Lys14	Glu10(D)	Lys95	Ser48(D)
Asn60	Glu76(C)	Glu10	Phe11(C)

Table 5.5: Top 5 interacting residues of the FtFur tetramer complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

The previous paragraphs gave a detailed description about the different methods used and showed how all the results are generated in the case of the FtFur tetramer complex. In what will follow, for each complex, the results will be listed without detailing the protocol. A small paragraph will highlight any important information. And the remaining results will be discussed as a whole in the end of this chapter.

5.3 The wild type tetramer from *P. aeruginosa*

For our study on Fur tetramers, we focused on three complexes: two tetramer from *P. aeruginosa* and one from *V. cholerae*.

The model of the PaFur tetramer was built from PDB structure 1MZB (Pohl et al., 2003) with the PISA program (Krissinel et al., 2007) and can be seen in Figure 5.14.

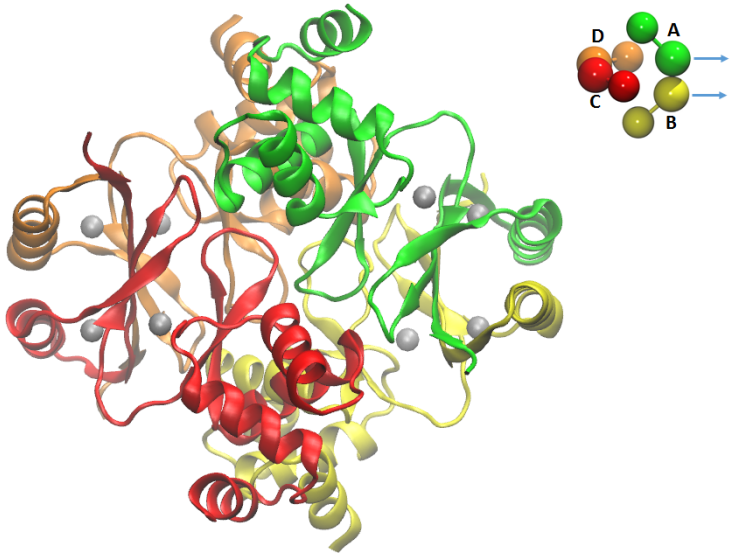


Figure 5.14: Calculated minimum energy structure of the PaFur-WT tetramer complex.

Figures 5.15, 5.16 and 5.17 show that 11 interacting residues are common for the two moving chains, with Arg56 and Glu62 being important for this interaction. It is interesting to note that Glu36 in the case of chain B is a major contributor to the interaction energy towards the end of the dissociation.

Tables 5.8 indicates that a strong interaction is taking place between Arg56(A,B) with Glu62(D,C), respectively. That indicates the importance of this interaction in tetramer stabilization.

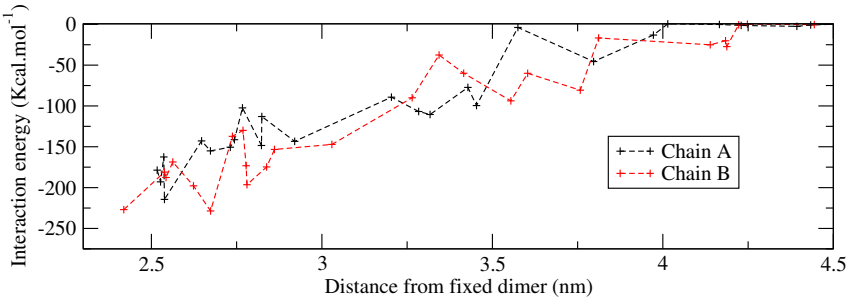


Figure 5.15: Comparison of average total interaction energies between PaFur chains A and B and the fixed subsystem made by chains C and D.

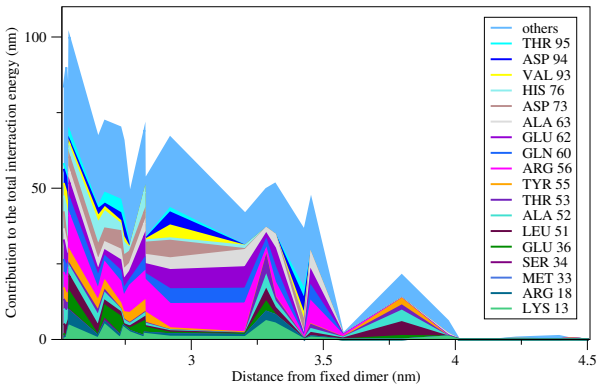


Figure 5.16: Major contributors of chain A.

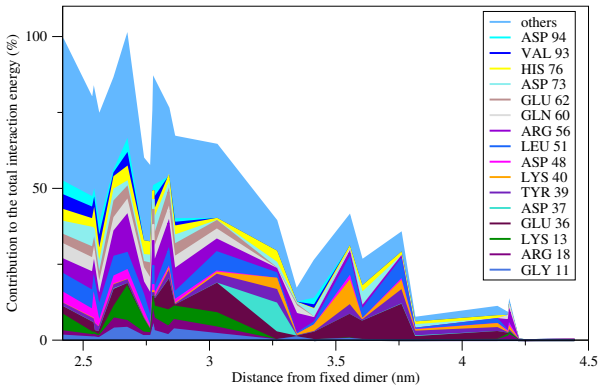


Figure 5.17: Major contributors of chain B.

Figure 5.16 and 5.17: Major contributors to the interaction energy between moving chains and the "fixed" dimer in the PaFur tetramer simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg56	-25.3	11.8	2.53
Glu62	-16.3	7.6	2.82
Gln60	-11.5	5.3	2.53
Ala63	-12.2	5.7	3.20
Asp73	-13.0	6.0	2.73

Table 5.6: Top 5 interacting residues of chain A in the PaFur tetrameric complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg56	-28.3	12.4	2.67
Glu36	-26.4	11.5	3.75
Leu51	-15.1	6.6	3.75
Gln60	-12.0	5.2	2.77
His76	-12.2	5.3	2.54

Table 5.7: Top 5 interacting residues of chain B in the PaFur tetrameric complex.

Chain A		Chain B	
Residue	Partner	Residue	Partner
Arg56	Glu62(D)	Arg56	Glu62(C)
Glu62	Arg56(C)	Leu51	Asp94(D)
Gln60	Ala63(D)	Glu36	Leu51(D), Ala52(D)
Ala63	Gln60(D)	Gln60	Ala63(C)
Asp73	Thr59(C), Arg56(C)	His76	His76(D)

Table 5.8: Top 5 interacting residues of the PaFur tetramer complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

5.4 The ΔS3 mutant from *P. aeruginosa*

The structure of the PaFurΔS3 mutant was obtained during this work and was used to build the model presented in Figure 5.18. The mutant lacks the S3 metal site and was studied in this simulation to find out if this mutation affects in any way the stability of the tetramer.

Figures 5.19, 5.20 and 5.21, show that during the dissociation, the two moving chains share 16 common residues, and 11 residues with the PaFur wild type tetramer. This indicates that the mutation of site S3 does afect dramatically the stability of the tetramer.

Similarly to the case of the wild type protein, Glu36 of chain B is a major contributor towards the end of the dissociation. In addition, Tables 5.9 and 5.10, indicate the importance of Arg56 that interacts with Gln62 (5.11)

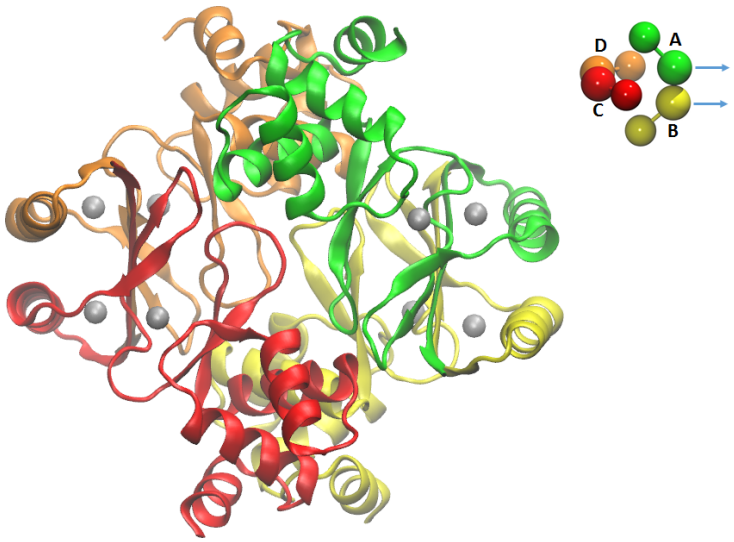


Figure 5.18: Calculated minimum energy structure of the PaFurΔS3 tetramer complex.

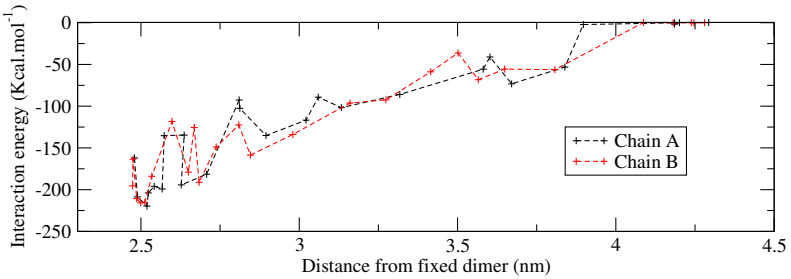


Figure 5.19: Comparison of average total interaction energies between PaFurΔS3 chains A and B and the fixed subsystem made by chains C and D.

16 residues in common between both chains. 11 residues in common between PaDS3 and PaWT. chain A and B arg 56 is important and glu36 is important at the end of the simulation.

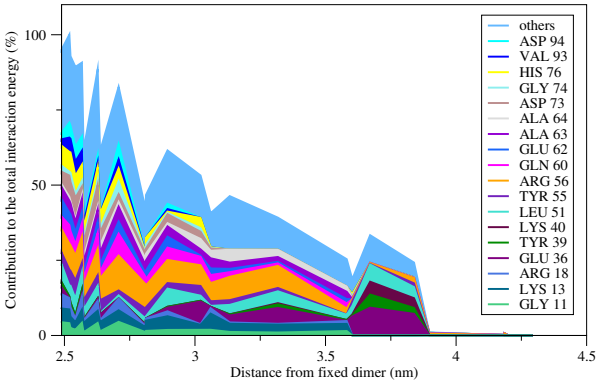


Figure 5.20: Major contributors of chain A.

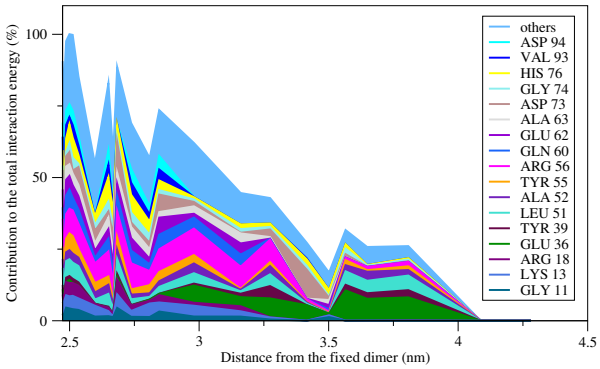


Figure 5.21: Major contributors of chain B.

Figure 5.20 and 5.21: Major contributors to the interaction energy between moving chains and the "fixed" dimer in the PaFurΔS3 tetramer simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg56	-25.4	11.5	2.70
Leu51	-13.1	5.9	2.89
Gln60	-16.0	7.3	2.70
Lys13	-11.8	5.3	3.06
His76	-14.8	6.7	2.52

Table 5.9: Top 5 interacting residues of chain A in the PaFurΔS3 tetrameric complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg56	-20.8	9.6	2.47
His76	-18.2	8.4	2.64
Leu51	-12.8	5.9	2.47
Asp73	-23.2	10.7	2.68
Gln60	-14.5	6.7	2.49

Table 5.10: Top 5 interacting residues of chain B in the PaFurΔS3 tetrameric complex.

Chain A		Chain B	
Residue	Partner	Residue	Partner
Arg56	Glu62(D)	Arg56	Glu62(C)
Leu51	Asp94(C)	His76	Glu36(D), His76(D)
Gln60	Ala63(D)	Leu51	Asp94(D)
Lys13	Gly11(D)	Asp73	Leu51(D), A52(D)
His76	Glu36(C), His76(C)	Gln60	Ala63(C)

Table 5.11: Top 5 interacting residues of the PaFurΔS3 tetramer complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

5.5 A "false" tetramer from *V. cholerae*

Although it does not form in the case of VcFur (Pérard et al., 2016), a tetramer was built as a negative control to test our protocol. The structure of the tetramer was built starting from a sequence alignment between VcFur (PDB ID: 2W57, Sheikh et al., 2009) and PaFur, followed by a series of superimposition between the two structure. Details on the construction can be found in section 7.6 and the model is shown in Figure 5.22.

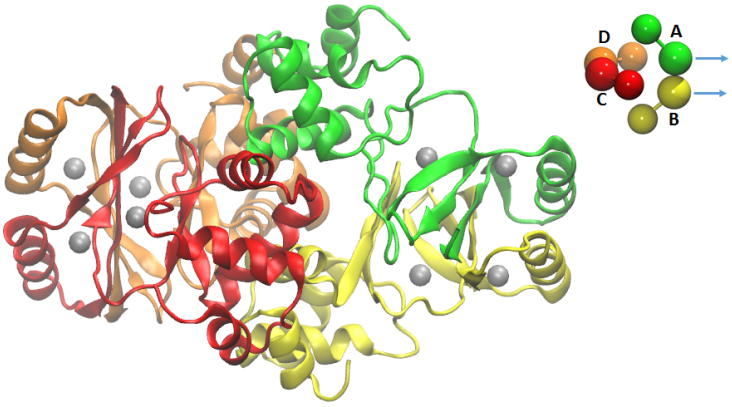


Figure 5.22: Calculated minimum energy structure of the VcFur tetramer complex.

Figures 5.23, 5.24 and 5.25 show that during the dissociation, the two chains exhibit a similar behaviour. They share 10 common interacting residues with Asp64 interacting the most with the fixed dimer subsystem (Tables 5.12 and 5.13).

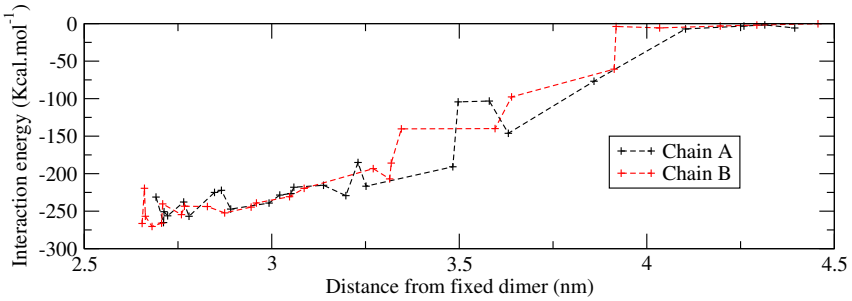


Figure 5.23: Comparison of average total interaction energies between VcFur chains A and B and the fixed subsystem made by chains C and D.

From table 5.14, we can see that the moving chain A mainly interacts with the fixed chain D, similarly, the moving chain B interacts with the fixed chain C.

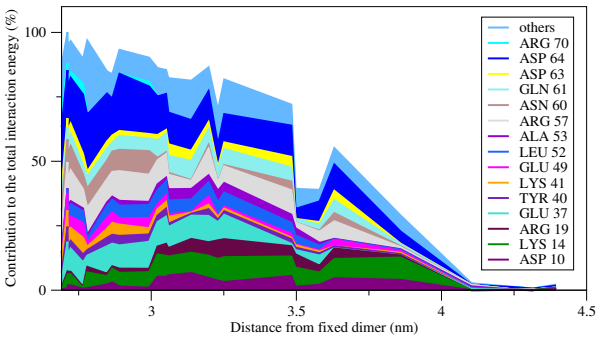


Figure 5.24: Major contributors of chain A.

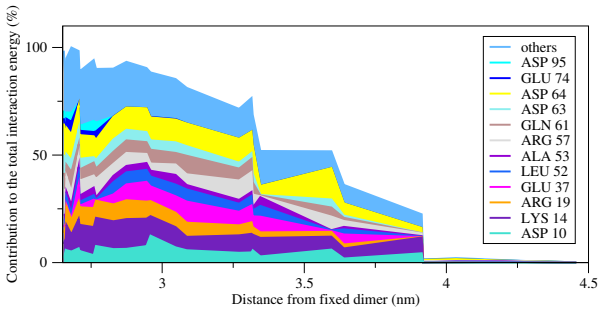


Figure 5.25: Major contributors of chain B.

Figure 5.24 and 5.25: Major contributors to the interaction energy between moving chains and the "fixed" dimer in the VcFur tetramer simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Asp64	-61.9	23.3	2.69
Arg57	-31.7	11.9	2.72
Glu37	-30.5	11.5	2.71
Lys14	-25.5	9.6	3.25
Gln61	-19.8	7.4	3.13

Table 5.12: Top 5 interacting residues of chain A in the VcFur tetrameric complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Asp64	-39.5	14.6	3.59
Lys14	-37.4	13.8	2.87
Arg19	-23.8	8.8	2.75
Arg57	-33.7	12.5	2.65
Asp10	-33.9	12.5	2.95

Table 5.13: Top 5 interacting residues of chain B in the VcFur tetrameric complex.

Chain A		Chain B	
Residue	Partner	Residue	Partner
Asp64	Arg19(D)	Asp64	Gln61(C)
Arg57	Asp63(D), Asp64(D)	Lys14	Asp64(C)
Glu37	Leu52(C), Ala53(C)	Arg19	Asp64(C)
Lys14	Asp10(D)	Arg57	Glu75(D)
Gln61	Asp64(D)	Asp10	Lys14(C)

Table 5.14: Top 5 interacting residues of the VcFur tetramer complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

5.6 Summary of PMFs for Fur tetramer models

Figure 5.26 and Table 5.15 compare the PMFs of all the tetrameric complexes that we simulated. By comparing the two profiles of PaFur and PaFurΔS3, we can see a difference of 1.25 kcal.mol^{-1}. This should be compared with physiological or in vitro stability of these two Fur tetramers.

Interestingly, the ΔG of binding of FtFur tetramer is lower than that of PaFur. This result should be compared to data collected *in vitro* in order to interpret our results. The crystal structures of PaFurΔS3 show more metal atoms bound to the protein than in the case of FtFur. This can be one possible explanation of the different behaviour observed here. PaFurΔS3 could potentially interact with the counter ions used in our simulation and gain in stability. This hypothesis needs to be confirmed when the unconventional metal sites in the structure are well characterised.

In the case of the VcFur tetramer, the fact that the minimum of the free energy profile is found at a COM-COM distance of 27 to 28 Å instead of 21 to 22 Å for other tetramers, could be due to a bad packing of the loops between the dimerization domain and the DNA binding domain. This region extends more easily during the pulling, leaving the moving DBD less affected by the translation and thus likely to maintain their interactions with the fixed DBD, or even enhance them. More work is necessary to check this model in depth. However, this complex has the lowest ΔG of binding and validates, as a negative control, our simulation protocol.

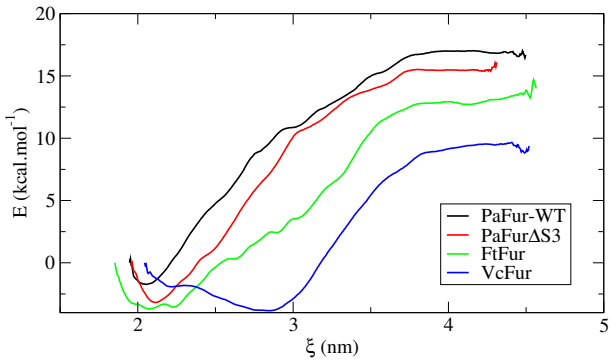


Figure 5.26: Plot showing the potential mean force curves of the tetrameric Fur complexes.

Tetramer complex	ΔG (kcal.mol^{-1})
FtFur	16.71 ±1.72
PaFur	19.25 ±2.20
PaFurΔS3	20.50 ±1.73
VcFur	13.95 ±2.32

Table 5.15: ΔG of binding of Fur tetramer complexes.

5.7 FtFur + Fur box

In our study on Fur complexes, apart from understanding tetramer stabilization and the involved residues, we were looking for mechanistic insights into Fur binding to DNA. This was done through the study of several Fur/DNA complexes, based on already obtained crystal structures (except FtFur/DNA).

To study the dissociation of FtFur from DNA, and in the absence of a structure for this complex, a model was created using the structure of MgFur with the Fur box. Details on model construction are given in section 7.6. The model used in our study is shown in Figure 5.27.

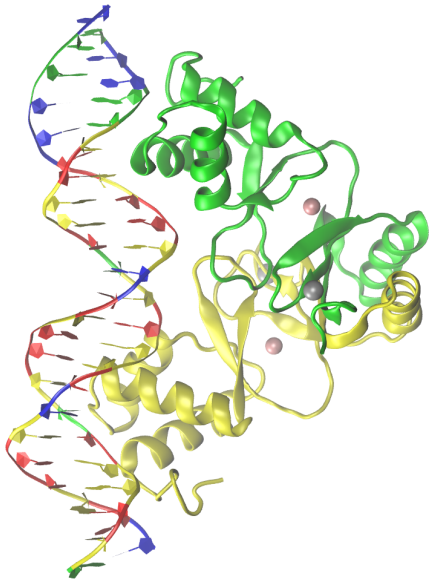


Figure 5.27: Calculated minimum energy structure of the FtFur DNA complex. Chain A is shown in green and B in yellow. Nucleotides : A(yellow), T(red), G(green), C(blue).

Figure 5.28 shows an asymmetric dissociation of FtFur from the Fur box of E. coli. Chain B starts with a higher interaction energy and a closer center of mass distance to the DNA than chain A. This will be discussed in more detail in section 5.10.

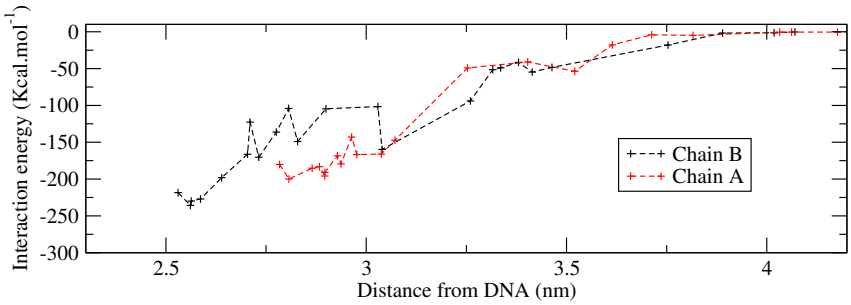


Figure 5.28: Comparison of average total interaction energies between FtFur chains A and B and the fixed DNA subsystem.

For both chains, 13 residues are in common and can be seen in Figures 5.29 and 5.30.

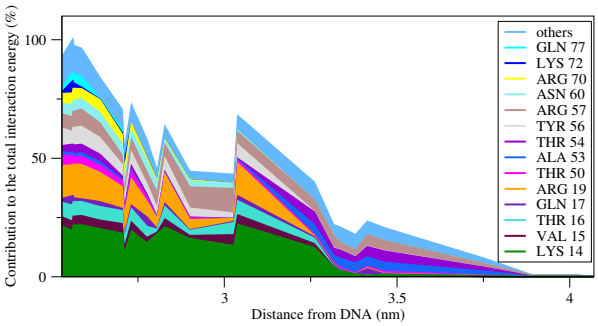


Figure 5.29: Major contributors of chain A.

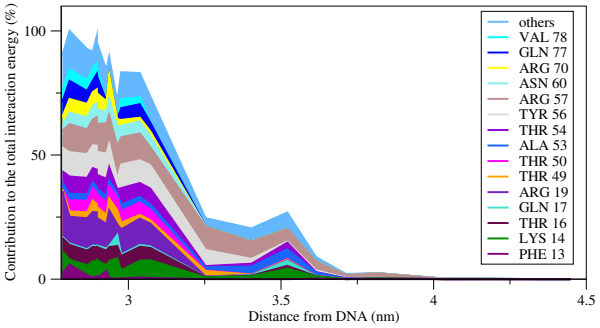


Figure 5.30: Major contributors of chain B.

Figure 5.29 and 5.30: Major contributors to the interaction energy between moving chains and the Fur box in the FtFur / Fur box simulation.

Two sets of adjacent residues, [14,16,17,19] and [50, 53, 56, 57, 60] reflect the helices in interaction with DNA. For both chains, the 4 most important contributors to the total interaction energy are the same: Lys14, Arg19, Tyr56 and Arg57, as shown in Tables 5.17 and 5.16. It is interesting to note that these four residues are also found in the contributors to the interaction energy in the FtFur tetramer simulation (Figures 5.12 and 5.13). Lys14 is a major contributor to the total interaction energy of chain A (22% at 3Å) with a specific interaction with A5(Z), as shown in Table 5.18.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg57	-26.1	13.0	2.96
Arg19	-36.9	18.4	2.78
Tyr56	-20.0	10.0	2.92
Lys14	-18.0	9.0	2.78
Thr54	-14.1	7.0	2.92

Table 5.16: Top 5 interacting residues of chain A in the FtFur Fur box complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys14	-52.0	22.8	3.03
Arg19	-34.2	14.5	2.58
Arg57	-24.3	10.3	3.02
Tyr56	-17.9	7.6	2.58
Thr16	-16.0	6.8	2.56

Table 5.17: Top 5 interacting residues of chain B in the FtFur Fur box complex.

Chain A			Chain B		
Residue	Partner	Type	Residue	Partner	Type
Arg57	A17(Z)	non-specific	Lys14	A5(Z)	specific
Arg19	G4(Y)	non-specific	Arg19	T6(Z)	non-specific
Tyr56	T15(Z)	non-specific	Arg57	G7(Z)	specific
Lys14	C3(Y)	non-specific	Tyr56	T13(Y)	non-specific
Thr54	G5(Y)	non-specific	Thr16	T6(Z)	non-specific

Table 5.18: Top 5 interacting residues of the FtFur DNA complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

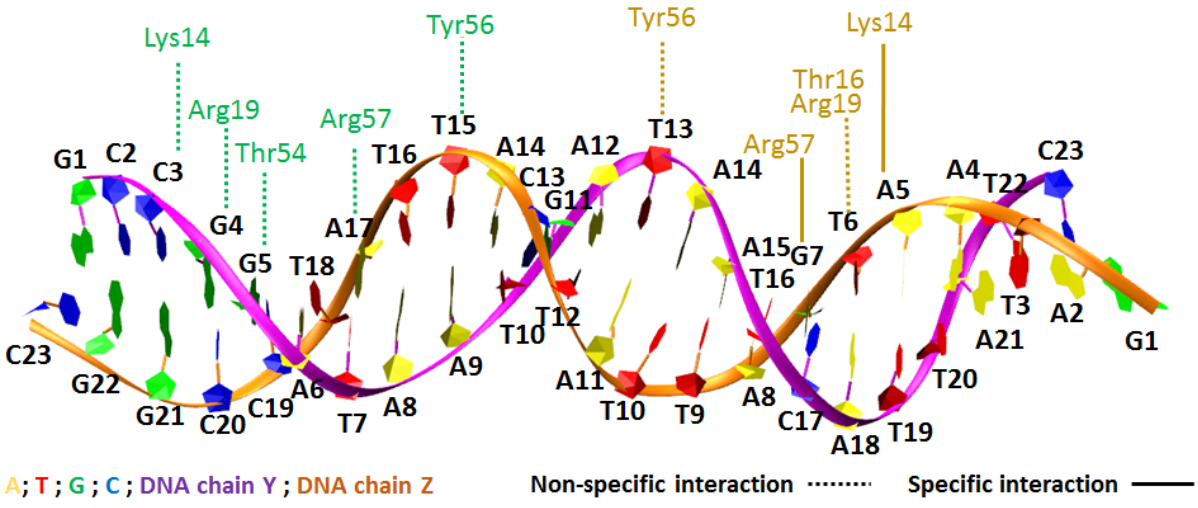


Figure 5.31: Interactions during the FtFur dimer DNA simulation. Green and yellow residues are from the protein chain A and B, respectively.

When our previous results obtained on both FtFur complexes Pérard et al., 2018, are compared to the ones described here, the major interacting residues involved in interaction with DNA did not change. The only difference is the Lys95 and Glu10 classified as fourth and fifth best contributors, respectively, in the tetramer simulation.

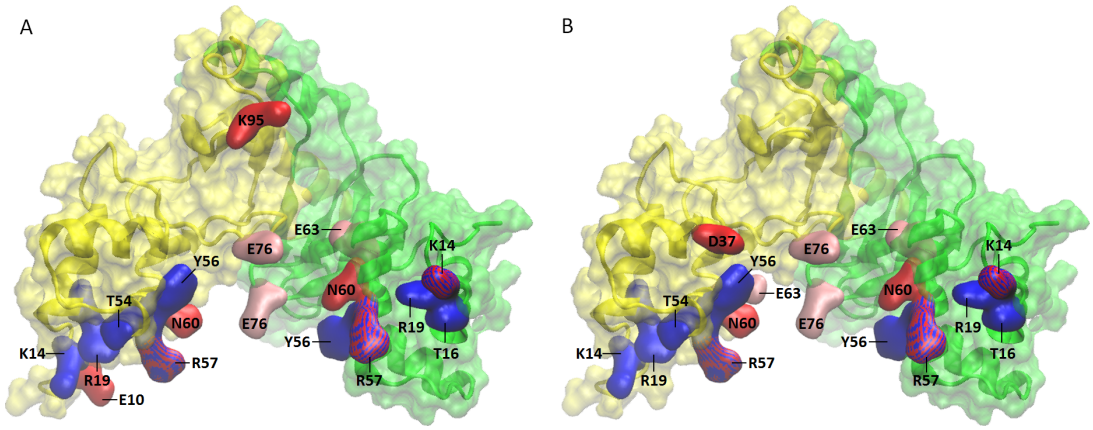


Figure 5.32: Visualization of the major interacting residues in the FtFur complexes. Blue residues are interacting with DNA, red and pink residues are involved in tetramer interactions. Pink residues were mutates *in vitro* and were shown to destabilize the tetramer. A: results obtained after 60 ns of simulation time per window. B: published results, found after 15 ns of simulation Pérard et al., 2018.

5.8 Fur from *M. gryphiswaldense* + Fur box

The structure of Fur from *Magnetospirillum gryphiswaldense* (MgFur) in the presence of the *E. Coli* Fur box (PDB ID: 4RB1 Deng et al., 2015) was used in this model presented in Figure 5.33. This 4RB1 structure is one of the few available Fur structures with DNA, and was included in this study to generate data on an active Fur conformation bound to the Fur box.

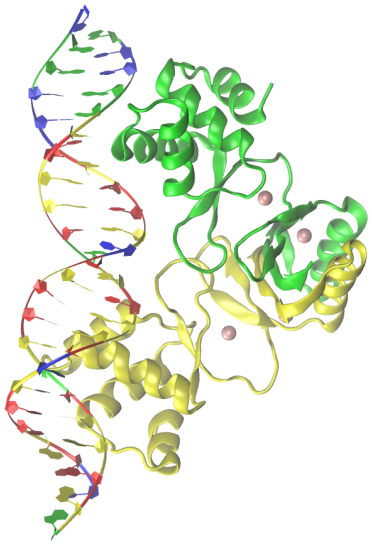


Figure 5.33: Calculated minimum energy structure of the MgFur DNA complex.

Figures 5.34 and 5.35 show that both protein chains have the same interacting residues with DNA, when a 4% threshold of total interaction energy is used. For example, in both chains, Arg49 contributes mainly towards the end of the dissociation .Even if the general interaction behaviour is similar, the contribution of some residues are different in each chain.

From Figure 5.35 and Tables 5.19 and 5.20, Lys15 is the residue with the maximum contribution to the total interaction energy between Fur and the DNA sequence. In fact, Lys15 of chain B, part of the L1 loop, inserts into the minor groove with specific interactions with T3-O2 (Z) and T22-O2 (Y). This same residue in chain A makes non specific interactions with the phosphate groups of C2(Y) and C3(Y). All charged arginines in Tables 5.19 and 5.20, also strongly interact with the phosphate groups of their respective partners.

If we analyse more residues with lower total interaction energies in chain B (ranked 6^{th}, 7^{th}, ... in Table 5.20), we find that the phenyl ring of Tyr56, part of the α4 helix, forms van der Waals interactions with T13(Y) as seen by Deng et al., 2015, and with T16(Z) in the major groove. In addition, Arg57 inserts into the major groove forming hydrogen bonds between its guanidinium group and atom O6 of G7(Z). Moreover, Ser51 hydroxyl group interacts specifically with A6(Y).

In protein chain A, the side chains of residues Thr17 and Gln19 from loop L1 interact with phosphate groups in the minor groove (Thr17 with G4 (Y) and Gln19 with G5 (Y)). Note that Thr17 forms 5% of maximum contribution to the total interaction energy with DNA, above the 4% threshold limit, and thus is visible in Figure 5.34. This residue is not present in Table 5.19 because its contribution is not one of the fifth best. Similarly, these two residues (Thr17 and Gln19) can be seen in Figure 5.35 for chain B, but are not represented in the top five Table 5.20.

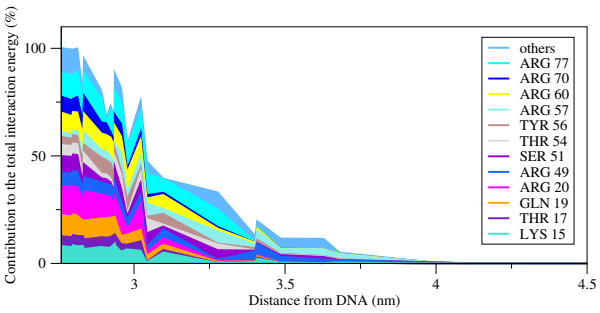


Figure 5.34: Major contributors of chain A.

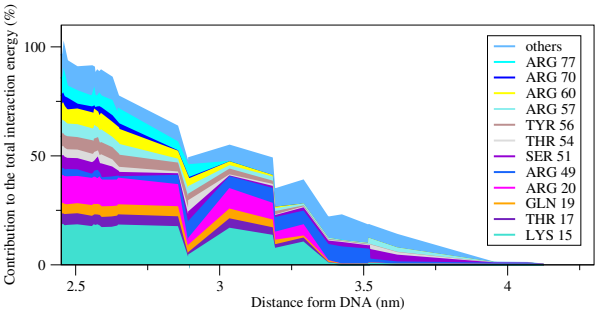


Figure 5.35: Major contributors of chain B.

Figure 5.34 and 5.35: Major contributors to the interaction energy between moving chains and the Fur box in the MgFur / Fur box simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg77	-30.9	11.8	2.81
Arg20	-34.8	13.5	2.83
Arg60	-24.3	9.4	2.97
Gln19	-25.2	9.7	2.75
Lys15	-23.7	9.2	2.93

Table 5.19: Top 5 interacting residues of chain A in the MgFur Fur box complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys15	-56.4	18.2	2.45
Arg20	-39.5	12.8	2.46
Arg60	-24.1	7.8	2.45
Arg49	-29.9	9.6	2.89
Arg77	-33.9	11.0	2.45

Table 5.20: Top 5 interacting residues of chain B in the MgFur Fur box complex.

Chain A			Chain B		
Residue	Partner	Type	Residue	Partner	Type
Arg77	A14(Z), T15(Z)	non-specific	Lys15	T3(Z), A4(Z), T22(Y), A21(Y)	specific
Arg20	G4(Y)	non-specific	Arg20	A5(Z)	non-specific
Arg60	T16(Z) A17(Z)	non-specific	Arg60	A14(Y), A15(Y)	non-specific
Gln19	G4(Y), G5(Y)	non-specific	Arg49	A8(Z)	non-specific
Lys15	C2(Y) C3(Y)	non-specific	Arg77	A12(Y), T13(Y)	non-specific

Table 5.21: Top 5 interacting residues of the MgFur Fur box complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

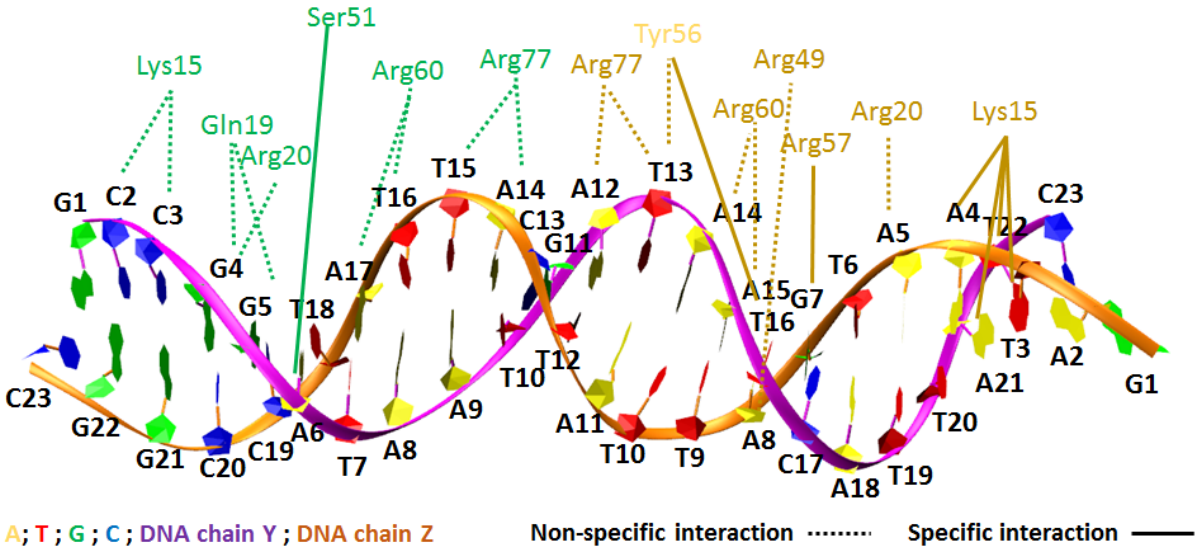


Figure 5.36: Interactions during the MgFur dimer DNA simulation. Green and yellow residues are from the protein chain A and B, respectively.

5.9 Fur from *M. gryphiswaldense* + *feoAB1*

The *feoAB1* gene codes for a Fe^{2+} transport protein involved in ferrous iron uptake (McHugh et al., 2003). To see the impact of a different DNA sequence in a Fur/DNA complex, the model shown in Figure 5.37 was built using the crystal structure 4RB3 obtained by Deng et al., 2015. The difference between the two DNA sequences is shown later in Table 5.25.

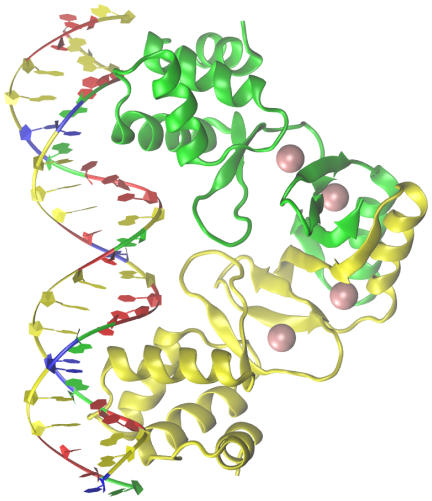


Figure 5.37: Calculated minimum energy structure of the MgFur *feoAB1* complex.

When bound to the operator of *feoAB1*, as in the case of MgFur + Fur box, Lys15 is the residue with the maximum contribution to the total interaction energy between Fur and DNA (Tables 5.22 and 5.23). Table 5.24 shows that in chain A, Lys15, from the L1 loop, inserts into the minor groove of the DNA with specific interaction with A3(Y), T4(Y) and T22(Z). In chain B, Lys15 interacts specifically with T22(Y) and G23(Y). In both tables, all the arginines described strongly interact with phosphate groups.

While analysing the results of residues with lower contribution energy, not shown in the tables, Tyr56, a residue from the α4 helix, is detected as in MgFur + Fur box simulation, in chains A and B. This residue forms van der Waals interactions, through its phenyl ring, with the methyl group of T14 (Z and Y). In addition, Arg57, from chain B, inserts into the major groove and forms hydrogen bonds between its guanidium group and atom O6 of G17(Z) in the first simulation windows. Similar interactions are seen for Arg57 from chain A in subsequent windows. Moreover, the hydroxyl group of Ser51 in chain B interacts specifically with G6(Z) and in chain A, the interaction of this residue is non specific with G6(Y). Tyr56, in both chains, interacts with the phosphate groups of T14(Y) and T14(Z). In this complex, similarly to the MgFur DNA complex, Thr17 and Gln19 also interact with phosphate groups of T5 in both DNA chains. Arg49 shows an increase in interaction towards the end of the dissociation in both chains. Globally the interactions between MgFur and the *feoAB1* operator are similar to those described above for MgFur bound to a Fur box (residues are the same in Figures 5.38, 5.39 and

5.34, 5.35).

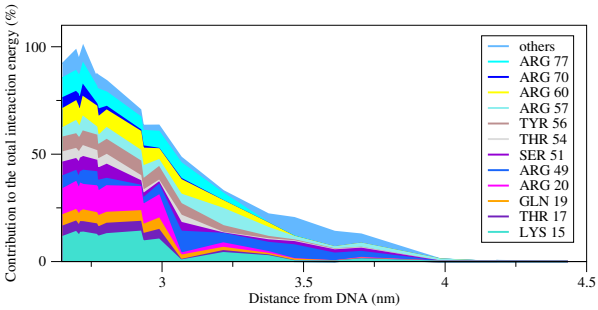


Figure 5.38: Major contributors of chain B.

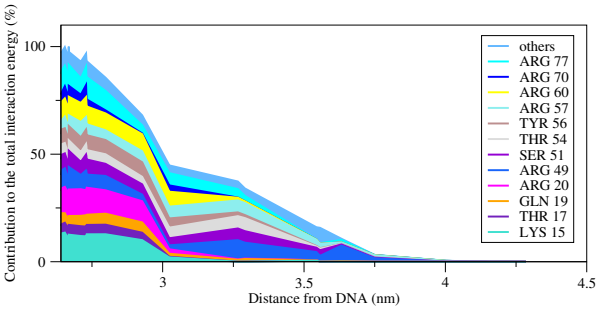


Figure 5.39: Major contributors of chain A.

Figure 5.38 and 5.39: Major contributors to the interaction energy between moving chains and the *feoAB1* box in the MgFur / *feoAB1* box simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys15	-40.4	13.9	2.92
Arg20	-37.7	13.0	2.69
Arg77	-31.7	10.9	2.71
Arg60	-28.1	9.7	2.76
Tyr56	-20.8	7.2	2.76

Table 5.22: Top 5 interacting residues of chain A in the MgFur *feoAB1* complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys15	-39.6	13.5	2.65
Arg20	-36.8	12.5	2.73
Arg49	-32.7	11.1	2.66
Arg77	-30.3	10.3	2.73
Arg60	-27.0	9.2	2.72

Table 5.23: Top 5 interacting residues of chain B in the MgFur *feoAB1* complex.

Chain A			Chain B		
Residue	Partner	Type	Residue	Partner	Type
Lys15	T22(Z), A3(Y), T4(Y)	specific	Lys15	T22(Y), G23(Y)	specific
Arg20	T4(Y), T5(Y)	non-specific	Arg20	T4(Z), T3(Z)	non-specific
Arg49	G6(Y), T22(Z)	non-specific	Arg77	A13(Y) T14(Y)	non-specific
Arg77	A13(Z)	non-specific	Arg60	T15(Y), T16(Y)	non-specific
Arg60	T15(Z), T16(Z)	non-specific	Tyr56	T14(Z)	non-specific

Table 5.24: Top 5 interacting residues of the MgFur *feoAB1* complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

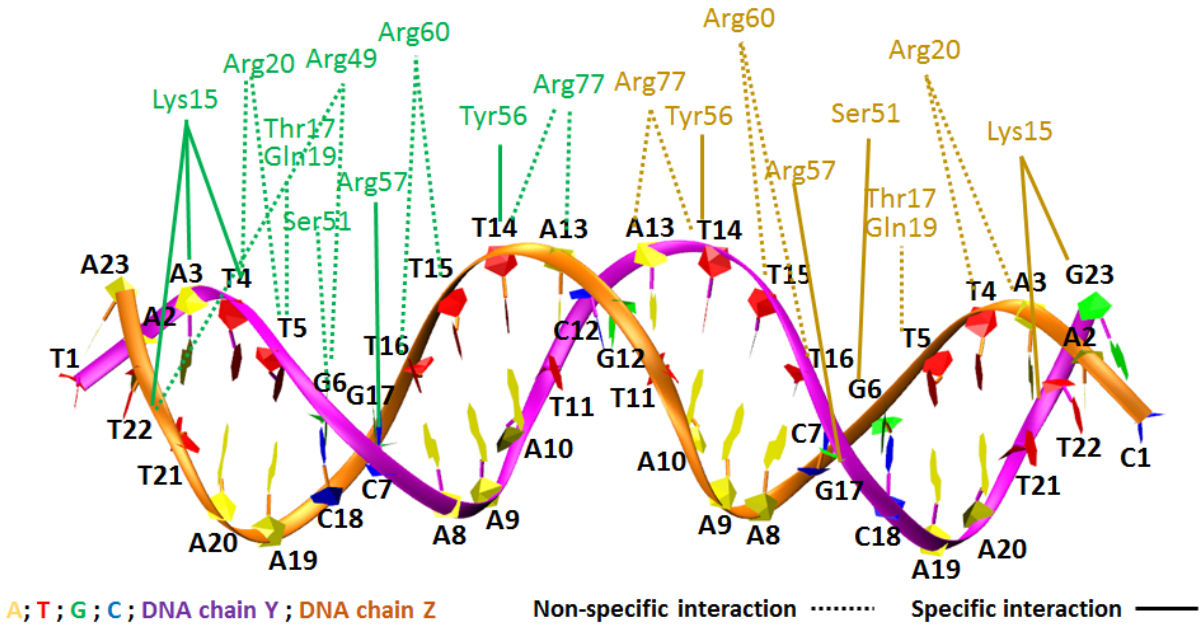


Figure 5.40: Interactions during the MgFur + *feoAB1* simulation. Green and yellow residues are from the protein chain A and B, respectively.

5.10 Comparing MgFur binding to different DNA sequences

The comparison of the two MgFur simulations shows that in each case, both chains have the same residues interacting with DNA. In addition, the contributions to total interaction energy are remarkably similar for both. When the average total interaction energy of each chain is plotted in the MgFur + Fur box simulation, an asymmetric dissociation is observed, especially at the start, with chains having different total interaction energies and distances from DNA (Figure 5.41). In contrast, in the case of the MgFur + *feoAB1*, the no asymmetric dissociation is observed between the two chains and a smooth decrease in interaction energy noted. Both chains start at relatively the same distance and energy (5.42).

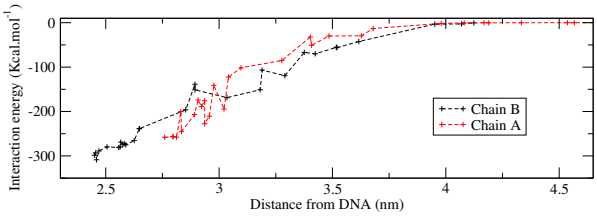


Figure 5.41: MgFur + Fur box simulation

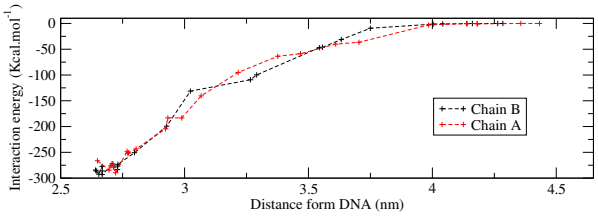


Figure 5.42: MgFur + feoAB1 simulation

Figure 5.41 and 5.42 : Comparison of average total interaction energies between MgFur chains A and B and the DNA fixed subsystem.

When free energy profiles of the dissociation of MgFur from both DNA sequences are compared, it is clear that a difference at close distances is present.

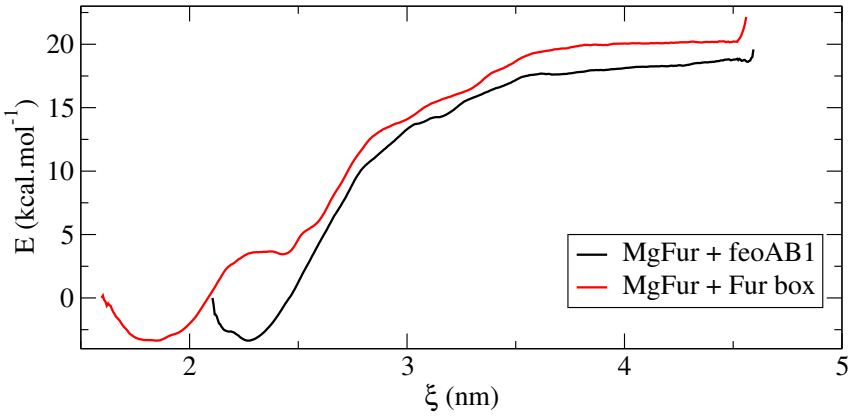


Figure 5.43: Potential of mean force for the separation of MgFur from Fur Box and *feoAB1*.

FtFur and MgFur simulations with the Fur box seem to dissociate asymmetrically, with one subunit detaching from the Fur box before the other. In the Fur dimer / DNA simulations, the structures were initially oriented so that the principal axis of the DNA duplex was oriented along the Y coordinate axis and that the Fur dimers were oriented with one of their principal axis (the one which approximately divides the dimers into

symmetric subunits) along the pulling direction X. Due to how the protein interacts with DNA, this had for consequence that one subunit of the dimer was initially closer to the center of mass of DNA in the X direction than the other. For instance after window 0 in the pulling simulations the center of mass of FtFur subunit A was 4.5 Å closer than subunit B to the center of mass of DNA. This has a direct effect in the shape of the interaction profiles (Figure 5.28). A second consequence is a kind of biphasic shape of the potential of mean force profile with two minima centred approximately 5 Å apart (Figures 5.4 and 5.5).

This effect is even more pronounced in the case of the binding of MgFur dimer to the Fur box: after window 0, the center of mass of subunit B is 6.7 Å closer than that of subunit A to the DNA dimer in the X direction. Again, this reflects both in the PMF profile of Figure 5.43 and the interaction profile of Figure 5.41. On the other hand, the binding of MgFur to the *feoAB1* promoter is almost symmetrical, with only 1 Å distance difference between the center of mass center of mass distances in early simulations, as seen in Figure 5.42.

Figure 5.44 shows the asymmetry in the case of MgFur simulations. In the case of MgFur bound to a Fur box and MgFur bound to *feoAB1*, the final protein conformations were extracted and superimposed to generate this image. The superimposition was done on all phosphorus atoms (spheres) with VMD (Humphrey et al., 1996) for a total rmsd of $\simeq 3$ Å. The figures are carefully aligned on the coordinate axis to highlight the difference in positioning of Fur on DNA.

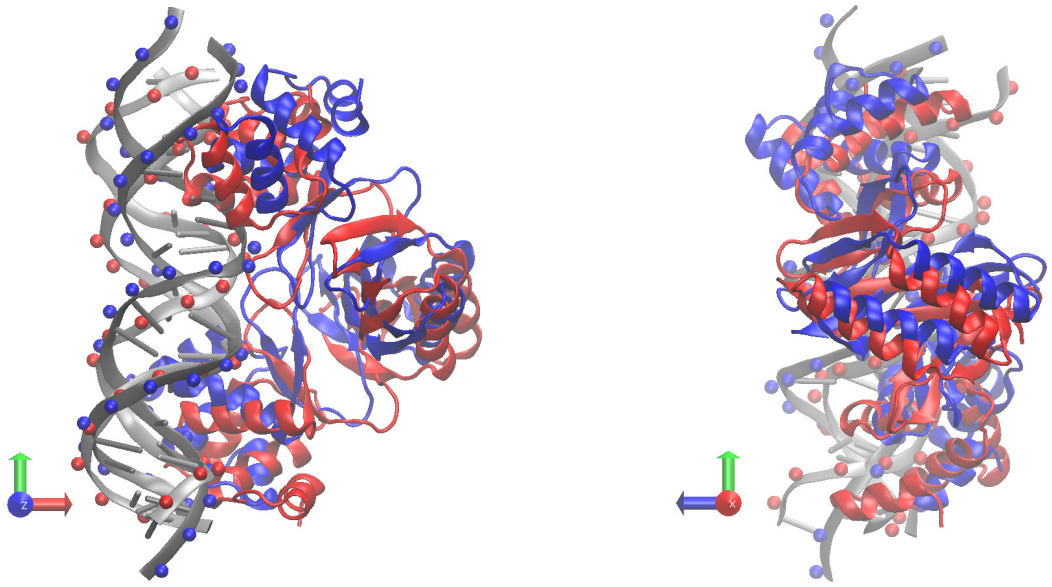


Figure 5.44: Front and side view of the structure of the first umbrella window, at the end of the simulation, for MgFur / Fur box (blue) and MgFur / *feoAB1* operator (red) complexes.

This pattern can be linked to the DNA sequence used in our simulations. Table 5.25 shows the difference between the two DNA sequences. More detailed investigation should be carried out to identify residues involved in this differential DNA binding mechanism.

Origin	Sequence	Size
feoAB1	(Y) 5'-TAATTGCAAATCATTTGCAATTG-3' (Z) 3'-ATTAACGTTTAGTAAACGTTAAC-5'	23 bp
EcFur box	(Y) 5'-GCCGGATAATGATAATCATTATC-3' (Z) 3'-CGGCCTATTACTATTAGTAATAG-5'	23 bp

Table 5.25: Sequence difference between the Fur box and the *feoAB1* operator. Green and red base pairs illustrate the palindromes present in each sequence.

5.11 EcZur with DNA

We included the Zinc Uptake Regulator (Zur), from *E. coli* (EcZur), in our study for its Fur-like properties and its resolved structure in the presence of a DNA duplex derived from the *znuABC* promoter. The structure was described by Gilston et al., 2014 (PDB ID: 4MTD) and was used to build our two EcZur models. The first one is discussed in this section and is shown in Figure 5.45.

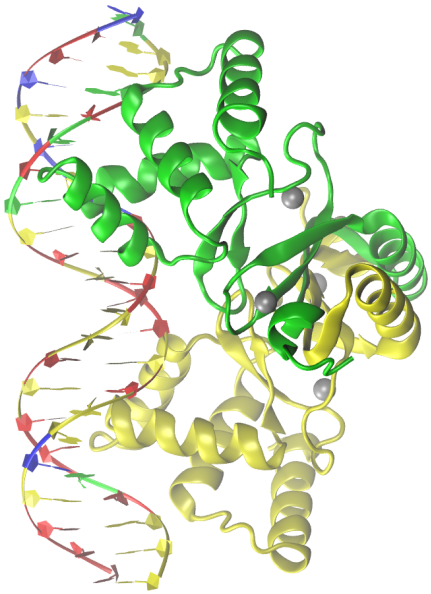


Figure 5.45: Calculated minimum energy structure of the EcZur DNA complex.

No asymmetric dissociation is observed in the case of EcZur, however, a slight difference in the total interaction energy is observed between chains A and B in Figure 5.46 at the beginning of the dissociation.

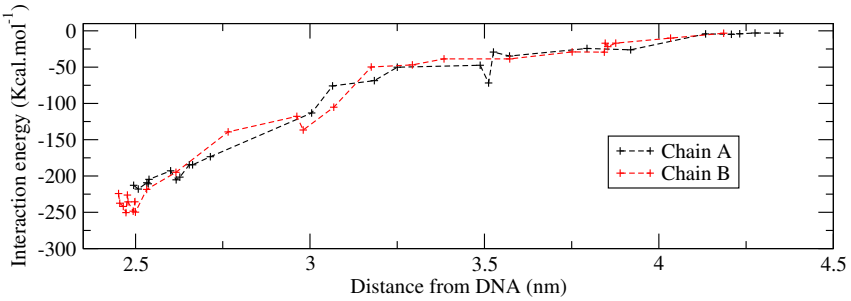


Figure 5.46: Comparison of average total interaction energies between EcZur chains A and B and the fixed DNA subsystem.

When the 4% threshold is applied, the two Zur subunits show the same interacting residues. Some residues like Arg23 and Arg28 are mainly important at the beginning of the dissociation, on the contrary, Arg52 appears toward the end. Interestingly, for both chains, Lys59 is interacting throughout the whole dissociation, as seen in Figure 5.47 and 5.48.

From Tables 5.26 and 5.27, residues Lys59, Arg23 and Arg28 show the largest interactions with DNA in our calculations, and all involve mainly non specific phosphate groups. Lys59 (A,B) interacts with T7(Y), G8(Y) and T6(Z), T7(Z), respectively; Arg28 (A,B) interacts with G4(Y), T5(Y) and A3(Z), T4(Z), respectively. Arg23, from chain A, forms specific interactions with A2(Y), A3(Y), C22(Z) and non specific interaction with phosphates of G4(Y). In chain B, Arg23 interacts with phosphate groups of A2(Z) and A3(Z). Both residues Tyr64 and Lys78 (chains A and B) make non specific interactions with the phosphate groups of A13(Z) and A14(Y), respectively. Other interactions with the phosphate groups involve Thr25, Gln57 and Gln27. These results are summarized in a common Table 5.31 with the other Zur simulation.

In their work, Gilston et al., 2014, emphasize interactions between Tyr45 or Arg65 and DNA; from our calculation Arg65(A,B) indeed makes specific H-bonds interactions with bases G6(Y), T7(Y) and G5(Z), A17(Y), respectively. These interactions are rather weak, ranked 10th in our simulations. During the first pulling window Tyr45 (Chain A) interacts with the base of A13(Z) and makes π-π stacking with T12(Z) before switching to non specific interactions with A11(Z) and T12(Z) for subsequent windows.

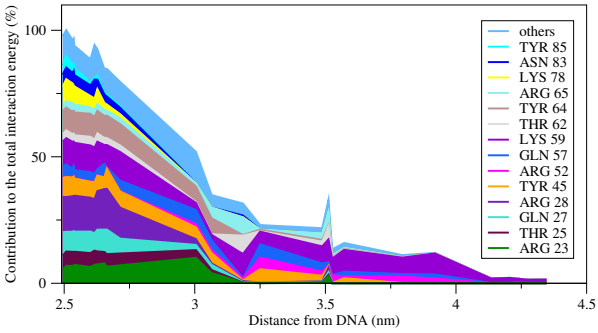


Figure 5.47: Major contributors of chain A.

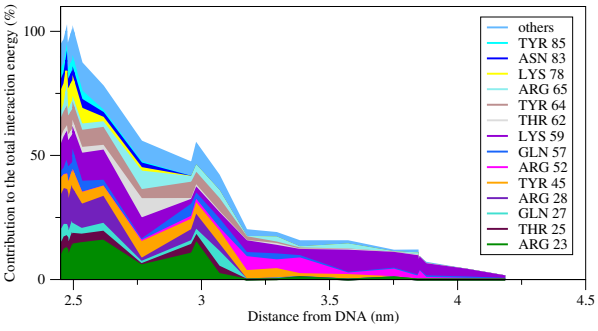


Figure 5.48: Major contributors of chain B.

Figure 5.47 and 5.48: Major contributors to the interaction energy between moving chains and the Zur box in the EcZur / Zur box simulation.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys59	-24.7	11.3	2.71
Arg28	-35.0	16.0	2.66
Tyr64	-19.7	9.0	2.60
Gln27	-21.1	9.7	2.66
Tyr45	-17.3	7.9	2.66

Table 5.26: Top 5 interacting residues of chain A in the EcZur DNA complex.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys59	-31.6	12.6	2.46
Arg23	-39.2	15.6	2.61
Arg28	-35.9	14.3	2.46
Tyr64	-19.6	7.8	2.47
Tyr45	-19.1	7.6	2.47

Table 5.27: Top 5 interacting residues of chain B in the EcZur DNA complex.

5.12 EcZur + chain C + DNA

The purpose of this second simulation was to check the importance of the interactions between the two EcZur dimers in the stability of the protein-DNA complex. Therefore, part of EcZur chain C (from the second dimer shown in red in Figure 5.49) interacting with chain B of the first dimer above, was added in this model. Precisely, residues Glu2 to Leu87 only.

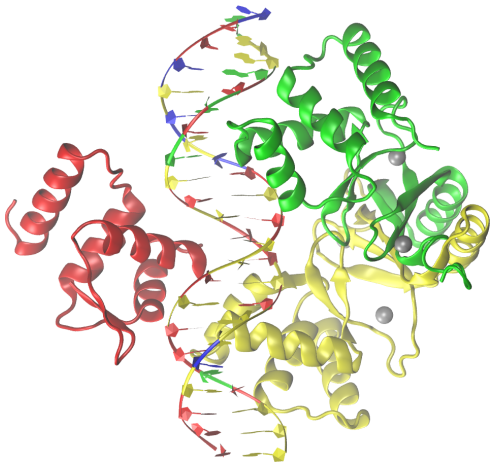


Figure 5.49: Calculated minimum energy structure of the EcZur + chain C + DNA complex.

5.12.1 Interaction of chain A and B with the DBD of chain C

Only chain B interacts through six residues with the DBD of chain C (Figure 5.50). Asp49 (B,C) forms a salt bridge with Arg52 (C,B) of high interaction energy (5.28). Tyr45 and Lys59, also found to interact with DNA, are amongst the six interacting residues shown in Figure 5.51.

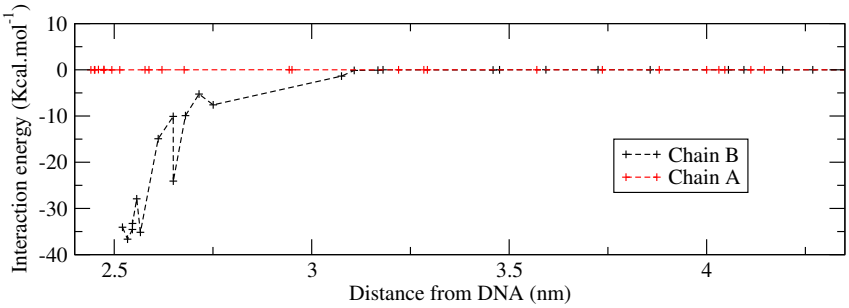


Figure 5.50: Comparison of average total interaction energies between EcZur chains A and B and the fixed DBD of chain C.

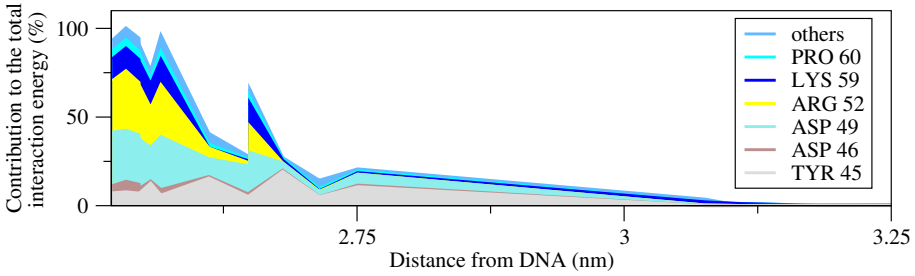


Figure 5.51: Major contributors to the interaction energy between moving chain A, and the DBD of chain C.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Asp49	-11.0	30.2	2.52
Arg52	-12.3	33.3	2.53
Tyr45	-7.1	19.3	2.68
Lys59	-5.2	14.3	2.56
Pro60	-1.7	4.7	2.56

Table 5.28: Top 5 interacting residues of chain B with Chain C in the EcZur + chain C + DNA complex.

5.12.2 Interaction of chain A and B with DNA

Figure 5.52 shows that the windows closest to the DNA have similar total interaction energies for both chains, in contrast to what was found for the same complex in the absence of the DBD of chain C, seen previously in Figure 5.46. This could highlight the stabilizing role, on the Fur dimer (A,B), that the DBD domain of chain C can have during DNA binding.

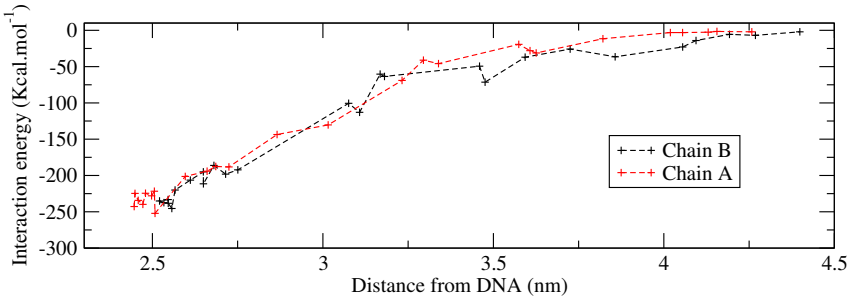


Figure 5.52: Comparison of average total interaction energies of EcZur chains A and B with the fixed DNA subsystem, in the presence of the DBD of chain C.

The presence of chain C increases the interaction of chain B with DNA as seen in Figure 5.54. Since chain A does not interact with chain C, its interaction with DNA is not affected by the presence of chain C as seen in Figure 5.53.

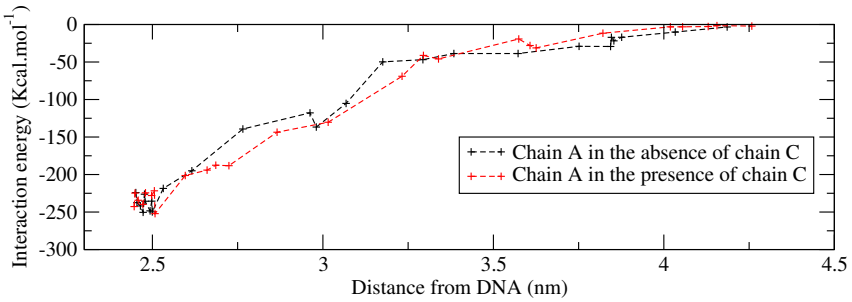


Figure 5.53: Comparison of average total interaction energies of EcZur chain A with DNA in the presence or absence of chain C

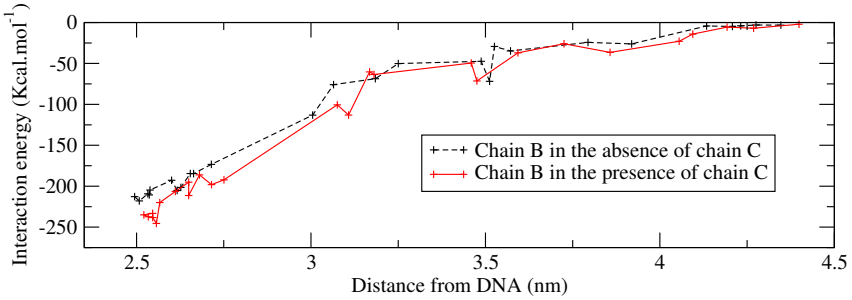


Figure 5.54: Comparison of average total interaction energies of EcZur chain B with DNA in the presence or absence of chain C

Figures 5.56 and 5.55 show the two chains A and B share the same interacting residues (except for Thr62 observed only in chain A). The results are similar to the ones obtained previously, for example, Lys59 is one of the major interacting residues throughout the whole dissociation and is the major interacting residue at greater distances.

Tables 5.29, 5.30 and 5.31 show that in addition to Lys59, Arg23 and Arg28 show the largest interactions with DNA in our calculations, and all involve mainly non specific

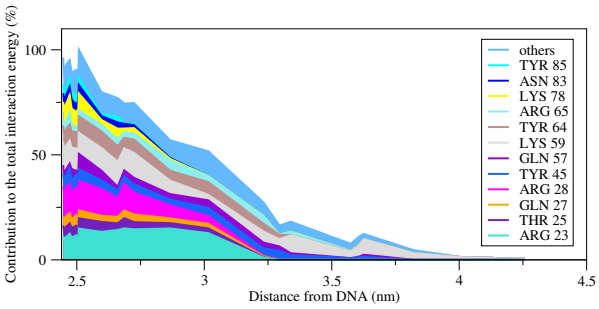


Figure 5.55: Major contributors of chain A.

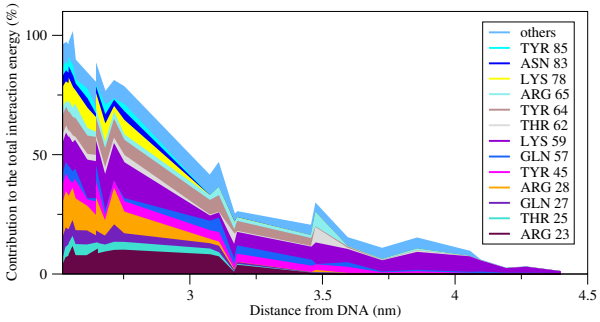


Figure 5.56: Major contributors of chain B.

Figure 5.55 and 5.56: Major contributors to the interaction energy between moving chains and the Zur box in the EcZur / Zur box simulation in the presence of the DNA binding domain of chain C.

phosphate groups: Lys59 (A,B,C) with T7(Y), T6(Z) and T12(Y), Arg28 (A,B,C) with G4(Y) and T5(Y), A3(Z) and T4(Z) and A9(Y) and T10(Y), respectively.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys59	-39.6	16.1	2.61
Arg28	-36.4	14.8	2.53
Arg23	-26.1	10.6	2.55
Tyr64	-19.6	8.0	2.68
Lys78	-22.6	9.2	2.54

Table 5.29: Top 5 interacting residues of chain A with DNA in the EcZur + chain C + DNA complex.

Arg23 is special in the fact that it interacts non specifically in the first simulations (and in the X-ray structure) but then forms specific interactions, around window 8 (distance of 2.68 nm) for instance, where Arg23 from chain B contributes to a maximum of 15% to the total interaction with DNA: for that window, we could observe interactions between Arg23 (A,B,C) and bases of T21(Z), C22(Z) and A1(Z), A2(Z) and Gua8 (Y), respectively. The two residues Tyr64 and Lys78 (A,B,C) make non specific interactions with the phosphate groups of A13(Z), A14(Y) and A8(Z), respectively. Other interactions with the phosphate groups involve Thr25 Gln57 and Gln27.

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Lys59	-30.0	11.9	2.660
Arg23	-37.8	15.0	2.68
Arg28	-35.5	14.1	2.47
Tyr64	-20.0	7.9	2.45
Tyr45	-18.3	7.2	2.47

Table 5.30: Top 5 interacting residues of chain B with DNA in the EcZur + chain C + DNA complex.

As previously mentioned, Gilston et al., 2014 emphasize on the interaction between Tyr45 or Arg65 and DNA; from our calculations Arg65 (B) indeed makes specific H-bond interactions with G5(Z) during the first pulling window while Tyr45 (A,B) interacts with bases of A13(Z) and A14(Y), respectively. The magnitude of these specific interactions are ranked 10^{th}, 6^{th} and 5^{th}, respectively (between 5 and 8 % of the total interactions). Tyr45 is not conserved across all Fur homologs, but is unique to gram negative Zur protein-DNA recognition and could, according to the authors, provide one possible discrimination between Zur- and Fur-regulated promoters.

Chain A			Chain B			Chain A with Chain C	
Residue	Partner	Type	Residue	Partner	Type	Residue	Partner
Lys59	T7(Y), G8(Y)	non-specific	Lys59	T6(Z)	non-specific	Asp49	Arg52
Arg23	T21(Z), C22(Z)	specific	Arg28	A3(Z), T4(Z)	non-specific	Arg52	Asp49
Arg28	G4(Y), T5(Y)	non-specific	Arg23	A1(Z), A2(Z)	specific	Tyr45	Lys59
Tyr64	A13(Z)	non-specific	Tyr64	A14(Y)	non-specific	Lys59	Asp49
Tyr45	A13(Z)	specific	Lys78	A14(Y)	non-specific	Pro60	Tyr45

Table 5.31: Top 5 interacting residues of the EcZur2 DNA complex and their respective partners. Blue residues are predicted by our simulations and black residues were determined from the simulation starting conformation.

As already mentioned, two salt bridges exist between adjacent dimers, between Asp49 from monomer B and Arg52 of monomer C, and vice versa; these intearctions are indeed ranked first in our calculations (Table 5.31 right).

Although we haven't studied the interactions between Fur subunit C (adjacent dimer) and DNA, some interactions have been described above after visual inspection. It is noteworthy, although not surprising, that there is a direct correspondence between interactions of residues in subunit B and DNA chain Z base i and same residues in subunit C DNA chain Y and base $i + 6$ (or Z base $i + 6$ and Y base i) which reminds us of hexameric motifs already described in section 3.3 for Fur DNA binding.

The authors conclude that cooperativity in these protein-DNA interactions for two dimers bound on adjacent faces of the duplex requires a pair of asymmetric salt bridges between Arg52 and Asp49 that connect otherwise independent dimers. They made mutant forms of Zur where the salt bridge is removed (D49A or R52A) and measured the different binding affinities of a Fur dimer for DNA. They found that the binding of one mutant dimer to the *znuABC* promoter site is significantly favored (Kd1 = 2.1-2.6 nM) over the binding of a second mutant dimer to an adjacent site (Kd2 = 65-220 nM).

The apparent macroscopic dissociation constant Kd-app for the cooperative binding of 2 Zur dimers to the *znuABC* promoter was estimated to 8.2 10^{-18} M^2. On the other hand, the Kd-app for the binding of mutant proteins was estimated between 140 and 570 10^{-18} M^2 which enabled Gilston et al to emphasize the role of the salt-bridge, described above,

and estimate the free energy penalty ($\Delta\Delta G$) for mutating either salt bridge linker to 2 kcal. mol^{-1}.

From our simulations the difference in binding free energy between the Zur dimer / DNA and 2 Zur dimer / DNA systems is another estimate of the strength of this linker: we estimated it to 1 $kcal.mol^{-1}$ (5.32), which is in good qualitative agreement with the experiments. Figure 5.57 compares the two PMF profiles of the two simulations.

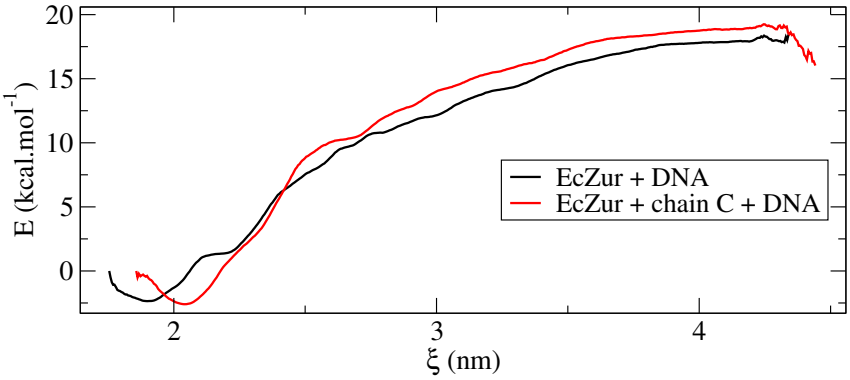


Figure 5.57: Potential of mean force for the separation of Zur from DNA in the absence (ZurDimer DNA) and presence (2 Zur dimer DNA) of a second Zur dimer

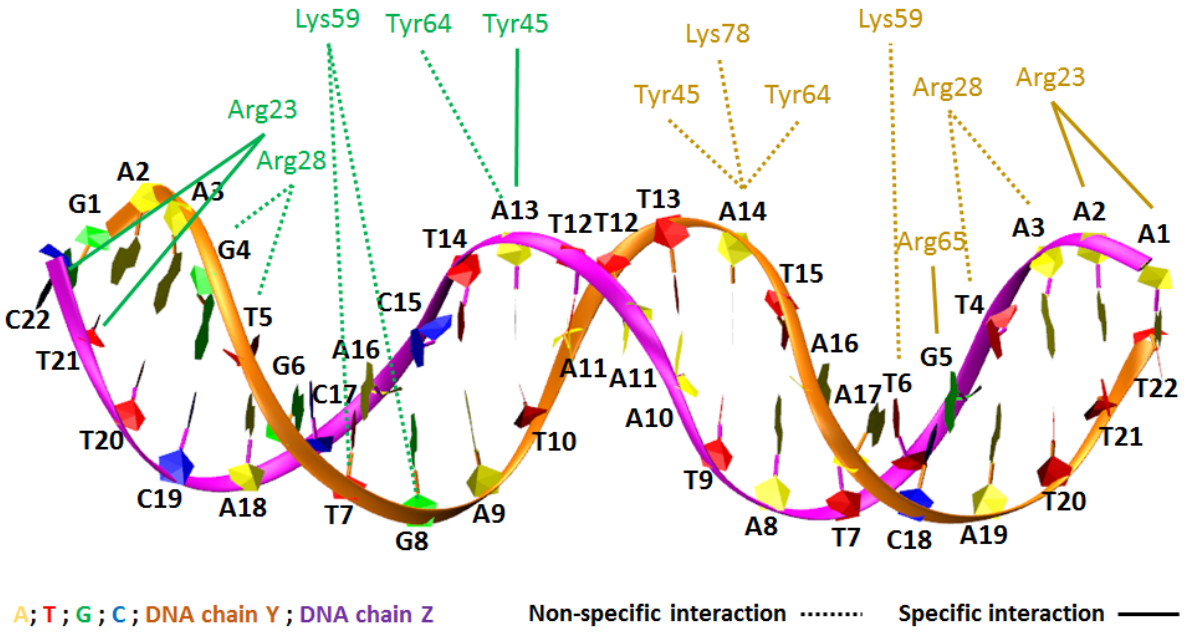


Figure 5.58: Interactions during the EcZur simulations. Green and yellow residues are from the protein chain A and B, respectively.

5.13 Summary of PMFs for Fur / DNA models

This section summarizes all potentials of mean force for Fur / DNA models studied in this work. The first important observation is the biphasic dissociation of MgFur, FtFur and to some extend EcZur from their DNA box, when only one dimer of Fur is present.

The curves shown in Figure 5.59, seem to become monophasic when MgFur dissociates from the mutated box *feoAB1* or when a second dimer is included in the simulation, as is observed in the case of EcZur.

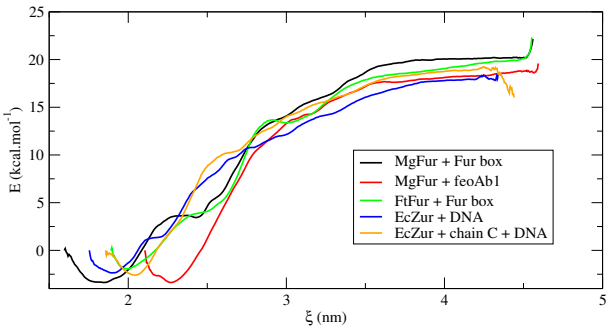


Figure 5.59: Plot showing the potential mean force curves of the models involving Fur (and Zur) interaction with DNA.

By determining the minimum and the maximum of each PMF, the ΔG of binding to DNA can be calculated (see section 7.1.8). Results are presented in Table 5.32. Our calculations show that:

1. The values of ΔG and the stability of the complexes with DNA can be ranked as the following: EcZur < FtFur < MgFur.

2. Binding of MgFur to the consensus Fur box is 1.4 kcal.mol^{-1} stronger than the binding to *feoAB1*. Unfortunately, there is no experimental value of binding affinity for MgFur bound to the Fur box.

3. In the case of EcZur, the presence of a second Zur dimer increases the binding affinity of the first by 1 kcal.mol^{-1}, pointing towards some sort of cooperative binding.

In addition, the MgFur and FtFur simulations, show that four important residues are always involved in DNA interaction: Lys14, Arg19, Tyr56 and Arg57, three of them able to form specific interactions with bases (Lys14, Tyr56 and Arg57). Moreover, in the case of diphasic dissociation, these specific interactions were present between one protein chain only and DNA, helping explain this behaviour.

Finally, after having done all the analysis, it appears that most highest contributions to the interaction energies include charged interaction between residues (Arg, Lys and phosphate groups). These contributions are non-specific, whereas specificity may be due to polar interactions like hydrogen bonds between protein NH_2 or OH groups and DNA

Complex	ΔG (kcal.mol^{-1})
FtFur + Fur box	22.01 ±1.81
MgFur + Fur box	23.70 ±2.40
MgFur + *feoAB1* box	22.31 ±1.37
EcZur + Zur box	20.05 ±1.49
EcZur + Zur box + Chain C	20.97 ±1.47

Table 5.32: ΔG of binding of the complexes involving Fur (and Zur) interaction with DNA.

bases. We thus notice that the complete analysis of specificity should involve residues ranked lower than 5th in their total energy contribution. This type of analysis should be continued on all complexes.

5.14 Chapter conclusion

The objective of this study was to characterise and understand the underlying mechanisms of Fur DNA interactions, in addition to getting insights into tetrameric Fur complexes. Figure 5.62, in the annex, summarizes the main interacting residues in each simulation.

By comparing the two systems, a striking difference is that during a tetramer dissociation, a lot of residues contribute in small amounts, compared to DNA interaction, where the interaction occurs through few residues that have important contributions. This could be explained by the need, in the case of tetramers, to cover a surface area with a lot of interactions, that can be triggered as a chain reaction when the tetramer starts interacting with DNA. Additionally, having distributed the contributions for tetramer stabilization on multiple residues, avoids mutations that can jeopardise the oligomeric state.

In the case of DNA interactions, and since Fur interacts with specific sequences, few residues have a great impact on the interaction energy to ensure specificity. Moreover, these residues are spatially constrained by the DBD conformation.

This information should be taken into account when developing new strategies to inhibit Fur from different pathogens. This will allow a better conception and understanding of interactions taking place at the molecular level.

Annex

COMMUNICATIONS BIOLOGY

ARTICLE

DOI: 10.1038/s42003-018-0095-6 OPEN

Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in *Francisella tularensis* virulence

J. Pérard[1], S. Nader[1], M. Levert[2], L. Arnaud[1], P. Carpentier[1], C. Siebert[2], F. Blanquet[2], C. Cavazza[1], P. Renesto[2], D. Schneider[2], M. Maurin[2], J. Coves[3], S. Crouzy[1] & I. Michaud-Soret[1]

Francisella tularensis is a Gram-negative bacterium causing tularaemia. Classified as possible bioterrorism agent, it may be transmitted to humans via animal infection or inhalation leading to severe pneumonia. Its virulence is related to iron homeostasis involving siderophore biosynthesis directly controlled at the transcription level by the ferric uptake regulator Fur, as presented here together with the first crystal structure of the tetrameric *F. tularensis* Fur in the presence of its physiological cofactor, Fe^{2+}. Through structural, biophysical, biochemical and modelling studies, we show that promoter sequences of *F. tularensis* containing Fur boxes enable this tetrameric protein to bind them by splitting it into two dimers. Furthermore, the critical role of *F. tularensis* Fur in virulence and pathogenesis is demonstrated with a *fur*-deleted mutant showing an attenuated virulence in macrophage-like cells and mice. Together, our study suggests that Fur is an attractive target of new antibiotics that attenuate the virulence of *F. tularensis*.

[1] Univ. Grenoble Alpes, CNRS, CEA, BIG-LCBM, 38000 Grenoble, France. [2] Univ. Grenoble Alpes, CNRS, CHU Grenoble Alpes, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France. [3] Univ. Grenoble Alpes, CNRS, CEA, IBS, 38000 Grenoble, France. Correspondence and requests for materials should be addressed to J.Pér. (email: julien.perard@cea.fr) or to S.C. (email: serge.crouzy@cea.fr) or to I.M-S. (email: isabelle.michaud-soret@cea.fr)

Francisella tularensis is a small, highly infectious Gram-negative bacterium, causing the zoonotic disease tularaemia[1]. This species is currently divided into three subspecies, including subsp. *tularensis* (type A strains), subsp. *holartica* (type B strains) and subsp. *mediasiatica*[2]. Only type A and type B strains of *F. tularensis* are known to cause tularaemia in humans. A large number of animal species can be infected with this pathogen, but lagomorphs and small rodents are considered the primary sources of human infections. The disease may also be transmitted through arthropod bites, mainly *Ixodidae* ticks and mosquitoes. *Francisella tularensis* also survives for prolonged periods in the environment, and humans can be infected through contact with contaminated soil or water. Because a few bacteria inhaled through aerosols may induce an acute severe pneumonia, with a mortality rate of 30% or more, *F. tularensis* has been classified as a potential category A biothreat agent by the Center for Disease Control and Prevention[3]. No effective vaccine is currently licenced for human or animal use, and a few antibiotic compounds are used as first-line drugs in tularaemia patients. Alternative treatments are urgently needed both to improve the prognosis of patients with severe diseases, and also to improve our preparedness to the intentional release of resistant strains of this pathogen in the context of bioterrorism[4,5]. Although numerous genes have been shown to be important for the pathogenesis and virulence of *F. tularensis*, there is still a blatant lack of knowledge about central biological functions such as iron homeostasis[6,7] and metalloregulators[8,9]. As a facultative intracellular bacterial pathogen, *F. tularensis* multiplication and virulence depend on the host cell iron pool[10]. Indeed, a major defence strategy used by infected eukaryotic organisms is to withhold this metal by sequestering free iron. In reaction to iron starvation, *F. tularensis* is able to secrete an iron chelator structurally similar to the polycarboxylate siderophore rhizoferrin[11,12]. The *figA* gene (also called *fslA*), involved in the siderophore synthesis, but also the *figE* gene (*fslE*), responsible for its uptake, have been characterized to play an important role in the virulence and/or intracellular replication of this pathogen[9,12,13]. These genes belong to the locus *figABCDEF* (*fig* for *Francisella* iron-regulated genes)[14,15] (Fig. 1). The *fig* operon is regulated by the ferric uptake regulator Fur, which is supposed to bind to the *fur-figA* intergenic region that contains a specific sequence called a FurBox (Fig. 1a), although such direct interaction has not yet been demonstrated. The Fur protein is a global transcriptional regulator that senses iron status and controls the expression of genes involved in iron homeostasis, virulence and oxidative stresses[16–18].

In the present study, to go further in the in vitro and in vivo characterization of the properties of the *F. tularensis* Fur (FtFur) protein, we used a virulent *F. tularensis* subsp. *holarctica* strain, a Type B biovar I, referred to as CHUGA-Ft6. This strain was isolated from a blood sample from a French patient suffering from a typhoidal form of tularaemia[19]. Interestingly, comparing FtFur to Fur from *Escherichia coli* (EcFur), *Pseudomonas aeruginosa* (PaFur), *Legionella pneumophila* (the agent of legionellosis, LpFur) and *Yersinia pestis* (the agent of plague, YpFur), we have evidenced that these proteins can be discriminated by their quaternary structure in solution[20]. EcFur and YpFur belong to the group of the commonly accepted dimers, while FtFur, PaFur and LpFur belong to a group of tetramers. A structural zinc in a cysteine-rich site (site 1) has been characterized in many Fur proteins including FtFur[16,21,22]. In addition, the Fur proteins need metallic dications such as Co^{2+}, Mn^{2+} or Fe^{2+} in a regulatory site (site 2) to be activated for the binding to DNA[20].

Here, we present, to our knowledge, the first crystal structure of a tetrameric Fur protein in the presence of its physiological cofactor, the ferrous ion. This structure sheds light on the metal-binding sites and corresponds to two intertwined pre-activated dimers. We demonstrate the direct interaction of the protein with the promoter region controlling expression of the genes involved in siderophore synthesis and identify essential residues in this interaction. In addition, owing to the coupling of computer models and free energy calculations with cross-link experimental studies, we bring evidence for a DNA-driven tetramer splitting mechanism mediated by specific promoter sequences, and leading

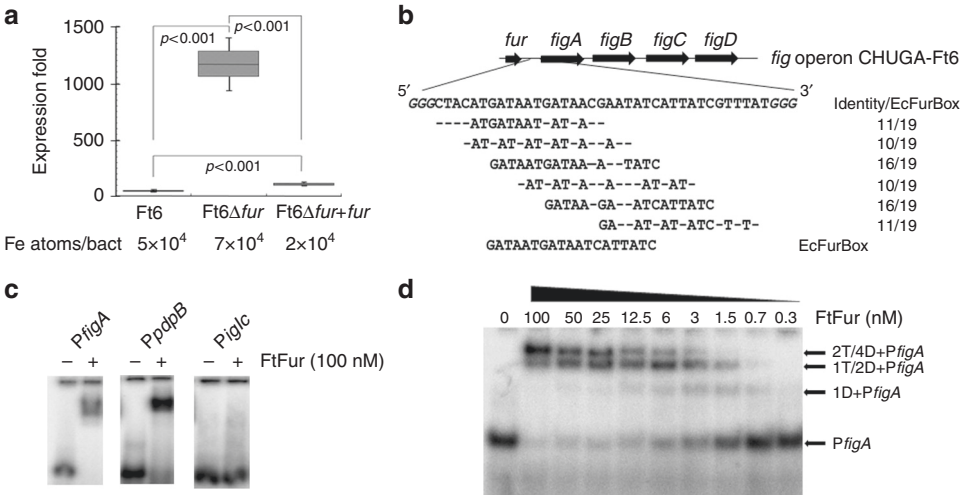


Fig. 1 FtFur regulates *fig* operon by recognition of DNA FurBox. **a** qRT-PCR showing the absence of *fur* transcripts in the CHUGA-Ft6Δ*fur* strain, with a 16S RNA standard as a control, Fur (Ft6 and Ft6Δ*fur* + *fur*) repressed the transcription of *figA*. This repression is abolished in the absence of Fur (Ft6Δ*fur*). The data correspond to two independent experiments made in triplicate. *P* values were calculated using the Student's *t* test. Iron concentration was measured by ICP-AES (error under 1%) from 2 mL bacterial culture (see Methods section) and number of Fe atoms per bacteria has been deduced. **b** Organization of the *fig* operon and sequence of the *fur-figA* intergenic region (P*fig*A). The identical bases between P*fig*A and EcFurBox are indicated underneath showing overlapping FtFur binding sites. **c** Evaluation of the ability of FtFur to bind identified or predicted Fur boxes and estimation of the apparent Kds (for DNA seq of each promoter see Supplementary Fig. 1). **d** EMSA of FtFur in the presence of the 43 bp P*fig*A sequence. The proposed stoichiometry is written on the figure: D corresponding to dimer and T to tetramer

to the formation of two Fur dimer–DNA complexes. Finally, the critical role of FtFur in bacterial virulence and pathogenesis is demonstrated using a *fur*-deleted CHUGA-Ft6 mutant (Ft6Δ*fur*), which shows an attenuated virulence, both in murine macrophage-like cells and in mice, reinforcing that FtFur can be thus defined as a crucial anti-virulence target.

Results

Fur is directly involved in *F. tularensis* iron homeostasis. A Δ*fur* mutant was already generated in the virulent Schu S4 strain (subsp. *tularensis*) to demonstrate that siderophore production is regulated by FtFur in *F. tularensis*[9]. However, the direct involvement of FtFur in virulence has never been reported to our knowledge. We have constructed the CHUGA-Ft6Δ*fur* strain by the allelic exchange method and deletion was confirmed by quantitative real-time polymerase chain reaction (qRT-PCR) and sequencing as we did not detect any *fur* transcript in Ft6Δ*fur* (Supplementary Fig. 1). Using this approach, we demonstrated that the siderophore synthesis is under the direct control of FtFur in CHUGA-Ft6 strain. CHUGA-Ft6Δ*fur* shows an approximately 25-fold higher level of *figA* transcript when cultured in iron-replete conditions compared to the wild-type (WT) strain. The WT phenotype, that is Fur transcriptional repression of the *fig* operon genes, is recovered when the WT *fur* is expressed in trans to complement the *fur* deletion (CHUGA-Ft6Δ*fur* + *fur*) (Fig. 1a). This means that siderophore production is repressed by FtFur in the presence of iron and derepressed in the absence of the protein. Inductively coupled plasma atomic emission spectroscopy (ICP-AES) quantification of the bacterial iron concentration showed that, under our culture conditions, the CHUGA-Ft6Δ*fur* strain accumulates 1.6-fold more iron than the WT (Fig. 1a). These data strongly suggest that FtFur can bind the *fur-figA* intergenic region that contains sequences closely related to the EcFurBox identified in *E. coli* (Fig. 1b). Only a few Fur boxes were identified in *Francisella* genome, compared to *E. coli*, in the promoter of *figA*, *pdpB* (coding for the pathogenicity determinant protein PdpB) and *iglC* (coding for the pathogenicity island protein IglC) both in Schu S4[14] and CHUGA-Ft6. Then, electrophoretic mobility shift assay (EMSA) with manganese-activated FtFur have been performed on consensus EcFurBox and on P*figA*, P*pbp*B and P*igl*C sequences (Fig. 1c and Supplementary Fig. 2).

FtFur binds with a very high affinity to EcFurBox when activated with Co(II) (Kdapp = 9 nM[20]) and to the P*figA* promoter (estimated Kdapp = 5 nM) and with a low affinity to P*pbpB* (averaged estimated Kdapp = 100 nM; Fig. 1c and Supplementary Fig. 2C). In contrast, and while *iglC* gene expression is also found up-regulated under iron-restricted conditions in *F. tularensis*[14], no binding is detected indicating the absence of direct regulation by Fur. The migration on EMSA gel of the FtFur/P*figA* complex shows a composite pattern with three successive bands assigned to the binding of one to two tetramers (or one to four dimers) (Fig. 1d and Supplementary Fig. 2B) which appear as the protein concentration is increased. This suggests that FtFur could bind to several predicted Fur boxes in the sequence (Fig. 1b). Indeed, using a shorter version (P*figA*S) of P*figA*L, the main species detected corresponds to one dimer bound to DNA (Supplementary Fig. 2B).

Western blot experiments suggest that the estimated 3000 protein subunits/bacteria may be mainly present as a tetramer in vivo (see Supplementary Fig. 2F, G) and that hydrogen peroxide (H_2O_2) treatment (1 mM for 4 h) does not impact this amount. This is not surprising considering the high stability of the tetramer in solution. This copy number is in the same range as described in *E. coli* or *Vibrio cholerae* (5000 and 2500 subunits/bacteria estimated, respectively, in normal growth conditions)[23,24]. Considering the number of 50,000 total iron atoms/bacteria quantified by ICP-AES (Fig. 1a) and the volume of CHUGA-Ft6 around 10^{-15} L, 5 µM of FtFur subunit and 80 µM total iron are expected. Assuming micromolar range Kd for (Fe-FtFur) as found in the literature for *E. coli* (1–10 µM for Fe-EcFur[25]), we can expect that a pool of metallated Fur tetramer exists in the cell prior to association with the few DNA target present in *F. tularensis*.

Fe-FtFur and Mn-FtFur contain intertwined pre-activated dimers. Recombinant FtFur was purified as a tetramer containing one equivalent of Zn(II) per subunit[20]. FtFur was crystallized in the presence of Mn(II) as $MnCl_2$ or Fe(II) as $(NH_4)_2Fe(SO_4)_2$, the latter under anaerobic conditions. The structure of Mn-bound FtFur was obtained from purified protein metalled with Mn at high concentration before crystallization in the presence of Mn (II) and was determined ab initio at 1.7 Å resolution by the single-wavelength anomalous diffraction (SAD) method. This structure was used to determine that of Fe(II)-bound FtFur at 1.8 Å resolution by molecular replacement (see X-ray data in Supplementary Fig. 3 and Supplementary Tables 1 and 2). Both Mn(II)-bound and Fe(II)-bound proteins have similar overall structures appearing as a compact tetramer made of a dimer of dimers per asymmetric unit (α-carbon root mean square deviation (r.m.s.d.) of 0.271 Å between the two structures). The main differences come from disordered N-terminal and C-terminal residues. Thus, the structure of Fur containing the physiological activator metal, namely Fe(II)-bound FtFur, the first one described to date, will be used for a detailed description (Fig. 2a). Among the total 140 residues of the protein, 131 to 133 were resolved per chain. Each subunit presents secondary structure elements similar to those found in other Fur structures. It consists of a N-terminal DNA-binding domain (residues 7–82) composed of a winged helix–turn–helix motif in which α4 is the DNA recognition helix. A short hinge connects the DNA-binding domain to the C-terminal dimerization domain (residues 89–138). The dimerization domain consists of three antiparallel β-strands (β3 to β4) and two α-helices (α5 to α6), α5 intersecting between β4 and β5 (Fig. 2a and Supplementary Fig. 4). The dimeric interface is mediated by β5 from each subunit forming an antiparallel β-sheet, part of a six-stranded β-sheet in the dimer (Fig. 2b).

The two dimers in the tetramer structure are nearly identical (α-carbons r.m.s.d. = 0.318 Å between the two dimers) with an almost perfect superposition of the secondary structure elements. The interaction between the two dimers through their DNA-binding domains is stabilized by H-bonds involving atoms of the DNA recognition helices (α4) of chains AB and CD (for chains A and C: $Gln61_C$ Hε/$Ser64_A$ O; $Ser64_C$ Hγ/$Ser64_A$ Oγ; $Ser64_C$ O/$Gln61_A$ Hε and $Arg57_C$ Hh/$Glu63_A$ Oε salt bridge, and equivalently for chains B and D). The two salt bridges between Arg57 and Glu63 constitute the most important interactions (Fig. 2c). These interactions combined with an interface area of 2830 Å^2 between the dimers AD and BC (PISA[26]) explain the high stability of the tetramer in solution (Fig. 2d). For comparison, in a previous work, we demonstrated that PaFur, initially described as a crystallographic dimer[27], was tetrameric in solution[20] with a substantially lower predicted interface area of 2120 Å^2.

Each dimer is in a closed conformation with the wing in 'inside' positions corresponding to the 'active' form in which the DNA-binding domains are prepared to bind target DNA[20]. However, the dimer–dimer interactions naturally prevent any kind of interaction with DNA through the recognition helices. Indeed, Tyr56 and Arg57 (the one involved in the salt bridge stabilizing the tetramer), both present in the recognition helix, are highly

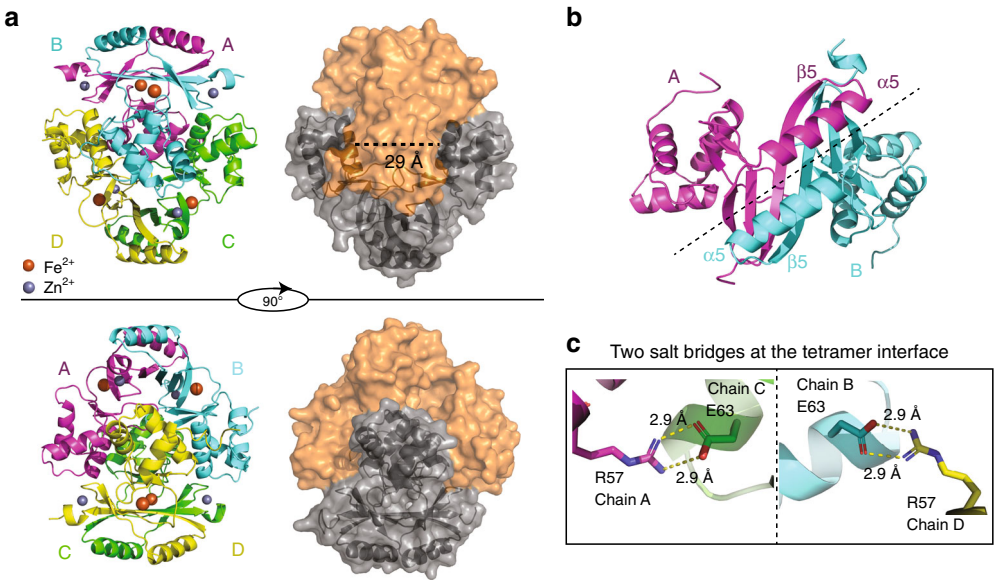


Fig. 2 Structure of FtFur at 1.8 Å resolution in the presence of physiological iron Fe^{2+}. **a** Fe-FtFur structure solved by SAD at 1.8 Å under anaerobic conditions in the presence of Fe^{2+}. The cartoon model presents the four chains, labelled A–B–C–D. Surface representation indicates the dimer/dimer interface and the distance between two recognition helices (29 Å). **b** Symmetry at the dimer/dimer interface between two monomers involving helix α5 and strand β5. **c** One of the most important interactions suggested by the structure is a salt bridge between Arg57 and Glu63

conserved residues known to have base-specific interaction with DNA[28–30]. We thus hypothesize that the metalled FtFur tetramer structure is a pre-activated form of the protein. The mechanism of pre-activated tetramer disruption driven by DNA is conceptually of interest.

First structural description of an iron substituted Fur. X-ray fluorescence spectra indicated that Fur crystals contain two metal species: one is the expected Zn and the second is the metal added during crystallization, that is, Mn(II) or Fe(II) (Supplementary Fig. 3A). The structures of Mn-FtFur and Fe-FtFur are similar and confirm the presence of one Zn^{2+} and one Mn^{2+} or Fe^{2+} per subunit. Zn^{2+} in structural site S1 is coordinated by four sulphur atoms from two pairs of cysteines in CX_2C motifs (Cys93-Cys96 and Cys133-Cys136) (Table 1). It connects the short C-terminal helix α5 to the β-sheet of the dimerization domain (Supplementary Fig. 4). The presence of S1 in Fe-FtFur but not in PaFur demonstrates that zinc is not a prerequisite for tetramer formation. The second site S2 binds either Mn^{2+} or Fe^{2+}. The metal ion adopts a distorted octahedral geometry with a 3N–3O coordination sphere (Fig. 3). S2 connects the DNA-binding domain (His33 and Glu81 (bidentate) and the dimerization domain which provides three ligands (His88, His90 and Glu101). It is described as the essential 'regulatory' site, present in all known activated Fur structures. H33A-H90A double mutations in FtFur S2 provoke a total inactivation of the protein in vitro (Supplementary Fig. 5C). The Fur structures containing an S2 site filled with Zn^{2+} show some variation in the coordination sphere. This flexibility may be explained by the preference of Zn^{2+} for a tetrahedral geometry compared to Fe^{2+}, found physiologically in S2, which favours a hexacoordinated octahedral environment with N/O ligands. Ab initio quantum chemical geometry optimizations of models of the S2 site using DFT with B3LYP hybrid functional and 6–31 G(d) basis set have been performed with bound Mn^{2+} or Fe^{2+}. The similar optimized geometries, with a larger coordination sphere for Mn^{2+} than for Fe^{2+}, validate the X-ray structures (Fig. 3, Supplementary Fig. 4A–D and Supplementary Table 3).

Some structures of Fur or Fur-like proteins (such as HpFur and PaFur) revealed the presence of a third metal-binding site (S3) involving four conserved residues 2 His, 1 Asp and 1 Glu[22,27]. In FtFur, Tyr125 is found in place of one very conserved His. The structure shows that the phenol group makes H-bonds with the other putative ligands preventing metal binding in the position where a metal ion was expected. Accordingly, the structures of FtFur do not display any S3 site (Fig. 3 and Supplementary Fig. 4).

In summary, the crystal structures of FtFur highlighted the presence of two metal ions per subunit: one structural Zn^{2+}, already present in the non-activated protein as purified, and either one Mn^{2+} or one Fe^{2+}, its physiological activator, with identical ligands in the regulatory site of similar geometry. Metalled FtFur behaves as a dimer of pre-activated dimers with the DNA-binding domains forming a kind of crown with interacting recognition helices through two salt bridges between Arg57 and Glu63 together with other weaker interactions.

Quaternary structure of FtFur in the presence of the FurBox. Size-exclusion chromatography coupled to multi-angle laser light scattering with online refractometer (SEC-MALLS-RI) was used to investigate the behaviour of FtFur in the presence of DNA. As

Table 1 Metal coordination from X-ray and DFT calculations

Atoms	Distance	Å
Site S1	**Zn**	**Zn**
Cys93/96/133/136S	2.3	2.3
Site S2	**Fe**	**Mn**
	X-ray / DFT	**X-ray / DFT**
His33 NƐ2	2.3 / 2.20	2.3 / 2.26
Glu81 OƐ1	2.2 / 2.25	2.3 / 2.43
Glu81 OƐ2	2.3 / 2.36	2.3 / 2.32
His88 NƐ2	2.3 / 2.22	2.3 / 2.25
His90 NƐ2	2.2 / 2.18	2.3 / 2.22
Glu101 OƐ1	2.0 / 2.05	2.3 / 2.10

Bond distances for sites S1 and S2 ligands deduced from the X-ray structures and calculated by DFT

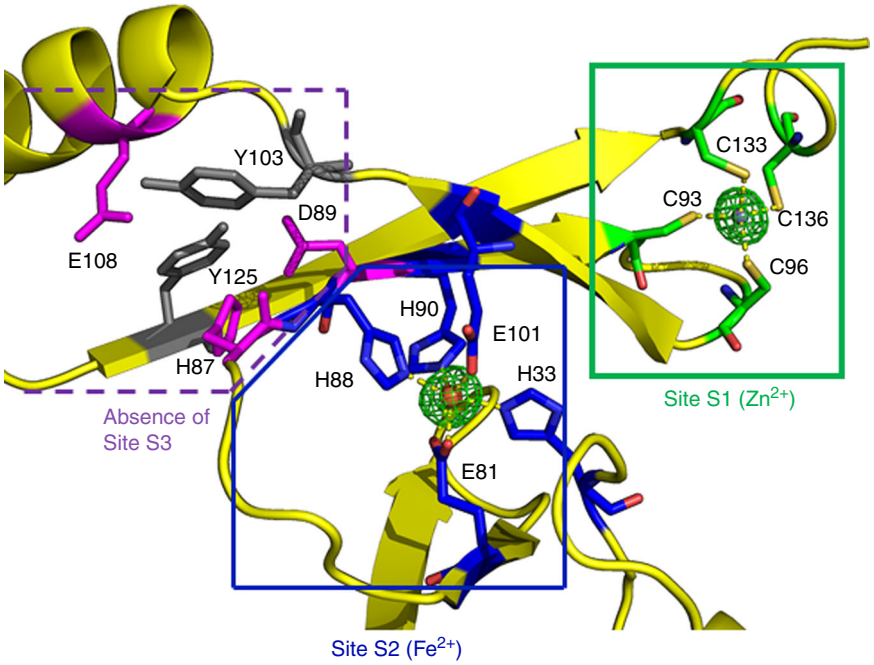


Fig. 3 Metal binding site in Fe-FtFur. Focus on one monomer extract on Fe-FtFur showing S1 and S2 sites simultaneously together with the unfilled S3 site

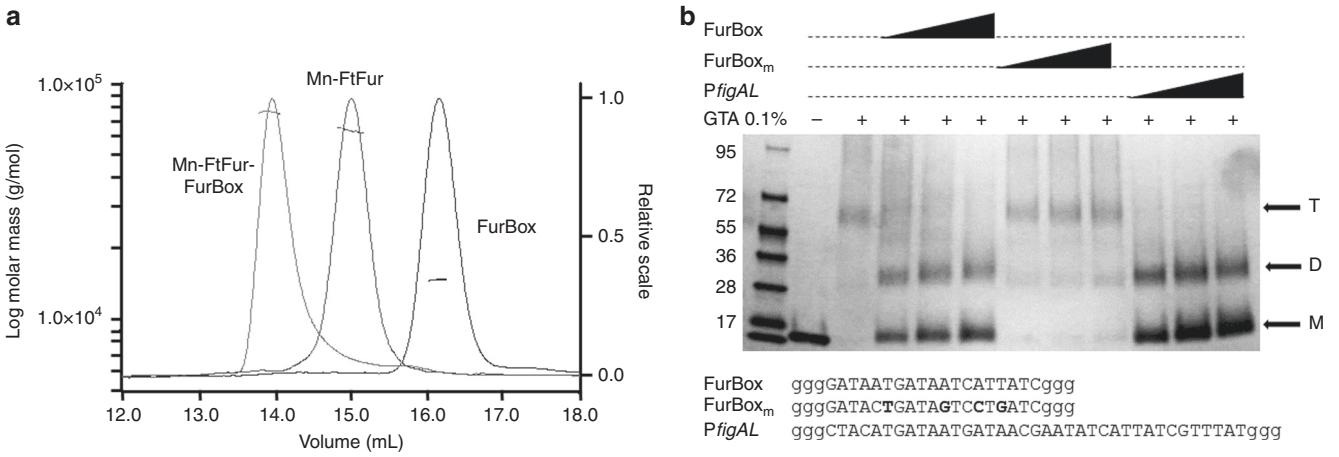


Fig. 4 Quaternary structure of Mn-FtFur in the presence of FurBox analysed by SEC-MALLS-RI and cross-link assay. **a** SEC-MALLS-RI data of Mn-FtFur and the FurBox analysed alone or as a complex in the presence of 1 mM $MnCl_2$. Data are normalized by RI scale with sample concentrations ranging between 4 and 6 mg mL^{-1}. FtFur/FurBox gives a MW of 78 ± 2 kDa, Mn-FtFur gives a MW of 64 ± 1 kDa and FurBox gives a MW of 15 ± 0.5 kDa. **b** Cross-link assay with GTA on SDS-PAGE 4–20% acrylamide in TGS buffer. A dimer (D) was trapped in the presence of specific Fur boxes (FurBox and P*figAL*), while FtFur exists as a tetramer (T) in the absence of DNA or in the presence of mutated $FurBox_m$

shown in Fig. 4a, in the presence of FurBox the protein eluted at a lower volume than the protein alone or the FurBox alone. The deduced molecular weight of the corresponding peak is 74 ± 2 kDa, fitting with a complex between tetrameric FtFur (64 kDa) and the FurBox duplex (15 kDa). This can be interpreted as the binding of FtFur to DNA as a tetramer or as two dimers.

The evolution of the purified tetrameric FtFur in the presence of DNA was then analysed by cross-link experiments using 0.1% glutaraldehyde (GTA). Under denaturing conditions, in the absence of DNA (Fig. 4b), the main detected band corresponds to a species with a molecular weight of approximately 62 kDa, in very good agreement with the size of a covalently bound FtFur tetramer. After cross-link in the presence of the EcFurBox, only two bands were detected corresponding to the monomer and to a dimeric form of the protein, respectively. Mutations of the FurBox ($FurBox_m$, see Supplementary Fig. 2) targeting four bases previously shown to be crucial for the specific Fur/DNA interactions[31], three of them being involved in interactions with Tyr56[28,29], resulted in the conservation of the tetramer without apparition of dimers.

The monomer (M)/dimer pattern was also obtained with P*figA*. These results demonstrate that tetrameric FtFur splits into dimers in the presence of specific DNA contrary to the dissociation of PaFur previously observed with non-specific DNA[20]. Besides, they strongly suggest that FtFur binds the FurBox as dimers in vitro.

MALLS and SAXS data validate a two-dimer/DNA complex. The activated Mn-bound FtFur form and the Mn-bound FtFur/EcFurBox complex were examined by small-angle X-ray scattering (SAXS). In both cases, at three different concentrations (1–10

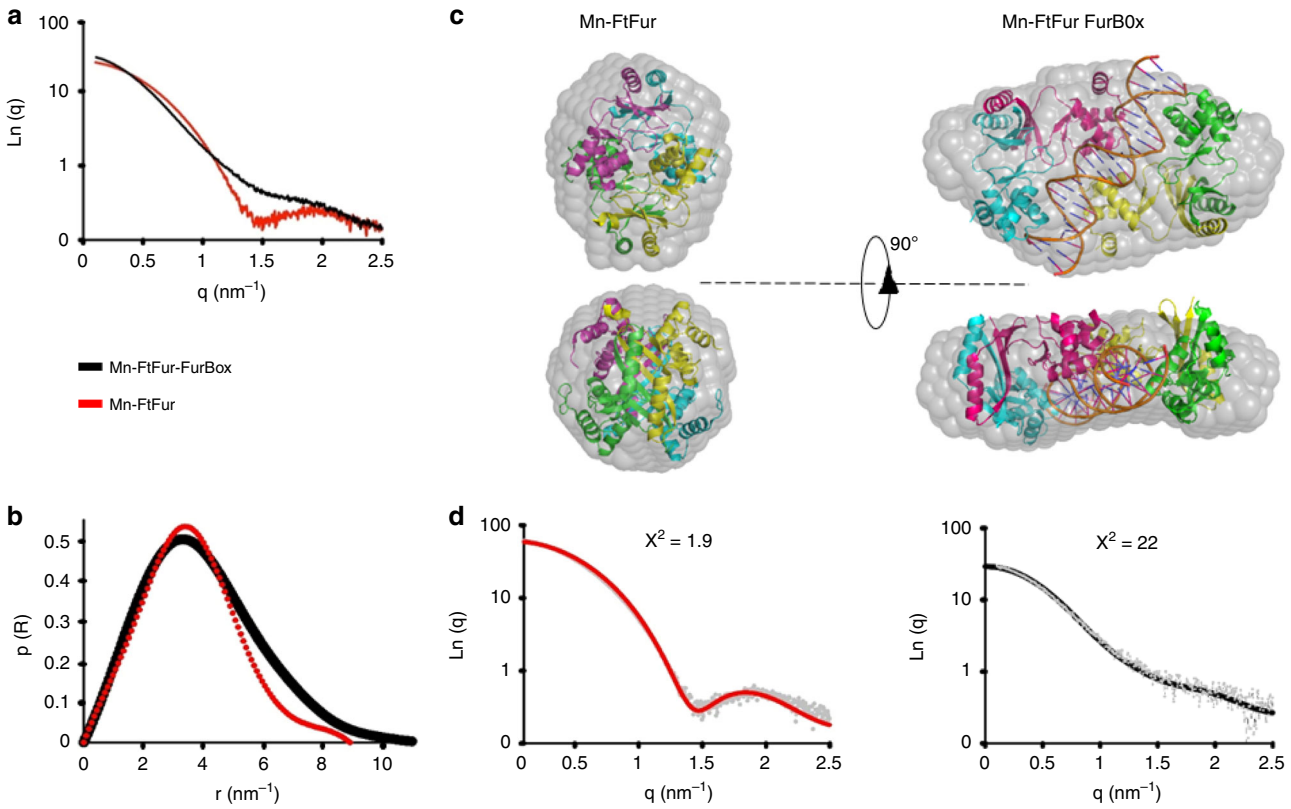


Fig. 5 Comparison of small-angle X-ray scattering curves of Mn-FtFur and Mn-FtFur/FurBox complex in solution. **a** Average scattering curves of Mn-FtFur and Mn-FtFur/FurBox complex in solution by small-angle X-ray scattering. **b** Pair distance distribution functions, p(R): Mn-FtFur/FurBox (MW = 81 kDa) and Mn-FtFur (MW = 64 kDa). **c** Molecular models of Mn-FtFur structure (left) and of Mn-FtFur/FurBox (right) fitted in the SAXS envelope. The model of the Mn-FtFur/FurBox complex is obtained, in the absence of X-ray structure, by superposition of Mn-FtFur on the Mn-MgFur/FurBox complex (see Supplementary information). **d** Fits of the scattering curves

mg mL^{-1}), the linearity of the Guinier plots indicates monodisperse samples. Average scattering curves of Mn-FtFur (red) and Mn-FtFur/FurBox complex (black) in solution were recorded (Fig. 5a). Pair distance distribution functions (Fig. 5b) point out an elongation of the protein/DNA system (Dmax = 112 Å for DNA/protein complex against 83 Å for the protein alone) and dramatic changes in the shape of the structure (Porod volume = 130 nm^3, against 100 nm^3, and radius of gyration = 32.5 Å, against 27.4 Å). Bead molecular models of Mn-FtFur alone and in complex with DNA complex, built by DAMMIF[32], show a globular Mn-FtFur and a thick pancake shape for the DNA complex (Fig. 5c). The X-ray structure of the protein determined in this study docks very well in the calculated envelope with a χ^2 of 1.9 (Fig. 5d). In the absence of high-resolution structure of the Mn-FtFur–DNA complex, a model was built based on Mn-MgFur/EcFurBox ternary complex[29] and fits well with the calculated envelope with a χ^2 of 2.2. These results are in agreement with the conclusions of the cross-linking experiments and support the DNA-driven split of the FtFur tetramer in two dimers sandwiching the FurBox. To better understand the mechanism of dimer–dimer and dimer–DNA dissociation, theoretical calculations were performed.

Dimer/dimer and dimer/DNA dissociation free energy profiles. The aim of this modelling was to evaluate precisely the difference in binding affinity between the FtFur dimers within the FtFur tetramers and between the FtFur dimers and DNA. Free energy (potential of mean force) profiles for the extraction (by translation along a fixed direction: Ox) of one FtFur dimer from the tetramer (dimer of dimers) and of FtFur from DNA were computed: the meticulous translation protocol is shown in Supplementary Fig. 6. The simulations include a 'moving' subsystem (FtFur dimer, chains A and D) and a 'fixed' subsystem (FtFur dimer, chains B and C, DNA) as shown in Supplementary Fig. 7. The profiles were built using the 'umbrella sampling' technique and result from the overlapping of 26 computation windows, one for each translation distance, and corresponding to 15 ns molecular dynamics simulation each. The results of the calculations are shown in Fig. 6a. Binding free energies are $\Delta G = 18.8$, 10.5 and 8.8 kcal mol^{-1} for dimer from FurBox, dimer from tetramer and dimer from mutated DNA (mutDNA containing FurBox$_m$), respectively. These binding free energies correspond to dissociation constants of 17 fM, 20 nM and 0.4 µM, respectively, allowing a thermodynamically easy separation of the tetramer into two dimers in the close proximity of DNA, deduced from the experiments. Statistical errors were estimated to be <1.5 kcal mol^{-1} with bootstrap analysis using the 'Bayesian bootstrap' method (b-hist option in g_wham).

According to Fig. 6b and Supplementary Fig. 8A, the residues mainly contributing to the stability of the tetramer are: E76, E63, N60, R57, D37 and K14, in agreement with the experimental results where the mutation of residues E76 and E63 into alanine leads to easier dissociation of the FtFur tetramer into two dimers. Close inspection shows that R57 interacts with E63, E76 with N60 and D37 with S35. For both A and D moving chains, these residues contribute to around 30% of the total interaction energy.

The residues with the strongest contribution to the FtFur/wtDNA complex stability are R57, Y56, T54, R19, T16 and K14, contributing more than 50% of the total interaction energy (Fig. 6c and Supplementary Fig. 8D). By homology with the

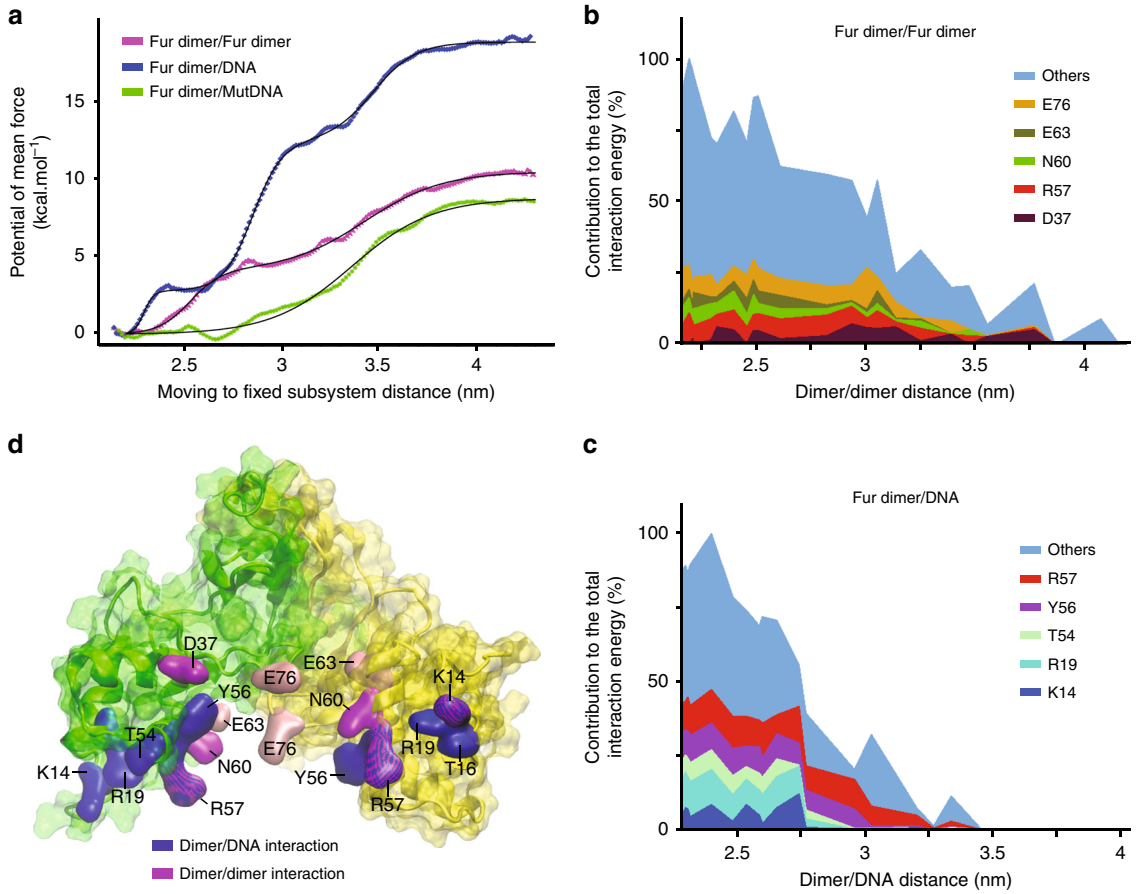


Fig. 6 Computation of free energy profiles for dimer/dimer and dimer/DNA dissociation. **a** Potentials of mean force for the extraction of FtFur dimer from the tetramer or from DNA. The *x*-axis, reaction coordinate, corresponds to the average centre of mass/centre of mass distance between the 'fixed' and the 'moving' subsystems. Data points corresponding to the outputs of Wham are fitted with 1, 2 or 3 sigmoid functions with R[55]. Statistical errors were estimated using the bootstrap method. **b** Major contributors to the average interaction energy between Fur chain D and the 'fixed' dimer in the tetramer simulation. On average, five residues contribute to around 30% of the total interaction energy. **c** Major contributors to the average interaction energy between Fur chain D and DNA in the FtFur/wild-type DNA simulation. The *x*-axis corresponds to the average centre of mass/centre of mass distance between Fur and DNA. On average, five residues contribute to 54.5% of the total interaction energy. **d** Structure of FtFur dimer showing chain A (yellow) and chain D (green). Residues shown in blue surface are the major contributors to the average interaction energy between FtFur and DNA in the FtFur/wild-type DNA simulation. Residues in magenta and pink surfaces are the major contributors to the interaction energy between the 'moving' dimer and the 'fixed' dimer in the FtFur tetramer simulation. The mutated residues E63 and E76 are in pink colour

Magnetospirillum gryphiswaldense (Mg) Fur-DNA structure[29], R57, Y56 and K14 are expected to make base-specific contacts, whereas T54, R19 and T16 interact with phosphates. Average interaction energy profiles between the 'moving' and 'fixed' subsystems are shown in Supplementary information. According to these profiles, the dissociation of the FtFur dimers from DNA would occur in two steps: slight unbinding of subunit D followed by unbinding of subunit A (Supplementary Fig. 8B, up and down). Noticeably, the mutations in the DNA FurBox drastically impede the binding of Fur to DNA with a 10 kcal mol^{-1} binding free energy decrease, explaining the selectivity of the binding of Fur to its FurBox sequence. More precisely, three of the four mutations face Fur chain D and their impact on the complex dissociation is visible in Supplementary Figure 8B where the initial average interaction energy of FtFur chain D with mutated DNA (-100 kcal mol^{-1}) is around half that with WT DNA (-200 kcal mol^{-1}). Below 2.8 nm the interaction of chain D with WT DNA remains stronger than with mutated DNA. (Supplementary Fig. 8C).

A summary of the FtFur dimer structure and the residues involved in its interactions with the other dimer within the tetramer or with DNA is shown in Fig. 6d.

New FtFur regulation mechanism from models and mutation data. Structural analysis suggests that the Arg57–Glu63 interaction plays a key role in the tetramer stabilization. Two of the four Arg57 (1 per dimer) are involved in such salt bridges and the two others are accessible to the solvent or to the DNA. Arg57 is predicted to be one of the most important residues for the interaction between Fur and bases in the specific DNA FurBox (Figs. 6, 7). This residue is highly conserved and its importance is in accordance with the Fur-DNA X-ray structure in *M. gryphiswaldense* where only few residues form base-specific interactions: Arg57-G7, Lys15-A24′ and Tyr56-T15′/T16′[29]. Similarly, Arg65 in EcZur, a Fur-like protein, interacts with a purine DNA base[30]. In FtFur, the four Lys14 (eq. Lys15 in MgFur) and two Arg57 are accessible for DNA interaction. Interestingly, the electrostatic potential around the tetramer shows a clear positive crown in the region of these residues (Fig. 7a) where the negatively charged DNA would be expected to first interact. We hypothesize that the specificity of the DNA-dependent tetramer dissociation could result from the interaction of DNA with Lys14 and the accessible Arg57, which would destabilize the tetramer by a progressive loss of the interaction of the two other Arg57 with Glu63. Mutations of Glu63 and/or Glu76 to Ala confirmed the

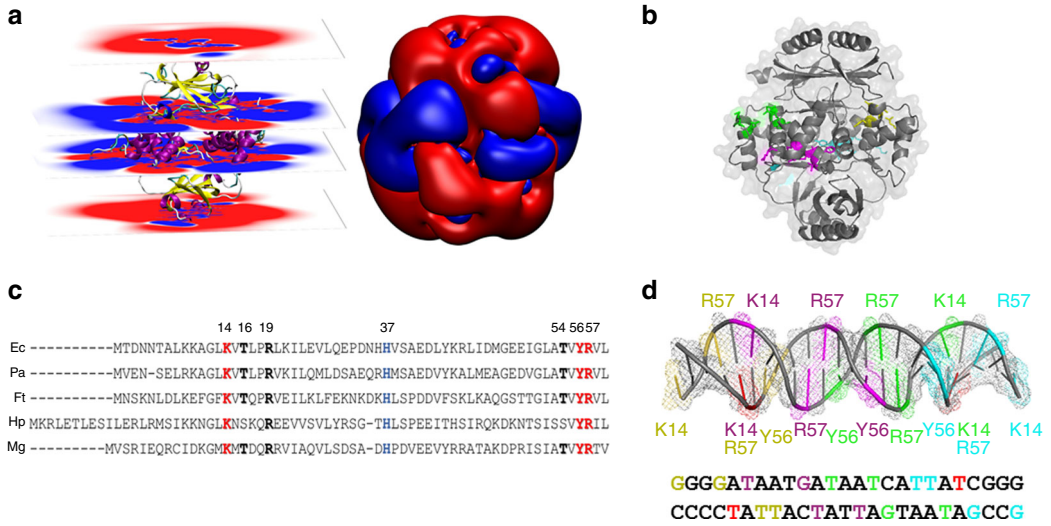


Fig. 7 DNA-driven FtFur tetramer dissociation mechanism. **a** Electrostatic potential around FtFur calculated on parallel planes (left) and on equipotential surfaces at −0.1 (red) and 0.1 V (blue) (right). **b** Mn-FtFur structure with its solvent accessibility surface. The residues predicted to be involved in the DNA interaction are coloured. **c** Sequence alignment of the DNA-binding domains of five Fur proteins with known structure EcFur (DNA-binding domain X-ray structures only)[56]; PaFur[27]; FtFur (this work); HpFur (*Helicobacter pylori*)[22] and MgFur[29]. The highly conserved amino acids implicated, in site S2 (blue) and in the interactions with DNA are in bold, coloured in red for those forming base-specific interactions and black for those having interactions with the phosphates (as evidenced in the structure of MgFur in complex with DNA). **d** Sketch of the DNA FurBox double-strand highlighting interactions with four Fur subunits (forming two dimers). Each of them is shown in a specific colour: yellow, purple, green and cyan, corresponding to the residues shown in **b**. Interactions between DNA bases and Fur residues are deduced from our results and the structures of the MgFur–DNA complex (T highlighted in red interact with two subunits)

importance of these residues in the stability of the tetramer since dimeric forms were obtained, partially for the single mutants and completely for the double mutant at high salt concentrations (Supplementary Fig. 5A, B). Both single mutants were still able to bind DNA in the presence of metal ions contrarily to the double mutant. Moreover, the Arg57–Glu63 salt bridge disruption should leave room to interactions of the crucial Tyr56 with DNA.

Fur is involved in *F. tularensis* virulence and pathogenicity. The critical role of Fur in pathogenicity and virulence of several pathogens is known[16,33]. To investigate the putative role of FtFur as a virulence factor we compared the phenotypes of CHUGA-Ft6 and CHUGA-Ft6Δ*fur* using in vitro or in vivo infection models. Three types of experiments were conducted: bacterial multiplication in J774-A1 murine macrophage-like cells, H_2O_2 sensitivity assay and in vivo virulence assays in mice.

A growth defect of the CHUGA-Ft6 mutant lacking *fur* in liquid medium was evidenced as shown by a longer lag time, a longer generation time and a lower optical density at the stationary phase as compared to the WT parental strain (Supplementary Fig. 1B). A similar phenotype was observed on solid medium with a delayed onset of visible colonies and a smaller size of colonies for CHUGA-Ft6Δ*fur* (Supplementary Fig. 1C). The ability of CHUGA-Ft6, CHUGA-Ft6Δ*fur* and CHUGA-Ft6Δ*fur* + *fur* to replicate within macrophages was then evaluated by infecting J774-A1 murine macrophage-like cells. One hour after infection, the host cells contained the same number of intracellular bacteria regardless of the infecting strain meaning that Fur is not required for macrophage infection. After 24 h incubation (Fig. 8a), the number of intracellular bacteria was markedly different as the WT cells were eight-fold more abundant compared to the CHUGA-Ft6Δ*fur*. The *fur*-complemented strain showed an intermediate level of intracellular macrophage multiplication. The ability of these bacterial strains to resist an oxidative stress corresponding to the respiratory burst set up by infected macrophages was also checked by growing bacteria previously exposed to 1 mM H_2O_2 during 4 h (Fig. 8b). CHUGA-Ft6 and CHUGA-Ft6Δ*fur* + *fur* displayed a similar percentage of survival while Ft6Δ*fur* was much more sensitive to the oxidative stress with about 50% of surviving cells.

The involvement of *fur* in the infectious process in vivo was then evaluated by using mice infected with *F. tularensis* by intranasal (IN) or intraperitoneal (IP) administration (Fig. 8c, d). The survival curves of the animals showed that regardless of the administration route, CHUGA-Ft6 and CHUGA-Ft6Δ*fur* + *fur* caused the mice death in approximately the same delay, which is 95 h post-infection for IP and 150 to 168 h for IN inoculation. On the other hand, mice infected with CHUGA-Ft6Δ*fur* survived a significantly longer time ($p < 0.001$ compared to CHUGA-Ft6 and CHUGA-Ft6Δ*fur* + *fur* whatever the route of infection), that is, 140 h and more than 200 h for the last animal infected by IP and by IN routes, respectively. These results define *fur* as an important virulence-associated gene in *F. tularensis* and are a further example that deletion of this gene leads to an attenuated phenotype in terms of virulence.

Discussion

Altogether, the involvement of Fur in the iron homeostasis and the virulence of *F. tularensis* have been demonstrated here as well as its direct interaction with the *figA* promoter region. FtFur belongs to the new family of tetrameric Fur proteins. It contains the structural zinc site S1 and the regulatory site S2 and lacks the third site S3, usually found in Fur proteins. S1 is not present in tetrameric PaFur and S3 is absent in FtFur, which still forms a tetramer upon deletion of S2, indicating that the tetrameric state of the protein does not rely on such sites. To our knowledge, the first published structure of FtFur containing the physiologically relevant ferrous iron is presented here with a ferrous ion in an octahedral geometry. Metalled FtFur behaves as a pre-activated tetramer with the DNA-binding domains forming a positively charged crown where the recognition helices interact through two stabilizing salt bridges between two Arg57 (out of four) and two

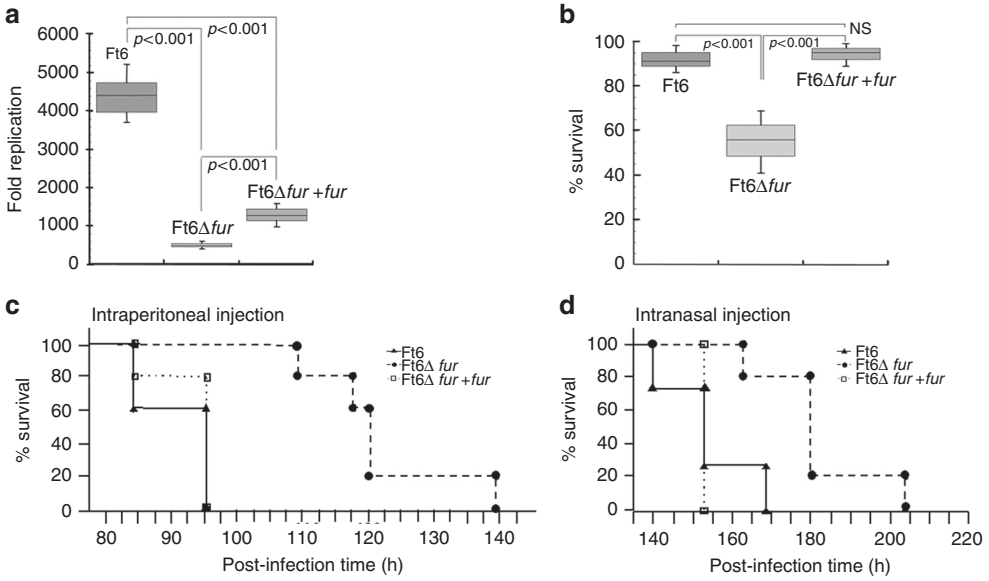


Fig. 8 Fur is directly involved in *F. tularensis* virulence and pathogenicity. **a** Bacterial multiplication in J774-A1 murine macrophage-like cells. The data correspond to two independent experiments made in triplicate. *P* values were calculated using the Student's *t* test. **b** Hydrogen peroxide sensitivity assay expressed as the percent survival of each strain exposed 4 h to 1 mM H_2O_2. Bacterial suspensions were incubated 4 h with 1 mM H_2O_2 before enumeration of the CFU onto chocolate agar plates. Data are expressed as mean ± SEM of three distinct experiments. The data correspond to two independent experiments made in triplicate. *P* values were calculated using the Student's *t* test. **c**, **d** In vivo virulence assay in mice inoculated IP with 5e2 CFU in 500 µL of physiological serum or IN with 2e3 CFU in 50 µL of physiological serum. Five mice were used for each condition and experiments were performed twice. Survival curves were compared using the Kaplan–Meier test. *P* values were <0.001 for CHUGA-Ft6 and CHUGA-Ft6Δ*fur* + *fur* vs. CHUGA-Ft6Δ*fur*, for both the IP and IN routes. No significant difference was found between CHUGA-Ft6 and CHUGA-Ft6Δ*fur* + *fur*

Glu63. The tetramer dissociation is driven by an interaction of the protein with a specific DNA sequence, suggesting the involvement of the two free Arg57 and Lys14, known to form base-specific contacts with DNA. We postulate that the two H-bound Arg57 would progressively lose their interaction with Glu63 replaced by interactions with DNA, leading to the breaking of the salt bridge, crucial for the stability of the tetramer and its dissociation into two dimers specifically bound to the FurBox. In vivo studies reveal that FtFur is important for the virulence of *F. tularensis*. Because there is no efficient vaccine and only few poorly efficient antibiotics available to fight tularaemia, this work shows that Fur is an attractive anti-virulence target for new inhibitors, whose design, starting from already known inhibitors against other Fur proteins[34], will be facilitated by the detailed structure and mechanism of interaction with DNA.

Methods

Bacterial strains and culture media. The biovar I strain of *F. tularensis* subsp. *holarctica* used in the virulence assay, referred as CHUGA-Ft6, was isolated at Verdun Hospital (France) from a blood sample collected during routine care of a patient with typhoidal tularaemia. Identification at the species and subspecies level was obtained by PCR amplification and sequencing of the intergenic region between the 16SrRNA and 23SrRNA encoding genes[35]. Bacterial cultures were performed either on chocolate agar plates supplemented with PolyVitex (CPV, Biomérieux, Lyon, France) or in liquid brain heart infusion medium supplemented with 2% PolyViteX (BHI-2%PV). When necessary, kanamycin (10 µg mL^{-1}) or sucrose (5% (w/v)) was added. Cultures were incubated at 37 °C, in a 5% CO_2-enriched atmosphere. Intracellular iron concentration was measured on stationary phase bacteria grown over 15 h in modified Mueller–Hinton medium into a shaking incubator (200 rpm at 37 °C). Briefly, the cells have been washed several times with phosphate-buffered saline-ethylenediaminetetraacetic acid (PBS-EDTA) 10 mM before hydrolysis with HNO_3 at 65% ON at 95 °C and measurements by ICP-AES (Shimadzu ICP 9000 instrument with Mini plasma Torch in axial reading mode[20]).

Construction and complementation of the Ft6 Δ*fur* mutant. A phase deletion of the *fur* gene was carried out in the CHUGA-Ft6 virulent isolate, by the method of allelic exchange (Supplementary Fig. 1), through the use of a suicide plasmid containing the *sacB* gene, pMP812[36]. Approximately 1000 bp adjacent regions of the Ft6 *fur* gene were amplified using the two primers LeftfurF-LeftfurR and RightfurF-RightfurR (Supplementary Table 4). The obtained PCR products were mixed and further submitted to a second PCR using the forward primer LeftfurF and the reverse primer RightfurR, generating a PCR fragment containing the two adjacent regions of *fur* flanked with the BamHI and EcoRV restriction sites. This fragment was digested with the two corresponding restriction enzymes and cloned into the plasmid pMP812[36] previously digested with the same enzymes. After electroporation, several selection (kanamycin, then sucrose) steps were performed to obtain a delta *fur* mutant devoid of antibiotic resistance. In order to complement the strain Ft6Δ*fur*, the *fur* gene and its promoter region were amplified using CompfurL and CompfurR primers and cloned the shuttle vector pMP828. The plasmid pMP828 containing the *fur* gene was then electropored into CHUGA-Ft6Δ*fur* and complemented colonies selected on agar plates supplemented with kanamycin. The *fur* expression in resulting transformants was checked using a specific qRT-PCR.

Evaluation of the gene expression by qRT-PCR. Gene expression was measured from strains grown for 16 h in BHI-2%PV. Approximately 10^7 cells were collected for total RNA extraction that was achieved using 1 mL TRIzol® reagent (Invitrogen) and following the manufacturer's instructions. Contaminating DNA was removed using the TURBO DNA-free™ Kit (Ambion, Life Technology). The first-strand complementary DNA (cDNA) synthesis reaction was carried out starting from 500 ng of purified RNA and using the SuperScript™ II Reverse Transcriptase Kit (Invitrogen, Life Technology). The resulting cDNA library was used as a template in combination with the specific primers for qRT-PCR, which was conducted using a Fast SYBR Green MasterMix in a StepOnePlus Real Time PCR Systems (Applied Biosystems). Cycling was 20 s at 95°C; 3 s at 95°C, 30 s at 60 °C, repeated for 40 cycles. The expression level of each target gene was calculated from three independent experiments and expressed as a ratio taking the expression of the housekeeping gene 16S RNA as the denominator. Primer sequences are indicated in Supplementary Table 5.

DNA sample preparation. DNA oligonucleotides were synthesized by MWG at high purity scale. DNA duplexes were first annealed in water at concentration of 20 mg mL^{-1} by heating the mixture at 95 °C for 5 min and rapid cooling on ice in buffer A (50 mM Tris-HCl, pH 8.8, 150 mM NaCl) and then stored at 4 °C. The formation and concentration of DNA duplexes were determined by SEC-MALLS-RI in binding buffer. DNA was used extemporary for biochemical experiment.

Protein expression and purification
Apo-FtFurWT. Recombinant FtFur of *F. tularensis* FSC198 ref: NC-008245.1, 100% identical in sequence to the CHUGA-Ft6 protein, was purified as a tetramer containing one equivalent of Zn per subunit, as previously described[20]. It was over-produced in BL21 DE3 R2 *E. coli* strain in Luria–Bertani (LB) medium after an overnight induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C. Cell were resuspended in buffer A, lysed by sonication and purified successively on several columns (GE Healthcare): (1) ion-exchange chromatography on DEAE Sepharose with linear gradient between buffer A and buffer A supplemented with 1 M NaCl, (2) hydrophobic interaction chromatography on Butyl fast flow Sepharose with linear gradient between buffer A containing 1 M of ammonium sulphate and buffer A and (3) size-exclusion chromatography Superdex-200 (10/60) equilibrated with buffer A.

Apo-FtFur mutants. E63A, E76A and E63A-E76A mutants were cloned in pET-TEV (based on pET28a) vector to produce N-terminal 6×HisTag cleavable TEV fusion proteins. PCR was done in the presence of appropriate primer with Phusion polymerase-HF at recommended T_m. PCR samples were incubated with a reaction buffer containing 2 UI of *Dpn*I, 10 UI of T4 DNA ligase, 1 mM of ATP and 2 UI polynucleotide kinase (PNK) in PNK buffer from NEB for 15 min at room temperature (RT) before transformation in Top10 ultracompetent cells. Each mutant has been DNA-sequenced before expression and purification as described. H33A-H90A double mutant (FtFurΔS2) was obtained from pET30b-FtFurWT before cloning in pET-TEV, over-produced and purified like FtFurWT. The other mutants were over-produced in in BL21 (DE3) R2 *E. coli* strain LB medium after induction with 1 mM IPTG at 37 °C for 4 h. Purification was done by using Ni-NTA resin (Qiagen) in batch mode in buffer A with 10 mM imidazole and 10% glycerol. Pure protein fractions were pooled and mixed with homemade HisTag-TEV protease (1% of the protein concentration to purify) together with 1 mM dithiothreitol (DTT) and 1 mM EDTA. The solution was then dialysed using a 3 kDa cut-off membrane against 2 L buffer A containing 1 mM DTT and 1 mM EDTA at 4 °C overnight, followed by a second 2 L dialysis against buffer A to remove DTT and EDTA. The protein sample was then passed through the Ni-NTA column in order to separate the pure protein from the HisTag-TEV protein and the HisTag itself. A final step of purification was performed by using Superdex-200 in buffer A supplemented with 500 mM NaCl at 4 °C. Collected fractions were concentrated on a 50 kDa cut-off Vivaspin from 20 to 40 mg mL^{-1} and used or frozen in liquid nitrogen in the presence of 10% of glycerol before storage at −80 °C.

Purification of the Mn-FtFur/Fur Box complex. The purification was done in buffer B (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM $MnCl_2$, 1 mM $MgCl_2$) at 20 °C. Molar equivalents (1.2) of FurBox Duplex (see DNA sample preparation) were incubated with Mn-FtFur before loading on a Superdex-200 increase 10/30 GE Healthcare column equilibrated with buffer B. Pooled fractions were analysed in SEC-MALLS-RI in the same buffer to check the integrity of the complex.

EMSA and nuclease activity assay. EMSA experiments were performed as previously described[20]. The formation of small-scale (under 1 µg) DNA duplexes was confirmed by native gel electrophoresis[20] on 10% acrylamide gel in 1× TAE buffer (40 mM Tris-acetate, pH 8.2, 1 mM EDTA). DNA radiolabelling was performed by incubating 20 nM DNA for 30 min at 37 °C in the presence of 1 U of T4 polynucleotide kinase (NEB) and 1 µL of γATP at 1 mCi mmol^{-1}. Labelled DNA was diluted 10 times in buffer A, desalted on G25 Mini Spin Column and stored at −20 °C. EMSA were performed with 250 pM of freshly prepared radiolabelled DNA incubated 30 min at 25 °C with different concentrations of protein in a binding buffer (20 mM Bis-Tris propane, pH 8.5, 100 mM KCl, 3 mM $MgCl_2$, 10 µM $MnCl_2$, 5% (v/v) glycerol and 0.01% Triton X-100). After 30 min incubation at room temperature, 10 µL of each sample were loaded on 10 % polyacrylamide (29/1) gel. The gel was pre-run for 30 min at 100 V in TA buffer (40 mM Tris-acetate, pH 8.2) supplemented with 100 µM of $MnCl_2$. Mobility shifts were revealed by exposing the gels on a storage phosphor screen (GE Healthcare) and quantified with a cyclone phosphoimager (Perkin Elmer). The nuclease activity assay was performed as previously described[20].

SEC-MALLS-RI experiments. Each sample was checked by size-exclusion chromatography coupled to multi-angle laser light scattering with online refractive index (SEC-MALLS-RI) as previously described[20] and using a standard procedure: 20 µL of sample with a 2 to 10 mg mL^{-1} concentration were loaded on an analytical Superdex-S200 increase pre-equilibrated at 0.5 mL min^{-1} with appropriate buffer (Apo-FtFur in buffer A, Mn-FtFur and Mn-FtFur/FurBox in 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM $MnCl_2$, 1 mM $MgCl_2$) and connected to an in-line DAWN HELEOS II spectrometer (Wyatt Instruments). An in-line refractive index detector (Optirex, Wyatt Instruments) was used to follow the differential refractive index relative to the solvent. After baseline subtraction of the buffer solution, all samples presented a single peak allowing the determination of absolute molecular masses with the Debye model using ASTRA6 software (Wyatt Instruments) and a theoretical dn/dc value of 0.185 mL g^{-1}. The final values correspond to the average of three independent experiments.

Crystallization. Protein crystallization conditions were obtained by using crystallization screens (Hampton Resarch Grid Screens™ and Qiagen protein crystallization suites) with the HTX Lab high-throughput robot screening (HTX Lab at EMBL-Grenoble). Diffracting crystals up to 8 Å were obtained in 50 mM MES, pH 5.6, 200 mM KCl, 5% (w/v) PEG 8000, 10 mM magnesium chloride hexahydrate. This crystallization condition was then manually optimized. Diffracting crystals up to 1.7 Å were obtained by 1 µL of a 16 mg mL^{-1} Mn-FtFur holoprotein solution with 1 µL of 50 mM MES, pH 5.8, 20% (w/v) PEG 3350, 200 mM $MgCl_2{\cdot}6H_2O$, 10 mM $MnCl_2$ reservoir solution, using the hanging drop vapour-diffusion method. Crystals of Fe-FtFur holoproteins were obtained in the same condition in the presence of 10 mM of $(NH_4)_2Fe(SO_4)_2{\cdot}6\ H_2O$ under anaerobic condition in a glove box. Crystals appeared in a few days and were back-soaked three successive times in a mother liquor containing 50 mM MES, pH 5.8, 20% (w/v) PEG 3350, 200 mM $MgCl_2{\cdot}6H_2O$, to remove the excess of free metal. All crystals were cryoprotected using a solution obtained by adding 25% (v/v) glycerol to the mother liquor containing 25% (w/v) PEG 3350 and flash-cooled in liquid nitrogen.

X-ray diffraction and structure resolution of Mn-FtFur and Fe-FtFur. Diffraction experiments were done on the beamline FIP-BM30-ESRF (European Synchrotron Radiation Facility, Grenoble, France).

For Mn-FtFur, a fluorescence spectrum was recorded to check the presence of Mn at 1.77 Å (right side of maximum Fd″ for manganese) and Zn at 0.97 Å cations (Supplementary Fig. 3A). A remote and an anomalous datasets were recorded at wavelengths 0.97 and 1.77 Å (right side of maximum F″ for manganese). Best dataset (0.97 Å) diffracted at 1.7 Å resolution. Diffraction data were integrated/scaled in the space group $P2_1$ using XDS Program Package version 15 October 2015[37]. The structure was solved by the SAD method using Phenix 1.10.1-2155/AutoSol[38] and 86% of the model was built automatically. The model was rebuilt/corrected manually and refined using alternatively COOT[39] and REFMAC[40,41]. Final refinement cycle was done in Phenix (Supplementary Table 1 and Supplementary Fig. 3B).

For Holo-Fe-FtFur, a fluorescence spectrum was recorded to check the presence of cations: Fe and Zn at 0.97 Å (Supplementary Fig. 3A). The datasets were collected at wavelengths 0.97 Å with a resolution of 1.8 Å, integrated/scale by XDS Program Package in the space group $P2_1$.

The structure was solved by molecular replacement using MOLREP[42] with the previous structure (Mn-FtFur) as the starting model. Indeed, we were not able to obtain a Fe-FtFur structure in the same conditions using iron Mohr salt in place of manganese even in a glove box. When trying to obtain the crystal from SEC-purified Mn-FtFur–DNA complex, only the same Mn-FtFur crystals were seen, but adding iron in the crystallization conditions, we were able to get crystals of Fe-FtFur diffracting at 1.8 Å and to solve the structure by molecular replacement with Mn-FtFur. The anomalous dataset was used to confirm the presence of Fe in the structure. The model was built and refined using REFMAC and COOT alternatively (Supplementary Table 2 and Supplementary Fig. 3B).

Both Holo-Mn-FtFur and Holo-Fe-FtFur final structures were validated by molprobity[43]. Protein Data Bank (PDB) redo[44] was used before deposition of the structures to the PDB. The PDB codes are 5NBC for Holo-Mn-FtFur and 5NHK for Holo-Fe-FtFur. King software (in Phenix) was also used to cross-validate the data.

SAXS experiments. Before each experiment, all samples were extemporaneously re-purified on SEC Superdex-200 increase (GE Healthcare) 10/300 equilibrated in an appropriate buffer. SAXS data were collected at the ESRF (Grenoble, France) on beamline BM29 BioSAXS. The scattering profiles were measured at several concentrations between 0.5 to more than 10 mg mL^{-1}. Data were processed using standard procedures with the ATSAS v2.5.1 suite of programs[45] as described[20]. The ab initio determination of the molecular shape of the proteins was done as previously described[20]. Briefly, radius of gyration (R_g), forward intensity at zero angle ($I(0)$), Porod volumes and Kratky plot were determined using the Guinier approximation and PRIMUS programs[46]. In order to build ab initio models, several independent DAMMIF[32] models were calculated in slow mode with pseudo-chain option and analysed using the program DAMAVER[47]. Docking of the tetrameric X-ray structure into the measured SAXS envelope was generated by SUPCOMB[48]. The model of the Mn-FtFur/FurBox complex was built from the Mn-MgFur/FurBox structure (PDB code 4RB1): after sequence alignment, the atom coordinates of the corresponding amino acids were directly copied from MgFur to FtFur and coordinates of missing side chain atoms were added from internal coordinates. The resulting model was energy minimized with CHARMM[49]. The program CRYSOL[32] was used to generate the theoretical scattering curves from the tetrameric structure of FtFur.

Cross-linking experiments. Cross-linking experiments between FtFur and Fur boxes were performed using 0.1% GTA. With a short spacer arm of approx: 5 Å, and when used at a low concentration, this cross-linker agent is well suited for intramolecular cross-linking and to specifically cross-link individual species in close interactions. Two micrograms of Mn-FtFur were used in each tube, fresh GTA was used at 0.1% and 25–50 and 100 ng of DNA oligonucleotides duplex was added sequentially. Incubation buffer are done with 1 mM of fresh $MnCl_2$ at RT

before being cross-linked by GTA 30 min at RT and loaded onto sodium dodecyl sulphate-polyacrylamide gel electrophoresis 4–20% acrylamide gradient.

Construction of the models. The X-ray structure of FtFur solved in this study (Asp7 to Arg137) was used as the initial model for the tetramer. The GROMACS program version 5.1.2[50] with the gromos54a7 united atom force field[51] was used to perform long molecular dynamics simulations needed to compute free energy profiles. Fe^{2+} and Zn^{2+} were modelled as simple Lennard Jones hard spheres with charge +2 with Zn coordinated to charged deprotonated cysteines (see for details Supplementary Methods).

In the absence of FtFur + DNA structure (FtFur/wtDNA), the structure of *M. gryphiswaldense* (4RB1)[29] in the presence of DNA was used to model the wtDNA FurBox and correctly position FtFur dimer on wtDNA (by least-square fit matching of atom positions). The 5′-GCCG**GATAATGATAATCATTATC**-3′ fragment (consensus FurBox in bold) and its complementary 3′–5′ sequence was used to model double-stranded wtDNA.

The mutDNA (equivalent to $FurBox_m$) sequence GCCGGATA**C**TGATA**G**TC **C**T**G**ATC contains four mutations with respect to the FurBox (A9 to C, A15 to G, A18 to C and T20 to G; see nucleotides set in bold font) which were constructed by simple matching of corresponding heavy atoms in the WT DNA model and building of missing hydrogens. The three above vacuum systems were immersed in parallelepipedic SPC[52] water boxes modelled with periodic boundary conditions after the addition of Na^+ and Cl^- counterions to ensure neutrality and a total ionic force of 0.1 mol L^{-1}. The solvated systems were energy minimized and equilibrated under NPT (constant Number of particles, Pressure and Temperature) conditions at 310 K and 1 atm.

Computation of free energy profiles. Free energy profiles for the extraction (by translation along a fixed direction: Ox) of one FtFur dimer from the tetramer (dimer of dimers) and of FtFur from DNA were computed: the meticulous translation protocol is shown in Supplementary Fig. 6.

The simulations include a 'moving' subsystem (FtFur dimer, chains A and D) and a 'fixed' subsystem (FtFur dimer, chains B and C, wtDNA or mutDNA) as shown in Supplementary Fig. 7.

The profiles were built using the 'umbrella sampling' technique and result from the overlapping of 26 computation windows, one for each translation distance.

Each window consisted of 100 ps NPT equilibration and 10 to 15 ns NPT production simulations. Position restraints on the 'fixed' subsystem and distance restraints on the whole protein, in the form of NOE-type restraints (nuclear Overhauser effect) between H-bonded H and O atoms to maintain its secondary structure, were applied. The 'moving' subsystem was subject to two harmonic biasing forces along the X direction only ('umbrella potential') applied between the centres of mass of the 2 Fur dimer subunits and the centre of mass of the 'fixed' subsystem.

After the dynamics runs, positions and forces were collected from the trajectories and the umbrella sampling harmonic potential was unbiased using the Wham algorithm[53] implemented in the 'g_wham' program[54] to yield the free energy profiles.

Computation of average interaction energy profiles. Interaction energy profiles were computed by extracting nonbonded interactions (electrostatic + Lennard Jones potential energies) from all the trajectories of the simulations and averaging for each window. Interaction energies were calculated between each residue in the Fur 'moving' dimer (chains A and D) and the 'fixed' subsystem (DNA or 'fixed' FtFur dimer).

H_2O_2 sensitivity assay. Three independent overnight cultures of each strain (CHUGA-Ft6, CHUGA-Ft6Δ*fur* and CHUGA-Ft6Δ*fur* + *fur*) in BHI-2%PV medium were diluted in PBS to obtain 1×10^7 colony-forming unit (CFU) mL^{-1} inocula. These bacterial suspensions were incubated 4 h with 1 mM H_2O_2 at room temperature without shaking. Culture samples taken before and after H_2O_2 exposure were serially diluted and spread onto CPV plates and incubated for 3 days at 37 °C. Counting of CFUs allowed to determine the percentage of surviving bacteria after H_2O_2 exposure. Data were expressed as the mean of three independent experiments.

In vitro macrophage infection. J774-A1 murine macrophage-like cells were grown in 24-well microplate using Dulbecco's modified Eagle's medium (DMEM) (Gibco) medium supplemented with 10% decomplemented fetal calf serum (FCS, Gibco) at 37 °C in 5% CO_2 enriched atmosphere. For infections assays, confluent J774-A1 monolayers were infected with a bacterium inoculum at a multiplicity of infection of 10:1 (approximately 5×10^6 bacteria for 5×10^5 cells), and reincubated for 1 h (37 °C, 5% CO_2). The cell monolayers were then washed with PBS (Gibco), and DMEM-10% FCS containing 5 μg mL^{-1} of gentamycin was added for 1 h to kill non-phagocytized bacteria. The cell supernatant was then replaced with MEM-10% FCS, and cell cultures were incubated for 24 h (37 °C, 5% CO_2). To evaluate bacterial multiplication within J774-A1 macrophages, infected cell monolayers treated with 0.1 % saponin immediately after gentamycin treatment (T0) or after 24 h incubation (T24) were lysed and serially diluted in PBS and plated onto chocolate agar plates enriched with PolyVitex for CFU numeration. Each point was performed in triplicate. Data obtained were compared using Student's *t* test and *P* values <0.05 were considered statistically significant.

In vivo virulence assay. Six-week- to eight-week-old BALB/c males were infected with overnight cultures of the strains CHUGA-Ft6, CHUGA-Ft6Δ*fur* and CHUGA-Ft6Δ*fur* + *fur* diluted in 0.9% NaCl. Experiments were performed in an animal biosafety level 3 laboratory. For each bacterial strain, groups of five mice were inoculated either intraperitoneally (500 CFUs in 500 μL) or intranasally (2000 CFUs in 50 μL) and infected animals were monitored several times a day, weighed every day, and euthanized when they had reached one of the following limit point: prostration, high piloerection, weight loss >15% of T0 weight and closed eyes. A group of unifected mice was used as control. All murine experiments were approved by our local ethics committee (ComEth, Grenoble, France). During the experiments, mice were monitored several times a day, weighed every day and euthanized when we felt they had reached our estimated limit point (prostrate, high piloerection, weight loss >15% of T0 weight and eyes closed). All these experiments were performed in compliance with the laws and regulations regarding animal experimentation in France.

Data availability. The datasets generated during the current study are available from the corresponding authors on reasonable request. Coordinates and structure factors for Mn-FtFur and Fe-FtFur have been deposited in the RCSB Protein Data Bank under accession codes 5NBC and 5NHK, respectively.

Received: 9 January 2018 Accepted: 14 June 2018
Published online: 17 July 2018

References

1. Maurin, M. & Gyuranecz, M. Tularaemia: clinical aspects in Europe. *Lancet Infect. Dis.* **16**, 113–124 (2016).
2. Meibom, K. L. & Charbit, A. The unraveling panoply of *Francisella tularensis* virulence attributes. *Curr. Opin. Microbiol* **13**, 11–17 (2010).
3. Conlan, J. W., Chen, W., Shen, H., Webb, A. & KuoLee, R. Experimental tularemia in mice challenged by aerosol or intradermally with virulent strains of *Francisella tularensis*: bacteriologic and histopathologic studies. *Microb. Pathog.* **34**, 239–248 (2003).
4. Maurin, M. *Francisella tularensis* as a potential agent of bioterrorism? *Expert Rev. Anti Infect. Ther.* **13**, 141–144 (2015).
5. Oyston, P. C., Sjostedt, A. & Titball, R. W. Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nat. Rev. Microbiol.* **2**, 967–978 (2004).
6. Perez, N. M. & Ramakrishnan, G. The reduced genome of the *Francisella tularensis* live vaccine strain (LVS) encodes two iron acquisition systems essential for optimal growth and virulence. *PLoS ONE* **9**, e93558 (2014).
7. Ramakrishnan, G., Sen, B. & Johnson, R. Paralogous outer membrane proteins mediate uptake of different forms of iron and synergistically govern virulence in *Francisella tularensis* tularensis. *J. Biol. Chem.* **287**, 25191–25202 (2012).
8. Lindgren, H. et al. Iron content differs between *Francisella tularensis* subspecies tularensis and subspecies holarctica strains and correlates to their susceptibility to H(2)O(2)-induced killing. *Infect. Immun.* **79**, 1218–1224 (2011).
9. Ramakrishnan, G., Meeker, A. & Dragulev, B. fslE is necessary for siderophore-mediated iron acquisition in *Francisella tularensis* Schu S4. *J. Bacteriol.* **190**, 5353–5361 (2008).
10. Pan, X., Tamilselvam, B., Hansen, E. J. & Daefler, S. Modulation of iron homeostasis in macrophages by bacterial intracellular pathogens. *BMC Microbiol.* **10**, 64 (2010).
11. Carrano, C. J. et al. Coordination chemistry of the carboxylate type siderophore rhizoferrin: the iron(III) complex and its metal analogs. *Inorg. Chem.* **35**, 6429–6436 (1996).
12. Sullivan, J. T., Jeffery, E. F., Shannon, J. D. & Ramakrishnan, G. Characterization of the siderophore of *Francisella tularensis* and role of fslA in siderophore production. *J. Bacteriol.* **188**, 3785–3795 (2006).
13. Wehrly, T. D. et al. Intracellular biology and virulence determinants of *Francisella tularensis* revealed by transcriptional profiling inside macrophages. *Cell Microbiol.* **11**, 1128–1150 (2009).
14. Deng, K., Blick, R. J., Liu, W. & Hansen, E. J. Identification of *Francisella tularensis* genes affected by iron limitation. *Infect. Immun.* **74**, 4224–4236 (2006).
15. Kiss, K., Liu, W., Huntley, J. F., Norgard, M. V. & Hansen, E. J. Characterization of fig operon mutants of *Francisella novicida* U112. *FEMS Microbiol. Lett.* **285**, 270–277 (2008).
16. Fillat, M. F. The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators. *Arch. Biochem. Biophys.* **546**, 41–52 (2014).

17. Seo, S. W. et al. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. *Nat. Commun.* **5**, 4910 (2014).
18. Touati, D. Iron and oxidative stress in bacteria. *Arch. Biochem. Biophys.* **373**, 1–6 (2000).
19. Ughetto, E. et al. An original case of *Francisella tularensis* subsp. holarctica bacteremia after a near-drowning accident. *Infect. Dis. (Lond.)* **47**, 588–590 (2015).
20. Perard, J. et al. Quaternary structure of Fur proteins, a new subfamily of tetrameric proteins. *Biochemistry* **55**, 1503–1515 (2016).
21. Vitale, S. et al. A ZnS(4) structural zinc site in the *Helicobacter pylori* ferric uptake regulator. *Biochemistry* **48**, 5582–5591 (2009).
22. Dian, C. et al. The structure of the *Helicobacter pylori* ferric uptake regulator Fur reveals three functional metal binding sites. *Mol. Microbiol.* **79**, 1260–1275 (2011).
23. Watnick, P. I., Eto, T., Takahashi, H. & Calderwood, S. B. Purification of *Vibrio cholerae* fur and estimation of its intracellular abundance by antibody sandwich enzyme-linked immunosorbent assay. *J. Bacteriol.* **179**, 243–247 (1997).
24. Zheng, M., Doan, B., Schneider, T. D. & Storz, G. OxyR and SoxRS regulation of fur. *J. Bacteriol.* **181**, 4639–4643 (1999).
25. Mills, S. A. & Marletta, M. A. Metal binding characteristics and role of iron oxidation in the ferric uptake regulator from *Escherichia coli*. *Biochemistry* **44**, 13553–13559 (2005).
26. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
27. Pohl, E. et al. Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol. Microbiol.* **47**, 903–915 (2003).
28. Tiss, A., Barre, O., Michaud-Soret, I. & Forest, E. Characterization of the DNA-binding site in the ferric uptake regulator protein from *Escherichia coli* by UV cross-linking and mass spectrometry. *FEBS Lett.* **579**, 5454–5460 (2005).
29. Deng, Z. et al. Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nat. Commun.* **6**, 7642 (2015).
30. Gilston, B. A. et al. Structural and mechanistic basis of zinc regulation across the *E. coli* Zur regulon. *PLoS Biol.* **12**, e1001987 (2014).
31. Stojiljkovic, I., Baumler, A. J. & Hantke, K. Fur regulon in Gram-negative bacteria—identification and characterization of new iron-regulated *Escherichia coli* genes by a Fur Titration Assay (Vol 236, Pg 531, 1994). *J. Mol. Biol.* **240**, 271–271 (1994).
32. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
33. Porcheron, G. & Dozois, C. M. Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity. *Vet. Microbiol.* **179**, 2–14 (2015).
34. Mathieu, S. et al. From peptide aptamers to inhibitors of FUR, bacterial transcriptional regulator of iron homeostasis and virulence. *ACS Chem. Biol.* **11**, 2519–2528 (2016).
35. Maurin, M., Pelloux, I., Brion, J. P., Del Bano, J. N. & Picard, A. Human tularemia in France, 2006–2010. *Clin. Infect. Dis.* **53**, e133–e141 (2011).
36. LoVullo, E. D., Molins-Schneekloth, C. R., Schweizer, H. P. & Pavelka, M. S. Jr. Single-copy chromosomal integration systems for *Francisella tularensis*. *Microbiology* **155**, 1152–1163 (2009).
37. Kabsch, W. Xds. *Acta Crystallogr D* **66**, 125–132 (2010).
38. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* **66**, 486–501 (2010).
40. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D* **60**, 2184–2195 (2004).
41. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235–242 (2011).
42. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
43. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12–21 (2010).
44. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
45. Petoukhov, M. V. et al. New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
46. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
47. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
48. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
49. Brooks, B. R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (1983).
50. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1* **11**, 19–25 (2015).
51. Schmid, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–856 (2011).
52. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Interaction Models for Water in Relation to Protein Hydration* (ed. Pullman, i. B.) 331–-342 (Reidel, Dordrecht, 1981).
53. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H. & Kollman, P.A. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
54. Hub, J. S., de Groot, B. L. & van der Spoel, D. GROMACS g_wham A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **6**, 3713 (2010).
55. R-Development-Core-Team. in *R: A Language and Environment for Statistical Computing* (ed. Computing, R.F.f.S) (R-Development-Core-Team, Vienna, Austria, 2008).
56. Pecqueur, L. et al. Structural changes of *Escherichia coli* ferric uptake regulator during metal-dependent dimerization and activation explored by NMR and X-ray crystallography. *J. Biol. Chem.* **281**, 21286–21295 (2006).

Acknowledgements

We acknowledge the use of resources of INEXT 2217 and the support of members of the HTX Lab (EMBL, Grenoble, France). We also thank the European Synchrotron Radiation Facility for access to beamlines FIP BM30A and ID30A-1. This work was funded by the CEA, the Laboratory of Excellence GRAL (grant ANR-11-LABX-49-01) and the LabEx ARCANE (grant ANR-11-LABX-0003-01). S.N. was supported by the Region Rhône-Alpes (ARC santé).

Author contributions

J.P., S.C., M.M., P.R., J.C. and I.M.-S. designed research; J.P., S.N., C.C., P.C. and L.A. carried out the biochemical and structural experiments, J.P., P.C. and C.C. resolved the structures; M.L., C.S., D.S., F.B. and P.R. carried out the in vivo experiments, S.N. and S. C. did the theoretical work; J.P., J.C., M.M., S.C. and I.M.-S. analysed data; J.P., S.C., J.C. and I.M.-S. wrote the paper.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-018-0095-6.

Competing interests: The authors declare no competing interests.

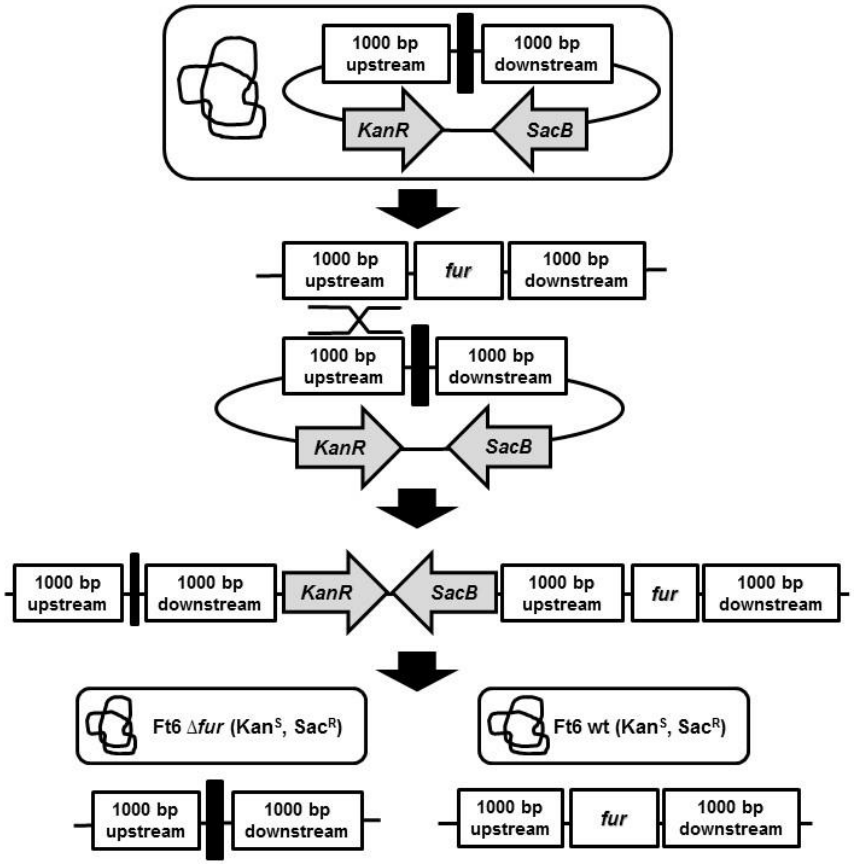
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

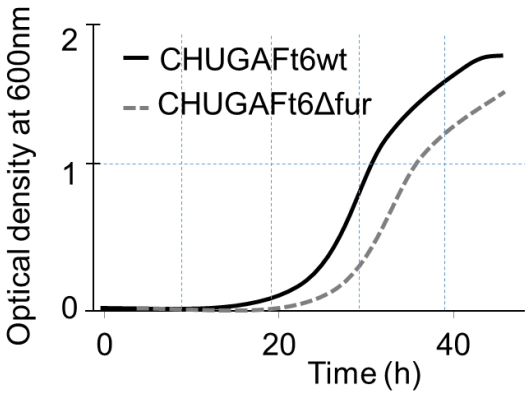
Supplementary information

Supplementary Figures

A

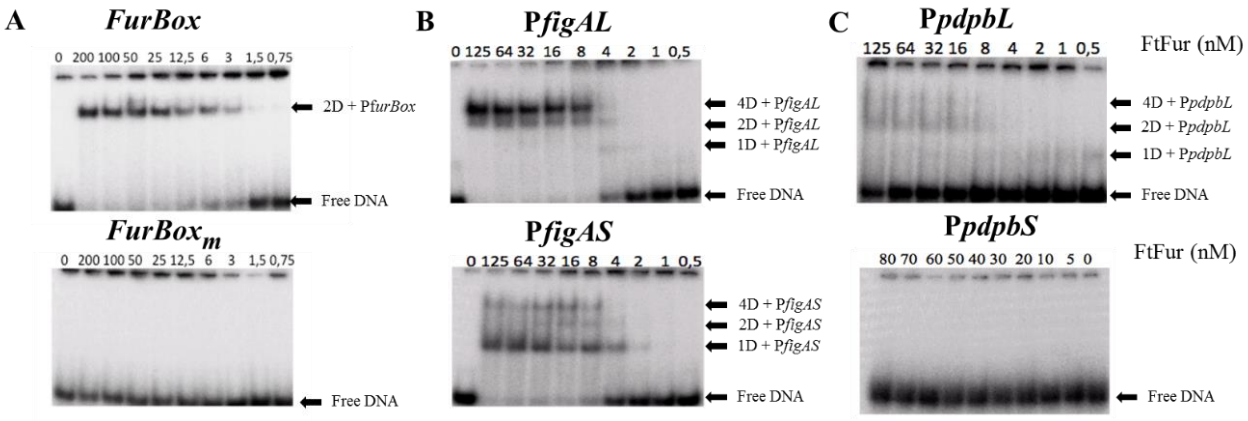


B



C

Supplementary Figure 1. Construction procedure of CHUGA-Ft6Δ*fur* and growth comparison with CHUGA-Ft6wt.
Panel A. Schematic representation of allelic change procedure. **Panel B.** Growth curve comparison of CHUGA-Ft6wt and CHUGA-Ft6Δ*fur* by monitoring the optical density at 600 nm as a function of time. **Panel C.** Colony morphology of CHUGA-Ft6wt and CHUGA-Ft6Δ*fur*. Bacteria grown for 72 h on Chocolate-agar plates (CPV) at 37°C with 5% CO_2 atmosphere.



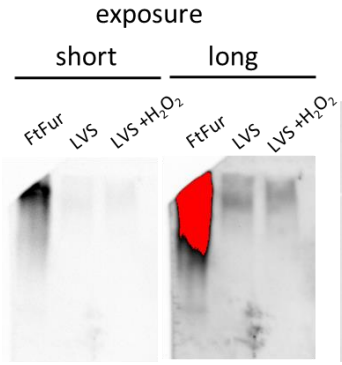
D

Name	Sequence
FurBox	GGGGATAATGATAATCATTATCGGG
FurBoxM	GGGGATACTGATAGTCCTGATCGGG
P*figAL*	GGGCTACATGATAATGATAACGAATATCATTATCGTTTATGGG
P*figAS*	GGGGATAATGATAACGAATATCATTATCGGG
P*pdpbL*	GGGCAAGTAAATGAAGATTGTGAGAATTCTTTTTTGATAAATGATAAAAAAGGG
P*pdpbS*	GGGGATTGTGAGAATTCTTTTTGGG

E

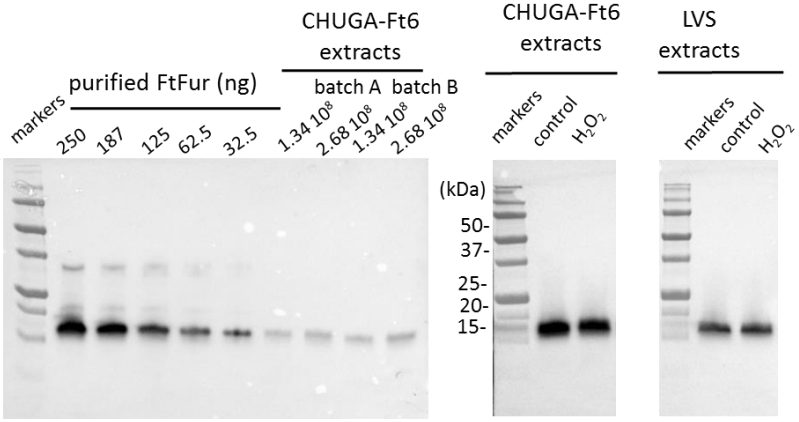
```
PdpbL  CHUGA-Ft6
5′ GGGCAAGTAAATGAAGATTGTGAGAATTCTTTTTTGATAAATGATAAAAAAGGG 3′
         G--AATGA--AT--T-A--                              10/19
            -AT-A-GAT--T-A-AT-                            10/19
               GA--AT--T-A--ATT-T-                        10/19
                  GAT--TGA-AAT--TT-T-                     12/19
                              --T--TGATAA----TA--         9/19
iglC Schu S4
         GGGGTACAAATAATAACTAAAGGG
           G-TA---ATAAT-A-TA--   11/19

figA Schu S4
5′ ATACTTAAATAAACGATAATGATATTCGTTATCATTATCATGTAGGGATT 3′
           ---AA-GATAAT-AT--TC                          12/19
                GATAATGATA-TC-TTATC                     17/19
                     GATA-T--T-ATCATTATC                15/19
                           --T-AT-AT-ATCAT----          10/19
```

F



same migration

G



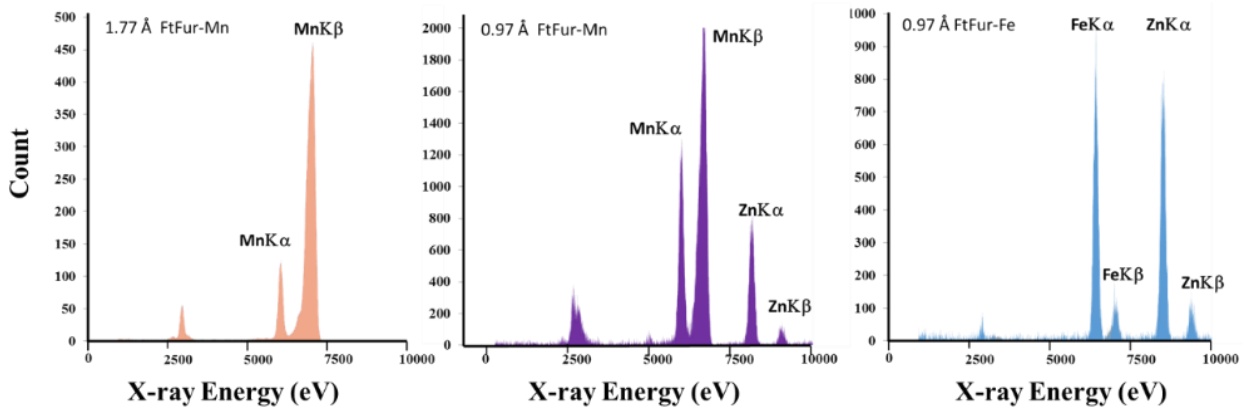
3000 subunits /CHUGA-Ft6 bacteria

	control	1 mM H_2O_2
CHUGA-Ft6	100%	93 % ±4
LVS	100%	105 % ±8

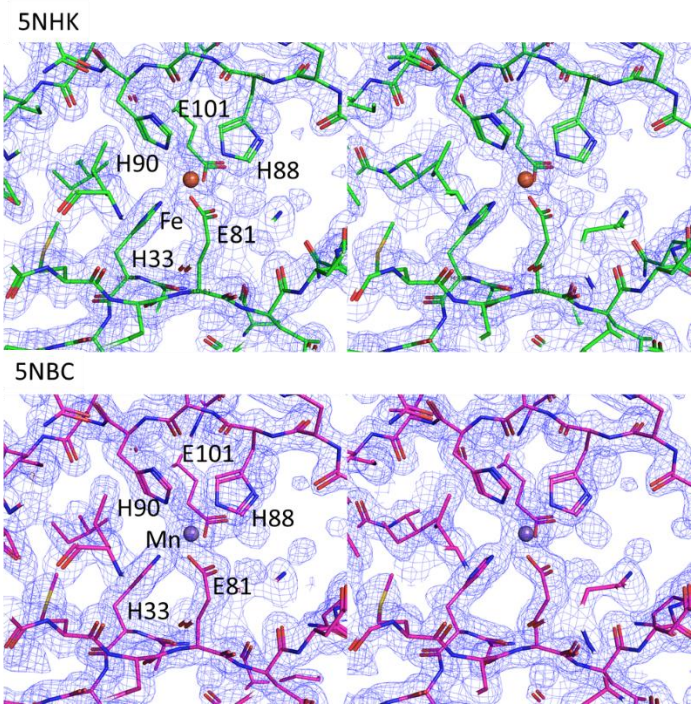
Supplementary Figure 2. DNA binding experiment by EMSA of FurBox containing promoters identified in *Francisella tularensis* genome and Western blot assays.

Panel A. FurBox corresponding to consensus sequence and a mutated FurBox as control ($FurBox_m$) previously described[4].
Panel B. P*fig*AL(long) and P*fig*AS (short) correspond to the genomic DNA sequence of CHUGA-Ft6 *Francisella tularensis*. The long version is already presented in Fig. 1 in the main text, the short version presents a different profile, with a major band corresponding to 1 dimer. DNA sequence analysis indicates the presence of 3 modifications compared to the FurBox, which could be important to the binding of FtFur on DNA.
Panel C. Result obtained with P*pdpb*L and P*pdpb*S are presented and indicate the low affinity of FtFur for this box. DNA sequence analysis indicates the presence of 6 modifications in nucleotide sequence and validates the low affinity obtained with this promoter.
Panel D. Sequence of forward primer used for EMSA experiment. The nucleotides corresponding to the consensus FurBox are underlined.
Panel E. promoters containing FurBox (with the homology with the consensus 19bp) are identified in *Francisella tularensis* genomes.
Panel F. Native Western blot on PAGE of *F. tularensis* live vaccine strain (LVS) extracts. For safety reasons, we are not allowed to manipulate the clinical CHUGA-Ft6 strain extracts outside the P3 laboratory without a previous denaturation step. We thus used LVS which contains an identical FtFur protein. Exponential growth phase bacteria were or not exposed for 4 h to 1 mM H_2O_2 at 37°C before lysis with BugBuster extraction reagent (Sigma). The recombinant purified FtFur (2.5 µg) and LVS extracts (10 µg) were loaded onto a 6% acrylamide gel and separated under non-denaturing conditions (100 V, 2 h). After transfer on nitrocellulose membrane, the samples were probed with anti-Ec Fur antibodies (1:15,000) and Fur was detected with HRP-conjugated goat anti-rabbit IgG antibody (Jackson Immunoresearch 1: 10,000) and visualized by chemiluminescence with the Biorad ECL substrate and using the Biorad chemidoc MP imaging system.
Panel G. Left panel: Quantification of the Fur proteins by the signal analysis obtained starting from various amount of recombinant protein and bacteria and leading to an average estimation of 3,000 Fur subunits per CHUGA-Ft6 bacteria. **Right** panel: Western blot of whole bacterial extracts of bacteria LVS and CHUGA-Ft6 grown in Modified Mueller Hinton medium supplemented with 0.0025 % ferric pyrophosphate and treated or not with H_2O_2 (1 mM for 4 h). Bacterial pellets resuspended in Laemmli buffer were separated by SDS-PAGE (14% acrylamide) and then processed as above.

A

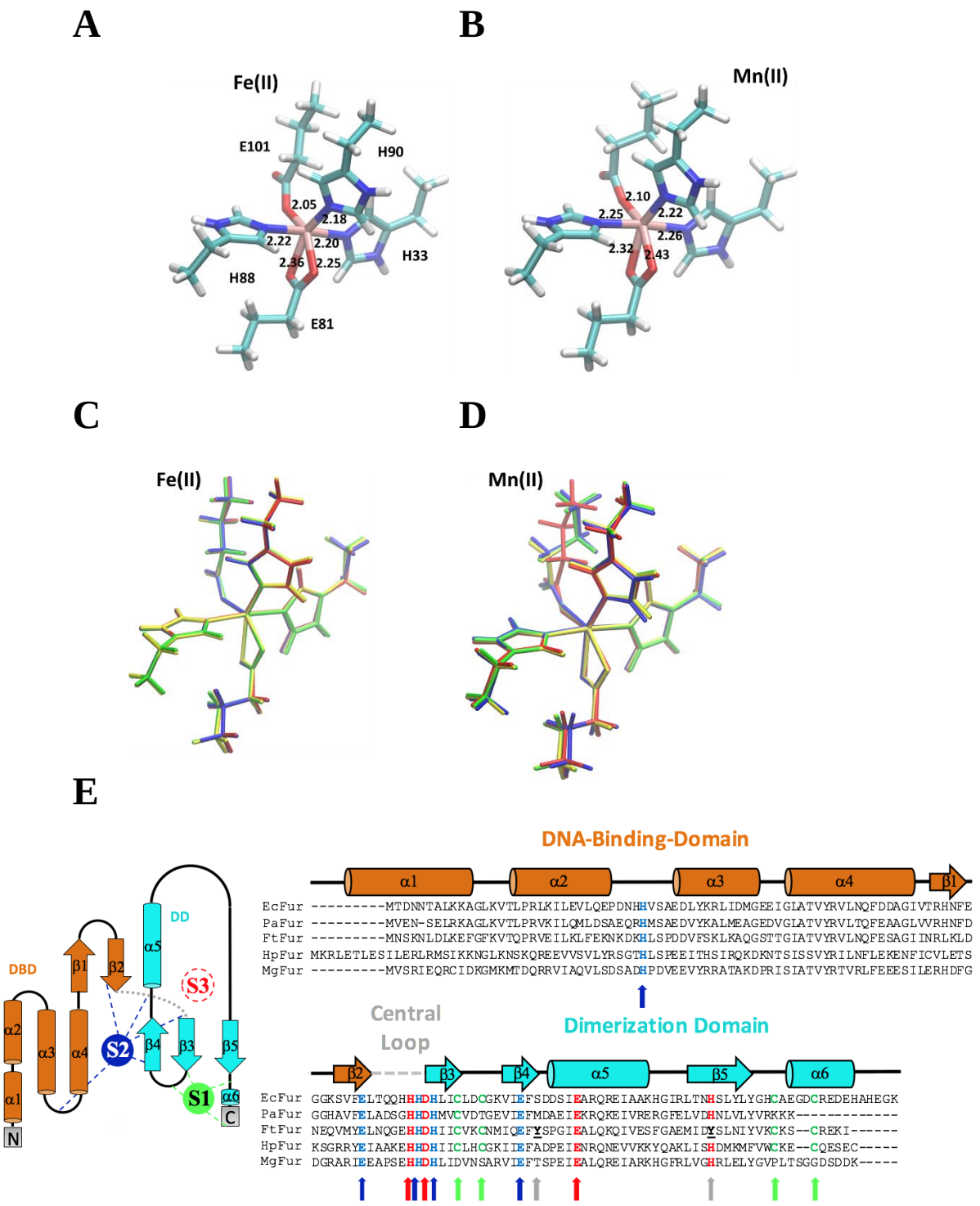


B

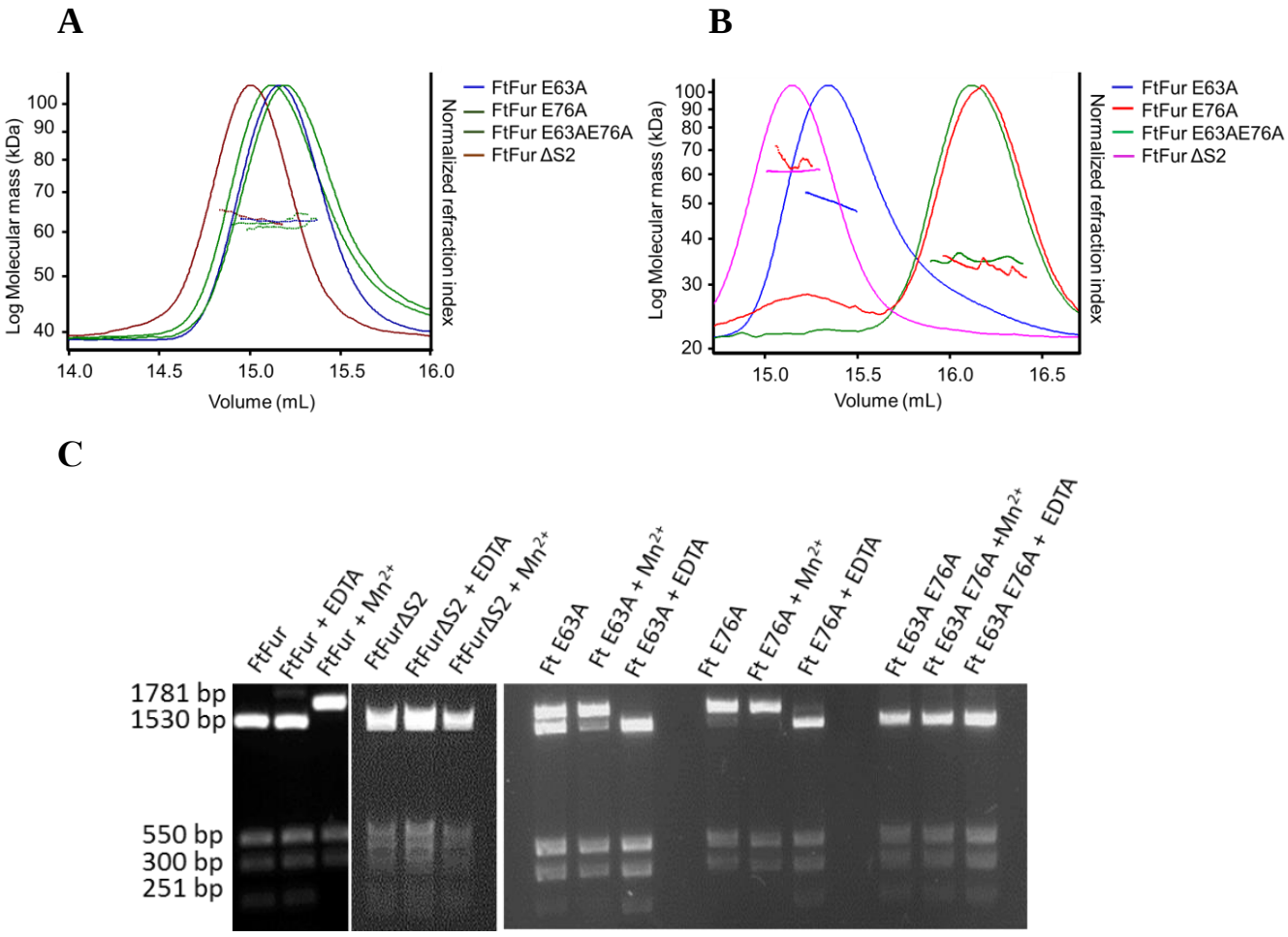


Supplementary Figure 3. Crystallographic data of FtFur structures and fluorescence spectra.

Panel A. Fluorescence spectra recorded at 1.77 Å and 0.97 Å to confirm the presence of Fe, Zn and Mn atom in protein crystals before X-ray diffraction.
Panel B. Stereo views of the Fe and Mn metal binding sites of the 5NHK and 5NBC structures, respectively. Stereo mode prepared with the Will Eye Stereo Program inside Pymol 2.1.0 (Sigma 1.5; Carve 2.0; mesh 0.1; Stereo shift -6; Stereo angle 2.1)



Supplementary Figure 4. Structures of metal binding sites and sequence alignments.
Panels A and B. FtFur Metal binding sites – DFT B3LYP Optimization under High spin for S2 sites filled with Fe(II) and Mn(II) respectively.
Panels C and D. Superimposition of DFT optimized structures of the metal sites in the 4 subunits, for S2 sites filled with Fe(II) and Mn(II) respectively.
Panel E. Sequence alignment of 5 Fur proteins of known structures and location of amino acids implicated in the S1 (Green), S2 (Blue) and S3 sites (red). Y103 and Y125 from FtFur are identified by grey arrows. Ec=*Escherichia coli* (DBD Xray structures only [5]; Pa: *Pseudomonas aeruginosa* [6]; Ft: *Francisella tularensis* (this work); Hp: *Helicobacter pylori* [7] and Mg: *Magnetospirillum gryphiswaldense* [8].

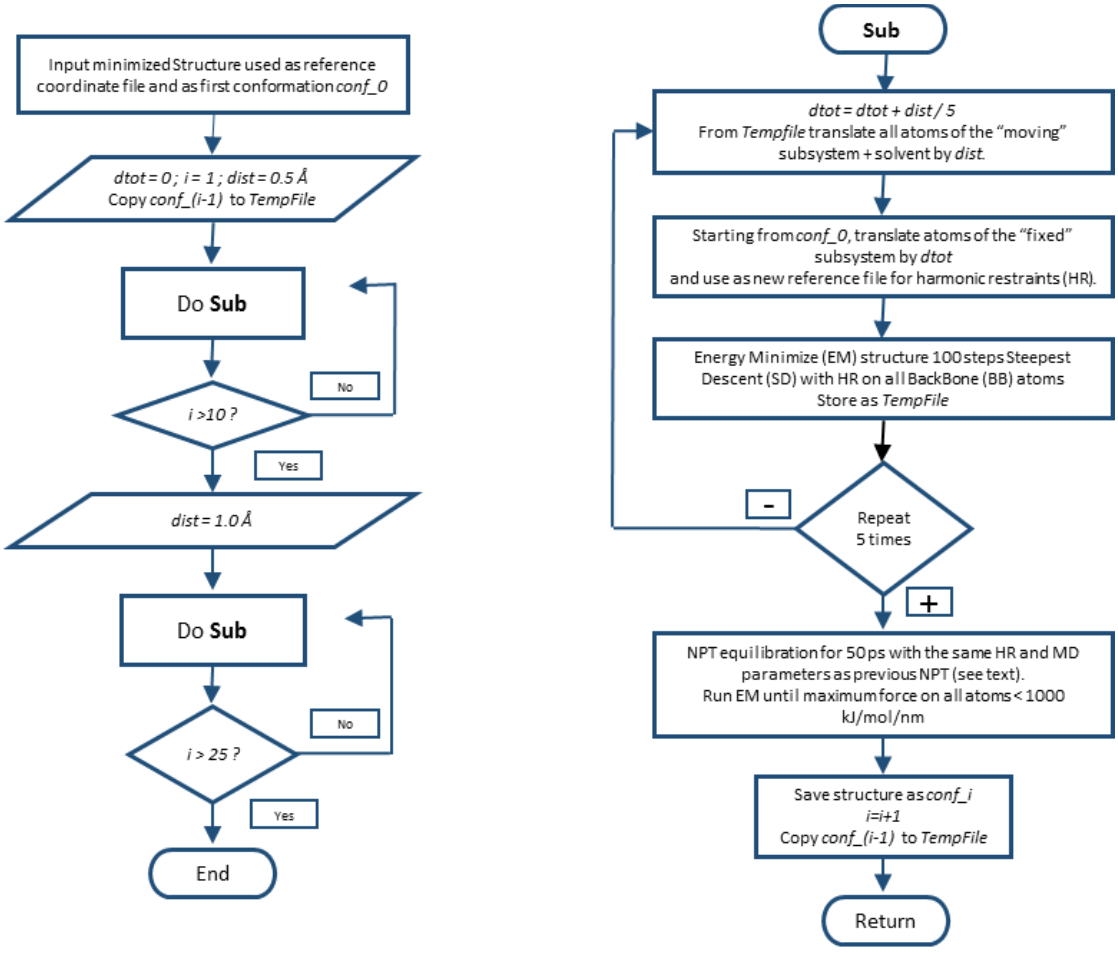


Supplementary Figure 5. Study of tetramer dissociation by SEC-MALLS-RI analysis on FtFur E63A, E76A and E63AE76A mutants and activity assays.

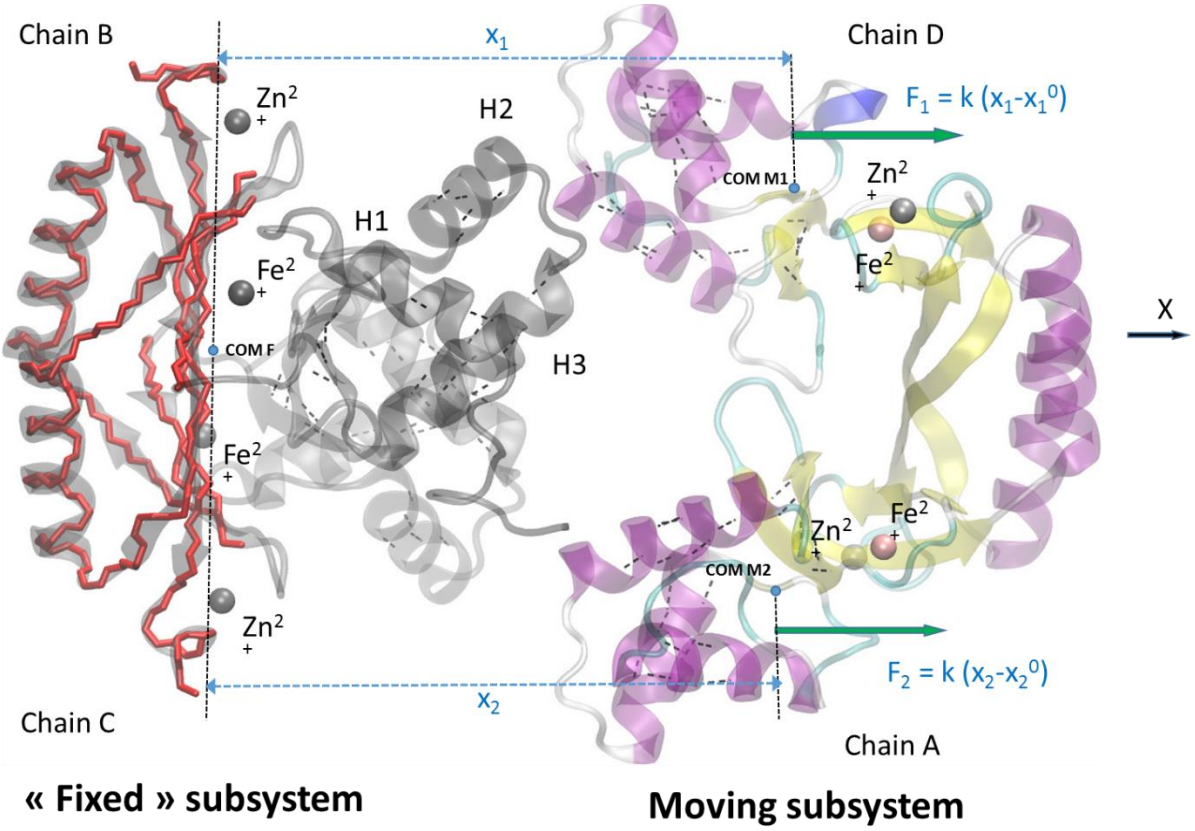
Panel A. SEC-MALLS analysis experiments are done in the presence of 150 mM NaCl, all constructs present the same MW (around 64 kDa) corresponding to a tetramer in solution.

Panel B. SEC-MALLS analysis experiments are done in the presence of 500 mM NaCl. FtFur-WT presents a MW of 64 kDa corresponding to a tetramer in solution. E63A mutant shows a non-symmetric peak at 48 kDa MW corresponding to a mixture of tetramer and dimer. E76A present two peaks. The major peak corresponds to a MW of 32 kDa for a dimer in solution, the minor peak presents a MW of 64kDa corresponding to a tetramer. Finally, the double mutant E63AE76A presents one single symmetric peak at 32 kDa corresponding to a stable dimer in solution. This experiment has been repeated three times and validates the major role of these two amino acids in the stabilization of the tetramer.

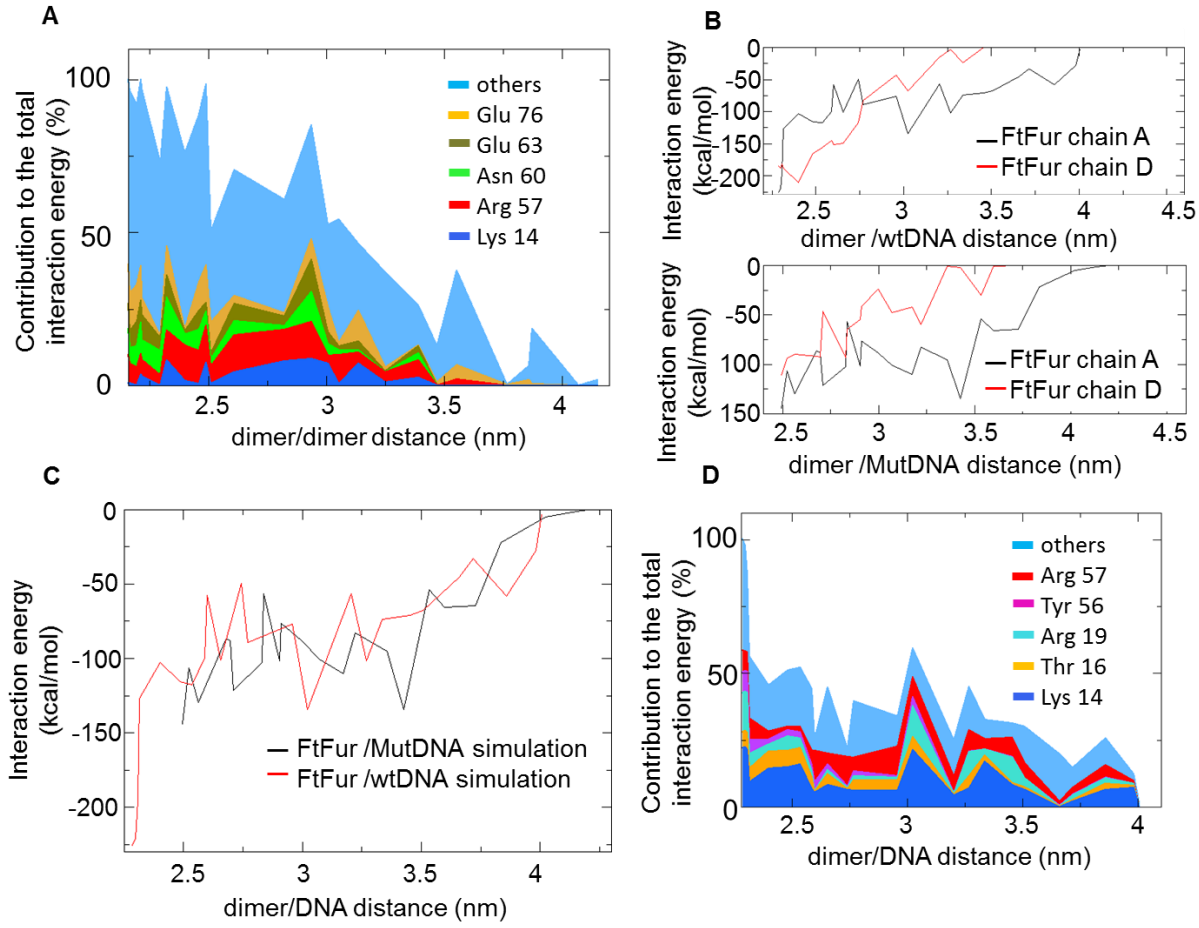
Panel C. Nuclease assay on FtFur, FtFurΔS2, FtFurE63A, FtFurE76A, FtFurE63AE76A performed as previously described[4]. The FtFurΔS2 and the FtFurE63AE76A mutants were not able to bind DNA contrarily to the WT, FtFurE63A and FtFurE76A mutants, in the conditions of the assay.



Supplementary Figure 6. Algorithm for the preparation of initial coordinates for the computation of free energy profiles.
The algorithm consists of a main subprogram doing the translations and pre-equilibrations, which is called twice: 1) 10 first windows corresponding to 0.5 Å translations followed by 2) 15 windows of 1 Å translations. Three variables are defined "dtot", the total translation distance for the "moving" subsystem, "i", the conformation counter and "dist", the current translation distance in Å.



Supplementary Figure 7. Structure of the FtFur tetramer corresponding to the last conformation in the pulling process.
The moving subsystem (Chains A and D) has been translated (10*0.5+15*1.0) =20 Å from the initial tetramer along the X direction. Centers of mass of the "fixed" subsystem and the two subunits of the moving Ft dimer are indicated. The moving subsystem is constituted of two Ft subunits: 1 and 2. Forces are applied on their centers of mass to pull them away or maintain them at a fixed distance of the "fixed" subsystem. Part of the backbone of the "fixed" system highlighted in red is subject to position restraints during the pulling process. Black dashed lines show hydrogen bonds in the secondary structures which are maintained using NOE type distance restraints. 8 metal cations are shown as spheres. Water and counter ions surrounding the system (in a 105.3*80.3*65.4 $Å^3$ periodic box) are not shown for clarity.



Supplementary Figure 8. Important residues in the interactions and their comparative energy contribution.

Panel A. The five major contributors to the interaction energy between chain A and the "fixed" dimer in the FtFur tetramer simulation. Same legend as in Fig. 2. On average, 5 residues contribute to around 30% of the total interaction energy (with a maximum of 56% at 2.93 nm). The maximum total interaction energy for all residues over all 26 windows is calculated first, then the contribution of each residue is normalized with respect to this maximum. The plot shows the corresponding percentage of the maximum interaction energy for selected residues. The blue surface corresponds to the sum of the contributions of all residues others than those selected.

Panel B. Up: Comparison of average interaction energies between Fur chains A and D and wtDNA during the translation of Fur away from DNA. Chain D loses its interaction with DNA faster than chain A. Down: Comparison of average interaction energies between Fur chains A and D and mutDNA during the translation of Fur away from DNA. Chain D loses its interaction with DNA very fast. (Averages were calculated over 15000 points).

Panel C. Comparison of average interaction energies between Fur chain A and DNA in the Fur/wtDNA and Fur/mutDNA simulations. Interaction energies between FUR chain A and wt or mut DNA are similar.

Panel D. The five major contributors to the average interaction energy between Fur chain A and DNA in the FtFur/wtDNA simulation. Same legend as in Fig. 2. On average, 5 residues contribute to 59% of the total interaction energy (with a maximum of 86% at 2.74 nm).

Supplementary Tables

Supplementary Table 1. Data collection, phasing and refinement statistics for the structure of Mn-FtFur.

		Mn-FtFur (5NBC)
Data collection		
Space group	P21	P21
Cell dimensions		
a, *b*, *c* (Å)	53.15, 89.93, 63.84	53.15, 90.0, 63.95
α, β, γ (°)	90.00, 93.59, 90.00	90.00, 93.48, 90.00
	Peak	*Remote*
Wavelength (Å)	1.7761	0.9797
Resolution (Å) [a]	50.0 - 2.1 (2.14 - 2.1)	50.0 - 1.7 (1.74 - 1.7)
R_{meas} [a, b]	17.6 (135.0)	8.9 (99.2)
$I/\sigma(I)$ [a]	10.2 (1.60)	11.6 (1.31)
$CC_{1/2}$ [a]	99.5 (57.8)	99.6 (64.4)
Completeness (%) [a]	99.3 (92.3)	99.7 (48.8)
Redundancy [a]	7.2 (6.4)	3.5 (2.4)
Refinement		
Resolution (Å)	*N.A.*[c]	42.1 - 1.7
No. reflections		64585
R_{work} / R_{free}		21.7/24.4
No. atoms		
Protein		4138
Ligand/ion		8 (4 Zn atoms and 4 Mn atoms)
Water		402
B factors		
Protein (Å^2)		24.1
Ligand/ion (Å^2)		20.9
Water (Å^2)		31.3
R.m.s deviations (Å^2)		
Bond lengths (Å)		0.007
Bond angles (°)		0.88

[a]Values in parentheses are for highest-resolution shell.

[b]R_{meas}, redundancy-independent merging R-factor

[c] Structure of Mn-FtFur has been solved by the MAD method using 1 crystal, and refined with the remote diffraction data only

Supplementary Table 2. Data collection and refinement statistics for the structure of Fe-FtFur.

	Fe-FtFur (PDB code 5NHK)
Data collection	
Space group	P21
Cell dimensions	
a, *b*, *c* (Å)	53.35, 90.04, 64.05
α, β, γ (°)	90.00, 93.53, 90.00
Resolution (Å) [a]	50.0 - 1.8 (1.9 - 1.8)
R_{meas} [a, b]	9.6 (75.2)
$I/\sigma(I)$ [a]	10.8 (2.1)
$CC_{1/2}$ [a]	99.6 (64.4)
Completeness (%) [a]	99.2 (99.6)
Redundancy [a]	3.0 (3.0)
Refinement	
Resolution (Å)	42.2 - 1.8
No. reflections	55477
R_{work} / R_{free}	20.4/24.3
No. atoms	
Protein	4237
Ligand/ion	8 (4 Zn atoms and 4 Fe atoms)
Water	507
B factors	
Protein (Å^2)	22.2
Ligand/ion (Å^2)	20.0
Water (Å^2)	29.4
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.89

[a]Values in parentheses are for highest-resolution shell.
[b]R_{meas}, redundancy-independent merging R-factor

Supplementary Table 3. Comparison of metal-ligand distances between FtFur-Mn and FtFur-Fe. Distances are calculated from the X-ray structures with Coot 0.8.2

Mn-FtFur / Fe- FtFur

	Site 1			
	Zn in dimer 1		Zn in dimer 2	
	Chain A	Chain D	Chain B	Chain C
C93	2.30/2.31	2.36/2.37	2.33/2.34	2.35/2.35
C96	2.31/2.32	2.30/2.32	2.31/2.26	2.32/2.31
C133	2.33/2.30	2.32/2.30	2.36/2.31	2.28/2.29
C136	2.32/2.30	2.37/2.30	2.31/2.31	2.30/2.30
	Site 2			
	Mn/Fe in dimer 1		Mn/Fe in dimer 2	
	Chain A	Chain D	Chain B	Chain C
H33 ($N_{\varepsilon}2$)	2.31/2.26	2.26/2.26	2.29/2.27	2.28/2.26
E81($O_{\varepsilon}1$)	2.32/2.19	2.30/2.20	2.27/2.10	2.35/2.21
($O_{\varepsilon}2$)	2.32/2.31	2.34/2.34	2.30/2.25	2.29/2.31
H88 ($N_{\varepsilon}2$)	2.31/2.29	2.36/2.26	2.31/2.30	2.32/2.27
H90 ($N_{\varepsilon}2$)	2.23/2.22	2.26/2.21	2.30/2.26	2.31/2.24
E101 ($O_{\varepsilon}2$)	2.35/2.03	2.20/2.00	2.32/2.11	2.16/1.97

Supplementary Table 4. Strains, plasmids, and primers used in this study.

Strain, plasmid, or primer	Genotype, description, or sequence	Source or reference
Strains		
CHUGA-Ft6	*F. tularensis subsp. holarctica*, human strain from blood sample (Verdun hospital)	This study
CHUGA-Ft6Δfur	CHUGAFt66 with in-frame deletion of FTA_1939 (fur)	This study
CHUGA-Ft6Δfur/pfur	CHUGAFt66Δfur trans complemented with fur gene and promotor region on plasmid pMP 828	This study
Plasmids		
Pmp812furpart	suicidal plasmid pMP 812 with 1000 pb regions upstream and downstream *fur g*ene	Lovullo *
Pmp 828fur+	shuttle plasmid containing *fur* gene of *F. tularensis subsp. holarctica* and its promotor region	Lovullo *
Primers		
LeftfurF	**GATCGGATCCTCTGCAGTTGTATAATGCCG**	This study
LeftfurR	**CTATTAGATAGTTTTATAAGTTC**	This study
RightfurL	**TATGCTTTTCGAATAATCTTTATATCAGAACTTATAAAACTATCTAATAGAC GCAATAATCACTATCCAG**	This study
RightfurR	**TAGCGATATCGTGATACCAAATTTTATATTGCTC**	This study
CompfurL	**GCTCATAAACACTTGGATCCTAATCAAGAAGA**	This study
CompfurR	**AAGCTTAAATTTAGAGATATCAGAAAAGCTGTT**	This study

CHUGA-Ft6 was isolated at Verdun Hospital (France) from a blood sample of a patient with typhoidal tularemia [1]. * Lovullo et al [2,3].

Supplementary Table 5. Oligonucleotide sequences used in Gene expression by qRT-PCR experiment.

Name	Sequence
fur: forward	GGAATCGAAGCTCTGCAAAA
fur: reverse	TTCACGACAGGATTTGCATT
figA: forward	AACTGCTCCCCATTGCTCTA
figA: reverse	TTGGCAATGGTTAACTGCAA
16S-RNA: forward	TTTCACCTTTGAGCTGTTGC
16S-RNA: reverse	CCTTTGGCAAATTCAATAGAAAC

Supplementary Methods

Molecular Modelling of the FtFur tetramer

The pdb file corresponding to the Xray structure of the FtFur tetramer was prepared for gromacs (GROningen MAchine for Chemical Simulations) version 5.1.2 [9] with pdb2gmx using gromos54a7 united atom force field [10]. All histidine residues were given type HISA with protonated Nδ1 atom. In all 4 protein subunits cysteines 93, 96, 133 and 136 were deprotonated and given a total charge of -0.75 (-0.05 for Cα, -0.15 for Cβ, -0.55 for S). This charge can be compared to that of the CYS residue in gromos54a7 force field, with total charge of -0.5. A value of -0.75 was adopted instead in this work, high enough to stabilize a tetracoordinated Zinc ion, providing correct orientation of the cysteines and proper metal environment during further simulations. (Use of a -0.5 charge led to the escape of the metal from its binding site).

Construction of the MgFur dimer DNA complex

The structure of *Magnetospirillum gryphiswaldense* (4RB1) was used as initial model of a Fur per DNA complex. The Fur dimer includes residues 20 to 133 and the DNA double strand is composed of CGGATAATGATAATCATTATCGC. First, an initial model (Mod1) of the MgFur dimer per DNA complex with 4 iron ions and 2 bound water molecules (included in structure 4RB1) was built and slightly energy minimized with CHARMM [11] using the all-atom "par_all27_prot_na" parameter file for proteins and nucleic acids. Second, another initial model (Mod2) of the protein dimer alone was built starting from the same PDB but with the extended atom "param19" force field for proteins. Then, an initial structure of the protein DNA complex compatible with the Gromacs force field was created from protein dimer from Mod2 and DNA from Mod1. The structure was oriented so that DNA was centered at the origin with its principal axis along Y. We noticed that the MgFur dimer in this structure sits asymmetrically on the DNA fragment with K15 in chain A interacting with CYT21 and K15 in chain B extending over CYT1. It was consequently decided to move the last two GC bases upstream leading to a GCCGGATAATGATAATCATTATC fragment. The new position of the 2 moved nucleotides was controlled with VMD and regularized through 1000 steps of energy minimization with CHARMM. This new DNA fragment offers the advantage of providing good interactions for all residues in the Fur dimer while still including the full length consensus Fur box GATAATGATAATCATTATC.

Construction of the FtFur dimer per wild type DNA complex

Structure of FtFur in the presence of iron (this work) from ASP7 to ARG137 was used for the Fur/wtDNA simulation. In the FtFur+DNA structure, the model built from MgFur was used as template to align FtFur. FtFur was superimposed on MgFur already sitting on the Fur box using least square fit of matching atom positions. Backbone atoms from corresponding structured parts (helices and sheets) of the two protein were selected for this structure superposition for a final rms difference of 4 Å.

Construction of the FtFur dimer per mutated DNA complex

The mutated DNA (mutDNA) sequence GCCGGATA**C**TGATA**G**TC**C**T**G**ATC contains four mutations with respect to the Fur-box (A9 to C, A15 to G, A18 to C and T20 to G). Initial coordinates for mutDNA were taken from the previously obtained structure of wild type DNA by simple matching of corresponding heavy atoms and building of missing hydrogens with CHARMM, followed by 1000 steps energy minimization. Initial direct superimposition of the Fur dimer onto this mutated DNA showed that relatively large structure modifications could occur in the region of the mutations. Consequently, a 100 ps NPT simulation of solvated

mutDNA was run to release the strains in the structure. This last structure was superimposed on the initial model built in the presence of FtFur to ensure optimal orientation. Finally, initial coordinates for the FtFur dimer per mutated DNA complex (FtFur/mutDNA) were obtained by concatenating coordinates of the protein in the FtFur/wtDNA complex and those of the reoriented mutDNA.

Computational details of the solvation and equilibration

For the tetramer, the maximum dimensions of the protein were calculated, 28 Å were added in the +X direction to allow for its further translation. Then, a simulation box 5 Å larger than these dimensions in all directions was created and the tetramer centered in the simulation box with "editconf". Initial dimensions of the box were 10.53*8.03*6.54 nm. "gmx solvate" was used to add solvent (SPC water) to the protein from a preequilibrated water box (spc216.gro) for a total of 15586 water molecules. 37 sodium and 33 chloride ions were added (replacing 70 water molecules) for a total ionic force of 0.1 $mol.L^{-1}$ and a zero total charge of the 52114 atom system.

For the FtFur/wtDNA complex, initial dimensions of the box were 9.28*8.91*6.35 nm. 15505 water molecules were added of which 77 were replaced by sodium and 32 by chloride ions to ensure neutrality and a total ionic force of 0.1 $mol.L^{-1}$ for a total of 50117 atoms in the system.

For the FtFur/mutDNA complex, initial dimensions of the box were 9.64*9.22*6.39 nm. 16680 water molecules were added of which 79 were replaced by sodium and 34 by chloride ions to ensure neutrality and a total ionic force of 0.1 $mol.L^{-1}$ for a total of 53642 atoms in the system.

All three systems (FtFurTetra, FtFur/wtDNA, FtFur/mutDNA) were energy minimized until the force on all atoms was less than 1000 $kJ.mol^{-1}.Å^{-1}$. Heavy atoms of the protein were restrained to their initial position to prevent big structural changes before the production run with a force constant of 10 $kJ.mol^{-1}.Å^{-2}$. Then we ran the system under NPT conditions at 310 K and 1 atm for 100 ps, with Berendsen temperature and isotropic pressure coupling [12]with τ_T = 0.1 ps, τ_P = 2 ps and compressibility = 4.5×10^{-5} bar^{-1} (same conditions in the NPT simulation used to equilibrate mutDNA).

Computation of Free energy profiles

Free energy profiles for the extraction (by translation along a fixed direction: Ox) of one FtFur dimer from the tetramer (dimer of dimers) and of FtFur from DNA were computed.

The FtFur tetramer is organized around 4 subunits with chain names B and C for the fixed dimer and A and D for the translated dimer. The simulations will thus include a "moving" subsystem (FtFur dimer, chains A and D) and a "fixed" subsystem (FtFur dimer, chains B and C, wtDNA or mutDNA).

The profiles were built using the "umbrella sampling" technique and result from the overlapping of 26 computation windows, one for each translation distance.

The simulation protocol for the equilibration of umbrella sampling windows is best described in supplementary Figure 6.

As a result of this initialization protocol, 26 structures "conf_i" were generated yielding initial positions for the translated dimer prior to the potential of mean force calculation.

Then the umbrella sampling calculation, itself, consisted of 26 repeats of:

- Reading of initial structure "conf_i" and reference structure "conf_0" for harmonic restraints;

- Running 100 ps NPT equilibration with position restraints on the "fixed" subsystem and distance restraints on the protein. The "moving" subsystem was subject to two harmonic biasing forces (umbrella potential) applied between the centers of mass of the 2 Fur

dimer subunits and the center of mass of the "fixed" subsystem with force constants of 500 kJ mol^{-1} nm^{-2}.

- Running 10 to 15 ns NPT production simulations with same restraints and biasing potential. These simulations were concatenated from series of 1 to 2 ns runs allowing us to check the convergence of the PMFs. Harmonic biasing forces were applied along the X direction only (direction of the translation).

Position restraints were applied on backbone atoms of the "fixed" subsystem for residues K28 to S35 (Loop between H1 and H2) and for all residues N83 to E138 (C-Terminal) except V94 to M98 possibly interacting with the "moving" subsystem during the translation. Force constants of 10 kJ $mol^{-1}Å^2$ in all directions were used.
NOE type distance restraints were added to maintain the secondary structure of the protein (both "fixed" and "moving" parts in the case of the tetramer). All backbone hydrogen bonds between N-H and O=C for residues T16 to K24 (H1), S35 to K44 (H2) and G51 to E63 (H3) plus the short beta-sheet (I67—N83, N69—E81, L71—M79) were restrained between 1.8 and 2.0 Å with force constants of 20 kJ $mol^{-1}Å^2$.

In both equilibration and production umbrella sampling MD simulations, the system was simulated under NPT conditions. Temperature was fixed at 310 K with Temperature coupling using a Nose-Hoover extended ensemble with τ_T = 0.5 ps, Pressure was controlled at 1 atm with extended-ensemble Parrinello-Rahman isotropic pressure coupling with τ_P =1 ps and compressibility = 4.5×10^{-5} bar^{-1} [13]. Again, a time step of 2 fs was used.

Fitting of the free energy profiles
Data points corresponding to the outputs of wham were fitted with a sum of 1, 2 or 3 sigmoid functions with R [14]:

$$S(x) = a + \frac{b}{1 + e^{(-c*(x-d))}}$$

After a first fit with the raw data, the energy offset (a in equation) was subtracted from the final energy profiles.

Supplementary References

1. Maurin, M., Pelloux, I., Brion, J.P., Del Bano, J.N. & Picard, A. Human tularemia in France, 2006-2010. *Clin Infect Dis* **53**, e133-41 (2011).
2. LoVullo, E.D., Molins-Schneekloth, C.R., Schweizer, H.P. & Pavelka, M.S., Jr. Single-copy chromosomal integration systems for Francisella tularensis. *Microbiology* **155**, 1152-63 (2009).
3. LoVullo, E.D., Sherrill, L.A. & Pavelka, M.S., Jr. Improved shuttle vectors for Francisella tularensis genetics. *FEMS Microbiol Lett* **291**, 95-102 (2009).
4. Perard, J. et al. Quaternary Structure of Fur Proteins, a New Subfamily of Tetrameric Proteins. *Biochemistry* **55**, 1503-15 (2016).
5. Pecqueur, L. et al. Structural changes of *Escherichia coli* ferric uptake regulator during metal-dependent dimerization and activation explored by NMR and X-ray crystallography. *J Biol Chem* **281**, 21286-95 (2006).
6. Pohl, E. et al. Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol Microbiol* **47**, 903-15 (2003).
7. Dian, C. et al. The structure of the *Helicobacter pylori* ferric uptake regulator Fur reveals three functional metal binding sites. *Mol Microbiol* **79**, 1260-75 (2011).
8. Deng, Z. et al. Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. *Nat Commun* **6**, 7642 (2015).
9. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1* **11**, 19-25 (2015).
10. Schmid, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**, 843-56 (2011).
11. Brooks, B.R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187-217 (1983).
12. Berendsen, H.J.C., Postma, J.P.M., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).
13. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys* **52**, 7182-7190 (1981).
14. R-Development-Core-Team. R: A language and environment for statistical computing. (ed. Computing, R.F.f.S.) (Vienna, Austria, 2008).

Fur and Fur like structure comparaison

This annex analyses Fur and Fur-like structures in a preliminary and complementary study to what was presented in this chapter.

28 structures of Fur or Fur like structures have been retrieved from the PDB and analyzed (the structure 2FU4 of EcFur has been omitted). The following table summarizes these structures with the type of metals involved, resolution, PDB-ID and associated publication.

Name	Structure	Resolution (Å)	PDB ID	Publication
Zn-PaFur	S2 Zn S3 Zn not dimer but tetramer	1.8	1MZB	Pohl 2003 MolMic
Zn-VcFur	S2 Zn S dimer	2.6	2W57	Sheikh 2009 MolMic
Zn-HpFur	S1 Zn S2 Zn S3 Zn dimer	1.85	2XIG	Dian 2011 MolMic
Zn CjFur	S1 Zn S3 Zn 2DBD in 180° rotation	2.1	4ETS	Butcher 2012 PNAS
Zn-CjFur	S1 only 2DBD in 180° rotation	1.8	6D57	Sarvan 2018 SciRep
Mn-MgFur Apo MgFur Mn-MgFur-DNA	feoAB1 and Fur box complexes	1.9 1.55 2.6 2.75	4RAZ 4RB0 4RAY 4RB1, 2, 3	Deng 2015 Nature
FeZn-FtFur MnZn-FtFur	S1 Zn S2 Fe tetramer S1 Zn S2 Mn tetramer	1.8 1.7	5NHK 5NBC	Perard 2018 NCB
Mn	Apo and Zn	2.48	5FD5, 5FD6	Bellini Never published ??
Zn-MtZur	FurB=Zur S1S2S3 open conf	2.7	2O03	Lucarelli 2007 JBC
ZnNi-ScNur	S2 ZnH4 NiH3 malonate Ethane diol	2.4	3EYY	An 2009 NAR
Zn-ScZur	S1 Zn S2 Zn S3 Zn	2.4	3MWM	Shin 2011 PNAS
Zn-EcZur-DNA	S1 Zn S2 Zn Znu ABC	2.5	4MTD & 4MTE	Gilston 2014 PlosOne
EcFur				This work
BsPerR 2OxoHis-Zn BsPerR	S1 Zn Cys4 open 2xoHis open conf dimer	1.75 2.0	2FE3 2RGV	Traore 2006 MolMic Traore 2009NatChemBiol

MnZn-BsPerR	closed dimer	3.15	3F8N	Jacquamet 2009 MolMic
ZnNi -SpPerR	open dimer	2.0	4I7H	Makthal 2013 JBC
ZnZn StPerR	streptococcus dimer	1.6	4LMY	Lin 2014 PlosOne
Zn LiPerR	S2 Zn no S1 asymetric DBD	1.9	5NL9	Kebouchi 2018 JBC
Mn Zn-CjPerR	S1 Zn S2 Mn asym DBD long helix	2.7	6DK4	Sarvan 2018 FEBS Letters

Table 1: Structures of full size Fur and Fur like proteins.

All the corresponding dimer structures have been built and energy minimized with CHARMM (Brooks et al., 1983) by doing the following steps:

- Building symmetries when indicated in the PDB files (using an in-house C program)
- Assigning chain names A and B to the protein subunits
- Changing a few atom names (ILE CD1 into CD) for compatibility with the force field
- Deleting N-ter or C-ter portions of the subunits when zero atom from the corresponding residues were resolved.
- Building other (than in N-ter and C-Ter) missing atoms and residues using CHARMM internal coordinates
- Optimizing the geometry of the dimers with ABNR minimizations down to a gradient of 0.1 kcal.mol^{-1}.Å^{-1} and subject to harmonic restraints on backbone atoms (N Cα, C) with a force constant of 5 kcal.mol^{-1}.Å^{-2}

The dimers were simulated in implicit water using the implicit solvent method EEF1 and the adapted extended atom CHARMM19_EEF1 force field (Lazaridis et al., 1999).

For each dimer, the N-Ter DNA binding domains were defined as extending from the first residue in the PDB to the last residue implied in the second β-strand (around residue 82). Then the distance between the centers of mass of the 2 N-Ter was calculated together with the rotation angle necessary to superimpose one on the other and the corresponding rmsd. Results are given in Table 2.

Distances between N-Ter vary from 27.2 (closed PerR) to 57.9 Å (wide open BsPerR) and a consensus distance between N-Ter for Fur dimers able to bind DNA seems to be 38 to 39 Å.

After superimposition, the rmsd between backbone atoms of these N-Ter portions varies from 0.0 Å (when built by crystallographic symmetry) to 5.09 Å for Zn-CjFur. The corresponding rotation angle is close to 180 degrees except for apo Mg-Fur with a strange 80 degree angle.

Name	NTer definition	Distance between NTer	Rotation angle (degree)	rmsd
Zn-PaFur	1-81	36.7	180.0	0.00
Zn-VcFur	3-82	32.7	164.7	0.46
Zn-HpFur	1-82	39.9	176.6	1.07
Zn CjFur	4-82	37.0	172.4	5.09
Zn-CjFur	8-99	40.7	175.8	2.17
Mn-MgFur	1-82	41.7	176.6	0.34
Apo MgFur	2-82 2-82	37.2 37.2	80.0 78.6	0.85 0.88
Mn-MgFur-DNA	2-82 2-82 2-82	38.6 38.9 39.1	171.0 175.7 173.1	1.28 0.76 0.45
FeZn-FtFur (tetra)	8-82	35.1	176.4	2.80
MnZn-FtFur (tetra)	15-82	34.2	179.6	1.60
Mn	5-85 6-85	37.4 40.9	179.2 172.0	0.60 0.74
Zn-MtZur	2-75	57.8	180.0	0.00
ZnNi-ScNur	3-82	37.4	180.0	0.00
Zn-ScZur	10-77	32.1	180.0	0.00
Zn-EcZur-DNA	4-87 4-87	38.0 38.0	175.2 175.3	1.35 1.36
EcFur	2-82	40.4	179.1	0.75
BsPerR	4-86	57.2	179.1	0.44
2OxoHis-Zn BsPerR	4-86	57.9	179.8	0.48
MnZn-BsPerR	4-86	43.5	179.0	0.33
ZnNi -SpPerR	2-94	42.2	179.0	0.52
ZnZn StPerR	1-94	42.2	179.1	1.45
Zn Li PerR	1-84	45.2	175.8	0.61
Mn Zn-CjPerR	19-79	27.2	178.8	1.08

Table 2: Information about the N-Ter domains of the dimers. The definition first residue – last residue of the N-ter is given in column 2. Then the distance between the centers of mass of the 2 N-Ter, the rotation angle necessary to superimpose one on the other and the corresponding rmsd are given in the subsequent columns.

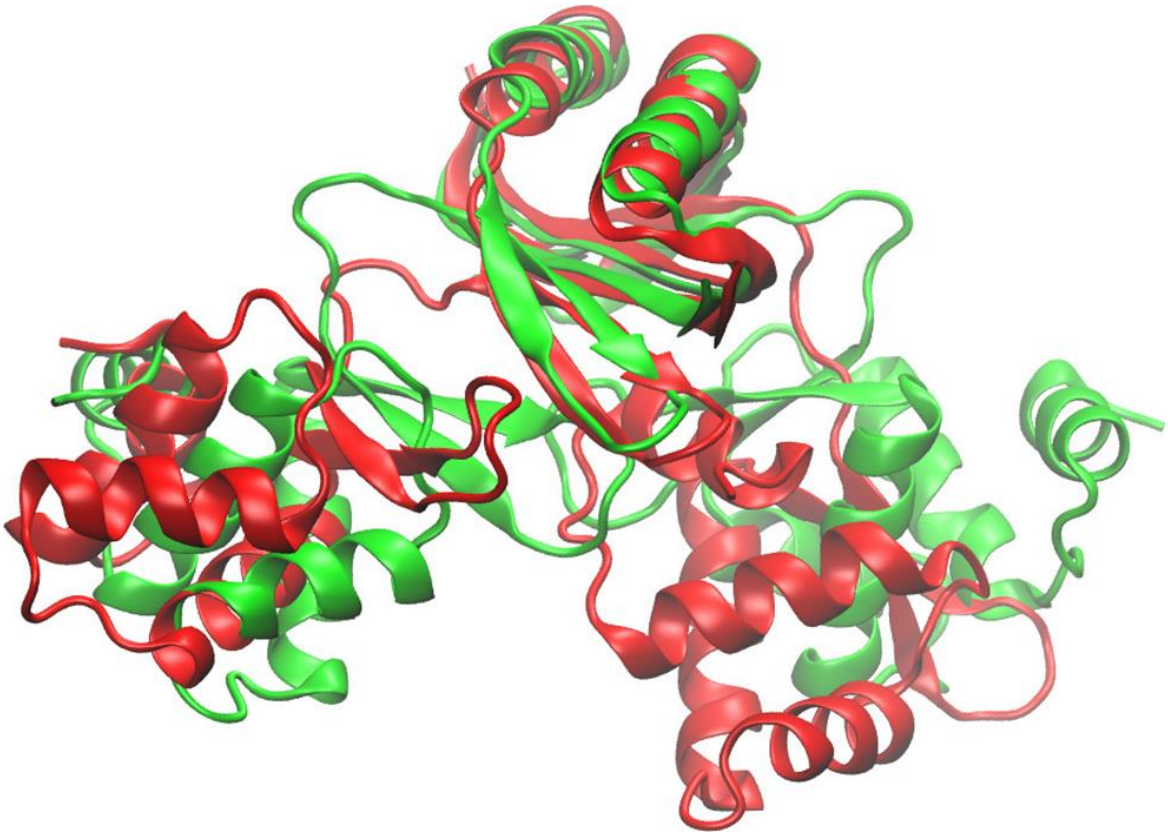


Fig.1: Superimposition of Apo-MgFur (4RB0) and Mn-MgFur (4RB1). Mn-MgFur (green) is a prototypical Fur dimer able to bind DNA. With a close to 180° degree angle between its 2 N-Ter domains. 4RB0 (red), on the contrary, presents a strange angle of 80° between its 2 N-Ter.

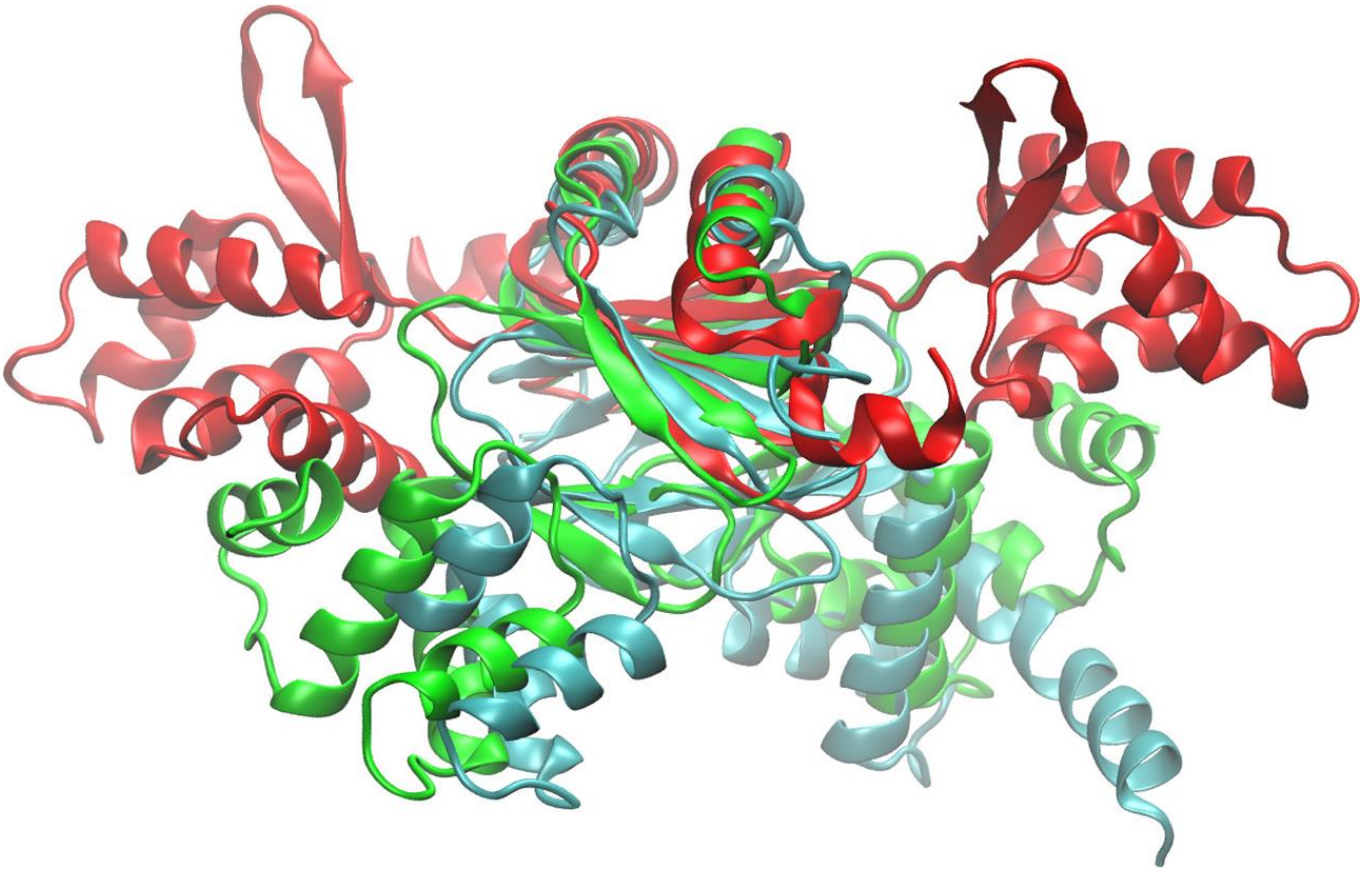


Fig.2: Superimposition of MnZn-CjPerR (6DK4) Mn-MgFur (4RB1) and 2OxoHis-Zn BsPerR (2RGV). Again, Mn-MgFur (green) has a typical distance between its N-Ter to allow binding to DNA whereas MnZn-CjPerR (cyan) seems in a too much close cornformation and BsPerR (red) is wide open.

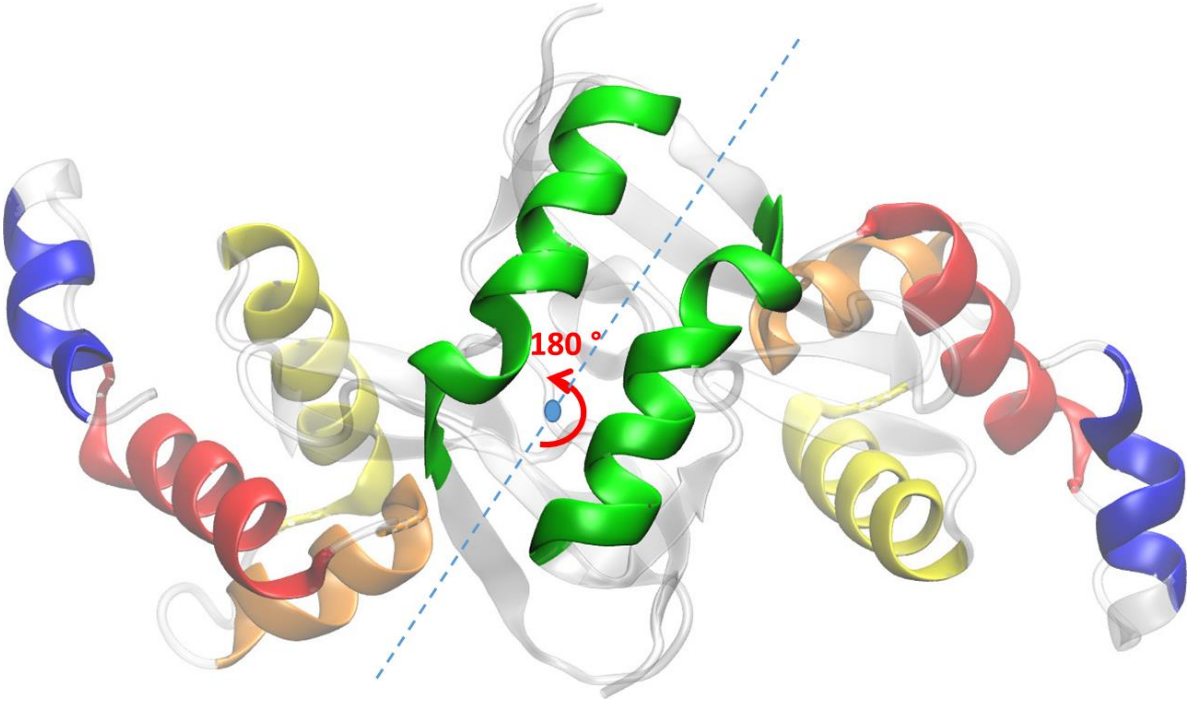


Fig. 3: Representation of ZnNi-ScNur (3EYY) where the 2 N-Ter are completely symmetrical by construction with a symmetry angle of exactly 180 degrees.

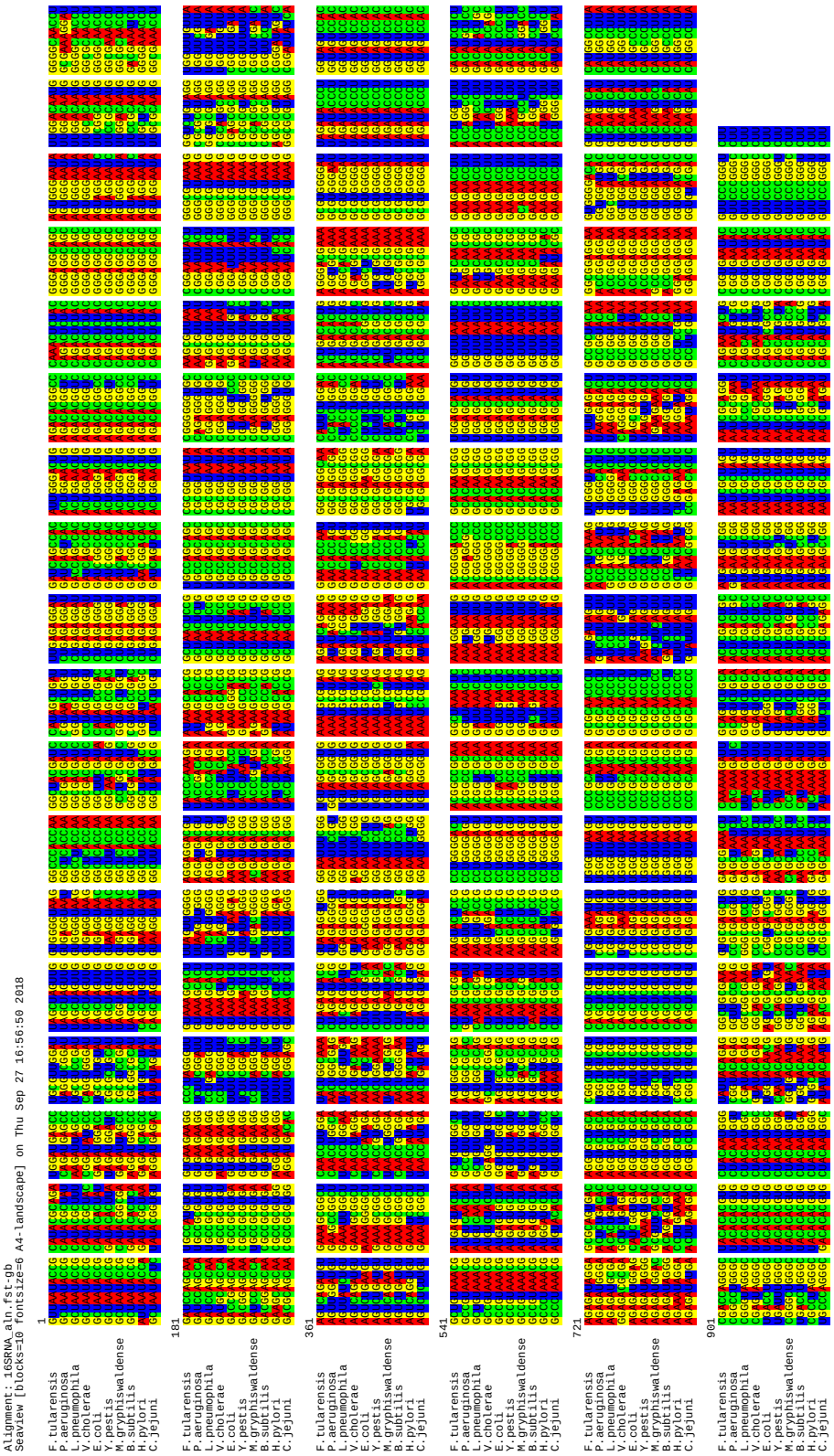


Figure 5.60: Sequence alignment used for the construction of the 16S rRNA tree.

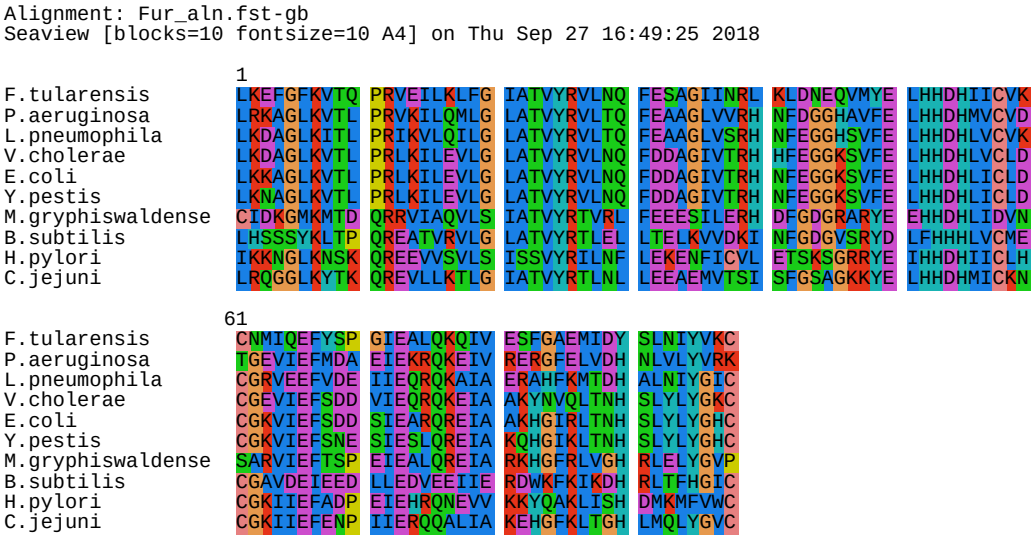


Figure 5.61: Sequence alignment used for the construction of the Fur proteins tree.

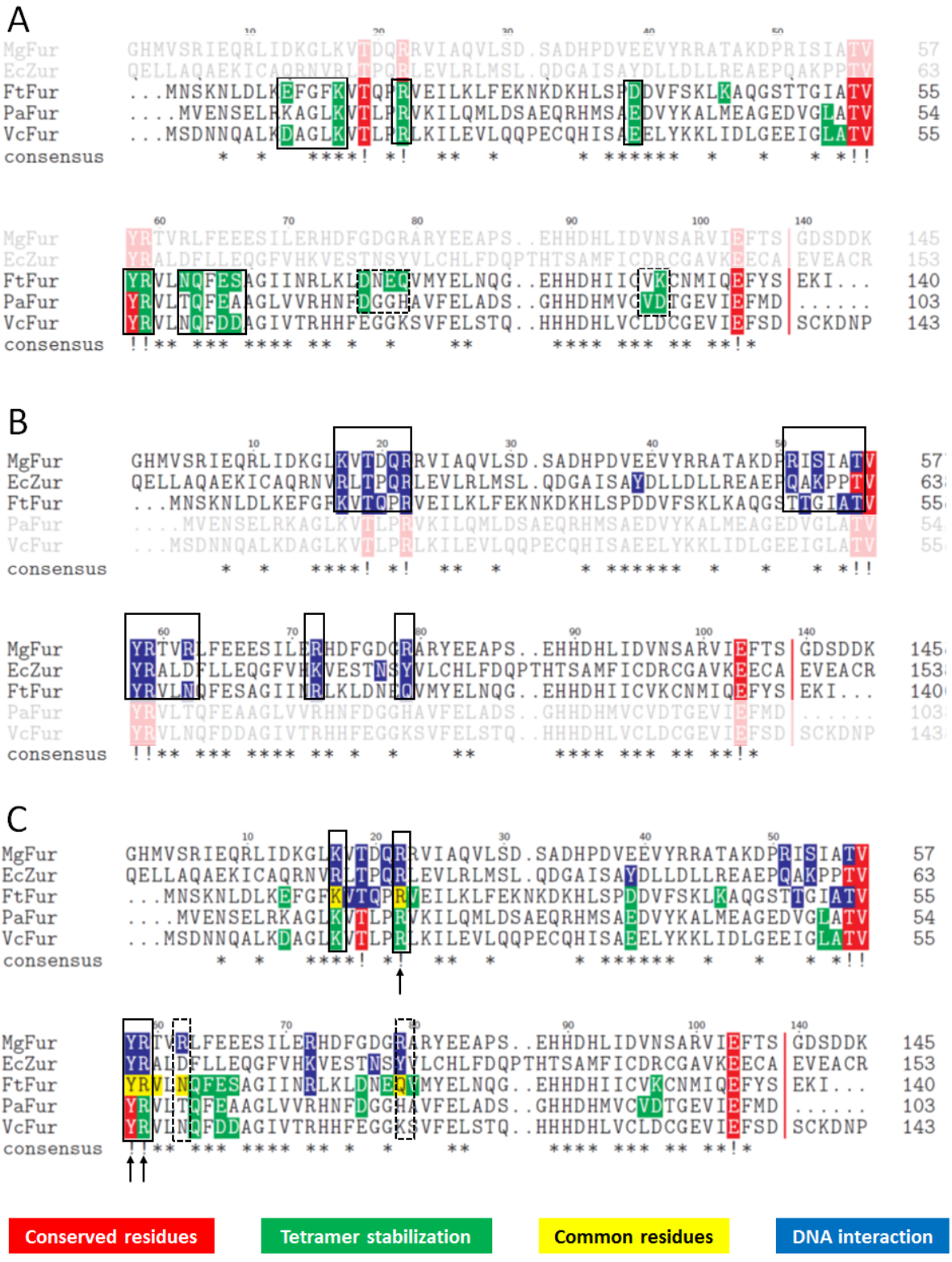


Figure 5.62: Sequence alignment of different Fur proteins showing the predicted residues involved in tetramer stabilization (Figure A, green) or DNA binding (Figure B, blue). Residues that are common for the two types of interactions are shown in yellow in Figure C. Boxes indicate similar results found for different Fur proteins. Residues colored in red are always conserved in Fur proteins (marked by "!"), residues marked by "*" in the consensus are highly conserved (<50%). For figure clarity, residues 107-139 are not shown.

Material and methods

This manuscript combines work from different theoretical and experimental approaches, used to study Fur proteins and their interactions with inhibitors. This section introduces, brief summaries of the methods behind both computational and experimental work, as well as the theory behind key principles. In addition, when necessary, wet laboratory experiment protocols will be detailed to be used as a reference in later work on this project.

Chapter 6

Experimental materials and methods

6.1 Biophysical characterisation

6.1.1 SEC-MALLS

Multi-angle laser light scattering (MALLS) is a technique that measures the scattering of laser light by a sample at different angles. This measurement determines the absolute molecular mass and the average size of molecules in the sample solution.

The calculations use the Rayleigh-Gans-Debye theory that describes the scattering of light by particles smaller than the incident wavelength Strutt, 1871. The intensity detected together with the concentration carry information about the molar mass and the measured scattering carries information about the size of the macromolecule.

In MALLS, a polarized laser beam, λ= 692 nm, is used to illuminate the sample solution. The oscillating electric field of the polarized light induces an oscillating dipole within the macromolecule that will radiate light. The bigger the macromolecule, the larger the induced dipole, and therefore, the greater the intensity of the scattered light (Figure 6.1A). This is why the scattered intensity detected carries information about the molar mass. To differentiate between the polarizability of the macromolecule and the surrounding medium, the change Δn of the solution's refractive index n with the change of the concentration of macromolecules Δc is measured : $dn/dc = \Delta n/\Delta c$ by a differential refractometer.

Light scattered from different parts of the macromolecule reaches the detector with different phases. When the angular dependence of the scattered light is measured in the horizontal plane the macromolecule size can be determined. It is commonly referred to as the mass-averaged root mean square radius r_g also known as the gyration radius (Figure 6.1B).

The equation that describes such variations is given by:

$$\frac{K^* \times c}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2 c \tag{6.1}$$

where $R(\theta, c)$ is the excess Rayleigh ratio of the solution as a function of scattering angle θ and concentration c. It is directly proportional to the intensity of the scattered light in excess of the light scattered by the pure solvent. M_w is the weight-averaged solute molar mass. A_2 is the second virial coefficient in the virial expansion of the osmotic pressure, it describes binary interactions which are proportional to the volume fraction squared. K^* is the vertically polarized incident light constant $4\pi^2(dn/dc)^2 n_0^2/N_a\lambda_0^4$ where n_0 is the refractive index of the solvant, N_a is Avogadro's number and λ_0 the vacuum wavelength of the incident light. $P(\theta)$ describes the angular dependence of scattered light, and can be related to the rms radius. The mean square radius $< r_g^2 >$ is calculated from the slope at $\theta = 0$ of the measured ratios $1/R(\theta, c)$ (Wyatt, 1993). It is determined form Guinier's law near $\theta = 0$ where $P(\theta)$ gives $1 - \mu^2 < r_g^2 > /3$ with $\mu = 2ksin\theta/2$ and $k = 2\pi n_0/\lambda_0$.

In our experiments, MALLS characterisation is used in conjugation with size exclusion chromatography (SEC) to ensure sufficient sample homogeneity and seperate different sizes when studying a mix of macromolecules. In our case it was done by High Performance Liquide Chromatography (HPLC) using a Superdex™ 200 Increase gel filtration column (GE Healthcare Life Sciences). Static light diffusion is measured at 17 different angles by the DAWN® HELEOS® II detector (Wyatt Technology Corp., Santa Barbara, CA) using a 692 nm laser. Then the differential refractometer (RI) Optilab T-rex (Wyatt Technology Corp., Santa Barbara, CA) measures the refraction difference between a control cell and the sample cell. Acquired data is analysed by the ASTRA VI software (Wyatt Technology Corp., Santa Barbara, CA). A schematic representation of a typical SEC-MALLS-RI setup in presented in Figure 6.1C.

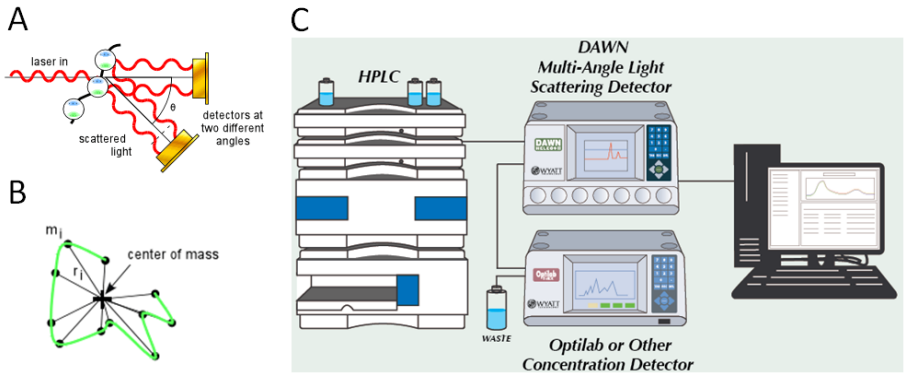


Figure 6.1: A: detection of scattered light by multiple detectors, the intensity reflects the molar mass of the macromolecule. B: size determination after measurement of the angular dependence of scattered light in the horizontal plane. C: a typical SEC-MALLS-RI setup with a fractionation module, in this case a size exclusion chromatography (SEC), and a Multi-Angle static Laser Light Scattering (MALLS) detector. (Figures are adapted from https://www.wyatt.com/library/theory/)

6.1.2 Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) is usually used to complement other techniques and provide information about the molecular weight and oligomeric states of proteins in addition to shape modifications, interaction with flexible ligands and polymerization. It is based on small-angle elastic scattering behaviour where radiation interacts with particles larger than the incident light wavelength and is deflected between 0.1 and 10°. For biological samples and because of their relative low concentration and weak scattering, SAXS measurements are performed at synchrotron radiation sources, with λ= 0.03-0.35 nm, a typical setup is shown in Figure 6.2A.

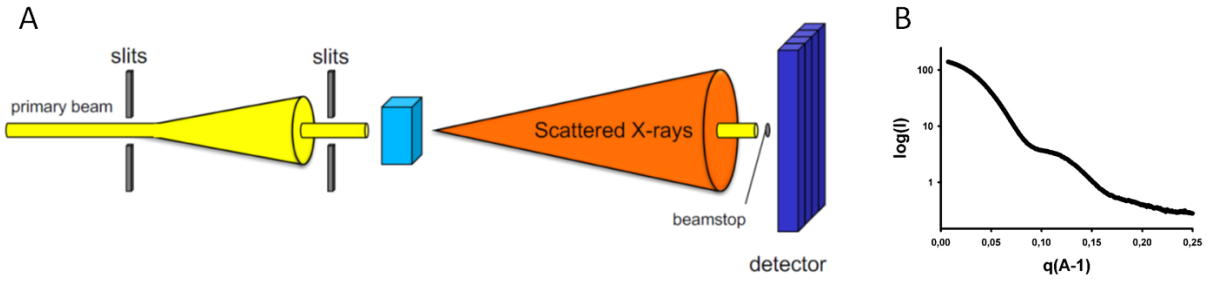


Figure 6.2: A: standard setup of a SAXS beamline, where a monochromatic and collimated X-ray irradiates the sample. Slits are used to reduce any parasitic scattering. Scattering waves are recorded on the 2D detector where a scattering pattern in built. B: 1D merge scattering curve, after buffer substraction, extracted from the pattern built on the detector.

When the X-ray beam hits the sample, an X-ray scattering curve is obtained and is used to generate a 3D structural envelope (SAXS envelope) of the protein that can be later fitted with X-ray or NMR structures. The scattering curve records the scattered intensity $I(s)$ as a function of momentum transfer $s = \frac{4\pi sin\theta}{\lambda}$ where 2θ is the angle between the incident and the scattered radiation.

To obtain the scattering curve of the macromolecule alone, the scattering of the solvent alone is acquired then subtracted from the final scattering curve. Due to susceptibility to radiation damage, scattering curves are obtained on a continuous flow of sample solution to ensure that the data reflects the state of a "fresh" sample. A typical acquisition protocol starts by measuring the sample's buffer alone followed by the sample before remeasuring the buffer alone, the capillary system is washed before and after each step. Usually, one scattering curve represents the mean of ten independent acquisitions.

The SAXS scattering profile of the difference between particle and matrix, shown in Figure 6.2B, is obtained after applying a direct Fourier Transform on $I(s)$ giving $I(q)$ that can be written as the following:

$$I(q) = \xi \frac{d\sigma(q)}{d\Omega} = \xi n \Delta\rho^2 V^2 P(q) S(q) \tag{6.2}$$

where ξ is the instrument constant, n the number of particles, $\Delta\rho^2$ the contrast ($\Delta\rho =<$

$\rho(r) - \rho_s >$, where ρ_s is the scattering density of the matrix), V^2 the particle's volume extrapolated for a globular model, $P(q)$ is the form factor and $S(q)$ is the structure factor.

From a scattering curve, at least four important biophysical parameters can be determined they are the radius of gyration R_g, the molar mass, Porod's volume V_P and the D_{max}.

Applying Guinier's law (Guinier, André, 1939) near $s = 0$ provides the Guinier plot that gives the radius of gyration (R_g). R_g is the quadratic mean of distances to the center of mass weighted by the contrast of electron density, it is given by:

$$I(s) \cong I(0)exp(-\frac{R_g^2 s^2}{3}); R_g^2 = \frac{\int_{V_r} \Delta\rho(r) r^2 dV_r}{\int_{V_r} \Delta\rho(r) dV_r} \tag{6.3}$$

where $I(0)$ is the extrapolated intensity at the origin.

Molar mass, the mass of one mole of a substance, can be determined by using a standard protein solution (usually BSA) and is given by:

$$MM_{sample} = N_a \frac{I(0)/c}{(\Delta\rho\nu)^2} = \frac{(I(0)/c)_{sample}}{(I(0)/c)_{standard}} \times MM_{standard} \tag{6.4}$$

where $I(0) = n\Delta\rho^2 V^2$, N_a is Avogadro's number, c the concentration, $\Delta\rho$ the contrast and ν the partial specific volume of the protein.

Applying Porod's law (Porod, 1951) on a higher s value range, allows the surface area of the particle to be determined. It is termed Porod's volume and is given by:

$$V_P = \frac{2\pi^2 I(0)}{\int_0^\infty I(s) s^2 ds} \tag{6.5}$$

Moreover, Porod's volume can be empirically related to the particle mass by $MM \sim \frac{V_P}{1.6}$

D_{max} is obtained by an indirect Fourier Transformation (IFT) of the scattering plot that gives the pair distribution function (also known as the real spaced distance distribution) $P(r)$ with an estimation of D_{max} that is determined for $P(r) = 0$ (Figure 6.3A and B). It describes probability of finding a point within the particle at a distance r from another given point. Specific shapes have their specific $P(r)$ distributions and can be used to interpret results as seen in Figure 6.3 A and B.

In addition, dimensionless (or globularity and unfoldedness) can be determined by plotting $q^2 I(q)$ versus q which gives the Kratky plot (Glatter O., 1982), Figure 6.3 C. Globular macromolecules follow Porod's law and have bell-shaped curves. Extended molecules, like unfolded peptides, lack this peak and have a plateau or are slightly increasing at large q values.

SAXS data processing uses the ATSAS 2.0 software suite (Franke et al., 2017), where PRIMUS is used to generate the scattering plot, Guinier's plot (R_g), Kratky's plot (dimensionless) and Porod's plot (V_P). GNOM is used to generate the IFT that evaluates

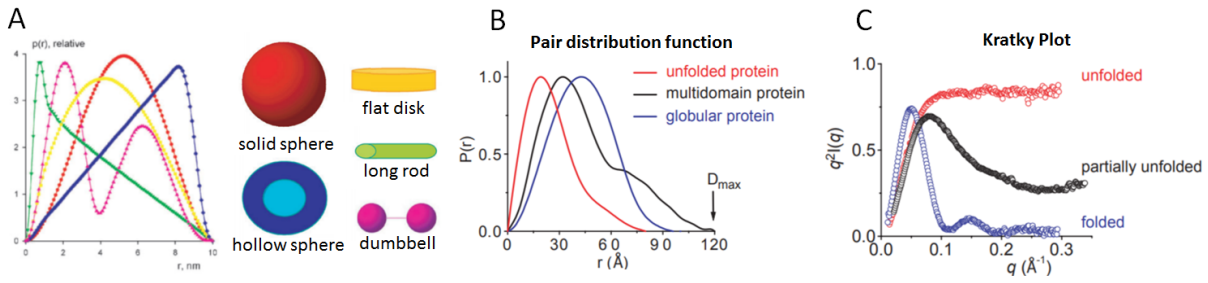


Figure 6.3: A: pair distribution functions of geometrical bodies. B: pair distribution functions of biological macromolecules with D_{max} determination (adapted from Svergun et al., 2003). C: Kratky plot of different folded or unfolded structures that help determine dimensionless (adapted from Putnam et al., 2007).

pair distribution function $P(r)$ and determines D_{max}. DAMMIF and DAMAVER can be used to generate *ab initio* models that fit the experimental scattering curve. Depending on the research strategy, other softwares can be used that combine *ab initio* and rigid body modelling.

6.1.3 X-ray absorption spectroscopy

X-ray absorption spectroscopy (XAS) is an element selective technique that can interrogate metalloproteins metal sites. It can probe a region of 5Å around the absorber, providing a high resolution structure of its binding environment (0.01 Å for interatomic distances). It is mainly applied to the study of detailed structure and dynamics of reactive centers in solution, both in the absence of X-ray structure or to complement it.

The energy of the X-rays in an XAS beamline are tuned in the range of 1 to 30 keV. For protein studies, frozen samples are used, both to prevent radiation damage and to trap the metal in a given state in the case of reactive centers. In these cases, glycerol is used as cryoprotectant. XAS dedicated beamlines are equiped with cryogenic setups cooled with liquid N_2 (-195.79 °C) or He (-270 °C). Due to the relatively low concentration of proteins in frozen samples, data collection uses fluorescence to get sufficient signal.

XAS is based on the photoelectric effect in the X-ray energy domain, where an incident photon beam induces excitation of core electrons in atoms. This photoexcitation creates a vacancy in the core energy level by ejecting a photoelectron from the atom (Figure 6.4A). The ejected photoelectron, due to the the wave–particle duality, can be considered as a spherical wave that will be scattered by neighbouring atoms, producing a backscattered wave, according to the Huygens-Fresnel principle (Huygens, 1690), that will interfere with the outgoing photoelectron (Figure 6.4B). This phenomenon will give rise to variations in absorbance depending on the nature of the neighbouring atoms, their distance, angles and numbers. XAS measures the X-ray absorption coefficient $\mu(E)$ (Figure 6.4C), whose energy dependence is given by the Lambert-Beer law:

$$I = I_0 e^{\mu(E)t} \tag{6.6}$$

Where I_0 is the incident photon beam intensity, I is its intensity after it passes through the sample (transmitted intensity) and E the incoming photon energy.

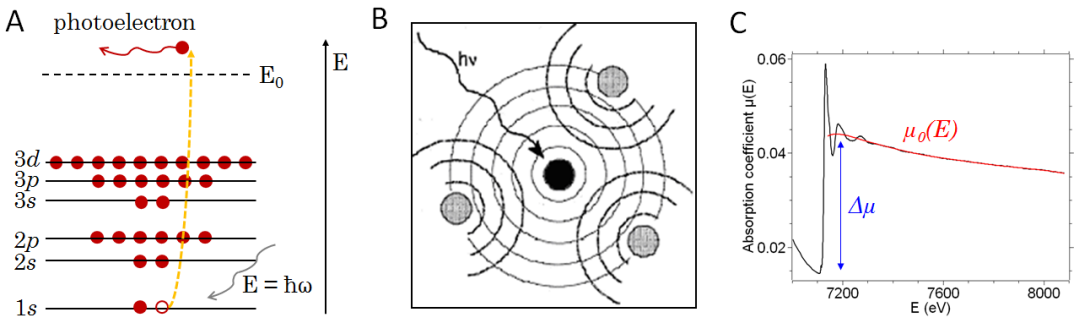


Figure 6.4: A: Schematic representation of the X-ray absorption process by an 1s electron. B: Schematic representation of the interference between the photoelectron ejected by an absorber and backscattered by neighbouring atoms .C: XAS spectra showing the modulation of the absorption coefficient, the measured jump $\Delta\mu_0$ and a function representing the absorption of an isolated atom $\mu_0(E)$ Newville, 2004; Veronesi, 2010.

The near-edge region of the spectra, called X-ray Absorption Near-edge Structure (XANES) contains information about the oxidation state and the coordination geometry. The higher energy region, called Extended X-ray Absorption Fine Structure (EXAFS), can give quantitative information about bond lengths, angles and dynamical parameters (Figure 6.5A). Both are based on the same physical phenomenon but have different mathematical descriptions.

The EXAFS sprectrum is given by:

$$\chi(E) = \frac{\mu(E) - \mu_0(E)}{\Delta\mu} \tag{6.7}$$

Where $\mu(E)$ is the measured absorption coefficient, $\mu_0(E)$ is a smooth background function representing the absorption of an isolated atom and $\Delta\mu$ is the measured jump in the absorption $\mu(E)$ at the threshold energy E_0

According to energy conservation, after absorption of a photon of energy E, the photoelectron final energy E_f is given by:

$$E_f = E - E_b \tag{6.8}$$

Where E_b is the electron binding energy. E_f can be expressed as a function of the photoelectron wave vector k.

$$\frac{\hbar^2 k^2}{2m} = E - E_b \tag{6.9}$$

where $\hbar$ is the reduced plank constant ($\frac{h}{2\pi}$) and m is the electron mass. To link χ to structural parameters, the wave vector is expressed as a function of the incoming photoelectron energy.

$$k = \sqrt{\frac{2m(E - E_b)}{\hbar^2}} \tag{6.10}$$

These EXAFS spectra $\chi(k)$ are shown in k-space, usually as a function of k^3 to get all the information from distant oscillations in the spectrum that are often dampened (Figure 6.5B). By Fourier-Transforming the data in a specific window in k-space gives an EXAFS spectrum in R-space with information about different coordination shells (Figure 6.5C).

The EXAFS formula can be derived from the expression of the quantum mechanical absorption coefficient (Fermi's Golden Rule) $\mu(E) \propto \mid \langle i \mid \hat{H} \mid f \rangle \mid^2$ where $\langle i \mid$ represents the initial state and $\mid f \rangle$ the final state of the photoelectron. The formula is a sum over all the scattering paths (j) of the photoelectron and is given is given by:

$$\chi(k) = S_0^2 \sum_j \frac{N_j f_j(k)}{k R_j^2} \sin(2kR_j + \phi_j(k)) e^{-2R_j/\lambda(k)} e^{-2k^2\sigma_j^2} \tag{6.11}$$

Where N_j is the number of atoms, R_j the distance form the absorber, σ_j^2 the Debye-Waller factor, f_j the backscattering amplitude, ϕ_j the phase shift, $\lambda(k)$ the photoelectron mean free path and S_0^2 the amplitude reduction factor.

Red terms are calculated by *ab initio* methods depending on the nature of the scattering atoms. Blue terms are fitted during data analysis, which provides coordination number, interatomic distances and dynamical parameters.

The EXAFS formula is derived by making use of some approximations. The one electron approximation assumes that only yhe photoelectron is excited by the photon beam, while the other electrons in the atom are not. The amplitude reduction factor S_0^2 is an empirical term that takes into account the relaxation of bound electrons and its value ranges between 0.7 and 1.0. The Muffin Tin approximation takes into consideration the anisotropic atomic potential that is spherically averaged in spheres centered on atoms, while a constant value is assigned to interstitial space. The dipole approximation assumes that there is no spatial variation of the electo-magnetic fields.

XANES analysis can be done using *ab initio* simulations (Feff Rehr et al., 2010, FDMNES Bunău et al., 2009,...) or fingerprint analysis where comparison with reference compounds of known geometry are used.

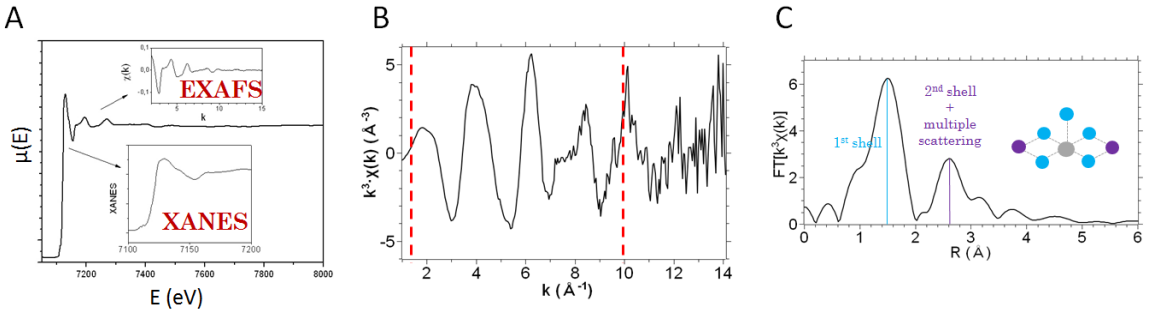


Figure 6.5: A: Iron X-ray absorption spectrum, the upper inset shows the EXAFS spectrum obtained after background subtraction, a magnification of the XANES region is visible in the lower inset. B: EXAFS spectrum in the k-space. C: EXAFS spectrum in the R-space (Newville, 2004; Veronesi, 2010).

In our studies, cobalt K-edge X-ray absorption spectra were acquired on the beamline CRG-FAME-BM30B at the European synchrotron Radiation Facility, ESRF, Grenoble, France (Proux et al., 2006). Drops of protein solution were deposited on the sample holder equipped with kapton windows and frozen in liquid nitrogen, then transferred into the He-cryostat and measured at 15 K. The Co K-edge energy region was scanned in the 7600-8700 eV range with a nitrogen-cooled Si(111) monochromator. The photon energy was calibrated over a metallic Co foil, by defining the first inflection point of its absorption spectrum at 7708.9 eV. X-ray absorption spectra were recorded in fluorescence mode with a 30-elements Ge solid-state detector (Camberra).

XAS data normalization and extraction was performed with standard methods using the Athena software (Ravel et al., 2005). Scattering amplitudes and phase shifts were calculated with the ab initio code FEFF9 (Rehr et al., 2010), including Self Consistent Field calculations of the scattering potential in a radius of 4.4 Å around the absorber. The amplitude reduction factor S_0^2 was calculated by the program from atomic overlap integrals: the value of 0.94 was kept constant during the fit. Input clusters for the *ab initio* calculations were built setting metal-ligand first-shell distances to average values calculated from a PDB structure survey within the MESPEUS database (Hsin et al., 2008), while interatomic distances and angles relative to each amino acid were set to values used for crystal structure refinement [ref structure ec]. The fits to the experimental data were performed with the Artemis program (Ravel et al., 2005) constraining amino acids to move around the absorber with two degrees of freedom (a shift and a tilt), applying the Rigid Body Refinement scheme (Binsted et al., 1992). The fits were performed in the k-space, in the range [2-10] Å^{-1}, with multiple k-weights (k=1, 2, 3). In order to disentangle structural and dynamical parameters, the fit was performed step-by-step, varying first shell distances and tilt angles (β parameter in Table 4.19) first, then fixing them to the best-fitting values and allowing Debye-Waller factors (σ^2) to vary. Dynamical parameters were grouped in order to reduce the number of free variables: a common value was assigned to Single Scattering paths (SS) involving first-shell atoms (σ_1^2), second-shell atoms (σ_2^2), or Multiple Scattering (MS) and outer shell SS paths (σ_{MS}^2). A shift in the threshold energy (ΔE_0) was always set as a global variable.

6.2 Crystallization assays

6.2.1 Manual crystallization assays

The hanging drop vapour diffusion technique is used in all the manual crystallization assays described in this work. This technique consists of a drop of protein mixed with a precipitant and placed on the inside of a lid over a reservoir filled with the precipitant (Figure 6.6A). Since the precipitant is diluted in the drop by the addition of the protein

solution, when the well is closed and the drop is "hanging" above the reservoir, equilibrium is reached by water vapour leaving the drop to get to the reservoir. By doing so, the protein and the precipitant concentration increase in the drop and supersaturation is reached (Figure 6.6C).

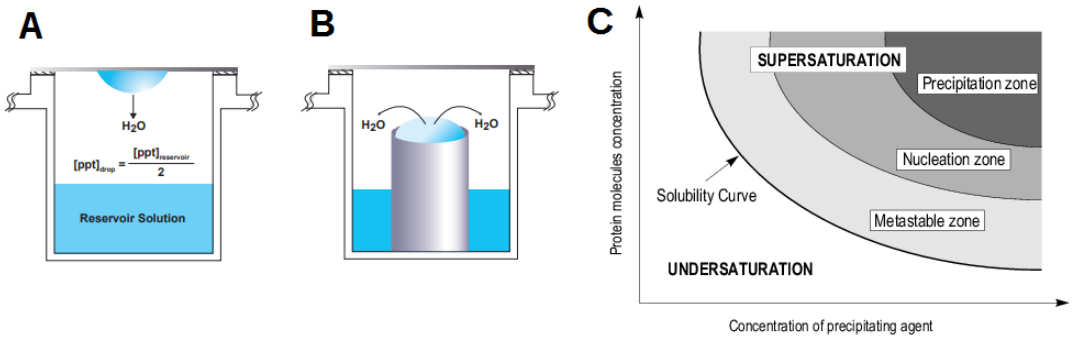


Figure 6.6: Crystallization techniques and crystallization phase diagram. Usually a crystallization drop consists of a mix of equal part (ratio 1:1) of protein and precipitant solution, diluting the precipitant concentration in the drop to half its concentration in the reservoir, this initiates water vapour diffusion. A: hanging drop technique, "ppt" is an acronym for precipitant. B: sitting drop technique. C: crystallization phase diagram showing the change of protein molecules concentration against precipitating agent concentration. For crystallization the ideal condition is at the interface between the nucleation zone and the metastable zone. Adapted from Hampton Research, 2017a; Hampton Research, 2017b; Chirgadze, 2001.

In all the manual crystallization assays described below, several protein:buffer were used and ranged from 1:0.5, 1:1, 1:1.5, 1:2 to 2:2.

6.2.2 Automated crystallization assays: HTXLab

The High Throughput Crystallization Laboratory (HTXLab) provides through the crystallization information management system (CRIMS©) an automated platform of crystallization screening. This platform was used for all the crystallization buffer and additives screens presented in this work. It uses CrystalDirect™plates with sitting drop (3 drops) and 45 µL wells, used for *in situ* characterization and automated crystal harvesting

For the standard crystallization buffer screening (six plates) a sample of 75 µL of protein is provided to HTXLab. For an additive screen (one plate) 20 µL of protein and 15 mL of buffer are needed. To avoid aggregation, protein and buffer sample were centrifuged at 10000 rpm and their supernatant sent to HTXLab.

6.2.3 Crystallization of PaFurΔS3

PaFurΔS3 is the metal site 3 mutant of Fur from *P. aeruginosa*, H86A - H125A, that crystallized in the following condition:

Zn Acetate 0.2 M ; MES 0.1 M pH 6 ; PEG 8000 5%

The stock solution protein concentration of 30 $mg.mL^{-1}$ was diluted to 10 $mg.mL^{-1}$ prior to any crystallization assay using the following protein dilution buffer:

HEPES 50 mM pH 7.5 ; NaCl 150 mM ; $MnCl_2$ 1 mM ; TCEP 0.1 mM

Cryo-conservation of crystals after fishing was done using the crystallization buffer to which 5% PEG 8000 and 20% glycerol were added. The crystals were soaked in this solution for approximately one minute before being flash frozen in liquid nitrogen. For further washing, the final step consisted in soaking the crystals in the same buffer but without Zn acetate to prevent excess Zn signal during SAD collection.

Diffraction experiments have been done on the beamline MASSIF ID30A-ESRF (European Synchrotron Radiation Facility in Grenoble, France). For Zn-PaFurΔS3, the best dataset (0.99 Å) diffracted at 2.7 Å of resolution. Diffraction data were integrated/scaled in the space group Symmetry space group P3221 (84.6 / 85.6/ 96.6 / 90 / 90 /120) using XDS-package Version October 15, 2015 37. The structure was solved by the molecular replacement method using Phenix 1.10.1-2155 with AutoSol and 95% of the model was built automatically. The model was rebuilt/corrected manually and refined using COOT (Emsley et al., 2010) and final refinement cycle was done in Phenix (Adams et al., 2010). For Mn-PaFurΔS3 remote and anomalous datasets were recorded at wavelengths 0.97 Å and 1.27 Å (right side of maximum F" for manganese). Best dataset (0.97 Å) diffracted at 2.2 Å resolution Symmetry space group P6122 R 2.2 Å (88.5 / 85 / 177 / 90 / 90). Diffraction data were integrated/scaled in the space group P6122 using XDS-package package Version October 15, 2015. The structure was solved by the molecular replacement method using Phenix 1.10.1-2155/AutoSol and 95% of the model was built automatically with Zn-PaDS3fur already solved. The model was rebuilt/corrected manually and refined using alternatively COOT and final refinement cycle was done in Phenix.

Anomalous map with collection data of 1.29A are generate by XDS and used by COOT to validate the localization of each Mn atom. Since Zn cannot be seen at 1.27A and the presence of Mn was validated with X-ray fluorescence spectra, Mn atom were placed in each peak superior to 5 (sigma intensity) in the anomalous map. Both final structures were validated by PHENIX and PDB redo was used before deposition of the structures to the Protein data Bank (PDB). In order to study the physiological tetramer of PaFurΔS3 protein the PISA web server (Krissinel et al., 2007, http://www.ebi.ac.uk/pdbe/pisa/) was used to generate the Tetramer.

6.2.4 Crystallization of PaFurΔS3 with inhibitors

Protein preparation for crystallization assays with the inhibitors is similar to its preparation for assays of the protein alone except a few changes that take into consideration

the addition of inhibitors in the crystallization drops. The same protein stock solution, with a 30 mg.mL^{-1} concentration, is diluted with the following buffer to obtain a solution with a concentration of 10.5 mg.mL^{-1} :

HEPES 50 mM pH 7.5 ; KCl 50 mM ; $MnCl_2$ 1 mM ; TCEP 0.5 mM

The final protein concentration is slightly higher (+0.5 mg.mL^{-1}) than in the preparation used for the protein alone to take into account the dilution resulting form the addition of the inhibitors in the crystallization drop. TCEP concentration was increased to ensure the reduction of cysteines in pL1 and pL2.

Inhibitors, solubilized in DMSO, were prepared to be ten times more concentrated than the protein in the crystallization drop. A maximum limit of 5% in final concentration of DMSO was respected to prevent any interference with the crystallization process but at the same time allow dissolution of the inhibitors. The inhibitors were also added to the crystallization buffer to prevent their dilution when the crystallization buffer is added to the protein inhibitor complex.

However, in the cases of pL1 and pL2, 5% DMSO were not enough to prevent the aggregation of the protein when the inhibitors were added. Slight shaking by finger tapping on the bottom of the eppendorf tube and incubation for 15 minutes on ice resolubilized the complex. No up and down mixing with the micropipette should be done since it renders the aggregation unrecoverable.

In the case of molecule B, its addition to the protein solution creates immediately a white precipitate, the only way to resolubilize it was by increasing the final DMSO concentration from 5% to 7.5%. Shaking of the eppendorf tube and incubation on ice alone do not solve the problem. It is important to note that the aspect of the precipitation of the protein inhibitor complex is not similar in the case of molecule B in comparison with pL1 and pL2. The addition of pL1 or pL2 to the protein solution create a foggy white aggregate that tends to stay somewhere around the middle of the solution volume. In contrast, the addition of molecule B creates a white flake like precipitate that falls directly to the bottom of the tube. This could be due to the difference in the chemical nature of the inhibitors, pL1 and pL2 being peptides and molecule B a small cyclic molecule.

Cryo-conservation of PaFurΔS3 with inhibitors was done in a similar way to the PaFurΔS3 crystals, however inhibitors were added to ensure that the cryo-conservation step did not change the chemical environment of the crystals. Typically, the cryo-buffer was the crystallization buffer supplemented with 5% PEG 8000, 20 % glycerol and inhibitors diluted to 1/4 of their concentration in the crystallization drop. For further washing, the final step consisted in soaking the crystals in the same buffer but without Zn acetate to prevent excess Zn signal during SAD collection.

For the crystallization assays where molecule B powder was used to soak PaFurΔS3 crystals, a 10 µL tip was dipped in molecule B powder then dipped in the crystallization

drop over the crystals. Fishing was carried out after 30 minutes of soaking and the last crystal was taken the next morning. Cryo-conservation was done similarly to what was described above.

6.2.5 Crystallization of EcFurΔS3

Crystallization trials of EcFurΔS3 used manual and automated screenings. Both methods are similar to what was described for PaFurΔS3. The protein concentration varied between test series, the lowest being 6 mg.mL^{-1} and the highest was 10 mg.mL^{-1}.

6.3 Cloning, expression and purification of Fur proteins

6.3.1 EcFur-WT, EcFurΔS2 and EcFurΔS3

The gene encoding EcFur described in this manuscript has already been cloned into pET30c vector as describe in D'Autreaux et al., 2002. Colonies were cultured in regular LB media at 37°C. The induction of protein expression was achieved by adding 0.5 M of IPTG to the cultures at OD=0.8, the ΔS2 mutant culture was stopped after 3 hours of induction at 37°C. The expression of the ΔS3 mutant was induced at 18°C overnight. Cultures were pelleted by centrifugation at 4°C for 20 minutes at 6000 rpm and resuspended in a lysis buffer containing : Tris 50 mM at pH 8.8, KCl 0.5 M, one "complete EDTA-free" pellet (Roche, ref: 11 873 580 001) and 0.5 mg of lysozyme (Euromedex, ref: 5934-C). Lysis by sonication (Bioblock Scientific Vibra Cell 72412) used a 2 seconds ON / 4 seconds OFF alternating protocol for a total of 30 minutes. The lysate was pelleted at 4°C for 30 minutes at 20000 rpm and the supernatant collected. After this step, the proteins were precipitated by a 80% solution of ammonium sulfate overnight at 4°C then pelleted at 4°C for 30 minutes at 10000g. Pellets were resuspended in 100 mL of Tris 25 mM at pH 8.8 and KCl 500 mM.

This volume was injected into an affinity chromatography chelating sepharose column, previously equilibrated with 0.1 M of Zn, followed by a series of washings containing independently KCl 0.5 M, amonium sulfate 1 M, glycine 25 mM and KCl 0.5 M. All washings were done in a solution of 25 mM Tris at pH 8.8. For the elution, a solution of Tris 25 mM at pH 8.8, KCl 500 mM and Imidazol 0.5 M was used.

This protocol was not effective for the purification of EcFurΔS2 and EcFurΔS3. To purify these two constructions, heparin sepharose ion exchange chromatography columns were used with an NaCl elution gradient. The final purification step for EcFur-WT, EcFurΔS2 and EcFurΔS3 was a size exclusion chromatography using a Superdex 75 column (GE Healthcare Life Sciences).

6.3.2 EcFur-140

The gene encoding EcFur described in this manuscript has already been cloned into pET30c vector as describe in D'Autreaux et al., 2002. C-terminal deletion mutants have been amplified by PCR from previous plasmid and cloned into pET-TEV vector (based on pET28a vector), at the EMBL-Grenoble, between NocI/ XhoI in fusion with 6xHis cleavable TAG with TEV protease in N-Terminal position.

Proteins expression was realized in BL21DE3 Rosetta 2. 1L of LB media (supplemented with 60 µg/L of Kanaymycin and 30 µg/L of Chloramphenicol) was inoculated with 10mL of a saturated overnight pre-culture at 37°C. The culture was left to grow at 37°C to an absorbance of 0.8 at 600 nm. The induction of protein expression was achieved by adding 0.5 M of IPTG to the cultures at OD=0.8, the culture was stopped after 3 hours of induction at 37 °C. Cultures were pelleted by centrifugation at 4°C for 15 minutes at 6000 rpm and resuspended in a lysis buffer containing : Tris 50 mM at pH 8.8, KCl 0.5 M, glycerol 10%, one "complete EDTA-free" pellet (Roche, ref: 11 873 580 001) and 0.5 mg of lysozyme (Euromedex, ref: 5934-C). Lysis by sonication (Bioblock Scientific Vibra Cell 72412) used a 2 seconds ON / 4 seconds OFF alternating protocol for a total of 30 minutes. The lysate was pelleted at 4°C for 45 minutes at 20000 rpm and the supernatant collected. For purification the supernatant was injected into an NiNTA column, previously equilibrated with nickel, then washed with a solution of Tris 50 mM at pH 8.8 and KCl 1 M. Elution was performed using a solution of Tris 50 mM at pH 8.8, KCl 0.5 M, glycerol 10% and an increasing concentration of imidazole. The excess imidazole was removed from the eluate by dialysis overnight at 4°C in a solution of Tris 25 mM at pH 8.8 and KCl 50 mM. After this step, the dialysate was precipitated using a 80% solution of ammonium sulfate overnight at 4°C then pelleted at 4°C for 30 minutes at 10000g. Pellets were resuspended and dialyzed in Tris 25 mM at pH 8.8 and KCl 250 mM overnight at 4°C. The next step in purification was a size exclusion chromatography, carried on twice to insure high purity, on HiLoad Superdex 200 pg preparative size exclusion chromatography columns (GE Healthcare Life Sciences) the protein was eluted with Tris 50 mM at pH 8.8, KCl 0.5 M and glycerol 10%.

6.4 Nuclease protection assay

The nuclease protection assay is an *in vitro* DNA binding assay where an active Fur bind its specific sequence and protects one cleaving site from being accessible to the restriction enzyme HinfI. This can be seen on the digestion gel where one band measures 1781 bp (1530 + 251) and the smallest band at 251 bp is no longer present in the case of an active Fur protein (Figure 6.7B).

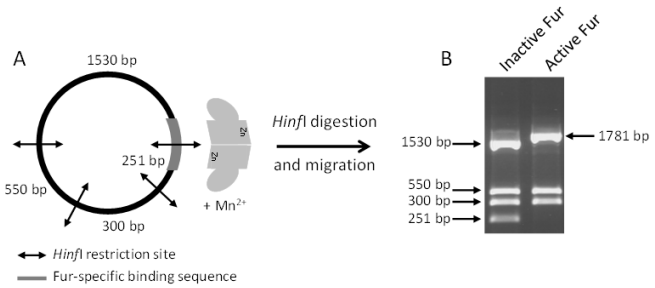


Figure 6.7: Principle of the nuclease protection assay: the Fur binding site is highlighted on the pDT10 plasmid. The HinfI restriction sites and the expected fragment sizes after digestion during one hour at 37°C are indicated.

6.5 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) take advantage of the differential migration speed of macromolecules of different size in an electrophoresis gel to indicate a binding to DNA. On the gel, a DNA sequence alone will migrate faster than when proteins are bound on it, this creates a shift in the migration of the band hence the name of the technique. Depending on the experiments, EMSA essays are used to detect DNA-protein or RNA-protein interactions, and can deliver the Kd of the observed complexes from the protein concentration used in the assay.

6.6 Inductively coupled plasma atomic emission spectroscopy

Inductively coupled plasma atomic emission spectroscopy (ICP-AES) technique is used to detect chemical elements in a sample. Elements are detected by their emitted electromagnetic radiation when they are excited by the inductively coupled plasma. For each element, the intensity of the emission indicates its concentration in the sample. In our laboratory, the setup consists of a Shimadzu ICP 9000 instrument with Mini plasma Torch in axial reading mode in addition to a CCD decetor with 1 Mpixel.

6.6.1 DTNB assay

DTNB stands for 5,5'-disulfanediylbis(2-nitrobenzoic acid) or Ellman's reagent (Ellman, 1959). DTNB assays are used to determine the quantity of thiols in a sample. In studying proteins, this method is used to detect reduced cysteines. Its use for measuring thiols is straight forward, DTNB reacts with thiols releasing TNB in a stoichiometric reaction (Figure 6.8). Knowing its extinction coefficient of 14150 $M^{-1}cm^{-1}$, the yellow TNB^{2-} is then quantified by spectrophotometry by measuring the absorbance of the sample at

412 nm. A standard range of free cysteines is used to calculate the concentration in the sample.

Figure 6.8: DTNB reaction with thiol. The reaction releases TNB^{2-} that can be quantified by spectrophotometry.

Chapter 7

Theoretical Methods

When studying biochemical processes, structural information such as NMR and X-ray crystallography structures are usually considered as key elements in the understanding of a specific process, and it is true. However, even after the laborious work to obtain a structure, the information that can be extracted from it remains specific to one "static" state. Studies on dynamic systems are required to complement data obtained from structural experiments. This can be done through molecular dynamics simulations that can determine entities that experiments can not obtain.

7.1 Theory

7.1.1 Classical molecular dynamics

Molecular Dynamics (md) models properties of a system of interacting particles by repeatedly calculating the interactions between the particles and integrating their equations of motion. As algorithms become more efficient and computers faster than before, this area of science is continuously developing and improves physical models of biomolecular systems.

Newton's second law

Molecular Dynamics simulations are based on Newton's second law, the equation of motion (Brooks et al., 1983; Leach, 2001) is :

$$F_i = m_i.a_i = m_i.\frac{dv_i}{dt} \tag{7.1}$$

It describes the motion of a particle of mass m_i, coordinates x_i and velocity v_i with F_i being the force on m_i and a_i its acceleration in that direction. This is used to calculate the motion of a finite number of atoms or molecules, respectively, under the influence

of a force field that describes the interactions inside the system with a potential energy function, $V(\vec{x})$, where $\vec{x}$ corresponds to the coordinates of all atoms in the system. The relationship of the potential energy function and Newton's second law is given by

$$F(\vec{x_i}) = -\nabla_i V(x_i), \tag{7.2}$$

with $F(\vec{x_i})$ being the force acting on a particle due to a potential, $V(\vec{x})$. Combining these two equaions gives

$$\frac{dV(\vec{x})}{dx_i} = -m_i.\frac{d^2x_i}{dt^2}, \tag{7.3}$$

which relates the derivative of the potential energy to the changes of the atomic coordinates in time. Due to the potential energy being a multidimensional function this equation can only be solved numerically with some approximations.

With the acceleration being $a = -\frac{1}{m}.\frac{dV}{dx}$ the changes of the system in time can be calculated knowing the potential energy $V(\vec{x})$, the initial coordinates x_{i0} and an initial distribution of velocities v_{i0}. This method is deterministic since it can predict the state of the system at any point of time in the future or the past.

The initial distribution of velocities is usually randomly chosen from a Gaussian or Maxwell-Boltzmann distribution (Leach, 2001), which gives the probability of an atom i having a velocity v_i in the direction of x at the temperature T by:

$$p(v_{i,x}) = \sqrt{\frac{m_i}{2\pi k_\mathrm{B} T}}.\exp\frac{-1}{2}\frac{m_i v_{i,x}^2}{2k_\mathrm{B}T}. \tag{7.4}$$

Velocities are then corrected so that the overall momentum of the system equals a zero vector:

$$P = \sum_{n=1^N} m_i.\vec{v_i} = \vec{O}. \tag{7.5}$$

Verlet algorithm

The solution of the equation of motion given above is only sufficient over a short period of time, where velocities and accelerations are regarded as constants. Algorithms were introduced to repeatedly perform small time steps to propagate the system's properties in time such as positions, velocities and accelerations. Time steps are typically chosen in the range of 1 fs (Leach, 2001). Small time steps are used since molecular processes occur in short periods of time and can not be resolved with larger time steps. A time series of coordinate sets calculated is referred to as a trajectory and a single coordinate set as a frame.

All algorithms assume that the system's properties can be approximated by a Taylor

series expansion (M. et al., 2003):

$$\begin{aligned}\vec{x}\,(t+\delta t) &= \vec{x}\,(t) + \delta t.\vec{v}\,(t) + \frac{1}{2}\delta t^2.\vec{a}\,(t) + ...\\ \vec{v}\,(t+\delta t) &= \vec{v}\,(t) + \delta t.\vec{a}\,(t) + \frac{1}{2}\delta t^2.\vec{b}\,(t) + ...\\ \vec{a}\,(t+\delta t) &= \vec{a}\,(t) + \delta t.\vec{b}\,(t) + \frac{1}{2}\delta t^2.\vec{c}\,(t) + ...\end{aligned} \tag{7.6}$$

with $\vec{x}$,$\vec{v}$ and $\vec{a}$ being the positions, the velocities and the accelerations of the system.The series expansion is usually truncated after the quadratic term. The Verlet algorithm (Verlet, 1967) is the most widely used algorithm for integrating the equations of motion in MD simulations (Brooks et al., 1983; Leach, 2001). It can be derived by summing the Taylor expressions for the coordinates at time $(t+\delta t)$ and $(t-\delta t)$:

$$\begin{aligned}\vec{x}\,(t+\delta t) &= \vec{x}\,(t) + \delta t.\vec{v}\,(t) + \frac{1}{2}\delta t^2.\vec{a}\,(t) + ...\\ \vec{x}\,(t-\delta t) &= \vec{x}\,(t) - \delta t.\vec{v}\,(t) - \frac{1}{2}\delta t^2.\vec{a}\,(t) - ....\\ \Rightarrow \vec{x}\,(t+\delta t) &= 2\vec{x}\,(t) - \vec{x}\,(t-\delta t) + \delta t^2.\vec{a}\,(t)\end{aligned} \tag{7.7}$$

The position $\vec{x}\,(t)$ and acceleration $\vec{a}\,(t)$ at time t and the positions from the previous step $\vec{a}\,(t-\delta t)$ are used to calculate new positions $\vec{x}\,(t+\delta t)$. In this algorithm velocities are not explicitly calculated but can be obtained in several ways. One is to calculate mean velocities between the positions $\vec{x}\,(t+\delta t)$ and $\vec{x}\,(t-\delta t)$.

$$\vec{v}\,(t) = \frac{1}{2\delta t}.[\vec{x}\,(t+\delta t) - \vec{x}\,(t-\delta t)] \tag{7.8}$$

Even if this algorithm does not require huge storing capacities by only saving the positions $\vec{x}\,(t)$ and $(t-\delta t)$ and the accelerations $\vec{a}(t)$, it has the disadvantage of being moderately precise since the new positions are obtained by adding a small term $[\delta t^2\vec{a}(t)]$ to the difference of two much larger terms $[2(t) - \vec{x}\,(t-\delta t)]$.

In addition, the Verlet algorithm is not a self-starting algorithm. New positions $\vec{x}\,(t+\delta t)$ are obtained from the current positions $\vec{x}(t]$ and the positions at the previous step $\vec{x}\,(t-\delta t)$. So, at $t = 0$ there are no positions for $(t-\delta t)$ and therefore, it is necessary to provide another way to calculate them. One way is to use the Taylor expansion truncated after the first term:

$$\vec{x}\,(t+\delta t) = x\,(t) + \delta t.\vec{v}\,(t) + \frac{1}{2}\delta t^2.\vec{a}\,(t) + ... \Rightarrow x(\delta t) = \vec{x}\,(0) + \delta t.\vec{v}\,(0) \tag{7.9}$$

Leap-Frog algorithm

Several variations of the Verlet algorithm exist and try to avoid its disadvantages. One example is the leap-frog algorithm (Leach, 2001) that uses the following equations:

$$\begin{aligned} \vec{v}\left(t+\frac{1}{2}\delta t\right) &= \vec{v}\left(t-\frac{1}{2}\delta t\right)+\delta t.\vec{a}t \\ \vec{x}\left(t+\delta t\right) &= \vec{x}\left(t\right)+\delta t.\vec{v}\left(t+\frac{1}{2}\delta t\right) \end{aligned} \tag{7.10}$$

First, the velocities $\vec{v}\left(t+\frac{1}{2}\delta t\right)$ are calculated from the velocities at $t-\frac{1}{2}\delta t$ and the accelerations $\vec{a}(t)$. Then the positions $\vec{x}\left(t+\delta t\right)$ are deduced from the velocities just calculated and the positions at time t . By doing so, the velocities first 'leap-frog' over the positions and then the positions leap over the velocities. The leap-frog algorithm's advantage over the Verlet algorithm is the inclusion of explicit velocities. However, an obvious disadvantage is that the positions and velocities are not synchronized.

7.1.2 Molecular mechanics force fields and parameters

The potential energy function $V(r)$ is described as a sum of individual energy terms and called the force field. An example is the CHARMM force field.

The CHARMM (Chemistry at HARvard Molecular Mechanics) program (Brooks et al., 1983) is a molecular mechanics simulation program used for minimization, dynamics simulations and thermodynamic calculations. The force field is given by:

$$\begin{aligned} V(r) = &\sum_{\text{bonds}} k_r(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\phi(1+\cos(n\phi-\delta)) \\ &+ \sum_{\text{impropers}} k_\omega(\omega-\omega_0)^2 + \sum_{\text{UB}} k_\xi(\xi-\xi_0)^2 \\ &+ \sum\sum_{\text{nonbonded}} \left\{ \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6} \right] + \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q_i q_j}{r_{ij}} \right\} \end{aligned} \tag{7.11}$$

Other variants of the potential energy function are also implemented in CHARMM and are used to treat long-range electrostatic interactions (extended electrostatics model (Stote et al., 1991) or the fast multipole method (Greengard et al., 1987)).

In summary, in molecular modelling, a force field is the functional form and parameter sets used to calculate the potential energy of a system of atoms. The energy functions

parameters are derived from experiments in physics or chemistry, calculations in quantum mechanics, or both.

Force fields enable the potential energy V of a molecular system to be calculated rapidly. The energy is determined by the positions of all the atoms in the system and more precisely as a function of the nuclear positions using the Born-Oppenheimer-Approximation (Born et al., 1927), unlike quantum mechanical methods that deal with the electrons of the system. This is why Molecular Mechanics is used on systems containing a significant number of atoms that would be impractical in quantum mechanical calculations.

Each atom in the system is represented as a single point and energies as a sum of atom interactions such as bond stretching and angle bending.

Force fields require the attribution of atom types to each atom in the system, this contains information about their hybridization state and their local environment. The equivalent of atom types in quantum mechanics calculation is the attribution of the charge of the nuclei, the geometry of the system, the overall charge and spin multiplicity.

The value of the potential energy V is calculated as a sum of internal or bonded terms, such as bonds, angles and bond rotations, and a sum of non-bonded terms, which account for interactions between non-bonded atoms or atoms separated by three or more covalent bonds. It is given by:

$$V\left(\vec{r}\right) = V_{\text{bonded}}\left(\vec{r}\right) + V_{\text{nonbonded}}\left(\vec{r}\right) \tag{7.12}$$

Each term of the force field is reviewd in details in the following section.

Bond Streching

The potential energy curve (Morse potential) for a typical bond has the shape shown in Figure 7.1 and has the following form as well:

$$V_B\left(t\right) = D_e\{1 - exp[-a\left(l - l_0\right)]\}^2 \tag{7.13}$$

D_e is the depth of the potential energy minimum and $a = \omega \times \sqrt{\frac{\mu}{2D_e}}$, where μ is the reduced mass and ω is the frequency of the bond vibration. The frequency ω is related to the streching constant of the bond k_l, by $\omega = \sqrt{\frac{k_l}{\mu}}$, where k_l determines the strength of the bond. The length l_0 is the reference bond length. Both l_0 and k_l are specific for each pair of bound atom. Values of k_l derive from experiments or from quantum mechanical calculations. Values of l_0 can be inferred from high resolution crystal structures or microwave spectroscopy data.

The Morse potential is computationally demanding, simpler expressions are often used in molecular mechanics force fields like CHARMM. The most elementary approach is to use a Hooke's Law formula in which the energy varies with the square of the displacement

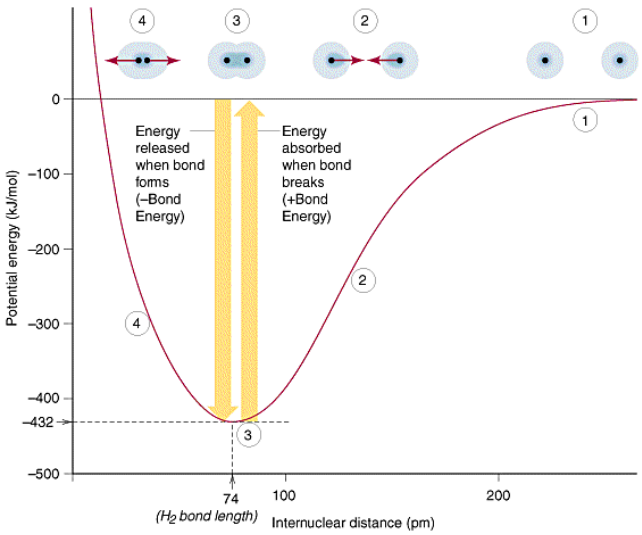


Figure 7.1: Variation in bond energy with interatomic separation applied to the case of an H_2 bond (*Chieh, 2018, Bond Lengths and Energies*). (1) At infinitely large distances no interaction exists between atoms. (2) Attraction between relatively close atoms. (3) Equilibrium distance. (4) Strong internuclear repulsion at short range.

from the reference bond length l_0:

$$V_B(l) = \frac{1}{2} k_l \times (l - l_0)^2 \tag{7.14}$$

Hooke's Law is a good approximation of the shape of the potential energy curve at distances that correspond to bonding in ground-state molecules but is less accurate away from equilibrium (Figure 7.2).

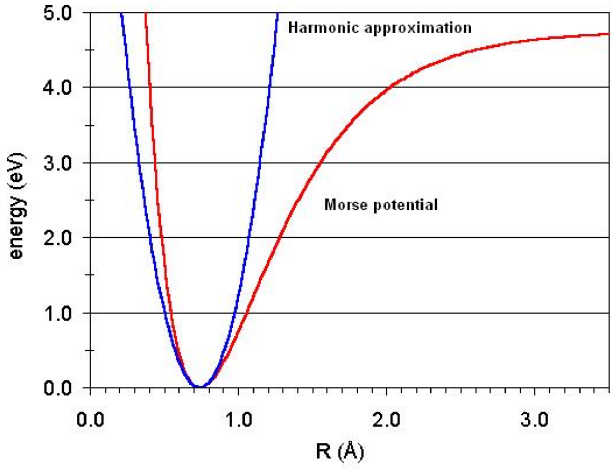


Figure 7.2: Comparison of the simple harmonic Potential (Hooke's Law) in blue with the Morse curve in red (*Tissue 2018, Harmonic Oscillator*).

Angle Bending

The deviation of valence angles from their reference values (Figure. 7.3) can be described by a harmonic potential:

$$V_A(\theta) = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{7.15}$$

The force constant k_θ and the reference value θ_0 depend on the chemical type of atoms forming the angle. Force constants are smaller than in the case of stretching or compressing bonds because less energy is required to distort an angle away from equilibrium.

The bond-stretching and angle-bending terms describe the deviatioin from an ideal geometry. In a way, they are penalty functions and their sum should be close to zero in an optimized structure.

Torsional Terms

Tortional terms represent the torsion angle potential function which models the presence of steric barriers between atoms separated by three covalent bonds. In conformational analysis, chemical bonds rotation barriers are essential to the understanding of the system.

Most force fields use explicit tortional potentials with contributions from every four bound atoms (Figure 7.3). Torsional potentials are often expressed as a cosine series expansion:

$$V_T(\phi) = \frac{1}{2} k_\phi.[1 + \cos(n\phi - \delta)] \tag{7.16}$$

k_ϕ is referred to as the barrier height and its value gives qualitative indication of the relative barriers to rotation. For example, k_ϕ for an amide bond (A-C=N-D) will be larger than that of a bond between two sp^3 carbon atoms (A-C-C-D). n is the periodicity, it gives the number of minimum points in the function as the bond is rotated through 360°. δ is the phase factor and sets the minimum energy angle.

Improper Torsions / Out-of-Plane Bending

For chemical groups that are made up of four or more atoms in a plane, it is advantageous to enforce their planarity using an additional term in the force field. For example, the oxygen of a cyclobutanone molecule remains in the plane of the ring since π-bonding energy is maximized in a planar configuration. A force field using the standard terms described previously would have the oxygen out of the plane. The desired geometry can be obtained by adding an out-of-plane bending term to the force field. It can be acheaved by trating the four atoms as an improper torsion angle or by calculating the angle between a bond from the central atom and the plane defined by the central atom and the other two atoms (Figure 7.3). A value of 0°corresponds to a planar arragement. Out-of-plane

coordinates can be modeled using the following harmonic potential:

$$V(\alpha) = \frac{1}{2} k_\alpha \times \alpha^2 \tag{7.17}$$

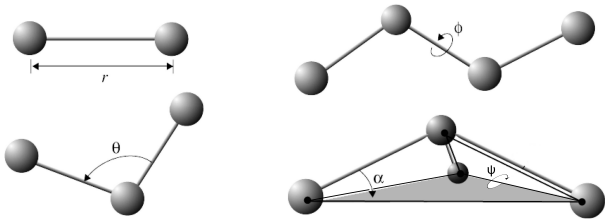


Figure 7.3: Internal coordinates for bonded interactions: r governs bond stretching; θ represents the bond angle term; ϕ gives the dihedral angle defined as the angle between the planes formed by two sets of three successive atoms; the small out-of-plane angle α is governed by the so-called "improper" dihedral angle ψ The improper dihedral term is designed to maintain planarity about certain atoms. The potential is described by a harmonic function. α is the angle between the plane formed by the central atom and two peripheral atoms and the plane formed by the peripheral atoms only. (*Internal coordinates for bonded interactons*).

7.1.3 Minimization algorithms

In computational chemistry, energy minimization reduces the energy of the system to the lowest possible point, by finding the correct arrangement of atoms in space. There are different methods used for calculating the lowest-energy conformation of a given structure. The Steepest Descent and Adopted Basis Newton-Raphson methods are the most commonly utilized with CHARMM.

Steepest Descent (SD)

The steepest descent is one of the simplest minimization algorithms, in which the coordinates are adjusted in the negative direction of the gradient iteratively. The step size, that determines how far the coordinates are shifted at each step, is the only adjustable parameter.

The step size is adjusted depending on whether a step results in a lower energy. If the energy drops, the step size is increased by 20% to accelerate the convergence. If the energy rises, overshooting a minimum, the step size is halved.

Adopted Basis Newton-Raphson Method (ABNR)

The adopted basis Newton-Raphson method performs energy minimization using a Newton-Raphson algorithm applied to a subspace of the coordinate vector spanned by the

displacement coordinates of the last positions. The second derivative matrix is numerically constructed from the change in the gradient vectors, and is inverted by an eigenvector analysis which avoids saddle points in the energy surface. At each step the residual gradient vector is calculated and used to add a steepest descent step onto the Newton-Raphson step, incorporating new direction into the basis set. This method is considered to suit most circumstances.

7.1.4 Non-bonded interactions

In addition to the bonded terms of the energy function described above, non-bonded terms are used to describe the energy between atoms which are not bonded together and consist at least of the electrostatic and the van der Waals interactions in the system. For atom separated by 3 bonds, nonbonded interactions are generally non-zero but are reduced compared to general nonbonded interactions (special 1-4 terms) involving atoms separated by 4 bonds or more (or from different molecules).

Electrostatic Interactions

Interactions between atoms due to their permanent dipole moments are described by treating the charged portions as point charges. The Coulomb potential is used for point charges to estimate the forces between charged atoms. It is given by:

$$V_E(i,j) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \tag{7.18}$$

where r_{ij} is the distance between q_i and q_j, the electric charges in coulombs carried by atom i and j respectively, and ϵ_0 is the electrical permittivity of space.

Van der Waals Interactions

The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The repulsive force is present at short distances where the electron-electron interaction is strong (Pauli's exclusion principle). The attractive force, also referred to as London dispersion forces, arises from fluctuations in the charge distribution in the electron clouds. A fluctuation in electron distribution of one atom creates an instantaneous dipole which induces a dipole in a neighbouring atom giving rise to an attractive interaction. These two effects are significant as the distance between atoms decreases since at infinite atomic separation both are equal to zero. The attractive interaction acts at longer ranges than the repulsive interaction. However, as the distance decreases the repulsive interaction becomes dominant giving rise to a minimum (see Figure 7.4).

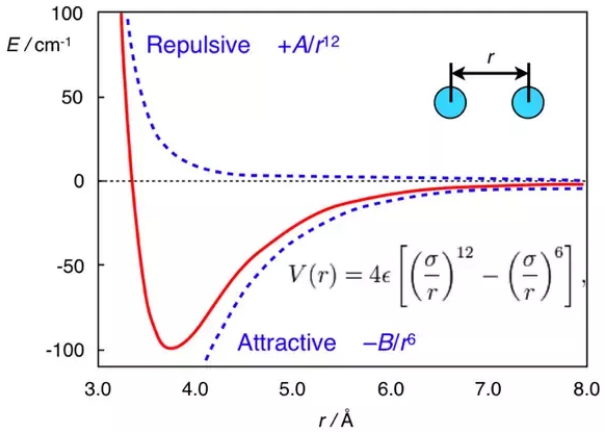


Figure 7.4: The Lennard-Jones potential. The repulsive r^{-12} term describes Pauli's repulsion at short ranges due to overlapping electron orbitals. The long range attractive term r^{-6} describes attraction at long ranges (van der Waals forces) (*Hanssen, Docking redone*).

In a force field, like CHARMM, the van der Waals potential is modeled using an empirical expression. The most widely used is the Lennard-Jones 12-6 function, which takes the following form for the interaction between two atoms:

$$V_{\text{vdw}}(r) = 4\epsilon. \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{7.19}$$

where σ is the collision diameter and ϵ the well depth (see Figure 7.4). Because of the large number of van der Waals interactions that must be determined in a system, especially large ones, the 12-6 potential is widely used since it is not computationally expensive. The r^{-12} term can be found by squaring the r^{-6} term that can be calculated from the square of the distance. Depending on the force field, other powers are also used (Halgren, 1992).

In molecular dynamics, the calculation of non-bonded terms in the potential energy function are the most time consuming task. Ideally, they are to be calculated for each pair of atoms, making their number increase as the square of the number of atoms in the model.

To avoid this computational penalty, two approaches are applied. In the first, a cut-off distance is defined above which the interactions between two atoms are ignored (truncation, shift or switch methods). The second approach reduces the number of interaction sites by merging some atoms into the atoms to which they are bonded (united atom method), this is usually applied to hydrogen atoms.

7.1.5 Periodic boundary conditions

Periodic boundary conditions are used to approximate an infinite system by using a small part called a unit cell. The minimum-image convention is used where short-range

non-bonded interaction terms are considered for the nearest image of each particle.

In molecular modelling the size of the unit cell is important to prevent artefacts. If the box is not large enough, a macromolecule might interact with its own image.

7.1.6 Statistical mechanics

Statistical mechanics is used to convert microscopic information, like molecular positions and motions gathered from molecular dynamics simulation, to macroscopic observables such as pressure, energy and heat. For a better understanding some definitions are reviewed here (McQuaerrie, 2000):

Definitions

The mechanical or microscopic state of a system is defined by the atomic positions x_i and the momenta $p_i = m_i v_i$. They can be considered as a multidimensional space, called phase space with 6N coordinates, for which they both contribute 3N coordinates.

The thermodynamic or macroscopic state of a system is defined by a set of parameters, like temperature, pressure and number of particles, that describes all its thermodynamic properties.

An ensemble is the collection of all possible systems which have different microscopic states but have the same macroscopic or thermodynamic state. Examples for ensembles with different characteristics are: NVE, NVT, NPT, μVT, (E = total energy, P = pressure, V = volume, μ = chemical potential)

Ensemble Averages and Time Averages

The measured thermodynamic properties of a macroscopic sample reflect the finite different conformations of the system present in this sample. Because an ensemble is the complete collection of microscopic systems, a sufficiently big sample can be seen as good approximation to an ensemble. This is why statistical mechanics defines averages corresponding to experimentally measured thermodynamic properties as ensemble averages (McQuaerrie, 2000). The ensemble average is given by:

$$\langle A \rangle_{ensemble} = \iint d\vec{p}^N d\vec{x}^N A\left(\vec{p}^N, \vec{x}^N\right) \rho\left(\vec{p}^N, \vec{x}^N\right) \tag{7.20}$$

where $\langle A \rangle$ is the measured observable, which is stated as a function of the momenta $\vec{p}_i$ and the positions $\vec{x}_i$. Quantity $\rho\left(\vec{p}^N, \vec{x}^N\right)$ is the probability density for the ensemble and the integration is performed over all momenta and positions of the system $d\vec{p}^N$, $d\vec{x}^N$. The ensemble average is the average value of an observable weighted with its probability.

In a MD simulation, an extremely large number of conformations is generated sequentially in time. To calculate an ensemble average the simulation has to cover all possible conformations corresponding to the ensemble, at which the simulation takes place. The time average is given by:

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(\vec{p}^N d\vec{x}^N\right) dt, \tag{7.21}$$

where τ is the simulation time. This expression is approximated by an average over a simulation performed for a sufficiently long period of time, representing a sufficient amount of phase space:

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(\vec{p}^N d\vec{x}^N\right) dt = \frac{1}{M} \sum_{i=1}^{M} A\left(\vec{p}^N d_M \vec{x}_M^N\right) \tag{7.22}$$

where M is the number of frames and $A\left(\vec{p}^N d_M \vec{x}_M^N\right)$ the values of the observable A in frame M. The idea is based on the Ergodic Hypothesis (McQuaerrie, 2000), which states that the time average equals the ensemble average.

In other words, simulating a system for a relatively long time, ensures that it will pass through an extremely high number of conformations, which can then be referred to as a representative subset of an ensemble.

Temperature and Pressure control in MD simulations

Protein simulations in solvents are run in a well defined statistical ensembles like NPT (constant number of particles, pressure and temperature). In MD simulations, the system has a specified total energy, a microcanonical ensemble (NVE = constant number of particles, volume and energy) is used to represent all its possible states. Additionally, the linear momentum $\vec{p}$ and the angular momentum L are conserved. When using periodic boundary conditions, the angular momentum is not conserved.

However, if the study focuses on the behaviour of the system at a specific temperature, an NVT (canonical ensemble) simulation using a thermostat is needed instead of the microcanonical ensemble. The temperature is controlled by multiplying the velocities at each time step by the factor:

$$\lambda = \sqrt{\frac{T_0}{T}}$$

where T(t) is the current temperature as calculated from the kinetic energy and T_0 is the desired temperature. One problem with this approach is, that it does not allow fluctuations in temperature which are present in the canonical ensemble.

Several algorithms have been developed to maintain Temperature and/or Pressure in a simulation, consistent with thermodynamic ensembles. Here, some of them are briefly cited.

Berendsen thermostat

In this method (Berendsen et al., 1984), the system is coupled to a thermostat that suppresses fluctuations of the kinetic energy of the system (trajectories are not consistent with the canonical ensemble). The temperature is corrected by a time constant τ which results in an exponential decay towards T_0:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}$$

In many applications, systems are initially equilibrated using the Berendsen scheme before calculating their properties with the Nosé-Hoover thermostat (Nose, 1984; Hoover, 1985), which correctly generates trajectories consistent with a canonical ensemble.

Nosé-Hoover temperature coupling

Originally introduced by Nosé and later developed by Hoover, the extended system method considers the heat bath as an integral part of the system. This is done by the addition of an artificial variable s, associated with a mass $Q > 0$ as well as a velocity p_s.Temperature fluctuations are influenced by the coupling between the reservoir and the system, determined by the value of Q. The Hamiltonian is given by:

$$H(P, R, p_s, s) = \sum_i \frac{\mathbf{p}_i^2}{2ms^2} + \frac{1}{2}\sum_{ij, i\neq j} U\left(\mathbf{r_i} - \mathbf{r_j}\right) + \frac{p_s^2}{2Q} + gkT\ln\left(s\right) \tag{7.23}$$

where g is the number of independent momentum degrees of freedom of the system, R and P represent all coordinates r_i and p_i. The coordinates R, P and t in this Hamiltonian are virtual. They are related to the real coordinates (with $'$):

$$R' = R,\ P' = \frac{P}{s} \text{ and } t' = \int^t \frac{\mathrm{d}\tau}{s}$$

Langevin dynamics

For simulations in implicit solvent, the Langevin equation is used, which derives from Newton's second law $F = m\ddot{x}$ where the force F on a particle consists of three parts 1) the external force F_e described by the force field, 2) a friction force F_f due to interaction with the medium and a molecular force F_m which characterizes the molecular bombardment of the particle by the solvent. With :

$$F_e(x) = -\nabla V(x)$$

For small velocities the friction force is given by stokes law and is linear in velocity

$$F_f = -\gamma\dot{x}$$

and F_m is random and corresponds to noise

$$F_m = R(t)$$

The resulting Langevin equation is

$$m_i \ddot{x}_i = F_e(x_i) - m_i \gamma_i \dot{x}_i + R_i(t)$$

γ is the viscosity, T is the temperature, k_B is Boltzmann's constant, and $R(t)$ is a Gaussian stochastic process with zero-mean, satisfying:

$$\langle R(t)R(t')\rangle = 2m_i\gamma_i k_{\mathrm{B}} T\delta(t-t')$$

where δ is the Dirac delta.

Langevin dynamics allows temperature to be controlled like with a thermostat, thus approximating the canonical ensemble.

The friction constant γ determines the strength of the coupling to the heat bath, and its inverse, τ, can be interpreted as a velocity relaxation time analogous to that used to choose the thermostat masses in the extended system methods.

Pressure control

The Berendsen barostat (Berendsen et al., 1984) adds an extra term to to the equations of motion which effects the pressure change:

$$\left.\frac{dP}{dt}\right|_{\mathrm{bath}} = \frac{P_0 - P}{\tau_P}$$

where P_0 is the reference pressure and P is the instantaneous pressure. τ_P is a time constant.

Using this method, the box sides and coordinates are rescaled every step. Assuming the system is isotropic and within a cubic box the scaling factor μ is given by:

$$\mu = 1 - \frac{\kappa_T \Delta t}{3\tau_P}(P_0 - P)$$

where κ_T is the isothermal compressibility.

Another method can be used where an extended dimension of Nosé-Hoover thermostat is applied to pressure to give a Nosé-Hoover barostat. The Parrinello-Rahman barostat (Parrinello et al., 1981), used in our GROMACS simulations, extends this by making each unit vector of the unit-cell independent. This way, the volume is variable during the simulation, as in the Nosé-Hoover barostat, allowing a dynamic shape change.

7.1.7 Implicit solvation

Implicit solvatation, or continuum solvatation (Roux et al., 1999), is a method used to represent solvent as a continuous medium instead of individual explicit molecules. It is often applied to the study of free energy of solute-solvent interactions in biological macromolecules. Its main interest is to save computational time usually "wasted" on recalculating the dynamics of water in explicit simulations.

In liquids, this solvation model is justified since the potential of mean force approximates the averaged behaviour of highly dynamic solvant molecules. Biological macromolecules or lipid membranes, can also be considered as continuous media with specific solvation properties. However, they are not necessarily uniform, since their properties can be described by analytical functions, such as "polarity profiles" of lipid bilayers (D, 2001).

The solvation free energy of a solute in conformation (X) can be written as:

$$\Delta G(X) = \Delta G_{np}(X) + \Delta G_{elec}(X) \tag{7.24}$$

with $\Delta G_{np}(X)$ being the free energy going from nothing to the non polar solute and $\Delta G_{elec}(X)$ the free energy for going from the non polar state to the polar solute.

Accessible surface area based method ; $\Delta G_{np}(X)$

The accessible surface area (ASA) method is based on experimental linear relations between Gibbs free energy of transfer and the surface area of a solute molecule (Richards, 1977). It directly uses the free energy of solvation to represent the solvent continuum, allowing for improvments in computational speed and decreasing errors in statistical averaging resulting from incomplete sampling of solvent conformation (Roux et al., 1999).

The free energy of solvation of a non polar solute molecule in the simplest ASA-based method is given by:

$$\triangle G_{np\ solv} = \sum_i \sigma_i ASA_i \tag{7.25}$$

where ASA_i is the accessible surface area of atom i, and σ_i is the solvation parameter of atom i determined experimentally.

Generalized Born ; $\Delta G_{elec}(X)$

The Generalized Born (GB) model (Constanciel et al., 1984; Still et al., 1990), is an approximation to the linearized Possion-Boltzmann equation. It generelizes the Born model used to calculate the free energy of placing a charge q at the center of a cavity in

a solvent of a dielectric constant ϵ. The model has the following form:

$$G_s = \frac{1}{8\pi}\left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon}\right)\sum_{i,j}^{N}\frac{q_i q_j}{f_{GB}} \tag{7.26}$$

where

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}} \tag{7.27}$$

and

$$D = \left(\frac{r_{ij}}{2a_{ij}}\right)^2 \quad a_{ij} = \sqrt{a_i a_j} \tag{7.28}$$

where ϵ_0 is the permittivity of free space, ϵ is the dielectric constant of the modeled solvent, q_i is the electrostatic charge on particle i, r_{ij} is the distance between particles i and j, and a_i is the effective Born radius (Still et al., 1990) that characterizes the burial of the atom inside the solute, its accurate estimation is critical for GB models (Onufriev et al., 2002).

GBSA

GBSA is communly used for implicit solvent models. It is basically a GB model augmented with the hydrophobic solvent accessible surface area (SA) term. In molecular mechanics this method is known as MM/GBSA.

Ad-hoc fast solvation models

Another method is to use ad-hoc quick strategies to estimate solvation free energy. One way to do so is by calculating the accessible surface area per atom through scaling the contribution to solvation of each atom type ("ASA-based model" described above) (Eisenberg et al., 1992).

Another method, EEF1 is based on a Caussian-shaped solvent exclusion and is implemented in the CHARMM19 force-field (Lazaridis et al., 1999). The solvation free energy is:

$$\Delta G_i^{solv} = \Delta G_i^{ref} - \sum_j \int_{Vj} f_i(r)dr \tag{7.29}$$

The reference solvation free energy of i corresponds to a suitably chosen small molecule in which group i is fully solvent-exposed. The integral is over the volume V_j of group j and the summation is over all groups j around i. EEF1 additionally utilizes a distance-dependent dielectric, and ionic sidechains of proteins are neutralized. This model was augmented with the hydrophobic effect in Charmm19/SASA (Ferrara et al., 2002), and is detailed in the next section.

7.1.8 Free energy calculations and potential mean force

The free energy of a system is the most important thermodynamic quantity directly related to experimental measurements. The definition of Helmholtz free energy is a direct application of the equation 7.20:

$$A = k_B T \; ln \iint dp^N dx^N exp[-\beta H(p^N, r^N)] \tag{7.30}$$

where H is the Hamiltonian. in practice, it is very difficult to evaluate this integral since high energy regions are not thouroughly sampled. Instead what is done is to calculate free enregy differences. The free energy difference between two states X and Y is:

$$\Delta A = -k_B T \; ln\langle exp[-\beta(H_X - H_Y)]\rangle_X \tag{7.31}$$

Where the ensemble average is taken, for instance, over state X and state Y becomes a "perturbation" of state X. State X is one point along the reaction coordinate which is often defined as a simple combination of inter- or intramolecular coordinates (distance, angle ...). In our case for example, we are interested in the binding of Fur proteins to DNA sequences or the interaction between Fur subunits in tetrameric complexes and the reaction coordinate is a distance between centers of mass. The free energy surface along the chosen coordinate is referred to as the potential of mean force (PMF). If the system of interest is in a solvent, the PMF also incorporates the solvent effects.

Weighted Histogram Analysis Method

The Weighted Histogram Analysis Method was developed by Ferrenberg and Swendsen (Ferrenberg et al., 1989). WHAM estimates the difference in free energy by considering several intermediate states along the reaction coordinate. Using a discrete number of states, WHAM creates an histogram that represents the probability of observing a specific state.

The WHAM method is shortly described by the two following equations:

$$P^b(\xi) = \frac{\sum_{i=1}^{N_{wind}} n_i(\xi)}{\sum_{i=1}^{N_{wind}} N_i exp\left((\mathcal{C}_i - U_{bias,i}(\xi))/k_\mathrm{B}T\right)} \tag{7.32}$$

$$\mathcal{C}_i = -k_\mathrm{B}T ln\left[\sum_{\xi_{bins}} P^b(\xi) exp\left(-U_{bias,i}(\xi)/k_\mathrm{B}T\right)\right] \tag{7.33}$$

where N_{wind} is the number of windows, $n_i(\xi)$ is the number of counts in histogram bin associated with ξ.

The usual technique for PMF determination is the Umbrella Sampling (Patey et al., 1975; Crouzy et al., 1994), which is a way to correctly overlap all ΔA between all states along a reaction coordinate.

Umbrella Sampling

Umbrella Sampling, developed by Torrie and Valleau (Torrie et al., 1974), uses an additional energy term, a bias, to ensure a sufficient sampling along the whole reaction coordinate. This can be achieved by one continuous simulation or by different simulations (windows), the distribution of which overlap. Connecting energetically seperated regions in phase space through the biasing potential gave rise to the name umbrella sampling.

When using different simulation windows, the idea is to devide the reaction path and constrain the system in each on of them using the biasing potential. Then the free energy is determined, corrected for the bias and all the windows are pasted together.

Theoretically, the free energy along a reaction coordinate ξ is given by:

$$\mathcal{W}_{(\xi=\xi_1)} = -\mathrm{k_B T} \ln \mathrm{P}_{(\xi=\xi_1)} = -\mathrm{k_B T} \ln \frac{\int \cdots \int \mathrm{e}^{-\beta \mathrm{E}(\mathbf{R})} \delta(\xi_{\mathbf{R}} - \xi_1) \mathrm{d}\mathbf{R}}{\int \cdots \int \mathrm{e}^{-\beta \mathrm{E}(\mathbf{R})} \mathrm{d}\mathbf{R}} \tag{7.34}$$

where $P_{(\xi=\xi_1)}$ corresponds to the probability to find the reaction coordinate near a given value, ξ_1, and β is equal to $\frac{1}{k_B T}$. To enhance the sampling in the neighborhood of a reference value ξ_0, a biasing harmonic potential centered on ξ_0, defined by $u_i(\xi) = \frac{k}{2}(\xi - \xi_0)^2$ is introduced. The unbiased potential of mean force is then:

$$\mathcal{W}_{i(\xi_1)} = -\mathrm{k_B T} \ln \mathrm{P}^{\mathrm{b}}_{\mathrm{i}(\xi=\xi_1)} - \mathrm{u_i}(\xi_1) + \mathcal{C}_\mathrm{i} \qquad \mathcal{C}_\mathrm{i} = \mathrm{k_B T} \ln \left\langle \mathrm{e}^{-\beta \mathrm{u_i}(\xi_1)} \right\rangle \tag{7.35}$$

where $\mathrm{P}^b_{i(\xi=\xi_i)}$ is the biased probability function of ξ under the biasing potential, $u_i(\xi)$. This derivation is exact, no approximation enters apart from the assumption that the sampling in each window is sufficient. This is facilitated by an appropriate choice of umbrella potentials $u_i(\xi)$.

The constant terms $\mathcal{C}_i$ can be calculated by combining the results from different simulation windows using the WHAM method (Kumar et al., 1992; Crouzy et al., 1994).

Once the full profile is obtained, the final free energy of binding is given by $\Delta G = PMF_{max} - PMF_{min}$, where PMF_{max} is estimated from the plateau obtained for large values of the reaction coordinate and PMF_{min} is the exact minimum of the PMF. The corresponding error is then:

$$\delta(\Delta G) = \sqrt{\delta(PMF_{min})^2 + \delta(PMF_{max})^2} \tag{7.36}$$

Error analysis on free energy profiles

Statistical errors on WHAM free energy profiles may be estimated with bootstrap analysis (Efron, 1979). Bootstrapping is a resampling technique applied to estimate the uncertainty of a quantity $Q(q_1, ..., q_n)$ which is computed from a large set of n observations q_k (k=1, ..., n). To calculate the uncertainty in Q, n observations can be made multiple times, giving several independent estimates for Q. The observations q_k are typically drawn

from an unknown probability distribution $P(q)$. The idea of boot-strapping is to estimate $P(q)$ using the n observations (the biased histograms) and subsequently generate new random sets of n hypothetical observations, based on the estimated distribution. Each of the sets of n hypothetical observations is used to calculate a hypothetical value for Q.

The WHAM procedure computes the PMF based on the N_w trajectories $\xi(t)$ along the reaction coordinate, each taken from one of the umbrella windows $i = 1, ..., N_w$. All positions i during the N_w simulations may thus be considered as the large set of observations, which were referred to as a k in the previous paragraph.

The probability distributions of ξ are available as the umbrella histograms. A new hypothetical observations can be generated, that is a "bootstrapped" trajectory b, $\xi_{b,i}(t)$ for each umbrella histogram $h_i(\xi)$, such that $\xi_{b,i}(t)$ is distributed according to the respective histogram. Each bootstrapped trajectory $\xi_{b,i}(t)$ yields a new histogram $h_{b,i}(\xi)$. The new set of N_w histograms $h_{b,i}$ is subsequently applied in WHAM to compute a bootstrapped PMF $\mathcal{W}_{\lfloor}(\xi)$. The whole procedure is repeated N_b times, typically 100, yielding a large set of N_b bootstrapped PMFs $\mathcal{W}_{b,l}(\xi), (l = 1, ..., Nb)$. The uncertainty of the PMF is then given by the standard deviation of the $\mathcal{W}_{b,l}$.

The umbrella histograms are considered as independent data points and the bootstrap is carried out by assigning random weights to the histograms ("Bayesian bootstrap"). Another approach is to generate new random trajectories, such as the generated data points that are distributed according to the given histograms: this is done by calculating the autocorrelation time (ACT) for each window (Kumar et al., 1992). The normalized autocorrelation function of umbrella window i is given by

$$AC_i(\delta t) = \frac{\langle (\xi_i(t) - \langle \xi \rangle)(\xi_i(t + \delta t) - \langle \xi \rangle) \rangle}{\sigma^2_{\xi,i}}$$

where $\xi_i(t)$ is the reaction coordinate during simulation i, $\sigma^2_{\xi,i} = \langle (\xi_i(t) - \langle \xi_i \rangle)^2 \rangle$ is the variance and $\langle ... \rangle$ represents the average over the simulation frames.

Following Kumar *et al.* (Kumar et al., 1992), the integrated autocorrelation time (IACT) of window i is defined by

$$\tau_{i,int} = \sum_{\delta t=1}^{\infty} AC_i(\delta t)$$

This sum is stopped, in practice, after $AC_i(\delta t)$ drops under a defined value, 0.05 for instance, to facilitate convergence. For correlated data the variance of the mean is by the factor $g = 1 + 2\tau_{int}$ larger than the corresponding naive variance for uncorrelated data σ^2/N.

7.2 GROMACS software and the GROMOS force field

GROMACS (GROningen MAchine for Chemical Simulations) is a molecular dynamics package originally developed in the Biophysical Chemistry department of University of Groningen Bekker et al., 1992; Van Der Spoel et al., 2005; Abraham et al., 2015. We used GROMacs to compute and analyse all the PMFs in this work because of its high speed through the use of sophisticated algorithms and its implementation on GPUs.

GROMACS supports all-atom, united atom and coarse-grained force-fields, in particular, the AMBER (Cornell et al., 1995), CHARMM, GROMOS and OPLS (Jorgensen et al., 1988) force fields. The GROMOS force field used in this work will be detailed in this section.

The GROMOS force fields are united atom force fields, without explicit aliphatic (non-polar) hydrogens. We used the recent 54A7 force field (Schmid et al., 2011), which derives from the previous 53A6 (Oostenbrink et al., 2004). The default MD integrator in GROMACS is the leap-frog algorithm. Bonded energy terms are similar to those used in the CHARMM force field except for the Covalent Bond-Angle Interactions which take the form:

$$V^{angle}(\vec{r};s) = V^{angle}(\vec{r};K_\theta,\theta_0) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} \left[cos\theta_n - cos\theta_{0_n}\right]^2 .$$

with θ_n being the actual value of the nth angle defined by atoms i, j, k. K_θ, and θ_0 are defined over bond-angle types.

$$K_{\theta_n} = \frac{2k_\mathrm{B}T}{\left[cos(\theta_{0_n} + \sqrt{k_\mathrm{B}T/K_{\theta_n}^{harm}} - cos\theta_{0_n}\right]^2 + \left[cos(\theta_{0_n} - \sqrt{k_\mathrm{B}T/K_{\theta_n}^{harm}} - cos\theta_{0_n}\right]^2}$$

7.2.1 Non-bonded interactions

In the GROMOS force field, nonbonded interactions are calculated over pairs of nonbonded atoms. As in the case of the CHARMM force field, not all atom pairs are included in this sum and it is generally calculated over a subset of atom pairs that have an interatomic distance shorter than a specific cutoff distance.

Three different pair distances are defined in GROMOS. Short cutoff length, R_p, are evaluated at every step in the simulation as well as pairs that are separated by distances between R_p and a long range cutoff length R_l. A reaction-field contribution from a homogeneous dielectric or ionic medium outside this large cutoff may be taken into account as a third, long-range electrostatic contribution but has not been used in this work.

The nonbonded van der Waals interactions are calculated as a classical sum over all interacting nonbonded atom pairs using a Lennard-Jones 12/6 interaction function with

parameters C12 and C6. The electrostatic interactions include the classical Coulomb potential.

For images created by the periodic boundary conditions GROMACS uses the minimum image convention. Only the nearest image of each particle is used for short-range non-bonded interaction terms. For long-range electrostatic interactions GROMACS uses lattice sum methods such as Ewald sum PME (Essmann et al., 1995) and the hybrid Particle-Particle/Particle-Mesh model (PPPM).

Long Range Electrostatics

The total electrostatic energy of N particles and their periodic images is given by:

$$V = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{\vec{r_{ij}}.\vec{n}}$$

where $f = 1/(4\pi\epsilon_0) = 138.935458$, $(n_x, n_y, n_z) = \vec{n}$ is the box index vector, the star indicates that terms with $i = j$ should be omitted when $(n_x, n_y, n_z) = (0, 0, 0)$. The distance r_{ij} is the real distance between the charges and not the minimum-image. This sum is conditionally convergent, but very slow.

This method was optimized by the Ewald summation that was first introduced as a method to calculate long-range interactions of the periodic images in crystals (Ewald, 1921). The idea is to convert the single slowly-converging previous sum into two quickly-converging terms, one in direct and one in reciprocal Fourier space, in addition to a constant term. Unfortunately, the computational cost of the reciprocal part of the sum increases as N^2 and it is therefore not realistic for use in large systems.

Particle-mesh Ewald

Proposed by Tom Darden (Darden et al., 1993), the Particle-mesh Ewald (PME) is a method used to improve the performance of the reciprocal sum. Instead of directly summing wave vectors, the charges are assigned to a grid using interpolation. The PME algorithm is now included in all MD packages. Its implementation in GROMACS uses B-spline interpolation referred to as smooth PME (SPME). The grid is then Fourier transformed with a 3D FFT algorithm and the reciprocal energy term obtained by a single sum over the grid in k-space. The PME algorithm scales as $Nlog(N)$, and is faster than ordinary Ewald summation on medium to large systems.

As an example for using Particle-mesh Ewald summation in GROMACS, the following should be specified in the parameter $*.mdp$ file:

coulombtype = PME
rvdw = 0.9

rlist = 0.9
rcoulomb = 0.9
fourierspacing = 0.12
pme-order = 4
ewald-rtol = 10^{-5}

In practice, the PME mesh terms are named "Coulomb reciprocal" and the short-range interactions (contained within rcoulomb, calculated by a modified switch potential) are named "Coulomb (SR)".

When Ewald summation or particle-mesh Ewald is used to calculate the long-range interactions, the short-range Coulomb potential is also modified. In this case, the short range potential is given by :

$$V(r) = f\frac{\mathrm{erfc}(\beta r_{ij})}{r_{ij}}q_i q_j$$

where β is a parameter that determines the relative weight between the direct space sum and the reciprocal space sum and erfc(x) is the complementary error function.

Long Range Van der Waals interactions Dispersion correction

Long Range Van der Waals interactions Dispersion correction (Shirts et al., 2007) are used in GROMACS. The assumption is made that the cut-off is long enough so that the repulsion is neglected, therefore only taking the dispersion term into account. The interaction energy between any two particles diminishes as r^{-6}, however, a single particle interacts with all of the other particles in a system. While the energy correction is usually small, it may be important for free energy calculations where differences between two different Hamiltonians are considered. In contrast, the pressure correction is very large and can not be neglected under any circumstances where a correct pressure is required, especially for any NPT simulations.

In the mdp file the option is:
; Long-range dispersion correction
DispCorr = EnerPres; Apply long range dispersion corrections for Energy and Pressure

Indeed, although the interaction energy between any two particles diminishes as r^{-6}), a single particle interacts with all of the other particles in a system. Thus, the neglected portion of this interaction energy involves an integral over all space where $r > R_c$, and hence (with uniform particle density) diminishes only as R_c^{-3}. While the energy correction is usually small, it may be important for free energy calculations where differences between two different Hamiltonians are considered. In contrast, the pressure correction is very large and can not be neglected under any circumstances where a correct pressure is required, especially for any NPT simulations.

In the mdp file the option is:
; Long-range dispersion correction
DispCorr = EnerPres; Apply long range dispersion corrections for Energy and Pressure

GROMACS temperature and pressure coupling

For temperature coupling, both the Berendsen-thermostat and Nose-Hoover Temperature coupling using a Nose-Hoover extended ensemble are implemented GROMACS. Similarly, two methods are implemented for pressure control and they are the Berendsen exponential relaxation pressure coupling and Parrinello-Rahman Extended-ensemble pressure coupling.

7.2.2 Potential of mean force calculation

In this work, potential of mean force are calculated for Fur dimer-dimer or dimer-DNA interactions. The system is built so that the reaction coordinate is the x-axis. To generate the configurations, we must pull the moving sub-system away from the fixed sub-system. We must capture enough configurations along the reaction coordinate to obtain regular spacing of the umbrella sampling windows, in terms of center-of-mass distances between the sub-systems.

Options in the mdp file for umbrella sampling of Fur dissociating from DNA would look like :

; COM PULLING
pull = yes
; Group and coordinate parameters
pull-group1-name = Protein
pull-group2-name = DNA
pull-coord1-type = umbrella; Center of mass pulling using an umbrella potential between the reference group and one or more groups
pull-coord1-geometry = distance; Pull along the vector connecting the two groups. Components can be selected with pull-coord1-dim.
pull-coord1-start = yes; Add the COM distance of the starting conformation to pull-coord1-init
pull-coord1-init = 0 ; The reference distance at t=0.
pull-coord1-rate = 0.0; [nm/ps] The rate of change of the reference position.
pull-coord1-k = 500; The harmonic force constant in kJ mol^{-1} nm^{-2}
pull-print-com yes; Print the COM of all groups for all pull coordinates
pull-nstxout (50); Frequency for writing out the COMs of all the pull groups (0 is never)
pull-nstfout (50); Frequency for writing out the force of all the pulled groups (0 is never)

Free energy profiles for the extraction (by translation along a fixed direction: Ox) of one Fur dimer from the tetramer (dimer of dimers) and of Fur from DNA were computed using the translation protocol shown in (5.3). The simulations include a "moving" subsystem (for instance FtFur dimer, chains A and B) and a "fixed" subsystem (for instance FtFur dimer, chains C and D or DNA). The profiles were built using the umbrella sampling technique and result from the overlapping of 26 computation windows, one for each translation distance.

As a result of this initialization protocol, 26 structures $conf_i$ were generated yielding initial positions for the translated dimer prior to the potential of mean force calculation. Then the umbrella sampling calculation, itself, consisted of 26 repeats of :

1. Reading of initial structure $conf_i$ and reference structure $conf_0$ for harmonic restraints.

2. Running 100 ps NPT equilibration with position restraints on the "fixed" subsystem and distance restraints on the protein. The "moving" subsystem was subject to two harmonic biasing forces (umbrella potential) applied between the centers of mass of the 2 FUR dimer subunits and the center of mass of the "fixed" subsystem with force constants of 500 kJ mol^{-1} nm^{-2}.

3. Running around 50 ns NPT production simulations with same restraints and biasing potential. These simulations were concatenated from series of 1 to 2 ns runs allowing us to check the convergence of the PMFs. Harmonic biasing forces were applied along the X direction only (direction of the translation).

Position restraints were applied on backbone atoms of the "fixed" subsystem with a force constant of 10 kJ mol^{-1}Å $^{-2}$ in all directions. NOE type distance restraints were added to maintain the secondary structure of the protein (both "fixed" and "moving" parts in the case of the tetramer). These restraints were applied between 1.8 and 2.0 Å with force constants of 20 kJ mol^{-1}Å $^{-2}$.

In both equilibration and production umbrella sampling MD simulations, the system was simulated under NPT conditions. Temperature was fixed at 310 K with Temperature coupling using a Nose-Hoover extended ensemble with time constant $\tau_T = 0.5$ ps, Pressure was controlled at 1 atm with extended-ensemble Parrinello-Rahman isotropic pressure coupling with $\tau_P = 1$ ps and compressibility $= 4.5 \times 10^{-5}$ bar^{-1} Parrinello et al., 1981. A time step of 2 fs was used.

After the dynamics runs, positions and forces were collected from the trajectories and the umbrella sampling harmonic potential was unbiased using the WHAM algorithm (Kumar et al., 1992) implemented in the g_{wham} program (Hub et al., 2010) to yield the free energy profiles.

Fitting of the free energy profiles

Data points corresponding to the outputs of wham can be fitted with a sum of sigmoid functions with R (R Core Team, 2013):

$$S(x) = a + b/(1 + exp((-c * (x - d))))$$

After a first fit with the raw data, the energy offset (a in equation) is subtracted from the final energy profiles.

WHAM equations

The WHAM equations are solved in GROMACS using the command *g_wham* (Hub et al., 2010). For statistical error estimation, four bootstrapping methods are supported and selected with *-bs-method* in *g_wham*.
(1) *b-hist* is the default: points in the time series are considered as independent, and the bootstrap is carried out by assigning random weights to the histograms ("Bayesian bootstrap").
(2) *hist*: again points are considered as independent. For each bootstrap, N_c histograms are randomly chosen from the N_g given histograms (allowing duplication).
(3) *traj*: The given histograms are used to generate new random trajectories, such that the generated data points are distributed according to the given histograms and properly autocorrelated. The autocorrelation time for each window is calculated via the *-ac* option.
(4) *traj-gauss*: The same as method traj, but the trajectories are not bootstrapped from the umbrella histograms but from Gaussians with the average and width of the umbrella histograms.

7.2.3 Computation of average interaction energy profiles

Interaction energy profiles were computed by extracting electrostatic potential energy and Lennard Jones potential energy, from all the trajectories of the simulation, between each residue in each chain (A or B) and the fixed fragment (DNA or fixed Fur dimer).

All interaction energies are obtained using the *gmx energy* command in GROMACS (version 2016.4). This command extracts energy components from energy files using an interactive interface to select the desired energy terms. To automate the process of energy term selection, several lists containing residues of interest are added to the index files (**.ndx*) and listed in the molecular dynamics parameters files (**.mdp*) using the *energygrps* term to define the interactions to be calculated (limited to 50 interaction pairs in one job).

Next *gmx grompp* and *gmx mdrun* are used to run the desired trajectory file using the *-rerun* option. At this step, *gmx energy* is called to write short-ranged non-bonded

potential energies to the energy file of specific pairs defined previously in the **.ndx* and **.mdp* files. After that, Lennard-Jones (LJ-SR) and electrostatic interaction (Coul-SR) are extracted and summed for each residue from each window.

It is worth mentioning that during the coding of these scripts, Justin Lemkul a GROMACS developer from Virginia Polytechnic Institute and State University, proposed the use of *energygrps* to speed up our calculations since we were, apparently, the first to look for an optimized way to screen for individual interactions. So if you are reading this in the not so near futur, make sure that *energygrps* still works, because we were told that it might be withdrawn from futur GROMACS versions.

7.3 CHARMM

The CHARMM program and force fields were already introduced and described in the previous sections. It was used in this work for all docking MD simulations and for its versatility and complete scripting language very useful for Fur-system model buildings.

7.3.1 CHARMM data structures

Data Structures include information about the molecule (composition, chemical connectivity, atomic properties, ...) used by the energy function. This information is contained in the topology file and the parameter file.

Residue Topology File (RTF)

The RTF contains local information about atoms, bonds, angles etc. for each protein residue for example. It defines the covalent structure of an atom by describing its connections to other protein residues.

Parameter File (PARAM)

The parameter File is associated with the RTF file as it contains all the necessary parameters for calculating the energy of the molecule. Information included in this file contains equilibrium bond lengths and angles for bond stretching, angle bending and dihedral angle terms in the potential energy function as well as the force constants and the Lennard-Jones 12-6 potential parameters.

Protein Structure File (PSF)

The PSF file gives the detailed composition and connectivity of a molecule, and describes its division into residues and molecular entities.

Coordinate File (CRD)

The CRD file contains the Cartesian coordinates of all atoms in the system usually obtained form experimental data (NMR, X-ray crystal structures). The *ICBUILd* command is often used to build missing atoms coordinates.

To illustrate the use of these files, here is an example of a minimization in CHARMM. The user creates an input file that reads the RTF and PARAM files then the PSF and CRD containing topology and coordinates.

```
! Read in Topology and Parameter files
OPEN UNIT 1 CARD READ NAME top_all27_prot_na.rtf READ RTF CARD UNIT
1
OPEN UNIT 1 CARD READ NAME par_all27_prot_na.prm
READ PARA CARD UNIT 1
! Read sequence from the PDB coordinate file
OPEN READ UNIT 27 CARD NAME full.psf
READ PSF CARD UNIT 27
OPEN READ UNIT 10 CARD NAME full.crd
READ COOR CARD UNIT 10
```

Hydrogen bond lengths can be fixed by the *SHAKeBONHPARAmeters* command to remove high frequency motion allowing the use of 2 fs time steps. The *MINI* command requests CHARMM to do a minimization, it should be followed by the algorithm to use and the maximum step number, *MINI SD NSTEp* 100 in the case of steepest descent with 100 steps.

7.3.2 EEF1 energy function

In our docking calculations, we use the key word INTE, to calculate interaction energies: a new energy term *ASP* appears which is the amount of solvation free energy that is excluded between the two atom selections. For example, the INTE between atom A and atom B will give the amount of solvation A loses due to B plus the amount B loses due to A.

The EEF1 energy function is a solvent exclusion model that provides an alternative formulation for the solvation free energy of a protein. It estimates the water effect on

a polypeptide from the solvation of each atom, modified by the presence of other solute groups that exclude solvent. The function is parametrized by experimental data on solvation and empirical corrections. Molecular dynamics simulations with the EEF1 water model are only 50% slower that simulations in *in vacuo*.

The solvation free energy of a given conformation $\vec{r}^M$, can be written as a volume integral of a density $f(\vec{r})$:

$$\triangle G_{solv} = \int_v f(\vec{r}) d\vec{r} \tag{7.37}$$

Solvent-solute and solvent-solvent interactions are described by the density $f(\vec{r})$. When a macromolecules changes its conformation, the solvation free energy of each group also changes, because the new conformation occupies a new volume and excludes the solvant from it, at the same time the solent density and the molecules orientation are modified in the remaining space. In EEF1, the solvation free energie for a polyatomic solute is given as a sum over atomic contributions:

$$\triangle G_{solv} = \sum_i \triangle G^i_{solv} \tag{7.38}$$

If the solvent exclusion effect is the only contributor to solvation energy, it can be writen:

$$\triangle G^i_{solv} = \triangle G^i_{ref} - \sum_j \int_{V_j} f_i(\vec{r}) d\vec{r} \tag{7.39}$$

where $\triangle G^i_{ref}$ is the reference solvation free energy, given by the solvation free energy of group i where the group is largely exposed to solvent. The volume V_j is occupied by group j and the sum runs over all groups j surrounding i. To simplify the calculation, the integral over V_j is approximated by the product $f_i(r_{ij})V_j$:

$$\triangle G^i_{solv} = \triangle G^i_{ref} - \sum_{j \neq i} f_i\left(r_{ij}\right) V_j \tag{7.40}$$

Therefore, for a group i, the free energy of solvation in a macromolecule is given by the reference value minus the reduction in solvation due to the surrounding groups. The function $f_i(r_{ij})$ is assumed to be Gaussian:

$$f_i(\vec{r}) = \frac{\alpha_i}{4\pi r^2} \exp\left[-\left(\frac{\mathrm{r} - \mathrm{R_i}}{\lambda_\mathrm{i}} \right)^2 \right] \tag{7.41}$$

where R_i is the van der Waals radius of atom i, λ_i corresponds to the length of the first solvation shell. α_i is related to the free energy of solvation of group i in isolation, $\triangle G^i_{free}$, given by the integral of $f_i(r)$ over the whole space, different from $\triangle G^i_{ref}$, derived from experimental measurements, which is affected by the presence of a small compound linked to the atom.

7.4 Docking with Autodock

in biology, docking simulations are often used in order to understand how a complex forms or where a ligand binds a specific receptor, enabling the users to state a hypothesis based on the high local affinity observed between ligand and receptor. In the search for therapeutic targets, docking simulations became a key element in the research pipeline for the optimization of leads and hits.

Usually, docking simulation are composed of two step, the exploration and scoring. The exploration step, enables the ligand to explore all accessible configurations on the target protein in order to find the complex with the lowest biding free energy. This is done in different ways, a systemic search uses the degrees of freedom of torsion angles to divide the ligand into fixed and rotable parts, that are used to reconstruct it progressively in the search for the best biding conformation. A stochastic search considers the ligand as a whole, where its rotation, translation and torsion angle variations are randomly generated. This random search can be done using Monte Carlo or genetic algorithms (both implemented in Autodock). The exploration step can also be achieved using deterministic exploration algorithms, where molecular dynamics and energy minimizations are used to find the best conformations used in turn to generate future conformations.

The second step in a docking simulation is the scoring step where previously generated conformations are evaluated to keep what is considered as the best. Scoring methods vary depending on the docking software used, however, they are all based on an energy function that takes into account all the interactions between the different partners.

In this work, Autodock 4.2 (Morris et al., 2009) was used for docking simulations, peptide structures were loaded as ligands and the number of 15 rotatable bonds were assigned using AutoDockTools. For each run of AutoDock 4.2, an initial population of 30 individuals was created, with an assignment of random torsions as specified in the peptide pdbqt file. The Lamarckian Genetic Algorithm (Morris et al., 1998) was run with a maximum of 6000 generations and 13 000 000 energy evaluations, an arbitrary maximum to ensure that the calculation was limited by the number of generations rather than the number of evaluations. The peptide with the lowest estimated free energy of binding was taken for further calculations. To compute the binding free energy, Autodock uses a semi-empirical force field based on protein-ligand complexes with known structures and binding energies (Morris et al., 2009).

7.5 Phylogeny: tree construction and software

Amino acid sequences or RNA sequences were aligned using Clustal Omega (Sievers et al., 2011). Gblocks (Castresana, 2000) was used on the aligned sequences to eliminate poorly aligned positions and divergent regions with a maximum number of contiguous

nonconserved positions equal to 8, minimum lenght of a block was set to 10 an no gaps were allowed. The generated blocks were used to build the phylogenetic trees in SeaView version 4.7 (Gouy et al., 2010) through the phyML method (Guindon et al., 2010), with 100 replicates, that uses Felsenstein's bootstrapping in phylogenetic trees (Felsenstein, 1985; Hillis et al., 1993).

Protein sequences were obtained from https://www.ncbi.nlm.nih.gov/protein/ and the 16S rRNA sequences were retrieved from the SILVA rRNA database on https://www.arb-silva.de/search/ as shown in Table 7.1.

Organism	GenBank ID	SILVA accession Nr
Pseudomonas aeruginosa	KXG13090.1	KX138389
Legionella pneumophila	PYB66067.1	JN983396
Francisella tularensis	KFJ77225.1	CP010287
Vibrio cholerae	KKP21436.1	X74695
Magnetospirillum gryphiswaldense	CAM76413.1	AM085146
Bacillus subtilis	P54574.2	GU339260
Escherichia coli	ADR26048.1	LFQP01000068
Yersinia pestis	CAL21256.1	AGKA01000172
Campylobacter jejuni	CAL34550.1	ANGY01000192
Helicobacter pylori	AVA26695.1	AP014710

Table 7.1: GenBank ID and SILVA rRNA database accession numbers used in this work.

7.6 Building structural models

CHARMM was used to build all initial models of Fur tetramer and Fur-DNA systems. Its scripting language was useful for building missing coordinates from internal coordinates (IC PARAm, IC BUILd), initial orientation and energy minimization of vacuum systems.

The GROMACS program version 5.1.2 (Abraham et al., 2015), with the gromos54a7 united atom force field (Schmid et al., 2011), was used to perform long Molecular Dynamics simulations needed to compute free energy profiles. Fe^{2+} and Zn^{2+} were modelled as simple Lennard Jones hard spheres with charge +2,further refered to as di-cations, with Zn coordinated to charged deprotonated cysteines. The pdb file corresponding to the X-ray structures were prepared for GROMACS with pdb2gmx.

7.6.1 FtFur complexes

FtFur Tetramer

The Xray structure of FtFur resolved recently in our laboratory (Asp 7 to Glu 138) was used as initial model for the tetramer (Pérard et al., 2018). All histidine residues were given type HISA with protonated Nδ1 atom. In all 4 protein subunits, cysteins 93, 96, 133 and 136 were deprotonated and given a total charge of -0.75 (-0.05 for Cα, -0.15 for Cβ, -0.55 for S). This charge can be compared to that of the CYS residue in gromos54a7 force field, set to -0.5. A value of -0.75 was adopted instead in this work high enough to stabilize a tetracoordinated Zinc ion providing correct orientation of the cysteins and proper metal environment during further simulations. (Use of a -0.5 charge led to the escape of the metal from its binding site). In addition to zinc, four Fe^{2+} ions were included in the simulation in FtFur site S2.

FtFur-dimer and DNA

In the absence of FtFur+DNA structure, the structure of MgFur in the presence of DNA was used to model the DNA Fur box and correctly position FtFur dimer on DNA (by least square fit matching of atom positions). The 5'-GCCGGATAATGATAATCATTATC-3' fragment and its complementary 3'-5' sequence was used, as described in the MgFur-Furbox paragraph above, to model double-stranded DNA.

Subunits A and B in the tetramer of FtFur from Asp 7 to Arg 137, in the presence of 2 Zn^{2+} and 2 Fe^{2+} ions, were chosen for this Fur/DNA simulation. FtFur was superimposed on MgFur already bound to the Fur box using least square fit of matching atom positions. Backbone atoms from corresponding structured parts (helices and sheets) of the two protein were selected for this structure superposition for a final rms difference of 4 Å. Then the FtFur/DNA complex was created by concatenating the aligned FtFur dimer and the MgFur DNA box and rapidly energy minimized to release bad contacts.

7.6.2 PaFur complexes

PaFur-Tetramer

The model of the PaFur tetramer was built from PDB structure 1MZB (Pohl et al., 2003) with the PISA program (Krissinel et al., 2007). The PISA tetramer structure with largest internal free energy (-304.5 kca/mol) was selected and 8 zinc cations were constructed with CHARMM in the metal sites described by Pohl (Pohl et al., 2003).

PaFurΔS3-Tetramer

A model of the PaFur tetramer without S3 metal site by mutating Histidine 86 and Hisitidine 124 into Alanine was obtained from the crystal structure recently resolved in our laboratoryi in the presence of zinc. Again, a PISA tetramer was selected with 8 Zn cations chosen from the 21 initial metal atoms in the PISA model. The zinc cations fill in the S2 metal site (H32, E80, H89 and E100) and the incomplete S3 site reduced to D88 and E107. The model includes residues Met 1 to Lys 132 (the last 2 residues being unrsolved in the X-ray structure).

7.6.3 VcFur tetramer

Although it does not form in the case of VcFur, a tetramer was built from the structure of PaFur, as a negative control to this study. VcFur contains 150 amino acids from Met 1 to Lys 150 but the crystal structure was resolved for residues Asp 3 to Cys 133, only. A VcFur dimer model (chains A, B) extending from residues Met 1 to Cys 133 was built and energy minimized with CHARMM from PDB structure 2W57 (Sheikh et al., 2009). Coordinates for Met 1 and Ser 2 were constructed from internal coordinates with CHARMM.

The structure of the tetramer was built starting from a sequence alignment between VcFur and PaFur. The coordinates of the PaFur tetramer built with PISA and described previously was used as template for the building of a potential VcFur tetramer. The coordinates of all backbone atoms (N, CA, C) of residues of the VcFur dimer and of PaFur (chains A, B) showing a sequence identity were superimposed with a home-made program. 432 atoms corresponding to 144 sequence identities were superimposed with a rms deviation of 3.75 Å . Then the PaFur dimer (A,B) was superimposed with the other PaFur dimer (Chains C, D) (rmsd = 0) to yield a superposition matrix which was applied on VcFur (A,B) to yield a second VcFur dimer (chains C,D) in the new tetramer.

7.6.4 MgFur complexes

The basis of the study of *Magnetospirillum gryphiswaldense* MSR-1 Fur (MgFur) is the high-resolution structures MgFur in four different states: apo-Fur, holo-Fur, the Fur-feoAB1 promoter complex and the Fur-*Pseudomonas aeruginosa* (or consensus) Fur box complex (Deng et al., 2015).

MgFur recognizes the feoAB1 promoter whose coding strand includes a 25-base pair region with the sequence 5'-TTAATCGCAACTCATTCGCAATTGC-3' and *P. aeruginosa* Fur box with the sequence 5'-CGCGATAATGATAATCATTATCCGC-3'. The sequence in the co-crystal of the Fur-Mn^{2+}-feoAB1 promoter ternary complex is 5'-TTAATTGCAAATCATTTGCAATTGC-3' carrying three single-base-pair mutations. The

feoAB1 -coding strand showed a 25-base pair (bp) main protected region with the sequence 5'-TTAATCGCAACTCATTCGCAATTGC-3', referred to as the ' feoAB1 promoter'.

MgFUR + Fur box

The structure of *Magnetospirillum gryphiswaldense* (4RB1) (Deng et al., 2015) (MgFur) in the presence of the *Escherichia Coli* Fur box was used in this model. The sequence of MgFur comprises 143 residues but the PDB structure (4RB1) only includes residues Val 2 to Pro 133 in chain A and residues Met 1 to Pro 133 in chain B. Moreover residues Cys 9 and Met 14 are mutated into Leu and residue Met 16 is mutated into Val. Our model of MgFur thus includes residues Val 2 to Pro 133 of both chains with the 3 above mentioned mutations. The structure further contains one strand of the Fur-box sequence including 24 base pairs 5'-CGCGATAATGATAATCATTATCCG-3'. The two Mg dimers bound to the DNA double strand were constructed using the symmetries given in the PDB file. Since the initial sequence is not totally palindromic, the symmetrized strand 3'-GCCTATTACTAATAGTAATAGCGC-5' is not totally complementary with the 5' to 3' part.

To reach complementarity and keep the central Fur-box signature, we have removed from the model the first cytosine and inverted 5'-GC and CG-3' terminal base pairs to model the 23 base pair sequence 5'-CGGATAATGATAATCATTATCGC-3' and its complement 3'-GCCTATTACTATTAGTAATAGCG-5'. The missing coordinates for the 5'-C(1)G(2) G(22)C(23) and the 3'-T(12) bases have been built from internal coordinates.

As a last refinement, and to better center MgFur dimer on its DNA we have extended the DNA double strand on its 5' terminal part by moving the last GC-3' base pairs in 5' position. Our final model thus includes 5'-GCCGGATAATGATAATCATTATC-3' and its complement 3'-CGGCCTATTACTATTAGTAATAG-5'.

MgFur + feoAB1

Similarly to what was previously described, the sequence of MgFur comprises 143 residues but the PDB structure (4RB3) now includes residues Met 1 to Ser 136 in chain A and residues Val 2 to Leu 134 in chain B. Moreover residues Cys 9 and Met 14 are mutated into Leu and residue Met 16 is mutated into Val. In the X-ray structure, the feoAB1 promoter sequence includes coordinates for a double stranded DNA dimer of 25 base pairs: 5'-TTAATTGCAAATCATTTGCAATTGC-3'. To keep the same DNA box size as in previous simulation, the first Thy and last Cyt were removed from this model, without loss of interaction with the protein.

Our model of MgFur thus includes residues Val 2 to Pro 133 of both chains with the 3 above mentioned mutations, 4 metallic dications per dimer and 2 X-ray water molecules bound to iron, in interaction with the 23 base pair sequence:

5'-TAATTGCAAATCATTTGCAATTG-3' after trimming the two terminal nucleotides.

7.6.5 EcZur + DNA

The structure of *Escherichia Coli* Zur (EcZur) protein in complex with a 33 bp duplex derived from the znuABC promoter (P znuABC) and determined by X-ray crystallography: PDB code (4MTD) (Gilston et al., 2014), was used in this model. The structure corresponds to two EcZur dimers staggered on the 5'-AGAAGTGTGATATTATAACATTTCATGACTATG-3' DNA duplex, on two opposite sides, with the recognition helix sitting in the major groove.

The sequence of EcZur comprises 171 residues but the PDB structure (PDB ID: 4MTD) only includes residues Thr 4 to Cys 152 in chain A and Glu 2 to Cys 152 in chain B for the dimer used in this model. Glu 2 to Cys 152 in both chains were used (building missing atoms from internal coordinates) including 5 deprotonated (charged) cysteines: C88 binding Zn^{2+} in one site and C103, C106, C143, C146 binding zinc in the second metal site. Out of 33 DNA base pairs, a 22 base pair sequence 5'-GAAGTGTGATATTATAACATTT-3' is sufficient to accomodate 1 EcZur dimer in this first model, saving computer ressources.

7.6.6 EcZur + chain C + DNA

The purpose of this second simulation was to check the importance of the interactions between the two EcZur dimers in the stability of the protein-DNA complex. Therefore, part of EcZur chain C (from the second dimer) interacting with chain B of the first dimer above, was added in this model. Precisely, residues Glu 2 to Leu 87 only, constituting the DNA binding domain of chain B were added, to limit the computational cost of this new simulation while keeping the important interactions. The same 22 base pair DNA sequence and 4 Zn2+ ions were used in this model as in the previous simulation.

7.6.7 Solvation and equilibration

All the above vacuum systems were immersed in parallelepipedic SPC (Berendsen et al., 1987) water boxes modelled with periodic boundary conditions after addition of Na^+ and Cl^- counterions to ensure neutrality and a total ionic force of 0.1 mol/liter. The size and ion composition of all systems are summarized in Table 7.2.

The solvated systems were energy minimized and equilibrated under NPT conditions at 310 K and 1 atm.

System	Box size nm^3	Num. Metal M^{2+}	Num. Na^+	Num. Cl^-	Num. atoms	PDB ID
EcZUR dimer + DNA	8.9*9.2*7.7	4	97	38	60119	4MTD
Ec 2 ZUR dimer + DNA	11.6*8.9*8.1	4	112	51	80887	4MTD
MgFUR-Furbox	9.4*9.2*5.9	4	76	30	51948	4RB1
MgFUR-feoAB1	9.3*9.8*5.9	4	78	32	48073	4RB3
PaFur-Tetramer	10.2*8.7*6.9	8	44	36	57337	1MZB
PaFurΔS3-Tetramer	10.4*8.6*6.8	4	53	37	58232	-
FtFur-Tetramer	10.5*8.0*6.5	8	37	33	52114	5NHK
FtFur-dimer and DNA (model)	9.3*8.9*6.4	4	77	32	50117	-
VcFur tetramer (model)	10.4*8.9*7.3	8	55	41	64287	-

Table 7.2: Size and ion composition of all systems studied in this work. Metal dications are respresented as charged van der Waals spheres that do not discriminate between different atoms in our simple models of the metal binding sites. Number of counterions and total number of atoms are indicated.

Chapter 8

General conclusion and perspectives

From the day our specie started evolving complex societies, human activity has dramatically affected earth's biosphere. One example is the role our modern science and technology has played in the distribution of drug resistant bacteria in hospitals, animals and the environment. With these radical changes, alarming threats to human health are emerging. Therefore, one of our modern challenges as a scientific community, is to invest human and financial resources into the search for new atimicrobial strategies.

Currently available molecules are effective in general, however, due to their mode of action, they indirectly selected for resistant pathogens. Theoretically, one can argue that no matter what the action mechanism is, resistance will always emerge according to the Red Queen hypothesis proposed by Leigh Van Valen. After all, evolvability is the key element of living systems. Nevertheless, alternative strategies should slow down this evolutionary race, helping us understand the world we live in, and ideally lead the way towards a more harmonious philosophy of life where humanity rethinks its place in nature.

The work presented here tackles the antimicrobial resistance problem by targeting iron homeostasis in bacteria. Throughout earth's history, iron played a major role in the emergence and early evolution of life. As iron became a key element for all living organisms, its regulation and control developed into essential pathways, leading to the complex iron sensing mechanisms we find today.

This research project, focuses on the Ferric Uptake Regulator protein (Fur), a key iron sensor and transcription regulator involved in iron homeostasis. In addition, Fur is involved in bacterial virulence, which is a microbe's ability to infect a host with actions like adhesion, colonisation or invasion. This makes Fur an interesting target for the development of new therapeutic strategies to fight pathogens.

When I began my work in 2015, the objective was to use a combined *in silico* / *in vitro* approach to develop and study previously discovered Fur inhibitors. These inhibitors were identified by former researchers in our laboratory from screening experiments, and later investigated in more detail using a similar combined approach. The key findings were

mainly on Fur from the model organism *E. coli*. The main idea behind my PhD project was to expand the knowledge gained on EcFur, in order to use it on other pathogens.

However, when I first started, the structure of EcFur was not available and models were used to understand how inhibitors work. The model of EcFur built by homology to VcFur enabled the docking of peptide inhibitors and the proposition of a possible inhibition mechanism. Despite the fact that several other techniques were used, limited information about the inhibition was available due to the absence of the crystal structure of EcFur.

One of my first tasks was to use the EcFur model to dock small cyclic peptides. This was done in an attempt to create a second generation of inhibitory peptides. The results showed that, in contrary to what we were looking for, small cyclic peptides did not bind the model as well as the linear ones. Further investigation should be done to determine the scientific potential that these cyclic peptides have and if they can be used in other sections of the Fur project.

Independently, X-ray absorption spectroscopy experiments were carried out to check if the inhibition of Fur proteins involved a change in its metal sites. The example of EcFur presented in this work, clearly shows that this is not the case. The inhibition by the pF2 inhibitor does not take place due to an interaction with the metal sites. This information was crucial, because it validated the previously proposed inhibition mechanisms.

It was not before the end of my PhD thesis, that we were able to obtain, for the first time, the full structure of EcFur without its last eight amino acids (EcFur-140). Interestingly, in this structure, we were able to identify an unusual S2 metal site that contains 4 histidines and 1 glutamate, instead of 3 histidines and 1 glutamate. This observation helped fit and validate XAS data on the wild type protein. This shows that by two different experimental methods, the S2 site of EcFur exhibits a new metal site configuration, not described in the literature and could be unique to EcFur. More information should be gathered to check if this site can be obtained in other proteins when their structures are resolved in the presence of nickel.

Moreover, the structure of EcFur-140 revealed a disulphide bridge in site S1, similar to the structure of VcFur. It would be interesting to crystallize EcFur-140 in reducing conditions, to study site S1 which should bind zinc, as described in the literature. With our current diffraction data set, the electron density map of site S3 is not well defined. We were able to propose a model where the site is composed of 2 histidines, one glutamate and one aspartate. More crystallization trials need to be done in order to collect new data, and resolve site S3.

After the structure resolution of EcFur-140, the protein was characterised. We were able to show that, similarly to the wild type protein, this protein is active. SAXS experiments showed that the structure can be used to fit experimental data obtained from the wild type protein. This allowed us to propose a model of the two EcFur dimers binding a

Fur box sequence. This structure and the DNA binding model will be used in the future, as a tool to understand the inhibition by all the inhibitors developed in our laboratory, and for the screening of new ones.

In addition to the structure of EcFur, this work also lead to the structure of Fur from *P. aeruginosa*. In fact, the mutant PaFurΔS3 crystallized and its structure was solved in the presence of zinc and manganese. The two structures showed that metal atoms are bound not only to the usual S2 and S3 metal sites present in PaFur, but also elsewhere on the protein. These metal sites are still being investigated and could be important in the understanding of tetrameric Fur proteins.

Previous research identified pL1, pL2 and molecule B to be specific inhibitors of PaFur. During my PhD, docking of pL1 and pL2 were obtained. They showed a similar binding pocket to what was obtained for pF2 on EcFur. To better understand how these inhibitors work, several months of crystallization trials were carried out in order to obtain crystals with a protein-inhibitor complex. The crystallization conditions were optimized and crystals were obtained, however structure resolution was not possible. In the future, work should be done to enhance the diffraction quality of these crystals. The resolution of a Fur structure in the presence of an inhibitor is extremely important to start optimizing and enhancing the inhibitors we already have.

In parallel, after the publication of the structure of FtFur, the question of tetramer dissociation and dimers binding to DNA was raised. To answer it, I spent several months implementing and optimizing the protocol used for potential of mean force calculations. In addition, the search for major interacting residues was made automatic, using scripts that help visualize the results and retrieve information faster. Very long MD simulations were necessary to tackle the PMFs calculations with good accuracy. During my last year, these calculation which amount to more than 20 µs in total were made possible by the use of optimized algorithms and GPUs which are becoming the Graal in MD simulations. On average, with the current scripts and job submission protocols, 1.5 µs of molecular dynamics, for a system with 70,000 atoms, takes one month on 20 processors. Two more weeks should be taken into consideration for result analysis. Before the use of GPU, the same calculations took up to two months to converge. When all the results were obtained, they showed similarities in the mechanisms, but at the same time highlighted the differences between each protein. Some results were compared to the literature when available. It was interesting to see biphasic dissociations in the cases of MgFur, FtFur and EcZur; this indicates that Fur DNA binding mechanisms can be complex and vary depending on the system, and gives ideas on the recognition steps betweeen protein and DNA.

This being said, it is clear that Fur proteins are not as similar to each other as they seem. With differences in oligomeric states, amino acid composition and nature of metal sites, the initial project idea of expanding EcFur inhibitors to other proteins does not

seem to be the optimal strategy. In addition, available structures are often solved in the presence of zinc, which is not the physiological metal of Fur proteins, making the analysis of metal sites in the structures more complicated. The two structures obtained from this work, in addition to the proposed models and simulation protocols, should be considered as tools and used to better understand Fur proteins, and adapt every research approach to each Fur protein.

Bibliography

[1] N. Abed, M. Bickle, B. Mari, M. Schapira, R. Sanjuan-España, K. Robbe Sermesant, O. Moncorgé, S. Mouradian-Garcia, P. Barbry, B. B. Rudkin, M.-O. Fauvarque, I. Michaud-Soret, and P. Colas. "A Comparative Analysis of Perturbations Caused by a Gene Knock-out, a Dominant Negative Allele, and a Set of Peptide Aptamers". In: *Molecular & Cellular Proteomics* 6.12 (2007), pp. 2110–2121. DOI: 10.1074/mcp.M700105-MCP200. eprint: http://www.mcponline.org/content/6/12/2110.full.pdf+html. URL: http://www.mcponline.org/content/6/12/2110.abstract.

[2] M. J. Abraham, T. Murtola, R. Schulz, S. Pll, J. C. Smith, B. Hess, and E. Lindahl. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 12 (2015), pp. 19 –25. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2015.06.001. URL: http://www.sciencedirect.com/science/article/pii/S2352711015000059.

[3] L. A. Achenbach and W Yang. "The fur gene from Klebsiella pneumoniae: characterization, genomic organization and phylogenetic analysis." In: *Gene* 185 (2 Feb. 1997), pp. 201–207. ISSN: 0378-1119.

[4] P. D. Adams et al. "PHENIX: a comprehensive Python-based system for macromolecular structure solution." In: *Acta crystallographica. Section D, Biological crystallography* 66 (Pt 2 Feb. 2010), pp. 213–221. ISSN: 1399-0047. DOI: 10.1107/S0907444909052925.

[5] P. Alamuri, N. Mehta, A. Burk, and R. J. Maier. "Regulation of the Helicobacter pylori Fe-S cluster synthesis protein NifS by iron, oxidative stress conditions, and fur." In: *Journal of bacteriology* 188 (14 July 2006), pp. 5325–5330. ISSN: 0021-9193. DOI: 10.1128/JB.00104-06.

[6] H. K. Allen, J. Donato, H. H. Wang, K. A. Cloud-Hansen, J. Davies, and J. Handelsman. "Call of the wild: antibiotic resistance genes in natural environments." In: *Nature reviews. Microbiology* 8 (4 Apr. 2010), pp. 251–259. ISSN: 1740-1534. DOI: 10.1038/nrmicro2312.

[7] E. W. Althaus, C. E. Outten, K. E. Olson, H Cao, and T. V. O'Halloran. "The ferric uptake regulation (Fur) repressor is a zinc metalloprotein." In: *Biochemistry* 38 (20 May 1999), pp. 6559–6569. ISSN: 0006-2960. DOI: 10.1021/bi982788s.

[8] R. I. Aminov and R. I. Mackie. "Evolution and ecology of antibiotic resistance genes." In: *FEMS microbiology letters* 271 (2 June 2007), pp. 147–161. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.2007.00757.x.

[9] A. D. Anbar. "Oceans. Elements and evolution." In: *Science (New York, N.Y.)* 322 (5907 Dec. 2008), pp. 1481–1483. ISSN: 1095-9203. DOI: 10.1126/science.1163100.

[10] S. C. Andrews, A. K. Robinson, and F. Rodríguez-Quiñones. "Bacterial iron homeostasis." In: *FEMS microbiology reviews* 27 (2-3 June 2003), pp. 215–237. ISSN: 0168-6445.

[11] E. S. Arnér and A Holmgren. "Physiological functions of thioredoxin and thioredoxin reductase." In: *European journal of biochemistry* 267 (20 Oct. 2000), pp. 6102–6109. ISSN: 0014-2956.

[12] S. S. Athavale, A. S. Petrov, C. Hsiao, D. Watkins, C. D. Prickett, J. J. Gossett, L. Lie, J. C. Bowman, E. O'Neill, C. R. Bernier, N. V. Hud, R. M. Wartell, S. C. Harvey, and L. D. Williams. "RNA folding and catalysis mediated by iron (II)." In: *PloS one* 7 (5 2012), e38024. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0038024.

[13] A Bagg and J. B. Neilands. "Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in Escherichia coli." In: *Biochemistry* 26 (17 Aug. 1987), pp. 5471–5477. ISSN: 0006-2960.

[14] N. Baichoo and J. D. Helmann. "Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence." In: *Journal of bacteriology* 184 (21 Nov. 2002), pp. 5826–5832. ISSN: 0021-9193.

[15] K. S. Baker, E. Burnett, H. McGregor, A. Deheer-Graham, C. Boinett, G. C. Langridge, A. M. Wailan, A. K. Cain, N. R. Thomson, J. E. Russell, and J. Parkhill. "The Murray collection of pre-antibiotic era Enterobacteriacae: a unique research resource." In: *Genome medicine* 7 (Sept. 2015), p. 97. ISSN: 1756-994X. DOI: 10.1186/s13073-015-0222-7.

[16] M. F. Barber and N. C. Elde. "Nutritional immunity. Escape from bacterial iron piracy through rapid evolution of transferrin." In: *Science (New York, N.Y.)* 346 (6215 Dec. 2014), pp. 1362–1366. ISSN: 1095-9203. DOI: 10.1126/science.1259329.

[17] M. F. Barber and N. C. Elde. "Buried Treasure: Evolutionary Perspectives on Microbial Iron Piracy." In: *Trends in genetics : TIG* 31 (11 Nov. 2015), pp. 627–636. ISSN: 0168-9525. DOI: 10.1016/j.tig.2015.09.001.

[18] M. Barlow and B. G. Hall. "Phylogenetic analysis shows that the OXA beta-lactamase genes have been on plasmids for millions of years." In: *Journal of molecular evolution* 55 (3 Sept. 2002), pp. 314–321. ISSN: 0022-2844. DOI: 10.1007/s00239-002-2328-y.

[19] J. A. Baross and S. E. Hoffman. "Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life". In: *Origins of life and evolution of the biosphere* 15.4 (Dec. 1, 1985), p. 327. DOI: 10.1007/BF01808177. URL: http://dx.doi.org/10.1007/BF01808177.

[20] H. Bekker, H. Berendsen, E. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma, and M. Renardus. "GROMACS: A parallel computer for molecular dynamics simulations." In: *Conference proceedings - Physics Computing 92*. Ed. by R. A. de Groot and J. Nadrchal. 1992.

[21] K. S. Bender, H.-C. B. Yen, C. L. Hemme, Z. Yang, Z. He, Q. He, J. Zhou, K. H. Huang, E. J. Alm, T. C. Hazen, A. P. Arkin, and J. D. Wall. "Analysis of a ferric uptake regulator (Fur) mutant of Desulfovibrio vulgaris Hildenborough." In: *Applied and environmental microbiology* 73 (17 Sept. 2007), pp. 5389–5400. ISSN: 0099-2240. DOI: 10.1128/AEM.00276-07.

[22] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The missing term in effective pair potentials". In: *The Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271. DOI: 10.1021/j100308a038. eprint: http://dx.doi.org/10.1021/j100308a038. URL: http://dx.doi.org/10.1021/j100308a038.

[23] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak. "Molecular dynamics with coupling to an external bath". In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.

[24] S Bereswill, F Lichte, T Vey, F Fassbinder, and M Kist. "Cloning and characterization of the fur gene from Helicobacter pylori." In: *FEMS microbiology letters* 159 (2 Feb. 1998), pp. 193–200. ISSN: 0378-1097.

[25] M. B. T. Bickle, E. Dusserre, O. Moncorgé, H. Bottin, and P. Colas. "Selection and characterization of large collections of peptide aptamers through optimized yeast two-hybrid procedures." In: *Nature protocols* 1 (3 2006), pp. 1066–1091. ISSN: 1750-2799. DOI: 10.1038/nprot.2006.32.

[26] N Binsted, R. W. Strange, and S. S. Hasnain. "Constrained and restrained refinement in EXAFS data analysis with curved wave theory." In: *Biochemistry* 31 (48 Dec. 1992), pp. 12117–12125. ISSN: 0006-2960.

[27] J. H. Blum, S. L. Dove, A Hochschild, and J. J. Mekalanos. "Isolation of peptide aptamers that inhibit intracellular processes." In: *Proceedings of the National Academy of Sciences of the United States of America* 97 (5 Feb. 2000), pp. 2241–2246. ISSN: 0027-8424. DOI: 10.1073/pnas.040573397.

[28] C. Bonfio, E. Godino, M. Corsini, F. Fabrizi de Biani, G. Guella, and S. S. Mansy. "Prebiotic iron-sulfur peptide catalysts generate a pH gradient across model membranes of late protocells". In: *Nature Catalysis* 1.8 (2018), pp. 616–623. ISSN: 2520-1158. DOI: 10.1038/s41929-018-0116-3. URL: https://doi.org/10.1038/s41929-018-0116-3.

[29] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 84 (1927), p. 457.

[30] E. S. Boyd, K. M. Thomas, Y. Dai, J. M. Boyd, and F. W. Outten. "Interplay between oxygen and Fe-S cluster biogenesis: insights from the Suf pathway." In: *Biochemistry* 53 (37 Sept. 2014), pp. 5834–5847. ISSN: 1520-4995. DOI: 10.1021/bi500488r.

[31] V. Braun, S. Schäffer, K. Hantke, and W. Tröger. "Regulation of Gene Expression by Iron". In: *The Molecular Basis of Bacterial Metabolism*. Ed. by G. Hauska and R. K. Thauer. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 164–179. ISBN: 978-3-642-75969-7.

[32] P. Brescia and P. Banks. *Multi-Volume Analysis of Nucleic Acids Using the Epoch™ Spectrophotometer System*. Tech. rep. BioTek Instruments Inc., Winooski, VT, 2009.

[33] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. " In: *J. Comput. Chem.* 4 (1983), pp. 187–217.

[34] A. Brunning. *A Brief Guide to the Twenty Common Amino Acids*. https://www.compoundchem.com/wp-content/uploads/2014/09/20-Common-Amino-Acids-v3.png. Accessed: 2018-09-10.

[35] A. Brunning. *Different classes of antibiotics - An overview*. http://www.compoundchem.com/wp-content/uploads/2014/09/A-Guide-to-Different-Classes-of-Antibiotics-Aug-15.png. Accessed: 2018-06-19.

[36] N Bsat, A Herbig, L Casillas-Martinez, P Setlow, and J. D. Helmann. "Bacillus subtilis contains multiple Fur homologues: identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors." In: *Molecular microbiology* 29 (1 July 1998), pp. 189–198. ISSN: 0950-382X.

[37] O Bunău and Y Joly. "Self-consistent aspects of x-ray absorption calculations." In: *Journal of physics. Condensed matter : an Institute of Physics journal* 21 (34 Aug. 2009), p. 345501. ISSN: 1361-648X. DOI: 10.1088/0953-8984/21/34/345501.

[38] J. Butcher, S. Sarvan, J. S. Brunzelle, J.-F. Couture, and A. Stintzi. "Structure and regulon of Campylobacter jejuni ferric uptake regulator Fur define apo-Fur regulation." In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (25 June 2012), pp. 10047–10052. ISSN: 1091-6490. DOI: 10.1073/pnas.1118321109.

[39] S. B. Calderwood and J. J. Mekalanos. “Iron regulation of Shiga-like toxin expression in Escherichia coli is mediated by the fur locus.” In: *Journal of bacteriology* 169 (10 Oct. 1987), pp. 4759–4764. ISSN: 0021-9193.

[40] J Castresana. “Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.” In: *Molecular biology and evolution* 17 (4 Apr. 2000), pp. 540–552. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a026334.

[41] C. Chieh. *Chieh, 2018, Bond Lengths and Energies*. http://www.science.uwaterloo.ca/~cchieh/cact/c120/bondel.html. Accessed: 2018-09-26.

[42] D. Chirgadze. *Protein Crystallisation in Action*. 2001. URL: http://www.xray.bioc.cam.ac.uk/xray_resources/whitepapers/xtal-in-action/node3.html.

[43] C. Cissé, S. V. Mathieu, M. B. O. Abeih, L. Flanagan, S. Vitale, P. Catty, D. Boturyn, I. Michaud-Soret, and S. Crouzy. “Inhibition of the ferric uptake regulator by peptides derived from anti-FUR peptide aptamers: coupled theoretical and experimental approaches.” In: *ACS chemical biology* 9 (12 Dec. 2014), pp. 2779–2786. ISSN: 1554-8937. DOI: 10.1021/cb5005977.

[44] P Colas, B Cohen, T Jessen, I Grishina, J McCoy, and R Brent. “Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2.” In: *Nature* 380 (6574 Apr. 1996), pp. 548–550. ISSN: 0028-0836. DOI: 10.1038/380548a0.

[45] J.-F. Collet and J. Messens. “Structure, function, and mechanism of thioredoxin proteins.” In: *Antioxidants & redox signaling* 13 (8 Oct. 2010), pp. 1205–1216. ISSN: 1557-7716. DOI: 10.1089/ars.2010.3114.

[46] R. Constanciel and R. Contreras. “Self consistent field theory of solvent effects representation by continuum models: Introduction of desolvation contribution”. In: *Theoretica chimica acta* 65.1 (1984), pp. 1–11. ISSN: 1432-2234. DOI: 10.1007/PL00020119. URL: https://doi.org/10.1007/PL00020119.

[47] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman. “ A second generation force field for the simulation of proteins and nucleic acids. ” In: *J. Am. Chem. Soc.* 117 (1995), pp. 5179–5197.

[48] M Coy, C Doyle, J Besser, and J. B. Neilands. “Site-directed mutagenesis of the ferric uptake regulation gene of Escherichia coli.” In: *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 7 (4 Oct. 1994), pp. 292–298. ISSN: 0966-0844.

[49] S. Crouzy, T. Woolf, and B. Roux. “A molecular dynamics study of gating in dioxolane-linked gramicidin A channels.” In: *Biophys. J.* 67 (1994), pp. 1370–1386.

[50] M. D. “Polarity and permeation profiles in lipid membranes”. In: *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001), pp. 7777–7782.

[51] T. Darden, D. York, and L. Pedersen. “Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems”. In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092. DOI: 10.1063/1.464397. eprint: https://doi.org/10.1063/1.464397. URL: https://doi.org/10.1063/1.464397.

[52] N Datta and V. M. Hughes. “Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics.” In: *Nature* 306 (5943 1983), pp. 616–617. ISSN: 0028-0836.

[53] B. D’Autreaux, D. Touati, B. Bersch, J.-M. Latour, and I. Michaud-Soret. “Direct inhibition by nitric oxide of the transcriptional ferric uptake regulation protein via nitrosylation of the iron.” In: *Proceedings of the National Academy of Sciences of the United States of America* 99 (26 Dec. 2002), pp. 16619–16624. ISSN: 0027-8424. DOI: 10.1073/pnas.252591299.

[54] B. D'Autréaux, L. Pecqueur, A. Gonzalez de Peredo, R. E. M. Diederix, C. Caux-Thang, L. Tabet, B. Bersch, E. Forest, and I. Michaud-Soret. "Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein." In: *Biochemistry* 46 (5 Feb. 2007), pp. 1329–1342. ISSN: 0006-2960. DOI: 10.1021/bi061636r.

[55] L. A. David and E. J. Alm. "Rapid evolutionary innovation during an Archaean genetic expansion." In: *Nature* 469 (7328 Jan. 2011), pp. 93–96. ISSN: 1476-4687. DOI: 10.1038/nature09649.

[56] J Davies. "Inactivation of antibiotics and the dissemination of resistance genes." In: *Science (New York, N.Y.)* 264 (5157 Apr. 1994), pp. 375–382. ISSN: 0036-8075.

[57] J. Davies. "Specialized microbial metabolites: functions and origins." In: *The Journal of antibiotics* 66 (7 July 2013), pp. 361–364. ISSN: 0021-8820. DOI: 10.1038/ja.2013.61.

[58] J. Davies and D. Davies. "Origins and evolution of antibiotic resistance." In: *Microbiology and molecular biology reviews : MMBR* 74 (3 Sept. 2010), pp. 417–433. ISSN: 1098-5557. DOI: 10.1128/MMBR.00016-10.

[59] V. M. D'Costa, C. E. King, L. Kalan, M. Morar, W. W. L. Sung, C. Schwarz, D. Froese, G. Zazula, F. Calmels, R. Debruyne, G. B. Golding, H. N. Poinar, and G. D. Wright. "Antibiotic resistance is ancient." In: *Nature* 477 (7365 Aug. 2011), pp. 457–461. ISSN: 1476-4687. DOI: 10.1038/nature10388.

[60] V De Lorenzo, M Herrero, F Giovannini, and J. B. Neilands. "Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of fur gene in Escherichia coli." In: *European journal of biochemistry* 173 (3 May 1988), pp. 537–546. ISSN: 0014-2956.

[61] I. Delany, G. Spohn, R. Rappuoli, and V. Scarlato. "An anti-repression Fur operator upstream of the promoter is required for iron-mediated transcriptional autoregulation in Helicobacter pylori." In: *Molecular microbiology* 50 (4 Nov. 2003), pp. 1329–1338. ISSN: 0950-382X.

[62] I. Delany, R. Rappuoli, and V. Scarlato. "Fur functions as an activator and as a repressor of putative virulence genes in Neisseria meningitidis." In: *Molecular microbiology* 52 (4 May 2004), pp. 1081–1090. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2004.04030.x.

[63] I. Delany, R. Ieva, A. Soragni, M. Hilleringmann, R. Rappuoli, and V. Scarlato. "In vitro analysis of protein-operator interactions of the NikR and fur metal-responsive regulators of coregulated genes in Helicobacter pylori." In: *Journal of bacteriology* 187 (22 Nov. 2005), pp. 7703–7715. ISSN: 0021-9193. DOI: 10.1128/JB.187.22.7703-7715.2005.

[64] I. Delany, R. Grifantini, E. Bartolini, R. Rappuoli, and V. Scarlato. "Effect of Neisseria meningitidis fur mutations on global control of gene transcription." In: *Journal of bacteriology* 188 (7 Apr. 2006), pp. 2483–2492. ISSN: 0021-9193. DOI: 10.1128/JB.188.7.2483-2492.2006.

[65] Z. Deng et al. "Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator." In: *Nature communications* 6 (July 2015), p. 7642. ISSN: 2041-1723. DOI: 10.1038/ncomms8642.

[66] C. Dian, S. Vitale, G. A. Leonard, C. Bahlawane, C. Fauquant, D. Leduc, C. Muller, H. de Reuse, I. Michaud-Soret, and L. Terradot. "The structure of the Helicobacter pylori ferric uptake regulator Fur reveals three functional metal binding sites." In: *Molecular microbiology* 79 (5 Mar. 2011), pp. 1260–1275. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2010.07517.x.

[67] A. C. Dlouhy and C. E. Outten. "The iron metallome in eukaryotic organisms." In: *Metal ions in life sciences* 12 (2013), pp. 241–278. ISSN: 1559-0836. DOI: 10.1007/978-94-007-5561-1_8.

[68] T. Dobzhansky. "Nothing in Biology Makes Sense except in the Light of Evolution". In: *The American Biology Teacher* 35.3 (1973), pp. 125–129. ISSN: 00027685, 19384211. URL: http://www.jstor.org/stable/4444260.

[69] C. L. Dupont, A. Butcher, R. E. Valas, P. E. Bourne, and G. Caetano-Anollés. "History of biological metal utilization inferred through phylogenomic analysis of protein structures." In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (23 June 2010), pp. 10567–10572. ISSN: 1091-6490. DOI: 10.1073/pnas.0912491107.

[70] B. Efron. "Bootstrap Methods: Another Look at the Jackknife". In: *Ann. Statist.* 7.1 (Jan. 1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: https://doi.org/10.1214/aos/1176344552.

[71] D. Eisenberg and L. Wesson. "Atomic solvation parameters applied to molecular dynamics of proteins in solution". In: *Protein Sci* 1 (1992), p. 227235.

[72] G. L. Ellman. "Tissue sulfhydryl groups." In: *Archives of biochemistry and biophysics* 82 (1 May 1959), pp. 70–77. ISSN: 0003-9861.

[73] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. "Features and Development of Coot". In: *Acta Crystallographica Section D - Biological Crystallography* 66 (2010), pp. 486–501.

[74] F. D. Ernst, S. Bereswill, B. Waidner, J. Stoof, U. Mäder, J. G. Kusters, E. J. Kuipers, M. Kist, A. H. M. van Vliet, and G. Homuth. "Transcriptional profiling of Helicobacter pylori Fur- and iron-regulated gene expression." In: *Microbiology (Reading, England)* 151 (Pt 2 Feb. 2005), pp. 533–546. ISSN: 1350-0872. DOI: 10.1099/mic.0.27404-0.

[75] J. F. Ernst, R. L. Bennett, and L. I. Rothfield. "Constitutive expression of the iron-enterochelin and ferrichrome uptake systems in a mutant strain of Salmonella typhimurium." In: *Journal of bacteriology* 135 (3 Sept. 1978), pp. 928–934. ISSN: 0021-9193.

[76] L Escolar, J Pérez-Martín, and V de Lorenzo. "Binding of the fur (ferric uptake regulator) repressor of Escherichia coli to arrays of the GATAAT sequence." In: *Journal of molecular biology* 283 (3 Oct. 1998), pp. 537–547. ISSN: 0022-2836. DOI: 10.1006/jmbi.1998.2119.

[77] L Escolar, J Pérez-Martín, and V de Lorenzo. "Opening the iron box: transcriptional metalloregulation by the Fur protein." In: *Journal of bacteriology* 181 (20 Oct. 1999), pp. 6223–6229. ISSN: 0021-9193.

[78] L Escolar, J Pérez-Martín, and V de Lorenzo. "Evidence of an unusually long operator for the fur repressor in the aerobactin promoter of Escherichia coli." In: *The Journal of biological chemistry* 275 (32 Aug. 2000), pp. 24709–24714. ISSN: 0021-9258. DOI: 10.1074/jbc.M002839200.

[79] U Essmann, L Perera, M. Berkowitz, T Darden, H Lee, and L. Pedersen. "A smooth particle mesh Ewald method". In: *Journal of Chemical Physics* 103 (1995), pp. 8577–8593.

[80] P. P. Ewald. "Die Berechnung optischer und elektrostatischer Gitterpotentiale". In: *Annalen der Physik* 369.3 (1921), pp. 253–287. DOI: 10.1002/andp.19213690304. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19213690304. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19213690304.

[81] Falkowski, Barber, and Smetacek. "Biogeochemical Controls and Feedbacks on Ocean Primary Production". In: *Science (New York, N.Y.)* 281 (5374 July 1998), pp. 200–207. ISSN: 1095-9203.

[82] J. Felsenstein. "CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP." In: *Evolution; international journal of organic evolution* 39 (4 July 1985), pp. 783–791. ISSN: 1558-5646. DOI: 10.1111/j.1558-5646.1985.tb00420.x.

[83] P. Ferrara, J. Apostolakis, and A. Caflisch. "Evaluation of a fast implicit solvent model for molecular dynamics simulations." In: *Proteins* 46 (2002), pp. 24–33.

[84] A. M. Ferrenberg and R. H. Swendsen. "Optimized Monte Carlo data analysis". In: *Phys. Rev. Lett.* 63 (12 1989), pp. 1195–1198. DOI: 10.1103/PhysRevLett.63.1195. URL: https://link.aps.org/doi/10.1103/PhysRevLett.63.1195.

[85] S Fields and O Song. "A novel genetic system to detect protein-protein interactions." In: *Nature* 340 (6230 July 1989), pp. 245–246. ISSN: 0028-0836. DOI: 10.1038/340245a0.

[86] M. F. Fillat. "The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators." In: *Archives of biochemistry and biophysics* 546 (Mar. 2014), pp. 41–52. ISSN: 1096-0384. DOI: 10.1016/j.abb.2014.01.029.

[87] D Franke, M. V. Petoukhov, P. V. Konarev, A Panjkovich, A Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. "ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions." In: *Journal of applied crystallography* 50 (Pt 4 Aug. 2017), pp. 1212–1225. ISSN: 0021-8898. DOI: 10.1107/S1600576717007786.

[88] M. Fuangthong, A. F. Herbig, N. Bsat, and J. D. Helmann. "Regulation of the Bacillus subtilis fur and perR genes by PerR: not all members of the PerR regulon are peroxide inducible." In: *Journal of bacteriology* 184 (12 June 2002), pp. 3276–3286. ISSN: 0021-9193.

[89] H. Gancz, S. Censini, and D. S. Merrell. "Iron and pH homeostasis intersect at the level of Fur regulation in the gastric pathogen Helicobacter pylori." In: *Infection and immunity* 74 (1 Jan. 2006), pp. 602–614. ISSN: 0019-9567. DOI: 10.1128/IAI.74.1.602-614.2006.

[90] H. Gao, D. Zhou, Y. Li, Z. Guo, Y. Han, Y. Song, J. Zhai, Z. Du, X. Wang, J. Lu, and R. Yang. "The iron-responsive Fur regulon in Yersinia pestis." In: *Journal of bacteriology* 190 (8 Apr. 2008), pp. 3063–3075. ISSN: 1098-5530. DOI: 10.1128/JB.01910-07.

[91] J. A. Garibaldi and J. B. Neilands. "Formation of iron-binding compounds by micro-organisms." In: *Nature* 177 (4507 Mar. 1956), pp. 526–527. ISSN: 0028-0836.

[92] V. Gasser, L. Guillon, O. Cunrath, and I. J. Schalk. "Cellular organization of siderophore biosynthesis in Pseudomonas aeruginosa: Evidence for siderosomes." In: *Journal of inorganic biochemistry* 148 (July 2015), pp. 27–34. ISSN: 1873-3344. DOI: 10.1016/j.jinorgbio.2015.01.017.

[93] B. A. Gilston, S. Wang, M. D. Marcus, M. A. Canalizo-Hernández, E. P. Swindell, Y. Xue, A. Mondragón, and T. V. O'Halloran. "Structural and Mechanistic Basis of Zinc Regulation Across the E. coli Zur Regulon". In: *PLOS Biology* 12.11 (Nov. 2014), pp. 1–16. DOI: 10.1371/journal.pbio.1001987. URL: https://doi.org/10.1371/journal.pbio.1001987.

[94] K. O. Glatter O. *Data treatment*. Academic Press, 1982. ISBN: 0-12-286280-5. URL: http://inis.iaea.org/search/search.aspx?orig_q=RN:16012436.

[95] M. Gouy, S. Guindon, and O. Gascuel. "SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." In: *Molecular biology and evolution* 27 (2 Feb. 2010), pp. 221–224. ISSN: 1537-1719. DOI: 10.1093/molbev/msp259.

[96] L. Greengard and V. Rokhlin. "A Fast Algorithm For Particle Simulations". In: *J. Comput. Chem.* 73 (1987), p. 325.

[97] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." In: *Systematic biology* 59 (3 May 2010), pp. 307–321. ISSN: 1076-836X. DOI: 10.1093/sysbio/syq010.

[98] Guinier, André. "La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques". In: *Ann. Phys.* 11.12 (1939), pp. 161–237. DOI: 10.1051/anphys/193911120161. URL: https://doi.org/10.1051/anphys/193911120161.

[99] T. Halgren. "Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters". In: *Journal of the American Chemical Society* 114 (1992), p. 7827.

[100] A. Hamdi and P. Colas. "Yeast two-hybrid methods and their applications in drug discovery." In: *Trends in pharmacological sciences* 33 (2 Feb. 2012), pp. 109–118. ISSN: 1873-3735. DOI: 10.1016/j.tips.2011.10.008.

[101] K Hantke. "Regulation of ferric iron transport in Escherichia coli K12: isolation of a constitutive mutant." In: *Molecular & general genetics : MGG* 182 (2 1981), pp. 288–292. ISSN: 0026-8925.

[102] K Hantke. "Cloning of the repressor protein gene of iron-regulated systems in Escherichia coli K12." In: *Molecular & general genetics : MGG* 197 (2 1984), pp. 337–341. ISSN: 0026-8925.

[103] K Hantke. "Iron and metal regulation in bacteria." In: *Current opinion in microbiology* 4 (2 Apr. 2001), pp. 172–177. ISSN: 1369-5274.

[104] K. Hantke. "Selection procedure for deregulated iron transport mutants (fur) in Escherichia coli K 12: fur not only affects iron metabolism". In: *Molecular and General Genetics MGG* 210.1 (Nov. 1, 1987), p. 135. ISSN: 1432-1874. DOI: 10.1007/BF00337769. URL: http://dx.doi.org/10.1007/BF00337769.

[105] D. M. Hillis and J. J. Bull. "An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis". In: *Systematic Biology* 42.2 (1993), pp. 182–192. DOI: 10.1093/sysbio/42.2.182. eprint: /oup/backfile/content_public/journal/sysbio/42/2/10.1093/sysbio/42.2.182/2/42-2-182.pdf. URL: http://dx.doi.org/10.1093/sysbio/42.2.182.

[106] H. D. Holland, B Lazar, and M McCaffrey. "Evolution of the atmosphere and oceans." In: *Nature* 320 (6057 Mar. 1986), pp. 27–33. ISSN: 0028-0836. DOI: 10.1038/320027a0.

[107] L Holm, C Sander, H Rüterjans, M Schnarr, R Fogh, R Boelens, and R Kaptein. "LexA repressor and iron uptake regulator from Escherichia coli: new members of the CAP-like DNA binding domain superfamily." In: *Protein engineering* 7 (12 Dec. 1994), pp. 1449–1453. ISSN: 0269-2139.

[108] A. H. Holmes, L. S. P. Moore, A. Sundsfjord, M. Steinbakk, S. Regmi, A. Karkey, P. J. Guerin, and L. J. V. Piddock. "Understanding the mechanisms and drivers of antimicrobial resistance." In: *Lancet (London, England)* 387 (10014 Jan. 2016), pp. 176–187. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(15)00473-0.

[109] K. Holmes, F. Mulholland, B. M. Pearson, C. Pin, J. McNicholl-Kennedy, J. M. Ketley, and J. M. Wells. "Campylobacter jejuni gene expression in response to iron limitation and the role of Fur." In: *Microbiology (Reading, England)* 151 (Pt 1 Jan. 2005), pp. 243–257. ISSN: 1350-0872. DOI: 10.1099/mic.0.27412-0.

[110] G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". In: *Phys. Rev. A* 31 (1985), p. 1695.

[111] M. J. Horsburgh, E Ingham, and S. J. Foster. "In Staphylococcus aureus, fur is an interactive regulator with PerR, contributes to virulence, and Is necessary for oxidative stress resistance through positive regulation of catalase and iron homeostasis." In: *Journal of bacteriology* 183 (2 Jan. 2001), pp. 468–475. ISSN: 0021-9193. DOI: 10.1128/JB.183.2.468-475.2001.

[112] K. Hsin, Y. Sheng, M. M. Harding, P. Taylor, and M. D. Walkinshaw. "MESPEUS: a database of the geometry of metal sites in proteins". In: *Journal of Applied Crystallography* 41.5 (2008), pp. 963–968. DOI: 10.1107/S002188980802476X. URL: https://doi.org/10.1107/S002188980802476X.

[113] J. S. Hub, B. L. de Groot, and D. van der Spoel. "g_wham: A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates". In: *Journal of Chemical Theory and Computation* 6.12 (2010), pp. 3713–3720. DOI: 10.1021/ct100494z. eprint: http://dx.doi.org/10.1021/ct100494z. URL: http://dx.doi.org/10.1021/ct100494z.

[114] W. Humphrey, A. Dalke, and K. Schulten. "VMD: visual molecular dynamics". In: *Journal of molecular graphics* 14.1 (1996), pp. 33–38.

[115] C. Huygens. *Traité de la lumière*. chez Pierre Vander Aa marchand libraire, 1690.

[116] H. Irving and R. J. P. Williams. "Order of Stability of Metal Complexes". In: *Nature* 162 (Nov. 1948), p. 746. URL: http://dx.doi.org/10.1038/162746a0.

[117] IUPAC. *Periodic Table of elements*. https://iupac.org/what-we-do/periodic-table-of-elements/. Accessed: 2018-09-10.

[118] L Jacquamet, D Aberdam, A Adrait, J. L. Hazemann, J. M. Latour, and I Michaud-Soret. "X-ray absorption spectroscopy of a new zinc site in the fur protein from Escherichia coli." In: *Biochemistry* 37 (8 Feb. 1998), pp. 2564–2571. ISSN: 0006-2960. DOI: 10.1021/bi9721344.

[119] B. H. S. Jans. *Hanssen, Docking redone*. http://www.cmbi.ru.nl/redock/. Accessed: 2018-09-26.

[120] W. Jorgensen and J. Tirado-Rives. "The OPLS Potential Functions for Proteins. Energy Minimization for Crystal of Cyclic Peptides and Crambin". In: *J. Am. Chem. Soc.* 110 (1988), pp. 1657–1666.

[121] G. F. Joyce. "RNA evolution and the origins of life." In: *Nature* 338 (6212 Mar. 1989), pp. 217–224. ISSN: 0028-0836. DOI: 10.1038/338217a0.

[122] W. Kabsch. "XDS." In: *Acta crystallographica. Section D, Biological crystallography* 66 (Pt 2 Feb. 2010), pp. 125–132. ISSN: 1399-0047. DOI: 10.1107/S0907444909047337.

[123] E. Krissinel and K. Henrick. "Inference of Macromolecular Assemblies from Crystalline State". In: *Journal of Molecular Biology* 372.3 (2007), pp. 774 –797. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2007.05.022. URL: http://www.sciencedirect.com/science/article/pii/S0022283607006420.

[124] S. Kumar, D. Bouzida, R. Swendsen, P. Kollman, and J. Rosenberg. "The Weighted Histogram Analysis Method for free-energy calculations on biomolecules. I. The method." In: *J. Comput. Chem.* 13 (1992), pp. 1011–1021.

[125] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, and S. Kmiecik. "CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site." In: *Nucleic acids research* 43 (W1 July 2015), W419–W424. ISSN: 1362-4962. DOI: 10.1093/nar/gkv456.

[126] M Kuroda et al. "Whole genome sequencing of meticillin-resistant Staphylococcus aureus." In: *Lancet (London, England)* 357 (9264 Apr. 2001), pp. 1225–1240. ISSN: 0140-6736.

[127] J. L. Lavrrar, C. A. Christoffersen, and M. A. McIntosh. "Fur-DNA interactions at the bidirectional fepDGC-entS promoter region in Escherichia coli." In: *Journal of molecular biology* 322 (5 Oct. 2002), pp. 983–995. ISSN: 0022-2836.

[128] T. Lazaridis and M. Karplus. "Effective energy function for proteins in solution". In: *Proteins* 35.2 (1999), pp. 133–152.

[129] A. Leach. "Molecular Modelling". In: *Pearson Prentice Hall, Harlow,GB* (2001).

[130] H. Lee, L. Heo, M. S. Lee, and C. Seok. "GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization." In: *Nucleic acids research* 43 (W1 July 2015), W431–W435. ISSN: 1362-4962. DOI: 10.1093/nar/gkv495.

[131] H.-J. Lee, K.-J. Park, A. Y. Lee, S. G. Park, B. C. Park, K.-H. Lee, and S.-J. Park. "Regulation of fur expression by RpoS and fur in Vibrio vulnificus." In: *Journal of bacteriology* 185 (19 Oct. 2003), pp. 5891–5896. ISSN: 0021-9193.

[132] J. W. Little, D. W. Mount, and C. R. Yanisch-Perron. "Purified lexA protein is a repressor of the recA and lexA genes." In: *Proceedings of the National Academy of Sciences of the United States of America* 78 (7 July 1981), pp. 4199–4203. ISSN: 0027-8424.

[133] C. M. Litwin and S. B. Calderwood. "Analysis of the complexity of gene regulation by fur in Vibrio cholerae." In: *Journal of bacteriology* 176 (1 Jan. 1994), pp. 240–248. ISSN: 0021-9193.

[134] N. London, B. Raveh, E. Cohen, G. Fathi, and O. Schueler-Furman. "Rosetta FlexPepDock web server–high resolution modeling of peptide-protein interactions." In: *Nucleic acids research* 39 (Web Server issue July 2011), W249–W253. ISSN: 1362-4962. DOI: 10.1093/nar/gkr431.

[135] V de Lorenzo, S Wee, M Herrero, and J. B. Neilands. "Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor." In: *Journal of bacteriology* 169 (6 June 1987), pp. 2624–2630. ISSN: 0021-9193.

[136] V de Lorenzo, F Giovannini, M Herrero, and J. B. Neilands. "Metal ion regulation of gene expression. Fur repressor-operator interaction at the promoter region of the aerobactin system of pColV-K30." In: *Journal of molecular biology* 203 (4 Oct. 1988), pp. 875–884. ISSN: 0022-2836.

[137] D. M. and F. T. "Skript zur Mathematik fuer die Molekulare Biotechnologie". In: *Universitaet Heidelberg,* 3.8 (2003).

[138] MacKerell. *Internal coordinates for bonded interactons.* https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node25.html. Accessed: 2018-09-26.

[139] J. L. Martinez. "The role of natural environments in the evolution of resistance traits in pathogenic bacteria." In: *Proceedings. Biological sciences* 276 (1667 July 2009), pp. 2521–2530. ISSN: 0962-8452. DOI: 10.1098/rspb.2009.0320.

[140] J. L. Martinez, M. B. Sánchez, L. Martínez-Solano, A. Hernandez, L. Garmendia, A. Fajardo, and C. Alvarez-Ortega. "Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems." In: *FEMS microbiology reviews* 33 (2 Mar. 2009), pp. 430–449. ISSN: 1574-6976. DOI: 10.1111/j.1574-6976.2008.00157.x.

[141] E. Massé and S. Gottesman. "A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli." In: *Proceedings of the National Academy of Sciences of the United States of America* 99 (7 Apr. 2002), pp. 4620–4625. ISSN: 0027-8424. DOI: 10.1073/pnas.032066599.

[142] E. Massé, C. K. Vanderpool, and S. Gottesman. "Effect of RyhB small RNA on global iron use in Escherichia coli." In: *Journal of bacteriology* 187 (20 Oct. 2005), pp. 6962–6971. ISSN: 0021-9193. DOI: 10.1128/JB.187.20.6962-6971.2005.

[143] S. Mathieu, C. Cissé, S. Vitale, A. Ahmadova, M. Degardin, J. Pérard, P. Colas, R. Miras, D. Boturyn, J. Covès, S. Crouzy, and I. Michaud-Soret. "From Peptide Aptamers to Inhibitors of FUR, Bacterial Transcriptional Regulator of Iron Homeostasis and Virulence." In: *ACS chemical biology* 11 (9 Sept. 2016), pp. 2519–2528. ISSN: 1554-8937. DOI: 10.1021/acschembio.6b00360.

[144] P Mazodier and J Davies. "Gene transfer between distantly related bacteria." In: *Annual review of genetics* 25 (1991), pp. 147–171. ISSN: 0066-4197. DOI: 10.1146/annurev.ge.25.120191.001051.

[145] G. W. McCoy and C. W. Chapin. "Further Observations on a Plague-Like Disease of Rodents with a Preliminary Note on the Causative Agent, Bacterium tularense". In: *The Journal of Infectious Diseases* 10.1 (1912), pp. 61–72. ISSN: 00221899. URL: http://www.jstor.org/stable/30071893.

[146] J. P. McHugh, F. Rodríguez-Quinoñes, H. Abdul-Tehrani, D. A. Svistunenko, R. K. Poole, C. E. Cooper, and S. C. Andrews. "Global iron-dependent gene regulation in Escherichia coli. A new mechanism for iron homeostasis." In: *The Journal of biological chemistry* 278 (32 Aug. 2003), pp. 29478–29486. ISSN: 0021-9258. DOI: 10.1074/jbc.M303381200.

[147] D. McQuaerrie. "Statistical Mechanics". In: *University Science Books* Sansalito (2000).

[148] E. Merritt. *X-ray Anomalous Scattering*. http://skuld.bmsc.washington.edu/scatter/. Accessed: 2018-07-18.

[149] A. R. Mey, E. E. Wyckoff, V. Kanukurthy, C. R. Fisher, and S. M. Payne. "Iron and fur regulation in Vibrio cholerae and the role of fur in virulence." In: *Infection and immunity* 73 (12 Dec. 2005), pp. 8167–8178. ISSN: 0019-9567. DOI: 10.1128/IAI.73.12.8167-8178.2005.

[150] I Michaud-Soret, A Adrait, M Jaquinod, E Forest, D Touati, and J. M. Latour. "Electrospray ionization mass spectrometry analysis of the apo- and metal-substituted forms of the Fur protein." In: *FEBS letters* 413 (3 Aug. 1997), pp. 473–476. ISSN: 0014-5793.

[151] S. A. Mills and M. A. Marletta. "Metal binding characteristics and role of iron oxidation in the ferric uptake regulator from Escherichia coli." In: *Biochemistry* 44 (41 Oct. 2005), pp. 13553–13559. ISSN: 0006-2960. DOI: 10.1021/bi0507579.

[152] M. L. of Molecular Biology. *SCOP2*. http://scop2.mrc-lmb.cam.ac.uk/. Accessed: 2018-10-08.

[153] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function". In: *Journal of Computational Chemistry* 19.14 (1998), pp. 1639–1662. DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-987X%2819981115%2919%3A14%3C1639%3A%3AAID-JCC10%3E3.0.CO%3B2-B. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%2819981115%2919%3A14%3C1639%3A%3AAID-JCC10%3E3.0.CO%3B2-B.

[154] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility". In: *J Comput Chem* 30.16 (2009). 19399780[pmid], pp. 2785–2791. ISSN: 0192-8651. DOI: 10.1002/jcc.21256. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760638/.

[155] J. M. Munita and C. A. Arias. "Mechanisms of Antibiotic Resistance." In: *Microbiology spectrum* 4 (2 Apr. 2016). ISSN: 2165-0497. DOI: 10.1128/microbiolspec.VMBF-0016-2015.

[156] M. Newville. *Fundamentals of XAFS*. 2004. URL: http://xafs.org/Tutorials?action=AttachFile&do=get&target=Newville_xas_fundamentals.pdf.

[157] E. van Nood, A. Vrieze, M. Nieuwdorp, S. Fuentes, E. G. Zoetendal, W. M. de Vos, C. E. Visser, E. J. Kuijper, J. F. W. M. Bartelsman, J. G. P. Tijssen, P. Speelman, M. G. W. Dijkgraaf, and J. J. Keller. "Duodenal infusion of donor feces for recurrent Clostridium difficile." In: *The New England journal of medicine* 368 (5 Jan. 2013), pp. 407–415. ISSN: 1533-4406. DOI: 10.1056/NEJMoa1205037.

[158] S. Nose. "A unified formulation of the constant temperature molecular-dynamics methods". In: *J. Chem. Phys.* 81 (1984), p. 511.

[159] C. D. Okafor, K. A. Lanier, A. S. Petrov, S. S. Athavale, J. C. Bowman, N. V. Hud, and L. D. Williams. "Iron mediates catalysis of nucleic acid processing enzymes: support for Fe(II) as a cofactor before the great oxidation event." In: *Nucleic acids research* 45 (7 Apr. 2017), pp. 3634–3642. ISSN: 1362-4962. DOI: 10.1093/nar/gkx171.

[160] A. Onufriev, D. Bashford, and D. Case. "Effective Born radii in the generalized Born approximation: The importance of being perfect". In: *J Comp Chem* 23 (2002), pp. 1297–1304.

[161] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6". In: *J Comput Chem* 25.13 (2004), pp. 1656–1676.

[162] J. O'Neill. "Tackling drug-resistant infections globally: final report and recommendations". In: *The Review On Antimicrobial Resistance* (2016).

[163] L. Pallecchi, C. Lucchetti, A. Bartoloni, F. Bartalesi, A. Mantella, H. Gamboa, A. Carattoli, F. Paradisi, and G. M. Rossolini. "Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure." In: *Antimicrobial agents and chemotherapy* 51 (4 Apr. 2007), pp. 1179–1184. ISSN: 0066-4804. DOI: 10.1128/AAC.01101-06.

[164] L. D. Palmer and E. P. Skaar. "Transition Metals and Virulence in Bacteria." In: *Annual review of genetics* 50 (Nov. 2016), pp. 67–91. ISSN: 1545-2948. DOI: 10.1146/annurev-genet-120215-035146.

[165] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190. DOI: 10.1063/1.328693. eprint: http://dx.doi.org/10.1063/1.328693. URL: http://dx.doi.org/10.1063/1.328693.

[166] G. Patey and J. Valleau. " A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. " In: *J. Chem. Phys.* 63 (1975), pp. 2334–2339.

[167] L. Pecqueur, B. D'Autréaux, J. Dupuy, Y. Nicolet, L. Jacquamet, B. Brutscher, I. Michaud-Soret, and B. Bersch. "Structural changes of Escherichia coli ferric uptake regulator during metal-dependent dimerization and activation explored by NMR and X-ray crystallography." In: *The Journal of biological chemistry* 281 (30 July 2006), pp. 21286–21295. ISSN: 0021-9258. DOI: 10.1074/jbc.M601278200.

[168] J. Pérard and S. Ollagnier de Choudens. "Iron–sulfur clusters biogenesis by the SUF machinery: close to the molecular mechanism understanding". In: *JBIC Journal of Biological Inorganic Chemistry* 23.4 (2018), pp. 581–596. ISSN: 1432-1327. DOI: 10.1007/s00775-017-1527-3. URL: https://doi.org/10.1007/s00775-017-1527-3.

[169] A Gonzalez de Peredo, C Saint-Pierre, J. M. Latour, I Michaud-Soret, and E Forest. "Conformational changes of the ferric uptake regulation protein upon metal activation and DNA binding; first evidence of structural homologies with the diphtheria toxin repressor." In: *Journal of molecular biology* 310 (1 June 2001), pp. 83–91. ISSN: 0022-2836. DOI: 10.1006/jmbi.2001.4769.

[170] O. Q. Pich, B. M. Carpenter, J. J. Gilbreath, and D. S. Merrell. "Detailed analysis of Helicobacter pylori Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes." In: *Molecular microbiology* 84 (5 June 2012), pp. 921–941. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2012.08066.x.

[171] E. Pohl, J. C. Haller, A. Mijovilovich, W. Meyer-Klaucke, E. Garman, and M. L. Vasil. "Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator." In: *Molecular microbiology* 47 (4 Feb. 2003), pp. 903–915. ISSN: 0950-382X.

[172] G. Porod. "Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen". In: *Kolloid-Zeitschrift* 124.2 (1951), pp. 83–114. ISSN: 1435-1536. DOI: 10.1007/BF01512792. URL: https://doi.org/10.1007/BF01512792.

[173] C. W. Price, W. A. Randall, V. L. Chandler, and R. J. Reedy. "Observations on the in Vivo and in Vitro Development of Bacterial Resistance to Streptomycin." In: *Journal of bacteriology* 53 (4 Apr. 1947), pp. 481–488. ISSN: 0021-9193.

[174] R. W. Prince, C. D. Cox, and M. L. Vasil. "Coordinate regulation of siderophore and exotoxin A production: molecular cloning and sequencing of the Pseudomonas aeruginosa fur gene." In: *Journal of bacteriology* 175 (9 May 1993), pp. 2589–2598. ISSN: 0021-9193.

[175] O. Proux, V. Nassif, A. Prat, O. Ulrich, E. Lahera, X. Biquard, J. J. Menthonnex, and J. L. Hazemann. "Feedback system of a liquid-nitrogen-cooled double-crystal monochromator: design and performances." In: *Journal of synchrotron radiation* 13 (Pt 1 Jan. 2006), pp. 59–68. ISSN: 0909-0495. DOI: 10.1107/S0909049505037441.

[176] K. Prévost, H. Salvail, G. Desnoyers, J.-F. Jacques, E. Phaneuf, and E. Massé. "The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis." In: *Molecular microbiology* 64 (5 June 2007), pp. 1260–1273. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2007.05733.x.

[177] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer. "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." In: *Quarterly reviews of biophysics* 40 (3 Aug. 2007), pp. 191–285. ISSN: 0033-5835. DOI: 10.1017/S0033583507004635.

[178] J. Pérard, S. Nader, M. Levert, L. Arnaud, P. Carpentier, C. Siebert, F. Blanquet, C. Cavazza, P. Renesto, D. Schneider, M. Maurin, J. Coves, S. Crouzy, and I. Michaud-Soret. "Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in Francisella tularensis virulence". In: *Communications Biology* 1.1 (July 2018), p. 93. ISSN: 2399-3642. URL: https://doi.org/10.1038/s42003-018-0095-6.

[179] J. Pérard, J. Covès, M. Castellan, C. Solard, M. Savard, R. Miras, S. Galop, L. Signor, S. Crouzy, I. Michaud-Soret, and E. de Rosny. "Quaternary Structure of Fur Proteins, a New Subfamily of Tetrameric Proteins." In: *Biochemistry* 55 (10 Mar. 2016), pp. 1503–1515. ISSN: 1520-4995. DOI: 10.1021/acs.biochem.5b01061.

[180] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: http://www.R-project.org/.

[181] B Ravel and M Newville. "ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT." In: *Journal of synchrotron radiation* 12 (Pt 4 July 2005), pp. 537–541. ISSN: 0909-0495. DOI: 10.1107/S0909049505012719.

[182] R. B. Rea, C. G. M. Gahan, and C. Hill. "Disruption of Putative Regulatory Loci in Listeria monocytogenes Demonstrates a Significant Role for Fur and PerR in Virulence". In: *Infection and Immunity* 72.2 (2004), pp. 717–727. ISSN: 0019-9567. DOI: 10.1128/IAI.72.2.717-727.2004. eprint: https://iai.asm.org/content/72/2/717.full.pdf. URL: https://iai.asm.org/content/72/2/717.

[183] J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen. "Parameter-free calculations of X-ray spectra with FEFF9." In: *Physical chemistry chemical physics : PCCP* 12 (21 June 2010), pp. 5503–5513. ISSN: 1463-9084. DOI: 10.1039/b926434e.

[184] Hampton Research. *Hanging Drop Vapor Diffusion Crystallization*. 2017. URL: https://hamptonresearch.com/documents/growth_101/3.pdf.

[185] Hampton Research. *Sitting Drop Vapor Diffusion Crystallization*. 2017. URL: https://hamptonresearch.com/documents/growth_101/4.pdf.

[186] The Regional Office for Europe of the World Health Organization. "Central Asian and Eastern European Surveillance of Antimicrobial Resistance". In: *Annual report* (2016).

[187] The Regional Office for Europe of the World Health Organization. "Central Asian and Eastern European Surveillance of Antimicrobial Resistance". In: *Annual report* (2017).

[188] World Health Organization. "Global action plan on Antimicrobial Resistance". In: *WHO Library Cataloguing-in-Publication Data* (2015).

[189] World Health Organization. "Global Antimicrobial Resistance Surveillance System (GLASS) Report". In: (2015).

[190] F. Richards. "Areas, volumes, packing, and protein structure." In: *Ann. Rev. Biophys. Bioeng.* 6 (1977), pp. 151–176.

[191] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann. "Origins of specificity in protein-DNA recognition." In: *Annual review of biochemistry* 79 (2010), pp. 233–269. ISSN: 1545-4509. DOI: 10.1146/annurev-biochem-060408-091030.

[192] B. Roux and T. Simonson. "Implicit Solvent Models". In: *Biophys. Chem.* 78 (1999), pp. 1–20.

[193] D. M. Ruden, J Ma, Y Li, K Wood, and M Ptashne. "Generating yeast transcriptional activators containing no yeast protein sequences." In: *Nature* 350 (6315 Mar. 1991), pp. 250–252. ISSN: 0028-0836. DOI: 10.1038/350250a0.

[194] M. J. Russell, R. M. Daniel, A. J. Hall, and J. A. Sherringham. "A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life". In: *Journal of Molecular Evolution* 39.3 (1994), pp. 231–243.

[195] T Saito, M. R. Wormald, and R. J. Williams. "Some structural features of the iron-uptake regulation protein." In: *European journal of biochemistry* 197 (1 Apr. 1991), pp. 29–38. ISSN: 0014-2956.

[196] S. Sarvan, F. Charih, M. Askoura, J. Butcher, J. S. Brunzelle, A. Stintzi, and J.-F. Couture. "Functional insights into the interplay between DNA interaction and metal coordination in ferric uptake regulators." In: *Scientific reports* 8 (1 May 2018), p. 7140. ISSN: 2045-2322. DOI: 10.1038/s41598-018-25157-6.

[197] I. J. Schalk. "Innovation and originality in the strategies developed by bacteria to get access to iron." In: *Chembiochem : a European journal of chemical biology* 14 (3 Feb. 2013), pp. 293–294. ISSN: 1439-7633. DOI: 10.1002/cbic.201200738.

[198] I. J. Schalk and L. Guillon. "Pyoverdine biosynthesis and secretion in Pseudomonas aeruginosa: implications for metal homeostasis." In: *Environmental microbiology* 15 (6 June 2013), pp. 1661–1673. ISSN: 1462-2920. DOI: 10.1111/1462-2920.12013.

[199] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren. "Definition and testing of the GROMOS force-field versions 54A7 and 54B7". In: *European Biophysics Journal* 40.7 (2011), p. 843. ISSN: 1432-1017. DOI: 10.1007/s00249-011-0700-9. URL: http://dx.doi.org/10.1007/s00249-011-0700-9.

[200] S Schäffer, K Hantke, and V Braun. "Nucleotide sequence of the iron regulatory gene fur." In: *Molecular & general genetics : MGG* 200 (1 1985), pp. 110–113. ISSN: 0026-8925.

[201] M. A. Sheikh and G. L. Taylor. "Crystal structure of the Vibrio cholerae ferric uptake regulator (Fur) reveals insights into metal co-ordination." In: *Molecular microbiology* 72 (5 June 2009), pp. 1208–1220. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2009.06718.x.

[202] M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande. "Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations". In: *The Journal of Physical Chemistry B* 111.45 (2007). PMID: 17949030, pp. 13052–13063. DOI: 10.1021/jp0735987. eprint: https://doi.org/10.1021/jp0735987. URL: https://doi.org/10.1021/jp0735987.

[203] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." In: *Molecular systems biology* 7 (Oct. 2011), p. 539. ISSN: 1744-4292. DOI: 10.1038/msb.2011.75.

[204] E. P. Skaar. "The battle for iron between bacterial pathogens and their vertebrate hosts." In: *PLoS pathogens* 6 (8 Aug. 2010), e1000949. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1000949.

[205] T. M. Staggs and R. D. Perry. "Identification and cloning of a fur regulatory gene in Yersinia pestis." In: *Journal of bacteriology* 173 (2 Jan. 1991), pp. 417–425. ISSN: 0021-9193.

[206] P. S. Stewart and J. W. Costerton. "Antibiotic resistance of bacteria in biofilms." In: *Lancet (London, England)* 358 (9276 July 2001), pp. 135–138. ISSN: 0140-6736.

[207] W. Still, A. Tempczyk, R. Hawley, and T. Hendrickson. " Semianalytical treatment of solvation for molecular mechanics and dynamics. " In: *J. Am. Chem. Soc.* 112 (1990), pp. 6127–6129.

[208] I Stojiljkovic and K Hantke. "Functional domains of the Escherichia coli ferric uptake regulator protein (Fur)." In: *Molecular & general genetics : MGG* 247 (2 Apr. 1995), pp. 199–205. ISSN: 0026-8925.

[209] R. Stote and D. States. "On the Treatment of Electrostatic Interactions in Biomolecular Simulation". In: *Journal de Chimie Physique* 88 (1991), p. 2419.

[210] R. W. Strange, N. J. Blackburn, P. F. Knowles, and S. S. Hasnain. "X-ray absorption spectroscopy of metal-histidine coordination in metalloproteins. Exact simulation of the EXAFS of tetrabis(imidazole)copper(II) nitrate and other copper-imidazole complexes by the use of a multiple-scattering treatment". In: *Journal of the American Chemical Society* 109.23 (1987), pp. 7157–7162. DOI: 10.1021/ja00257a042. eprint: https://doi.org/10.1021/ja00257a042. URL: https://doi.org/10.1021/ja00257a042.

[211] H. J. Strutt. "LVIII. On the scattering of light by small particles". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41.275 (1871), pp. 447–454. DOI: 10.1080/14786447108640507. eprint: https://doi.org/10.1080/14786447108640507. URL: https://doi.org/10.1080/14786447108640507.

[212] D. I. Svergun and M. H. J. Koch. "Small-angle scattering studies of biological macromolecules in solution". In: *Reports on Progress in Physics* 66 (Oct. 2003), pp. 1735–1782. DOI: 10.1088/0034-4885/66/10/R05. URL: http://adsabs.harvard.edu/abs/2003RPPh...66.1735S.

[213] A Theisen. *Refractive increment data-book for polymer and biomolecular scientists*. Nottingham University Press, 2000. ISBN: 1-897676-29-8. URL: https://www.researchgate.net/profile/Stephen_Harding5/publication/300717603_Refractive_Index_Data_Book/links/5968ef88458515e9afa7a48b/Refractive-Index-Data-Book.pdf.

[214] B. M. Tissue. *Tissue 2018, Harmonic Oscillator*. https://www.tissuegroup.chem.vt.edu/chem-ed/quantum/harmonic-oscillator.html. Accessed: 2018-09-26.

[215] V. J. Torres, A. S. Attia, W. J. Mason, M. I. Hood, B. D. Corbin, F. C. Beasley, K. L. Anderson, D. L. Stauff, W. H. McDonald, L. J. Zimmerman, D. B. Friedman, D. E. Heinrichs, P. M. Dunman, and E. P. Skaar. "Staphylococcus aureus fur regulates the expression of virulence factors that contribute to the pathogenesis of pneumonia." In: *Infection and immunity* 78 (4 Apr. 2010), pp. 1618–1628. ISSN: 1098-5522. DOI: 10.1128/IAI.01423-09.

[216] G. M. Torrie and J. P. Valleau. "Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid." In: *Chem. Phys. Let.* 28 (1974).

[217] D Touati, M Jacques, B Tardat, L Bouchard, and S Despied. "Lethal oxidative damage and mutagenesis are generated by iron in delta fur mutants of Escherichia coli: protective role of superoxide dismutase." In: *Journal of bacteriology* 177 (9 May 1995), pp. 2305–2314. ISSN: 0021-9193.

[218] O. Trott and A. J. Olson. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." In: *Journal of computational chemistry* 31 (2 Jan. 2010), pp. 455–461. ISSN: 1096-987X. DOI: 10.1002/jcc.21334.

[219] B. Troxell, M. L. Sikes, R. C. Fink, A. Vazquez-Torres, J. Jones-Carson, and H. M. Hassan. "Fur negatively regulates hns and is required for the expression of HilA and virulence in Salmonella enterica serovar Typhimurium." In: *Journal of bacteriology* 193 (2 Jan. 2011), pp. 497–505. ISSN: 1098-5530. DOI: 10.1128/JB.00942-10.

[220] A. D. Tsaousis, E. Gentekaki, L. Eme, D. Gaston, and A. J. Roger. "Evolution of the cytosolic iron-sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes." In: *Eukaryotic cell* 13 (1 Jan. 2014), pp. 143–153. ISSN: 1535-9786. DOI: 10.1128/EC.00158-13.

[221] L. Valen. "Molecular evolution as predicted by natural selection". In: *Journal of Molecular Evolution* 3.2 (June 1, 1974), p. 89. ISSN: 1432-1432. DOI: 10.1007/BF01796554. URL: http://dx.doi.org/10.1007/BF01796554.

[222] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen. "GROMACS: fast, flexible, and free". In: *J Comput Chem* 26.16 (2005), pp. 1701–1718.

[223] L. Verlet. "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules." In: *Phys. Rev.* 159(1) (1967), pp. 98–103.

[224] G. Veronesi. "Advanced applications of X-ray absorption spectroscopy to the study of protein metal sites". PhD thesis. School of Physics, University of Bologna, 2010. URL: http://amsdottorato.unibo.it/2535/1/Veronesi_Giulia_Tesi.pdf.

[225] S. Vitale. "Inhibiteurs anti-Fur et caractérisation de Fur d'Helicobacter pylori". PhD thesis. UNIVERSITE JOSEPH FOURIER – GRENOBLE I, 2009.

[226] S. Vitale, C. Fauquant, D. Lascoux, K. Schauer, C. Saint-Pierre, and I. Michaud-Soret. "A ZnS(4) structural zinc site in the Helicobacter pylori ferric uptake regulator." In: *Biochemistry* 48 (24 June 2009), pp. 5582–5591. ISSN: 1520-4995. DOI: 10.1021/bi9004396.

[227] G. Wächtershäuser. "Pyrite formation, the first energy source for life: a hypothesis". In: *Systematic and Applied Microbiology* 10.3 (1988), pp. 207–210.

[228] K. J. Waldron and N. J. Robinson. "How do bacterial cells ensure that metalloproteins get the correct metal?" In: *Nature reviews. Microbiology* 7 (1 Jan. 2009), pp. 25–35. ISSN: 1740-1534. DOI: 10.1038/nrmicro2057.

[229] K. J. Waldron, J. C. Rutherford, D. Ford, and N. J. Robinson. "Metalloproteins and metal sensing." In: *Nature* 460 (7257 Aug. 2009), pp. 823–830. ISSN: 1476-4687. DOI: 10.1038/nature08300.

[230] F. Wang, S. Cheng, K. Sun, and L. Sun. "Molecular analysis of the fur (ferric uptake regulator) gene of a pathogenic Edwardsiella tarda strain." In: *Journal of microbiology (Seoul, Korea)* 46 (3 June 2008), pp. 350–355. ISSN: 1225-8873. DOI: 10.1007/s12275-008-0038-x.

[231] P. I. Watnick, T Eto, H Takahashi, and S. B. Calderwood. "Purification of Vibrio cholerae fur and estimation of its intracellular abundance by antibody sandwich enzyme-linked immunosorbent assay." In: *Journal of bacteriology* 179 (1 Jan. 1997), pp. 243–247. ISSN: 0021-9193.

[232] E. M. H. Wellington, A. B. Boxall, P. Cross, E. J. Feil, W. H. Gaze, P. M. Hawkey, A. S. Johnson-Rollings, D. L. Jones, N. M. Lee, W. Otten, C. M. Thomas, and A. P. Williams. "The role of the natural environment in the emergence of antibiotic resistance in gram-negative bacteria." In: *The Lancet. Infectious diseases* 13 (2 Feb. 2013), pp. 155–165. ISSN: 1474-4457. DOI: 10.1016/S1473-3099(12)70317-1.

[233] P. J. Wilderman, N. A. Sowa, D. J. FitzGerald, P. C. FitzGerald, S. Gottesman, U. A. Ochsner, and M. L. Vasil. "Identification of tandem duplicate regulatory small RNAs in Pseudomonas aeruginosa involved in iron homeostasis." In: *Proceedings of the National Academy of Sciences of the United States of America* 101 (26 June 2004), pp. 9792–9797. ISSN: 0027-8424. DOI: 10.1073/pnas.0403423101.

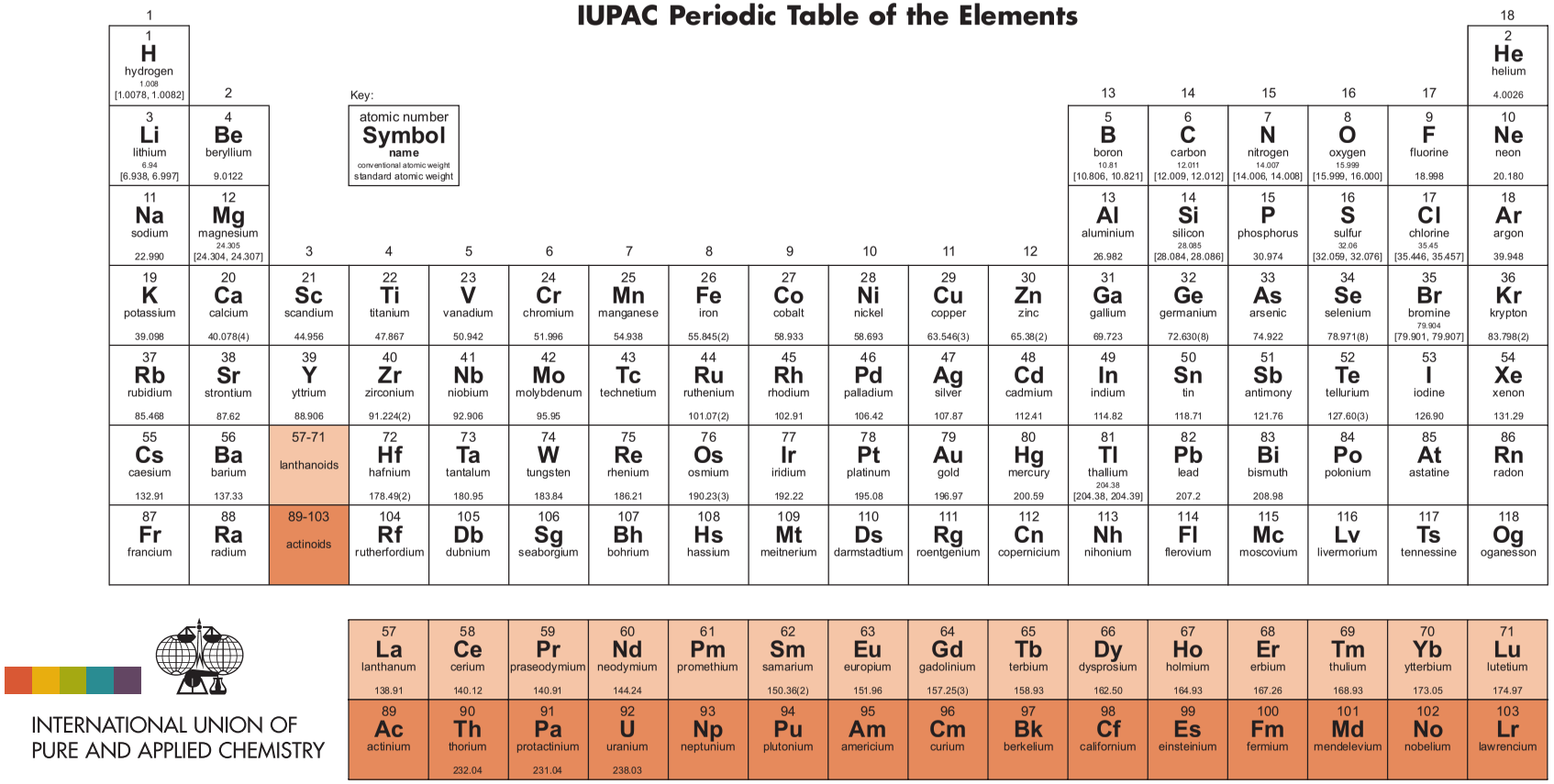
[234] R. J. P. Williams. "Iron in evolution." In: *FEBS letters* 586 (5 Mar. 2012), pp. 479–484. ISSN: 1873-3468. DOI: 10.1016/j.febslet.2011.05.068.

[235] R. J. P. Williams and J. J. R. Fraústo Da Silva. "Evolution was chemically constrained." In: *Journal of theoretical biology* 220 (3 Feb. 2003), pp. 323–343. ISSN: 0022-5193.

[236] C. R. Woese, O Kandler, and M. L. Wheelis. "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." In: *Proceedings of the National Academy of Sciences of the United States of America* 87 (12 June 1990), pp. 4576–4579. ISSN: 0027-8424.

[237] P. J. Wyatt. "Light scattering and the absolute characterization of macromolecules". In: *Analytica Chimica Acta* 272.1 (1993), pp. 1 –40. ISSN: 0003-2670. DOI: https://doi.org/10.1016/0003-2670(93)80373-S. URL: http://www.sciencedirect.com/science/article/pii/000326709380373S.

[238] E. E. Zheleznova, J. H. Crosa, and R. G. Brennan. "Characterization of the DNA-and metal-binding properties of Vibrio anguillarum Fur reveals conservation of a structural Zn2+ ion". In: *Journal of bacteriology* 182.21 (2000), pp. 6264–6267.



IUPAC Periodic Table of the Elements

1																	18
1 H hydrogen 1.008 [1.0078, 1.0082]	2		Key: atomic number **Symbol** **name** conventional atomic weight standard atomic weight									13	14	15	16	17	2 He helium 4.0026
3 Li lithium 6.94 [6.938, 6.997]	4 Be beryllium 9.0122											5 B boron 10.81 [10.806, 10.821]	6 C carbon 12.011 [12.009, 12.012]	7 N nitrogen 14.007 [14.006, 14.008]	8 O oxygen 15.999 [15.999, 16.000]	9 F fluorine 18.998	10 Ne neon 20.180
11 Na sodium 22.990	12 Mg magnesium 24.305 [24.304, 24.307]	3	4	5	6	7	8	9	10	11	12	13 Al aluminium 26.982	14 Si silicon 28.085 [28.084, 28.086]	15 P phosphorus 30.974	16 S sulfur 32.06 [32.059, 32.076]	17 Cl chlorine 35.45 [35.446, 35.457]	18 Ar argon 39.948
19 K potassium 39.098	20 Ca calcium 40.078(4)	21 Sc scandium 44.956	22 Ti titanium 47.867	23 V vanadium 50.942	24 Cr chromium 51.996	25 Mn manganese 54.938	26 Fe iron 55.845(2)	27 Co cobalt 58.933	28 Ni nickel 58.693	29 Cu copper 63.546(3)	30 Zn zinc 65.38(2)	31 Ga gallium 69.723	32 Ge germanium 72.630(8)	33 As arsenic 74.922	34 Se selenium 78.971(8)	35 Br bromine 79.904 [79.901, 79.907]	36 Kr krypton 83.798(2)
37 Rb rubidium 85.468	38 Sr strontium 87.62	39 Y yttrium 88.906	40 Zr zirconium 91.224(2)	41 Nb niobium 92.906	42 Mo molybdenum 95.95	43 Tc technetium	44 Ru ruthenium 101.07(2)	45 Rh rhodium 102.91	46 Pd palladium 106.42	47 Ag silver 107.87	48 Cd cadmium 112.41	49 In indium 114.82	50 Sn tin 118.71	51 Sb antimony 121.76	52 Te tellurium 127.60(3)	53 I iodine 126.90	54 Xe xenon 131.29
55 Cs caesium 132.91	56 Ba barium 137.33	57-71 lanthanoids	72 Hf hafnium 178.49(2)	73 Ta tantalum 180.95	74 W tungsten 183.84	75 Re rhenium 186.21	76 Os osmium 190.23(3)	77 Ir iridium 192.22	78 Pt platinum 195.08	79 Au gold 196.97	80 Hg mercury 200.59	81 Tl thallium 204.38 [204.38, 204.39]	82 Pb lead 207.2	83 Bi bismuth 208.98	84 Po polonium	85 At astatine	86 Rn radon
87 Fr francium	88 Ra radium	89-103 actinoids	104 Rf rutherfordium	105 Db dubnium	106 Sg seaborgium	107 Bh bohrium	108 Hs hassium	109 Mt meitnerium	110 Ds darmstadtium	111 Rg roentgenium	112 Cn copernicium	113 Nh nihonium	114 Fl flerovium	115 Mc moscovium	116 Lv livermorium	117 Ts tennessine	118 Og oganesson

57 La lanthanum 138.91	58 Ce cerium 140.12	59 Pr praseodymium 140.91	60 Nd neodymium 144.24	61 Pm promethium	62 Sm samarium 150.36(2)	63 Eu europium 151.96	64 Gd gadolinium 157.25(3)	65 Tb terbium 158.93	66 Dy dysprosium 162.50	67 Ho holmium 164.93	68 Er erbium 167.26	69 Tm thulium 168.93	70 Yb ytterbium 173.05	71 Lu lutetium 174.97
89 Ac actinium	90 Th thorium 232.04	91 Pa protactinium 231.04	92 U uranium 238.03	93 Np neptunium	94 Pu plutonium	95 Am americium	96 Cm curium	97 Bk berkelium	98 Cf californium	99 Es einsteinium	100 Fm fermium	101 Md mendelevium	102 No nobelium	103 Lr lawrencium

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

For notes and updates to this table, see www.iupac.org. This version is dated 28 November 2016.

Figure 8.1: The latest release of the Periodic Table (dated 28 November 2016) by the IUPAC (*Periodic Table of elements*).

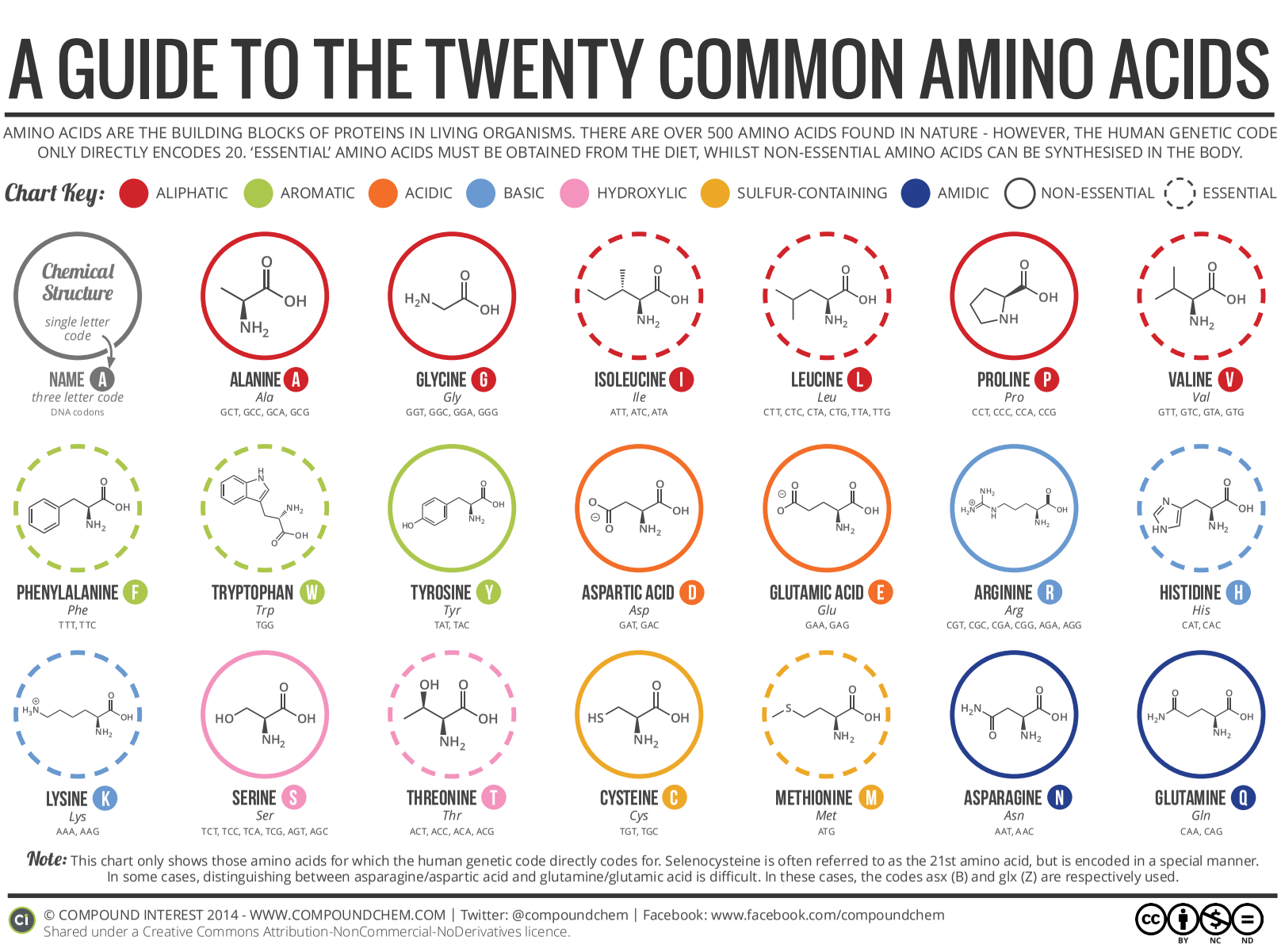


Figure 8.2: The common amino acids found in living organisms (*A Brief Guide to the Twenty Common Amino Acids*).

Notes

Notes

Notes

Notes

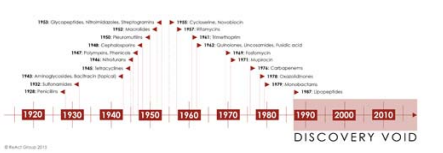
Structural studies on inhibition mechanisms, oligomerization and DNA binding of the transcription regulator Fur :
from *in silico* simulations to *in vitro* biological assays

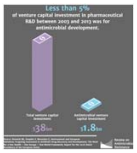
PhD supervisors: Serge Crouzy & Julien Pérard

Laboratoire de Chimie et Biologie des Métaux

PhD Grant

Why aren't we discovering new antibiotics ?





Ideal target = New therapeutic targets + Weak evolutionary pressure + Unique to bacteria + Virulence

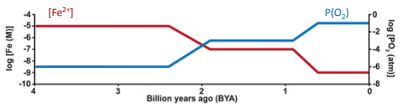
Antimicrobial Resistance : a worldwide threat

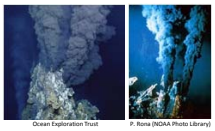
Europe : 25000 deaths/year ; **France** : Budget overrun of 1.5 billion €/year

"With increasing **resistance**, society could return to the conditions of a **pre-antibiotic era**" - WHO 2014

O'Neill, The Review On Antimicrobial Resistance, 2014

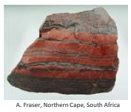
Iron and life on earth





Reducing environment
Abundant [Fe^{2+}]

Great oxygenation events

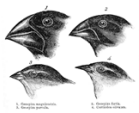


Oxidizing environment
Precipitation of [Fe^{3+}]

Natural selection acted in favor of tight iron regulation

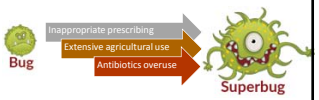
Boyd et al., Biochemistry, 2014

Resistance is a product of adaptation !



Darwin's finches

Uroplatus ebenaui



Holmes et al., Lancet, 2016

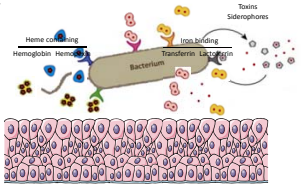
The battle for iron

Host extracellular [Fe] = 10^{-9} µM

Nutritional immunity

Bacteria needs [Fe] = 180 µM

Iron piracy



An interesting target is the Ferric Uptake Regulator protein

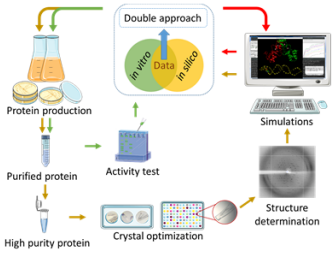
Barber et al., Trends Genet, 2015

The Ferric Uptake Regulator protein

Consensus Fur box : 5′-GATAATGATAATCATTATC-3′
3′-CTATTACTATTAGTAATAG-5′

M. F. Fillat. Archives of biochemistry and biophysics. 2014

Main focus and objectives

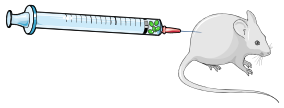


I. Structural studies on inhibition

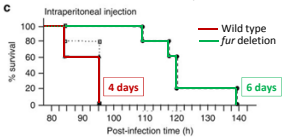
II. Structures as a tool to simulate Fur oligomers

Fur and virulence

- *H. pylori*
- *S. enterica*
- *C. jejuni*
- *S. aureus*
- *V. Cholerae*
- *L. Monocytogenes*
- *N. meningitides*
- ***F. tularensis***



CHU de Grenoble
In vivo virulence assay in mice

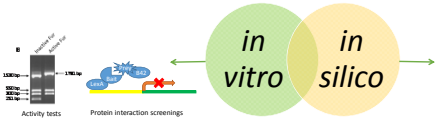


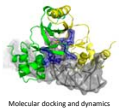
Fur inhibition, like *fur* deletion should decrease bacterial virulence

Dicard, Nader et al. Communications Biology. 2018

I. Structural studies on inhibition

Discovery of Fur inhibitors

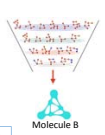




Library : 20 million peptides 13 amino acid long

In silico design

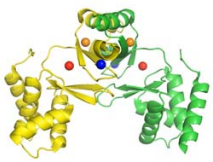
pL1 and pL2



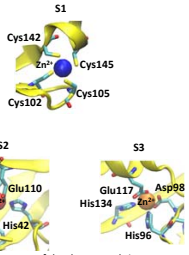
Identification of 6 peptide sequences and a chemical compound capable of interacting with Fur from *E.coli*

N. Abed et al., Molecular & Cellular Proteomics. 2007 ; C. Cissé et al., ACS chemical biology. 2014 ; S. Mathieu et al., ACS chemical biology. 2016

An example of Fur structure



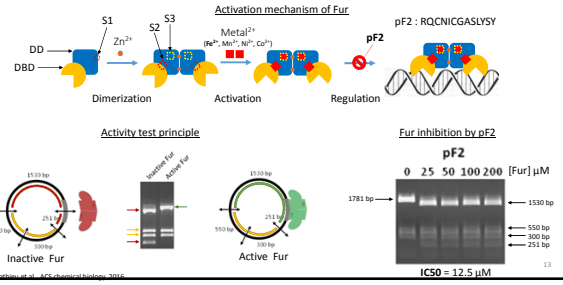
Structure of Fur from *H. pylori* (PDB ID: 2XIG), in the presence of zinc



Structure of the three metal sites

Zhao et al., Molecular microbiology. 2011

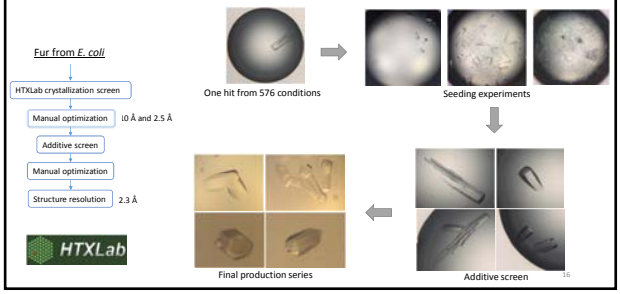
Activation mechanism and inhibition of Fur



Mathieu et al., ACS chemical biology, 2016

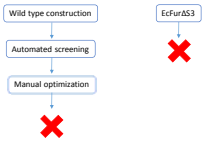
13

Crystallization of EcFur-140



16

Crystallization trials on Fur from *E.coli*



EcFurΔS3

20 years of failed crystallization attempts

Search for the most stable construction

Tested constructions

EcFur construction	C-terminal domain
EcFur wild type	. . . RLTNHSLYLYGHCAEGDCREDEHAHEGK
EcFur-125	. . . RLTNH**XXXXXXXXXX**REDEHAHEGK
EcFur-131	. . . RLTNHSLYLYG**XXXXXXXXX**EDEHAHEGK
EcFur-132	. . . RLTNHSLYLYGH**XXXXXXXXX**EHAHEGK
EcFur-133	. . . RLTNHSLYLYGHC**XXXXXXXXX**HAHEGK
EcFur-136	. . . RLTNHSLYLYGHCAEG**XXXXXXXXX**HEGK
EcFur-140	. . . RLTNHSLYLYGHCAEGDCRE**XXXXXXXXX**EGK

Thermal stability assays

EcFur construction	T_m (°C)
EcFur wild type	60
EcFurΔS3	59
EcFur-140	65

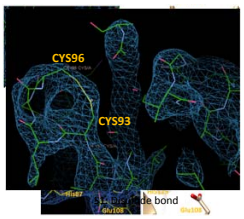
Removing the C-terminal flexible domain increases thermal stability

Is the EcFur-140 construction active ?

14

The first structure of Fur from *E. coli*

Structure of EcFur in the presence of Zn and Ni



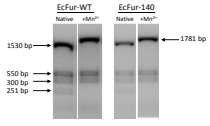
S3: 2 histidines, 1 glutamate and 1aspartate

Christine Cavazza

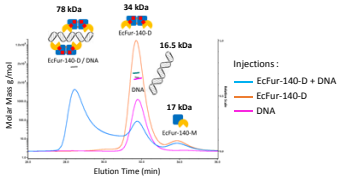
17

Is the EcFur-140 construction active ?

Activity test



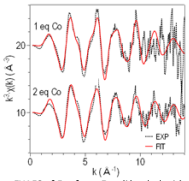
SEC-MALLS characterization



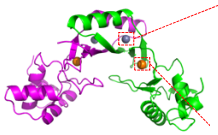
EcFur is active and bind DNA like the wild type protein

15

XAS studies on metal sites



EXAFS of Fur from *E. coli* loaded with **cobalt**



Structure of Fur from *E. coli*

XAS data confirm the unusual site S2 and proposes an additional histidine in site S3

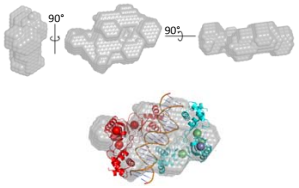
Guilia Veronesi

18

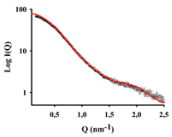
FAME-BM30B : J.-L. Hazemann, O. Proux, D. Testemale

In solution characterization of the EcFur-DNA complex by SAXS

EcFur-WT + DNA SAXS envelope



Scattering curve of EcFur-WT fitted with the curve of the model



Model of two EcFur140 dimers + DNA ; X^2 = 1.1

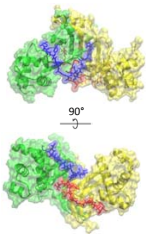
EcFur-140 is a good tool to study EcFur-WT inhibition

19

II. Structures as a tool to simulate Fur oligomers

22

Docking of pF2 on EcFur140



Strong binding affinity
ΔG = −19.8 kcal/mol

Docking (Autodock)
Fixed protein; Flexible peptide

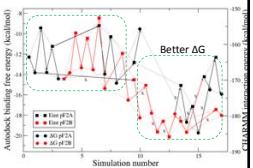
ΔG

Molecular Dynamics (CHARMM)
Protein restrained to crystal structure, flexible side chains; Flexible peptide

E_{int}

Binding pocket

Workflow

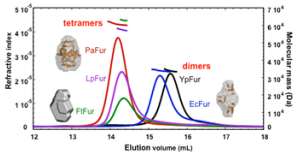


Convergence of docking cycles of pF2 on EcFur-140

Mathieu et al., ACS chemical biology, 2016

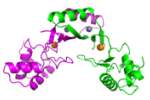
20

Rethinking the dimeric Fur dogma



SEC-MALLS-RI analysis of Fur proteins from pathogenes

Tetrameric Fur from *F. tularensis* (PDB ID: 5NHK)



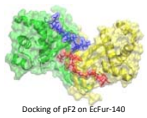
Dimeric Fur from *E. coli*

Tetrameric states exist in solution for certain bacterial species

Pérard *et al.*, Biochemistry, 2016; Pérard, Nader et al., Communications Biology, 2018.

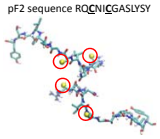
23

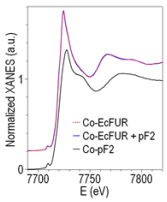
Validation of the inhibition pocket



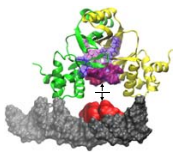
Docking of pF2 on EcFur-140

pF2 sequence RQCNICGASLYSY





XANES spectra of EcFur loaded with cobalt



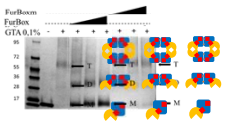
Proposed inhibition mechanism

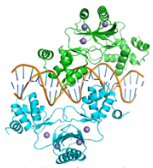
The inhibition by pF2 does not take place through perturbation of metal sites

21

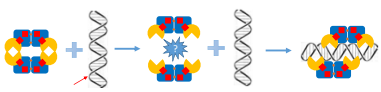
How do tetramers bind DNA ?

Cross-link assay with glutaraldehyde on SDS PAGE





M. gryphiswaldense: Fur/DNA complex (PDB ID: 4RB1)

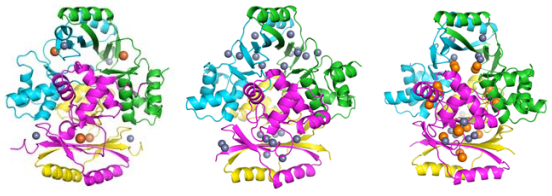


Deng et al., Nature Communications, 2015
Pérard, Nader et al., Communications Biology, 2018

Specific DNA

24

Tetrameric Fur structures solved by our teams



F. Tularensis (Fe), 1.7 Å (PDB ID: 5NHK)

P. aeruginosa (Zn), 2.2 Å PaFurΔS3

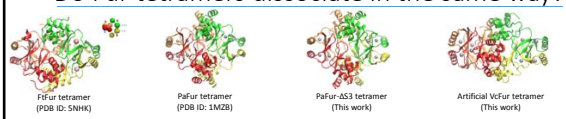
P. aeruginosa (Mn), 2.6 Å PaFurΔS3

Which complex is more stable ?

Christine Cavazza, Philippe Carpentier
Pérard, Nader et al., Communications Biology, 2018

25

Do Fur tetramers dissociate in the same way?



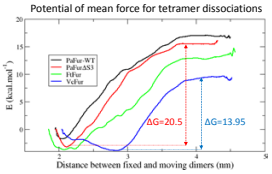
FtFur tetramer (PDB ID: 5NHK)

PaFur tetramer (PDB ID: 1MZB)

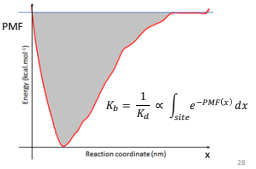
PaFur-ΔS3 tetramer (This work)

Artificial VcFur tetramer (This work)

Potential of mean force for tetramer dissociations



PMF

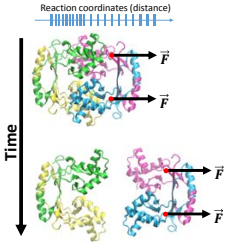


$$K_b = \frac{1}{K_d} \propto \int_{site} e^{-PMF(x)}\,dx$$

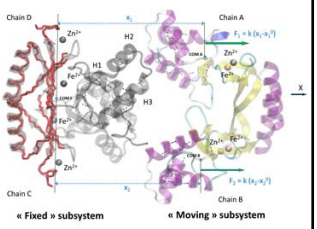
Pohl et al., Molecular microbiology, 2003; Pérard, Nader et al., Communications Biology, 2018

28

Simulating the dissociation



Free energy calculation by the "umbrella sampling" technique



Detailed forces and restraints

26

Are the interaction hotspots shared between species ?

Visualizing the results on sequence alignments

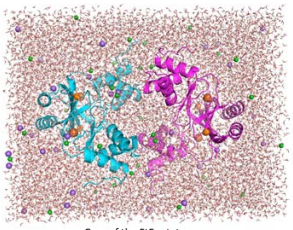
Conserved residues

Tetramer stabilization

Detected residues are shared between species

29

Building the systems

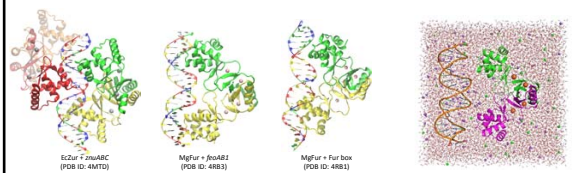


Case of the FtFur tetramer

Gromos force field I energy minimization and MD)
NPT equilibration (100 ps) and production (50 ns/window) : 1 atm, 310 K
Counter ions: NaCl, 0.1 mol.L⁻¹, system charge = zero
Solvent: explicit water (SPC)
Simulation box (periodic in all dimensions (PBC box)
Systems constructed using GROMACS tools (gmx editconf)
Box extended in Fe-allow penetration (20Å added in +X)

27

How do Fur proteins bind DNA ?



EcZur + *znuABC* (PDB ID: 4MTD)

MgFur + *feoAB1* (PDB ID: 4RB3)

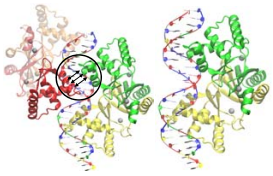
MgFur + Fur box (PDB ID: 4RB1)

Simulating a Fur/DNA complex

Gilston et al., PLOS Biology, 2014.
Deng et al., Nature Communications, 2015

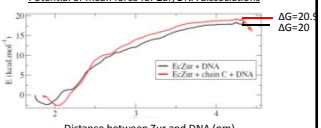
30

Impact of cooperative binding



Zur/DNA models used during the simulations.
(based on PDB ID: 4MTD)

Potential of mean force for Zur/DNA dissociations

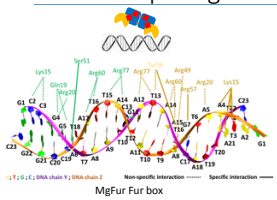


Slight difference between the two profiles

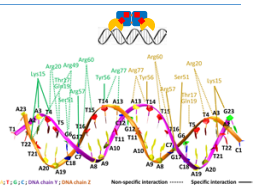
Gilston et al., PLOS Biology, 2014.

31

Comparing individual interactions



MgFur Fur box



MgFur *feoAB1*

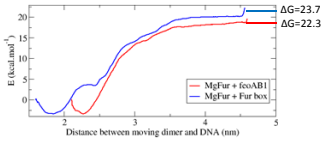
Origin	Sequence	Size
feoAB1	(Y) 5'-TAATTGCAAATCATTTGCAATTG-3'	23 bp
	(Z) 3'-ATTAACGTTTAGTAAACGTTAAC-5'	
EcFur box	(Y) 5'-GCCGGATAATGATAATCATTATC-3'	23 bp
	(Z) 3'-CGGCCTATTACTATTAGTAATAG-5'	

Sequence differences between *feoAB1* and the Fur box

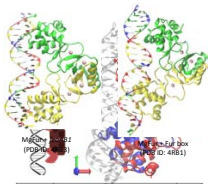
34

Do different DNA sequences affect binding ?

Potential of mean force for MgFur/DNA dissociations



- Binding of MgFur to the Fur box is stronger than the binding to *feoAB1*.
- Biphasic dissociation in the case of the Fur box.

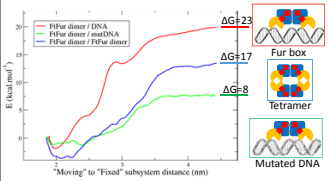


Origin	Sequence	Size
feoAB1	(Y) 5'-TAATTGCAAATCATTTGCAATTG-3'	23 bp
	(Z) 3'-ATTAACGTTTAGTAAACGTTAAC-5'	
EcFur box	(Y) 5'-GCCGGATAATGATAATCATTATC-3'	23 bp
	(Z) 3'-CGGCCTATTACTATTAGTAATAG-5'	

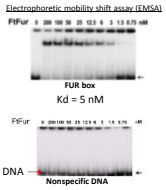
Sequence differences between *feoAB1* and the Fur box

Hong et al., Nature communications, 2015.

The case of FtFur



Electrophoretic mobility shift assay (EMSA)



Consensus sequence: 5'-**GATAATGATAATCATTATC**-3'
Mutated sequence : 5'-**GATACTGATAGTCCTGATC**-3'

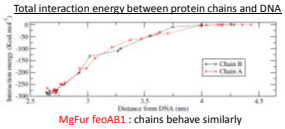
FtFur/DNA complex is more stable than the tetramer.
In agreement with specific DNA dependent tetramer dissociation *in vitro*

Pérard *et al.*, Biochemistry, 2016; Pérard, **Nader** et al., Communications Biology, 2018.

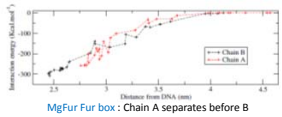
35

Do different DNA sequences affect binding ?

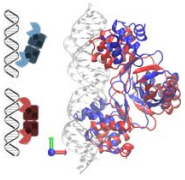
Total interaction energy between protein chains and DNA



MgFur feoAB1 : chains behave similarly



MgFur Fur box : Chain A separates before B



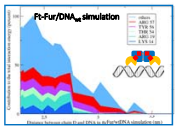
MgFur Fur box ; MgFur feoAB1
PDB ID: 4RB1; 4RB3

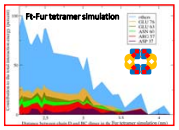
33

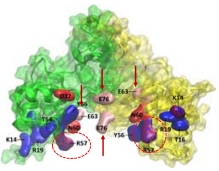
Defining interaction hotspots

Residue	E_{max} (kcal.mol^{-1})	Contribution (%)	Distance (nm)
Arg57	-26.1	13.0	2.96
Arg19	-36.9	18.4	2.78
Tyr56	-20.0	10.0	2.92
Lys14	-18.0	9.0	2.78
Thr54	-14.1	7.0	2.92

Top 5 interacting residues between FtFur and DNA in the FtFur/DNA simulation





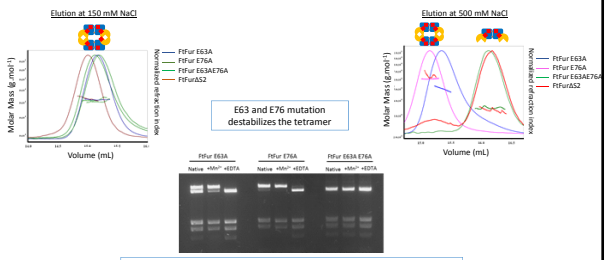


Major contributor to the interaction energy:
Blue: Fur/DNA^{wt} simulation
Red and Pink : Tetramer simulation

Pérard, **Nader** et al., Communications Biology, 2018

36

Experimental validation of theoretical hotspots



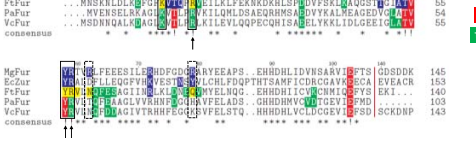
E63 and E76 mutation destabilizes the tetramer

E63 and E76 simple mutants are active, but the double mutant is not

Pérard, Nader et al., Communications Biology, 2018

37 Sarah Ancelet

Common residues between dimer/DNA and tetramer complexes

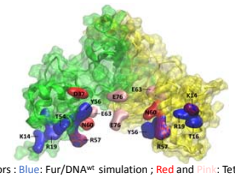


Conserved residues
Tetramer stabilization
DNA interaction
Common residues

Combined DNA interaction and tetramer stabilization fonctions are observed for conserved residues

40

Proposed DNA binding mechanism of FtFur



Major contributors : Blue: Fur/DNA^{wt} simulation ; Red and Pink : Tetramer simulation

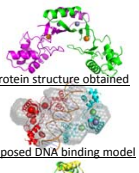
R57-DNA interaction triggers the tetramer dissociation

Pérard, Nader et al., Communications Biology, 2018

38

Conclusion and perspectives

- Structural insight into Fur from *E. coli*
  - Tool for future development of inhibitors
- Proposed model for EcFur/DNA complex
  - Starting point for future dissociation studies
- Better understanding of the inhibition by pF2
  - Good lead for the optimization of second generation peptides



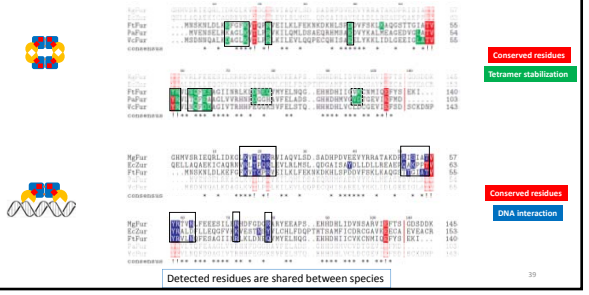
Protein structure obtained

Proposed DNA binding model

Proposed inhibition mechanism

41

Are the interaction hotspots shared between species ?



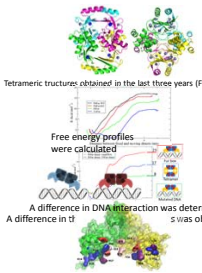
Conserved residues
Tetramer stabilization
Conserved residues
DNA interaction

Detected residues are shared between species

39

Conclusion and perspectives

- Identification of tetrameric Fur proteins
  - Additional structural characterization of tetramers
- Comparison of Fur tetramer dissociation
- Comparison of Fur/DNA interactions
  - Experimental validation of simulation results
- Understanding the case of FtFur
  - A double approach to study other Fur complexes
- Comparison of interaction hotspots
  - Future investigations for interaction fingerprints



Tetrameric tructures obtained in the last three years (FtFur & PaFur)

Free energy profiles were calculated

A difference in DNA interaction was determined

A difference in th s was observed

Interaction hotspots from FtFur simulatio

42

Acknowledgements

- MCT group:
  - Serge Crouzy
  - Giulia Veronesi
- BioMet team:
  - Isabelle Michaud Soret
  - Julien Pérard
  - Sarah Ancelet
  - Lucie Arnaud
  - Chloé Leprêtre
- BioCat team:
  - Christine Cavazza
  - Philippe Carpentier
  - Julien Pérard
- FAME – ESRF:
  - Jean Louis Hazemann
  - Olivier Proux
  - Denis Testemale
- IBS:
  - Eve de Rosny
  - Jacques Covès

ESRF
FIP-BM30A
FAME-BM30B
BIOSAXS-BM29

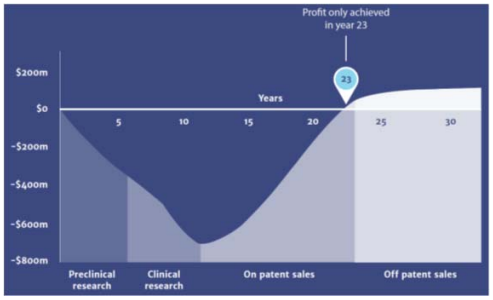
High Throughput Crystallization Laboratory (EMBL-Grenoble)
HTXLab

- Département de Chimie Moléculaire:
  - Didier Boturyn
  - Melissa Degardin
  - Maxim Früh

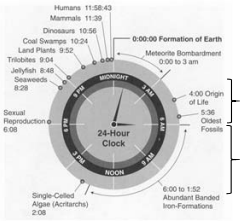
PhD Grant
La Région Auvergne-Rhône-Alpes
ARCANE

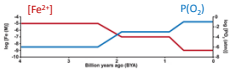
@CBMLab
Chemistry and Biology of Metals Laboratory

INEXT PID: 2217 - Inhibition of Fur from pathogenic bacteria: P. aeruginosa, F.tularensis and M. tuberculosis. LabEx **ARCANE** (ANR-11-LABX-0003-01) and **GRAL** (ANR-10-LABX-49-01)



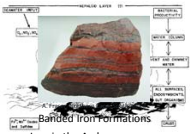
How iron became important





Reducing environment
Abundant [Fe²⁺]

Great Oxygenation events
Precipitation of [Fe²⁺]



Banded Iron Formations
Iron in the Archean ocean

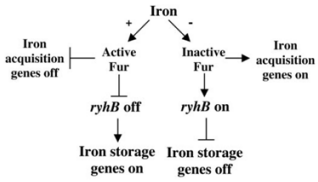
Life that used to have "easy" access to iron, now had to fight for it

Causes of GOE

Natural selection acted in favor of tight iron regulation

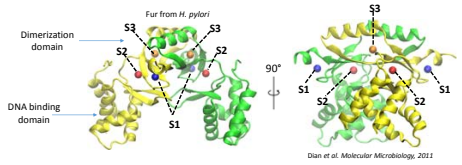
Boyd et al., "Interplay between oxygen and Fe-S cluster biogenesis: insights from the Suf pathway", Biochemistry, 2014.
Barosa et al., "Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life", Origins of life and evolution of the biosphere, 1985.

Gene regulation by Fur



E. Massé et al., "A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli.", PNAS, 2002.

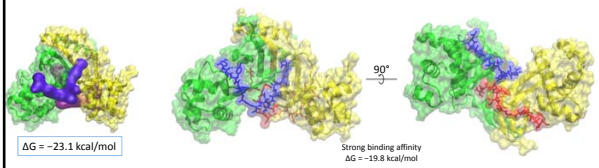
Activation mecanisme and inhibition of Fur



Activation mechanism of Fur

49

New Docking of pF2 on EcFur140



ΔG = −23.1 kcal/mol

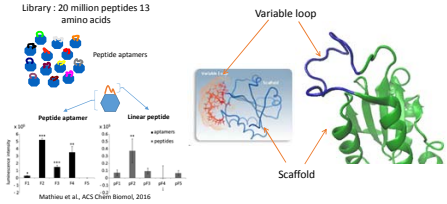
Strong binding affinity
ΔG = −19.8 kcal/mol

Peptide	FUR residues involved	Peptide residues involved
pF2 old	K77 N72 R70 Y130 E37 N60 S35 Y128 Y56 G75 H88 H125 S126 D38 H33	Q2 R1 S9 Y11 N4 W10 C6 C3 A8
pF2 new	N72 K77 E74 K98 Y130 S126 Y128 V99 K41 H125 G76 H88 E37 S35 G75 R42 D38	R1 Q2 Y13 N4 I5 Y11 S9 A8 S12

52

Optimizing first generation peptides : Cyclic peptides

Library : 20 million peptides 13 amino acids



Better interaction with peptide aptamers

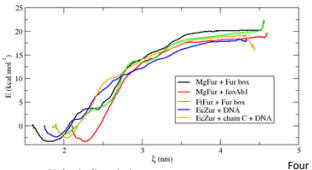
Mimicing the aptamer loop restraints

Increase stability and protection *in vivo ;* Resistance to hydrolysis by exopeptidases

DCM : Didier Boturyn, Melissa Degardin, Maxim Früh

50

DNA complexes



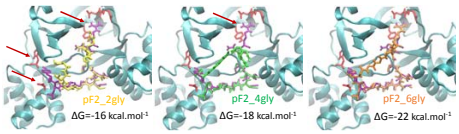
Dimer DNA complex	ΔG (kcal.mol^{-1})
MgFur + Fur box	23.70 ±2.4
MgFur + *feoAB1* box	22.31 ±1.3
FtFur + Fur box	22.01 ±1.8
EcZur + Zur box	20.05 ±1.4
EcZur + Zur box + Chain C	20.97 ±1.4

Biphasic dissociation
EcZur < FtFur < MgFur
MgFur feoAB1 < MgFur Fur box
Zur + chain C = + 1 kcal/mol → cooperative binding

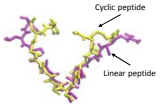
Four important residues are always present in DNA interactions
Lys14, Arg19 (non specific), Tyr56 and Arg57

53

Docking of cyclic pF2

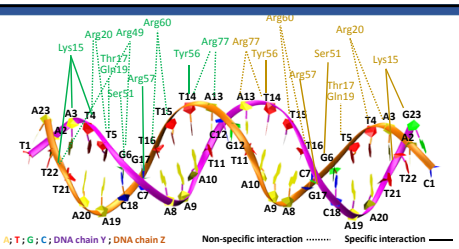


Optimized docking of pF2 -23 kcal/mol



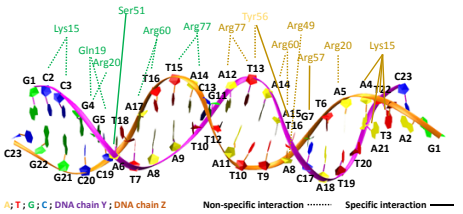
Cycle ↗ Affinity ↗
→ equivalent of the linear peptide

51



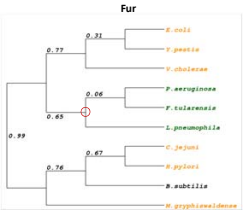
A, T, G, C, Z, Y

Feoab1 second picture



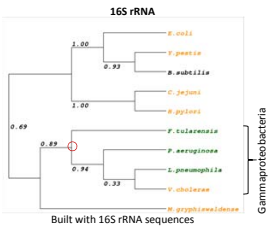
MgDNA case of MgFur

Phylogeny



Built with Fur protein sequences

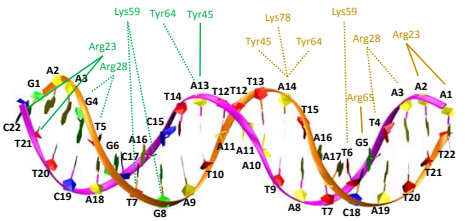
Tetrameric Fur proteins are grouped together



Built with 16S rRNA sequences

Bacterial phylogeny is in agreement with Fur proteins phylogeny

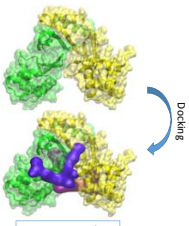
58



EcZur second picture

Docking pF2 on EcFur

FUR from *E. coli*



Docking

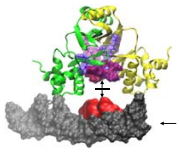
ΔG = −23.1 kcal/mol

Strong binding affinity

Isothermal titration calorimetry data

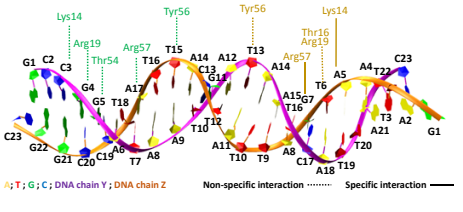
Ligand	n	Kd (μM)
pF2(1-13)	1.16 (0.18)	0.49 (0.10)
pF2(1-11)	1.17 (0.26)	0.33 (0.17)

Mathieu S *et al.* 2016 ACS Chem Biol



Proposed inhibition mechanism

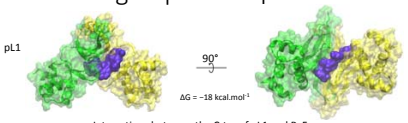
59



A, T, G, C, Z, Y

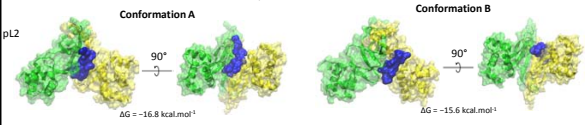
MgDNA case of FtFur

Docking of pL1 and pL2 on PaFur



Interactions between the C-ter of pL1 and PaFur

Inhibitor	Sequence
pL1	RQCRYCHW
pL2	WCRYCHPR
pF1	RLWCRYPHPPLTD



Interactions between the N-ter of pL2 and PaFur

60

XAS

J. Pérard, L. Arnaud

Fur of *E. coli*, *P. aeruginosa*, *F. tularensis*

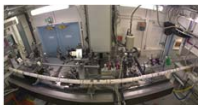
WT ΔS2 ΔS3 +/- inhibitor

30 mg/mL 25 samples

Protein production and purification

Activity tests MALLS XAS UV spectroscopy

10 days of continuous aquisition



FAME-BM30B : J.-L. Hazemann, O. Proux, D. Testemale.

E. coli	composition	F. tularensis	composition	P. aeruginosa	composition
EcFur-WT	+ Co ± pF2	FtFur-WT	+ Co	PaFur-WT	+ Co ± pL1
	+ Mn ± pF2		+ Mn		+ Mn
	+ Co + Mn ± pF2				+ Co + Mn + pL1
EcFurΔS2	+ Mn	FtFurΔS2	+ Co	PaFurΔS3	+ Mn
EcFurΔS3	+ Mn ± pF2				+ Co

61

Crystallogenesis of Fur proteins

Fur from *F. tularensis* | Fur from *P. aeruginosa* | Fur from *E. coli*

Published

Pérard et al., Communications Biology, 2018

PaFur ΔS3 → 2.4 Å

PaFur ΔS3 + inhibitors

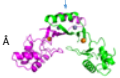
Co-crystallogenesis Soaking

HTXLab Buffer screening

Manual optimization

Additive screen

Manual optimization



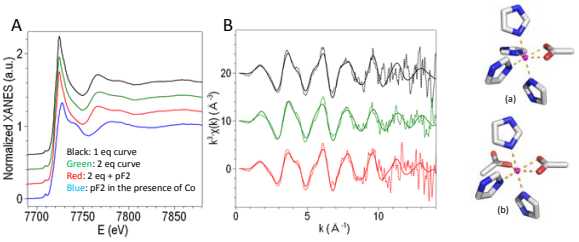
EcFur-140 : 2.3 Å

BioCat team :
C. Cavazza, P. Carpentier

FIP & ID30B

INEXT-PID: 2217 - Inhibition of Fur from pathogenic bacteria: P. aeruginosa, F.tularensis and M. tuberculosis

64

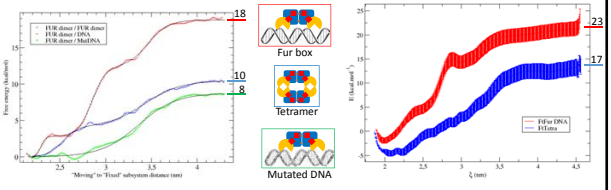


Model	R-factor (%)
3.5 His 1 Glu (bid) 0.5 Glu (mono)	7.3

G. Veronesi

62

The case of FtFur

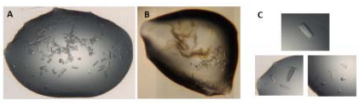


Consensus sequence: 5'-**GATAATGATAATCATTATC**-3'

Mutated sequence : 5'-**GATACTGATAGTCCTGATC**-3'

65

Obtaining crystals with inhibitors



3.9 Å

Figure 4.13: Co-crystallization of PaFurΔS3 with molecule B (F06 well of the PACT_MD screen). A: Crystals after the standard crystallization screens. B: Crystals after manual optimization. C: Crystals after the automated additive screen with 5.6 Å diffraction. A and C are 0.2 µL drops, whereas B is a 2 µL drop.

Protein complex	T_m (°C)
PaFurΔS3	66
PaFurΔS3 + pL1	66
PaFurΔS3 + pL2	66
PaFurΔS3 + molecule B	69

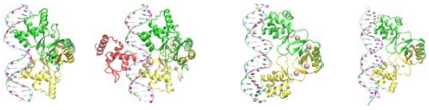
Table 4.6: Melting temperatures of PaFurΔS3 with or without inhibitors

PaFurΔS3 + pL2
in standard crystallization screens

63

Other complexes

PaFur tetramer | PaFur-ΔS3 tetramer | VcFur tetramer

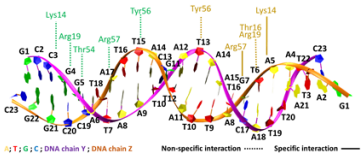


EcZur / DNA | MgFur / feoAB1 box | MgFur / Fur box

≈ 1,000,000 hours of simulation time → 1.5 µs/complex

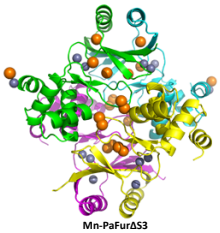
66

FtFur DNA, biphasic



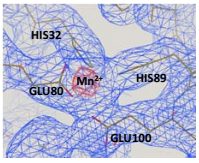
Structure of PaFurΔS3

Condition based on FtFur then optimized Washing

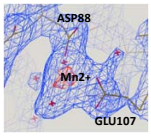


Mn-PaFurΔS3

2.6 Å

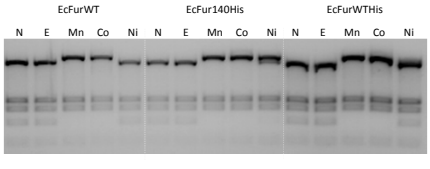


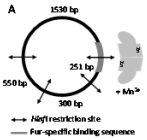
S2



Residual S3

Is EcFur-140 active ?

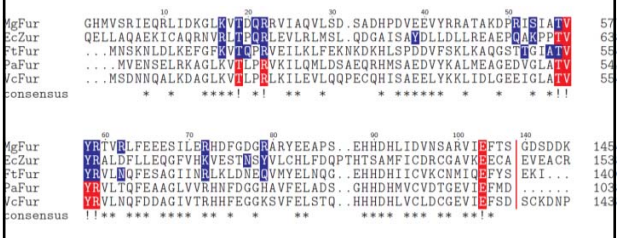




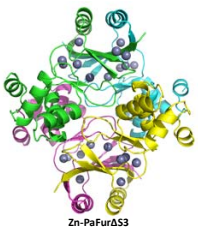
Remove nickel from gel
Keep ec140-his and wt-his

EcFur140 is activated by nickel

Residues important in DNA/dimer interactions

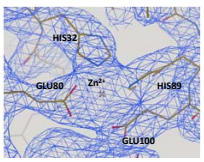


Structure of PaFurΔS3

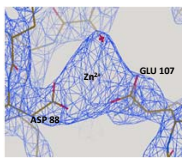


Zn-PaFurΔS3

2.2 Å

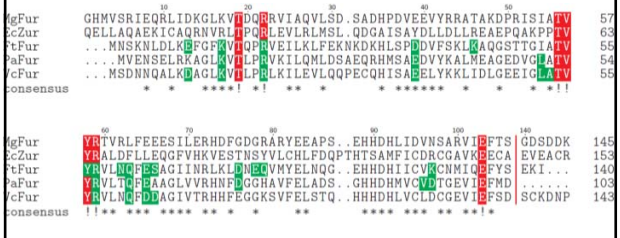


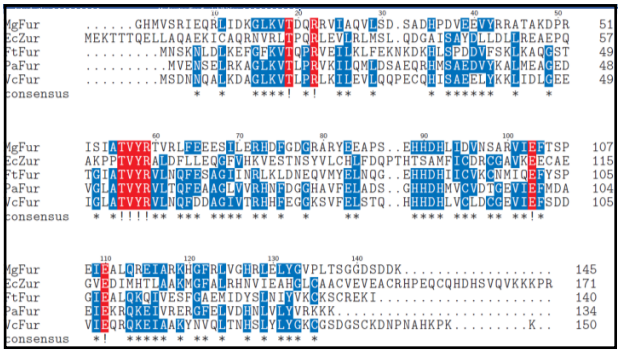
S2

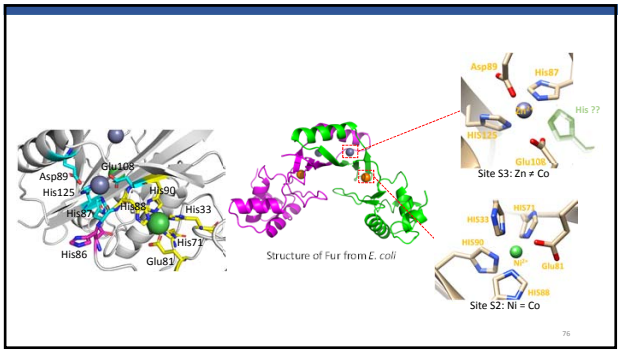


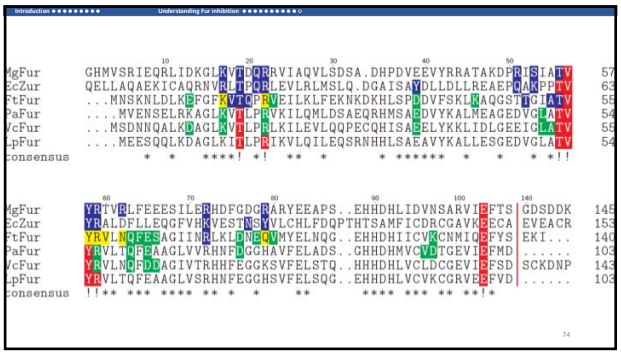
Residual S3

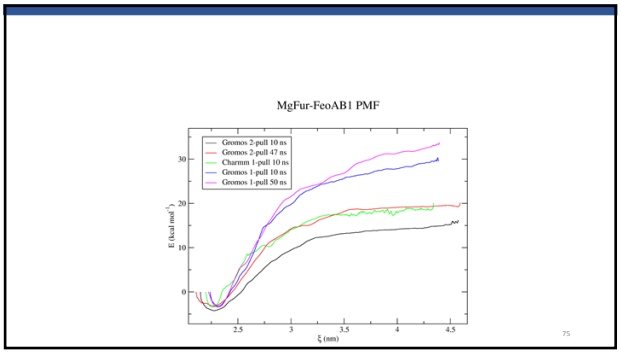
Residues important for tetramer dissociation




MgFur
EcZur
FtFur
PaFur
VcFur
consensus


Asp89
His87
His ??
His125
Glu108
Site S3: Zn ≠ Co
Glu108
Asp89
His90
His125
His88
His87
His33
His71
His86
Glu81
Structure of Fur from E. coli
HIS71
HIS33
HIS90
Glu81
HIS88
Site S2: Ni = Co


Introduction
Understanding Fur Inhibition
MgFur
EcZur
FtFur
PaFur
VcFur
LpFur
consensus


MgFur-FeoAB1 PMF
Geomss 2-pull 10 ns
Geomss 2-pull 47 ns
Charmm 1-pull 10 ns
Geomss 1-pull 10 ns
Geomss 1-pull 50 ns
E (kcal mol-1)
ξ (nm)

Thesis summary in French

Ce chapitre, rédigé en français, résume les travaux présentés dans ce manuscrit. Pour mieux comprendre l'intérêt de ces recherches ainsi que les résultats obtenus, ce résumé contient aussi certaines parties de l'introduction.

Cette thèse s'inscrit dans le cadre de la lutte contre la résistance bactérienne aux antibiotiques. La découverte des antibiotiques a révolutionné la médecine moderne grâce au traitement des infections préalablement mortelles. Mais à travers l'adaptation, une bactérie devient résistante aux antibiotiques. Toutefois, l'émergence de cette résistance, qui existait déjà dans la nature, est accélérée par la mauvaise utilisation des antibiotiques. Ces derniers sont les médicaments les plus utilisés dans la médecine moderne, toutefois, dans 50% des cas leur prescription est jugée non nécessaire. La période qui a suivi la seconde guerre mondiale marque l'âge d'or des antibiotiques, avec plusieurs nouvelles molécules mise sur le marché entre les années 1940 et 1970. Depuis, de moins en moins de nouvelles molécules sont découvertes. De nos jours, avec l'absence de développement de nouvelles molécules et la perte en efficacité des antibiotiques actuels, la résistance microbienne aux antibiotiques est considérée comme un risque majeur sur la santé publique. Ce risque concerne tout le monde et ne met pas uniquement en danger les pays pauvres et surpeuplés. En effet, dans les pays du "premier monde", le coût financier de ces épisodes sanitaires dépasse les milliards d'euros, accablant ainsi les gouvernements et les infrastructures hospitalières.

Pour les industries pharmaceutiques, les médicaments qui perdent leur efficacité perdent aussi leur valeur. D'où le besoin et l'intérêt de ces industries à chercher de nouvelles cibles pour traiter et prévenir la résistance aux antibiotiques. Les efforts principaux doivent se concentrer sur la compréhension des mécanismes de résistance et de sa transmission à travers l'eau, la nourriture et l'environnement. D'une façon parallèle, le diagnostic de la résistance et ses mécanismes d'émergence sont cruciaux pour préserver un niveau sain de surveillance à l'échelle mondiale.

Le développement de nouvelles stratégies pour combattre les pathogènes doit être une priorité. Les cibles idéales doivent exercer une faible pression évolutive, impacter la virulence et être uniques aux bactéries. Une façon d'atteindre cet objectif est d'interférer avec la régulation du fer et son homéostasie chez les bactéries. En effet, l'importance de cet élément pour tout type de vie est devenue une faiblesse qu'on peut exploiter dans notre combat contre l'infection.

Le chemin évolutif que la vie primitive a pris n'est pas un produit du hasard, il était contraint par les changements thermodynamiques de l'environnement qui ont forcé une progression dans un sens bien spécifique. En effet, la biodisponibilité des métaux a fortement influencé l'évolution des premières formes de vies, ainsi que leurs stratégies métaboliques utilisées durant cette période-là. La majorité des organismes vivant requièrent la première ligne des métaux de transitions du tableau périodique (fer, cobalt,

nickel, cuivre, zinc). La capacité de ces métaux à changer d'états d'oxydations a fait d'eux des cofacteurs essentiels pour les enzymes. Cela contribue ainsi à leur importance catalytique dans le vivant, mais aussi à leur toxicité. La régulation de ces éléments, si importants, se fait grâce à plusieurs mécanismes senseurs capables de réguler l'import et l'export des métaux entre la cellule et le milieu externe. Pour les bactéries pathogènes, un organisme hôte est une riche source de métal, ce qui a justifié pourquoi ces pathogènes ont développé des méthodes d'acquisition de métal assez sophistiquées. La présence de tels mécanismes a créé une compétition entre les organismes qui a conduit à ce qu'on nomme la bataille pour les métaux, qui a lieu entre hôtes et pathogènes. La compréhension de la perspicacité de cette bataille, peut être utilisée pour explorer de nouvelles opportunités thérapeutiques contre la virulence bactérienne.

Chez l'hôte, un système de défense très important contre les infections est la séquestration des nutriments, ce qui conduit à un ralentissement de la croissance bactérienne. Ce phénomène est nommée immunité nutritionnelle, où l'hôte séquestre notamment le fer. La plupart des pathogènes luttent contre cette privation de fer par des mécanismes ayant une haute affinité au fer, tel que les sidérophores. Ces mécanismes sont régulés par des senseurs de fer essentiel pour la virulence bactérienne. La recherche sur les métaux de transition en biologie, appliquée à l'interface hôte pathogène, présente plusieurs opportunités d'étudier un large potentiel thérapeutique à cause du travail interdisciplinaire dans diffèrents domaines de la science fondamentale et appliqués.

Chez les bactéries, toutes les modifications transcriptionnelles nécessaires à l'acquisition du fer sont dépendantes d'un senseur de cet élément, nommé "Ferric Uptake Regulator" (Fur). Comme la protéine Fur joue un rôle central dans l'homéostasie du fer, elle est une cible thérapeutique idéale dans la lutte contre l'infection et la virulence bactérienne. Depuis sa découverte dans les années 80, la protéine Fur a était minutieusement caractérisée comme régulateur de transcription métal dépendant. Grâce à différents mécanismes de régulation, Fur affecte l'expression d'un grand nombre de gènes, dont certains sont impliqués dans la virulence bactérienne. En effet, il a été démontré que la mutation du gène *fur* diminue la virulence de souches pathogènes. La protéine Fur a été caractérisée chez différentes bactéries et plusieurs structures PDB sont disponibles. Cette protéine est un homo-dimère dans certains cas et un homo-tétramère dans d'autre, sa partie N-terminale contient un domaine de liaison à l'ADN et sa partie C-terminale contient le domaine de dimérisation. De plus, vu son rôle dans la détection du fer, la protéine Fur contient des sites métalliques qui peuvent être différents d'un organisme à l'autre.

Les études précédentes ont permis l'identification, après un criblage d'une banque d'aptamères peptidiques, de quatre peptides inhibant Fur d'*E. coli* (EcFur). Leur caractérisation a été effectuée par une approche double qui combine des expériences *in vitro* et des simulations *in silico*. Les résultats présentés dans la partie suivante suivent la même approche combinée afin de mieux comprendre cette protéine clé.

Les travaux précédents ont pu montrer que des aptamères peptidiques anti-Fur nommés F1 à F4, constitués d'un squelette fixe formé de la thiorédoxine A d'*E. coli* et d'une boucle variable de 13 acides aminés, ainsi que leurs peptides linéaires correspondants (pF1 à pF4), pouvaient lier spécifiquement EcFur et inhiber sa liaison à l'ADN. Leur activité a été testée *in vitro* et *in vivo* où une diminution de la virulence a été démontrée. Par ailleurs, un criblage d'une librairie de composés chimiques et des simulations *in silico* ont permis la mise en évidence de nouveaux inhibiteurs capables d'inhiber Fur de *P. aeruginosa* (PaFur). Ces inhibiteurs sont deux peptides nommés pL1 et pL2 et une molécule chimique qu'on nommera « molécule B ». Une partie de ce travail de thèse était de déchiffrer les mécanismes d'inhibitions de ces inhibiteurs.

Pour atteindre cet objectif et en absence de structure cristallographique d'un complexe [protéine + inhibiteur], des simulations d'amarrage moléculaire ont été effectuées sur un modèle d'EcFur obtenu par homologie avec Fur de *V. cholerae* (VcFur) qui a une structure PDB. L'idée derrière ces simulations était d'optimiser les inhibiteurs déjà développés dans le laboratoire. Pour cela, des amarrages de peptides cycliques ont été effectués. Les résultats montrent que plus le peptide cyclique est petit, plus on perd en interaction avec la protéine. Cette information montre que des peptides cycliques pourraient inhiber EcFur, par contre l'idée de base était de trouver des peptides cycliques de petites taille qui en amarrage moléculaire n'avaient pas une bonne énergie libre de liaison. Une fois que le consensus sur la taille des peptides sera défini, les constructions testées *in silico* seront transférés à nos collaborateurs chimistes qui les synthétiseront pour réaliser des tests *in vitro*.

Des amarrages moléculaires ont également été effectués sur les peptides pL1 et pL2, en utilisant la structure de PaFur déterminée par Pohl et al., 2003. Les résultats montrent une interaction avec la protéine similaire au cas de pF2 avec EcFur, où dans les deux cas, les inhibiteurs interagissent avec la protéine au niveau de la région entre les deux sous unités. Pour avoir une idée plus complète de chaque mécanisme d'inhibition, et pour valider l'hypothèse que l'inhibition n'est pas due à une interaction avec les sites métalliques de la protéine, des expériences d'absorption aux rayons X ont été menées. Des échantillons de protéines Fur de différentes souches bactériennes ont été préparés en les métallant au manganèse ou au cobalt, et incubés en présence d'inhibiteurs. Les résultats obtenus dans le cas d'EcFur, montrent que l'inhibition ne change pas l'environnement chimique des métaux dans la protéine, ce qui indique que cette inhibition n'implique pas une interaction entre les inhibiteurs et les sites métalliques. De plus, l'analyse de ces données acquises expérimentalement avec des modèles de sites métalliques, a permis la proposition de sites composés de 6 ligands, en accord avec les données cristallographiques et les travaux précédents sur les protéines Fur.

Comme un de nos objectifs est l'étude des mécanismes d'inhibition, une façon directe de le faire est à travers la cristallogenèse. Après nos expériences en absorptions aux rayons

X, nous avions à disposition des solutions à concentrations élevées de protéines pures. Nous les avons utilisés dans de séries de cristallogenèses automatisées et manuelles. Ces essais ont réussi sur une construction de PaFur qui contient une mutation au niveau du site métallique S3, d'où son nom PaFurΔS3. D'une façon intéressante, la résolution de cette structure en présence de zinc ou de manganèse, a révélé un nombre considérablement élevé de métaux liés à la protéine. Cette information est intéressante parce que dans la structure de Fur de *F. tularensis* (FtFur), on ne trouve pas ces métaux surnuméraires. Ce qui pose la question du rôle de ces atomes dans la stabilisation du tétramère de PaFur, un rôle qui est toujours en train d'être étudié. Après la résolution de la structure de PaFurΔS3 des essais de cristallogenèse en présence d'inhibiteurs ont été effectués. Après plusieurs mois de tests automatisés et d'optimisation manuelle, des cristaux ont été obtenus mais leur qualité de diffraction n'était pas suffisante pour déterminer la structure du complexe. Dans le futur, la poursuite de ces études pourra se concentrer sur l'optimisation de la qualité de diffraction des cristaux obtenus durant cette thèse.

D'une façon similaire, des tests de cristallisation sur EcFur ont été conduits. Dans un premier temps nous avons essayé de cristalliser le mutant EcFurΔS3 sans succès. Dans un deuxième temps, les tests de cristallogenèse ont été effectués sur une construction d'EcFur sans la partie flexible située dans le domaine C-terminal (résidus 141 à 148). En même temps, l'étiquette polyhistidine utilisée lors de la purification a été conservée. Cette construction qu'on nomme EcFur-140 est stable en solution, et peut lier l'ADN d'une façon similaire à celle de la protéine sauvage. Comme dans le cas de PaFurΔS3, les tests de cristallogènes ont commencé par un criblage automatique qui a donné une seule et unique touche. Après plusieurs semaines d'optimisation, la qualité de diffraction était suffisante pour résoudre pour la première fois, après vingt ans d'essais, la structure complète d'EcFur sans ses huit derniers acides aminés.

Les spectres de fluorescence des cristaux d'EcFur-140 ont montré la présence de zinc et de nickel dans la protéine. Ces résultats sont validés par des expériences d'ICP-AES, qui montrent la liaison d'un atome de nickel et d'un atome de zinc par monomère d'EcFur-140. Le nickel n'a pas été rajouté à la protéine, sa présence peut être liée au protocole de purification qui utilise des colonnes d'affinité NiNTA.

L'analyse de la structure d'EcFur-140 montre une organisation structurelle similaire à celle d'autres protéines Fur dimériques. Les résidus 2 à 82 forment le domaine de liaison à l'ADN et les résidus 83 à 134 forment la partie du domaine de dimérisation. Sur la carte de densité, l'étiquette polyhistidine n'est pas visible. L'analyse des sites métalliques d'EcFur-140 montre l'absence d'un site S1 et la présence d'un pont disulfure entre les cystéines 93 et 96. Le site S2 présente une configuration non conventionnelle avec 4 histidines (33, 71, 88 et 90) et un glutamate (81) qui coordonnent un atome de nickel. L'histidine 71 est une nouveauté observée pour ce site S2. Dans d'autre protéines Fur, ce site est constitué de deux glutamates et de trois histidines. Les résultats obtenus à travers

la résolution de la structure d'EcFur-140 valident les modèles utilisés dans l'analyse des données expérimentale d'EXAFS où un modèle avec 4 histidines a été utilisé. De plus, cette structure a permis de monter que EcFur contient un site S3. Malheureusement, vu la qualité de la carte de densité actuelle a cet endroit de la protéine, on ne peut que proposer un modèle pour ce site qui serait formé par deux histidines (87 et 125), un aspartate (89) et un glutamate (108). Pour mieux définir ce site, une nouvelle campagne de cristallisation devrait être initiée pour obtenir de nouveaux cristaux et de nouvelles cartes de densités.

A part la cristallographie aux rayons X sur EcFur, des expériences en SAXS ont permis l'acquisition de données sur EcFur en complexe avec l'inhibiteur pF2. Les résultats ne montrent aucun changement dans la structure globale de la protéine, ce qui indique que l'inhibition par pF2 n'implique pas une réorganisation structurale d'EcFur. En parallèle, la structure d'EcFur-140 a permis d'analyser les données expérimentales obtenue en SAXS pour la protéine EcFur sauvage en présence d'ADN, permettant ainsi de proposer un modèle de deux dimères d'EcFur fixés sur une séquence d'ADN. Ce modèle sera utilisé comme outil dans la poursuite des recherches sur l'inhibition de cette protéine.

Avec la résolution de la structure d'EcFur-140, les anciens amarrages moléculaires effectués sur un modèle d'EcFur, obtenu par homologie avec VcFur, ont été reconduits sur la structure cristallographique. Les résultats obtenus valident les amarrages précédents et montrent que dans le cas de l'inhibiteur pF2, la poche d'inhibition prédite reste la même que celle trouvée dans les précédentes investigations. Les interactions entre inhibiteur et protéines restent majoritairement les mêmes avec quelques modifications ce qui mène à la proposition d'un mécanisme d'inhibition encore plus précis. En plus de révéler de nouveaux détails sur une protéine modèle très étudiée, la structure d'EcFur-140 servira comme plateforme d'amarrage moléculaire pour tous les inhibiteurs présents dans le laboratoire, ainsi que toute molécule découverte dans le futur.

La deuxième partie de ce travail étudie les différentes formes oligomériques des protéines Fur. En effet, suivant l'espèce bactérienne, Fur peut être un dimère ou un tétramère. Par contre les études expérimentales et les structures cristallographiques disponibles indiquent que toutes les protéines Fur lient l'ADN sous forme de dimères. A part les études expérimentales qui caractérisent les états oligomériques, aucun travail théorique n'a été effectué pour comprendre ce comportement.

Pour comprendre comment les tétramères se dissocient et lient l'ADN en dimères, des modèles basés sur des structures cristallographiques ont été construits. Ils ont permis la détermination des profils d'énergie libre de la dissociation d'un dimère de Fur de l'ADN et celle de la dissociation entre deux dimères d'un tétramère. Ces simulations avaient pour but de déterminer quel complexe était le plus stable et quels étaient les résidus impliqués dans chaque type d'interaction (dimère/dimère ou dimère/ADN).

Dans ce manuscrit, l'exemple de FtFur, qui est un tétramère, est détaillé pour décrire

les différentes étapes nécessaires à la génération des fenêtres de simulation, au calcul du profil d'énergie libre et à la détermination des résidus les plus impliqués dans la stabilisation de chaque complexe. Une partie de ces résultats est présentée dans un article paru cette année : Julien Pérard, Serge Nader *et al.* "Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in *Francisella tularensis* virulence". Dans: Communications Biology 1.1 (Juillet 2018), p. 93. issn: 2399-3642. url: https://doi.org/10.1038/s42003-018-0095-6 (Pérard et al., 2018). Les principaux résultats montrent que le dimère lié à son ADN spécifique est plus stable que le tétramère de FtFur. Ce qui est expérimentalement vrai, vu que le tétramère se dissocie uniquement en présence d'une séquence d'ADN spécifique. De plus, la détermination des résidus impliqués dans la stabilisation du tétramère a permis leur mutation, ce qui a validé les résultats théoriques.

Le même protocole développé pour étudier FtFur a été appliqué sur sept autres complexes, dont des Fur tetramériques ou des dimères de Fur liés à des séquences d'ADN différentes. En même temps, deux simulations sur la protéine Zur (Zinc uptake regulator) ont eu lieu.

Les résultats ont permis de comparer différentes protéines Fur entre elles. Dans le cas des tétramères, une différence est observable entre le tétramère de PaFur et celui de FtFur. Les deux sont assez stables en solution et nécessitent une concentration élevée de sel pour dimériser ; cette différence pourrait être lié à la capacité de PaFur a lier plus de métal que FtFur. Cette hypothèse reste à vérifier avec la finalisation de la structure de PaFurΔS3. De plus, dans ces simulations, un « faux tétramère » a été inclus. Il s'agit d'un tétramère de VcFur qui est un dimère en solution mais a été construit en tétramère comme contrôle négatif pour valider notre méthode. Effectivement cette simulation donne le plus faible ΔG parmi tous les tétramères. Concernant les simulations des dimères de Fur liés à l'ADN, les résultats montrent que comme dans le cas de FtFur, ces complexes sont plus stables que les tétramères. D'une façon intéressante, dans certains cas, on remarque des profils biphasiques quand des protéines Fur se dissocient de la séquence d'ADN de la Fur box. En même temps, ma comparaison des diffèrents complexes a permis la mise en évidence de résidus souvent impliqué dans l'interaction avec l'ADN. Ces simulations seront utilisées pour mieux comprendre les interactions moléculaires qu'effectuent les protéines Fur des différentes espèces.

Quand tous les résultats de ce manuscrit son pris en compte, il est clair que les protéines Fur de différentes espèces sont différentes. Cette différence peut être au niveau de l'état oligomérique, des acides aminés ou des sites métalliques. Ceci remet en cause l'idée initiale derrière ce projet qui cherchait a optimiser les inhibiteurs d'EcFur pour les utiliser sur d'autres protéines Fur. Les deux structures de protéines obtenues lors de cette thèse, ainsi que les modèles proposés et les protocoles de simulation devront être considérés comme des outils, utilisés pour mieux comprendre et adapter chaque sujet de recherche à chaque

protéine Fur.

Remerciements

Je tiens à remercier tout particulièrement mon directeur de thèse Serge Crouzy et mon co-encadrant Julien Pérard pour m'avoir encadré durant cette thèse. Au début de cette aventure nous avions beaucoup de pistes à explorer sans avoir nécessairement une bonne visibilité, finalement nous avons réussi a extraire un maximum de données et j'en suis vraiment fière. Serge, merci pour ta confiance, ton soutien et tout ce que tu m'as appris durant ces trois années. Julien, merci pour ton enthousiasme, ta présence au quotidien et surtout ta persévérance. Je remercie également Isabelle Michaud-Soret pour les discussions scientifiques autours du projet Fur.

De même, je tiens à exprimer mes remerciements aux membres du jury pour avoir accepté de critiquer ce manuscrit et de juger mon travail.

Merci à Stéphane Ménage, directeur du Laboratoire de Chimie et Biologie des Métaux, et à Patrice Catty, directeur adjoint, ainsi qu'à tous les membres du laboratoire pour m'avoir accueilli. Grace à vous j'ai pu apprécier la gestion de la composante majeure d'un laboratoire: l'humain.

Durant trois ans, j'ai partagé ma paillasse avec Lucie Arnaud, Sarah Ancelet et Melissa Degardin, sans oublier Aurélien Deniaud voisin de paillasse et collectionneur de mes blouses blanches. Merci d'être mon point de référence quand je suis arrivé au laboratoire. Merci également pour votre soutien au quotidien tout en partageant vos mésaventures avec les protéines Fur.

Je remercie aussi Giulia Veronesi, mon étalon interne au laboratoire, pour sa présence précieuse au groupe de Modélisation et de Chimie Théorique, ainsi que pour son apport scientifique à ce projet. Je tiens a remercié également Yohann Moreau pour les nombreuses discussions autour d'un café.

Les permanents du laboratoire ont aussi enrichi mon expérience, Mireille Chevallet et Peggy Charbonnier, les piliers du poulailler, merci pour vos conseils sur la vie. Bernard Sartor, Patrice Catty et Roger Miras, le trio favori du labo, merci pour les repas de midi que nous avons partagé ensemble. Je tiens a remercié aussi, les chimistes du K', notamment Victor Duarte, Christelle Caux-Thang, Stéphane Ménage et Vincent Artero pour leur soutien et les discussions qu'on a pu avoir autours des repas.

Un grand merci à tous les non-permanents du laboratoire avec qui j'ai partagé beaucoup d'aventures, que se soit pour du ski, des randonnées, du squash ou de l'escalade, sans oublier les soirées ! Merci Amel Kouadri, Marianne Marchioni, Johanna Cormenier, Jérémy Caburet, Steve Chiumento, Marila Alfano, Sarah Lopez, Xavier Maréchal, Micheline El Khoury, Jessy Hasna et Tim Rosser. Un remerciement spécial a Jordan Mangue et Jaione Saez Cabodevilla, nous étions les trois derniers à soutenir cette année, merci pour votre soutien et encouragement dans les moments les plus délicats.

I would also like to thank Caitlin Bresnahan and Jennifer Fry, for the nights in and

out, and for sharing a lot of common ideas.

Et bien sûr, Rolf David et Jérôme Laisney. Merci pour les bières et les souvenirs. A bientôt pour de nouvelles aventures et de la bonne science !

Ces remerciements ne seront jamais complets sans mentionner ma famille. Merci à mes parents et ma sœur de m'avoir fait confiance et permis de prendre le chemin que j'ai choisi. Un grand merci aussi à Naji, Mireia et Laïa de m'avoir soutenu depuis le début. Cette thèse n'aurait pas pu être possible sans votre soutien à tous.

Et finalement, je tiens à remercier tous les scientifiques qui ont contribué, de près ou de loin, à travers leurs savoirs, à la finalisation de ce projet. Cela peut être résumé par une citation d'Isaac Newton de 1675:

"If I have seen further it is by standing on the shoulders of Giants".

Abstract

The most commonly prescribed drugs in human medicine are antibiotics. Since their discovery, they have drastically impacted the way we treat infections. However, a bacterium eventually becomes resistant to antimicrobial treatment through the natural process of adaptative evolution. Even if resistant bacteria are omnipresent in the biosphere, their emergence rate is accelerated by the misuse of antimicrobial agents leading to the public health threat we are facing now. As currently available antimicrobial agents lose their effectiveness and very few new drugs are being developed, a breakthrough in new strategies to fight pathogens should be a priority. Ideal new therapeutic targets should exert weak evolutionary pressure, disarm or weaken the pathogen and be unique to microorganisms. One way to do so is by interfering with the iron regulation and its homeostasis within Bacteria. The bioavailability of iron strongly influenced early life and the metabolic strategies that sustained it. A central iron sensing mechanism evolved to ensure the regulation of such an important element. Sadly for bacteria this sensor became an exploitable weakness in our battle against infection. The "Ferric Uptake Regulator" is a metal dependent transcription regulator with a large regulatory network controlling iron homeostasis and bacterial virulence. This work continues previous investigations on Fur inhibitors using a combined experimental and theoretical approach by performing XAS, SAXS and MALLS experiments together with computer simulations. We describe for the first time the structures of Fur from E. coli in addition to a tetrameric Fur structure of a mutant from P. aeruginosa. Moreover, free energy profiles of Fur proteins, as tetramers or dimers bound to DNA, from different species were generated and key residues involved in the interactions determined, providing mechanistic insights into Fur complexes. The structural information gathered from this work will be used to better understand inhibition mechanisms of Fur proteins providing new opportunities to overcome drug development challenges.

Key words: Ferric uptake regulator ; molecular dynamics ; free energy profiles ; X-ray diffraction structures ; Biophysical characterisation

Résumé

Les antibiotiques sont les médicaments les plus utilisés dans la médecine moderne. Depuis leurs découvertes, ils ont drastiquement changé la façon dont les infections sont traitées. Toutefois, à travers le processus d'adaptation, les bactéries deviennent éventuellement résistantes aux antibiotiques. Malgré leur omniprésence dans la biosphère, l'émergence de souches résistantes est favorisée par le mauvais usage des antibiotiques, ce qui crée une menace importante pour la santé publique. Les antibiotiques actuelles perdent graduellement leur efficacité, et vue le faible nombre de nouvelles molécules développées, la priorité est donnée pour la découverte de nouvelles stratégies capables de combattre les pathogènes. Les nouvelles cibles thérapeutiques idéales doivent exercer une faible pression évolutive, diminuer la virulence et être unique aux microorganismes. Une façon d'atteindre cet objectif est d'interférer dans la régulation et l'homéostasie du Fer chez les bactéries. La biodisponibilité du Fer a fortement influencé l'émergence de la vie sur terre et les stratégies évolutives qu'elle a adoptée. Ce qui a mené à l'apparition d'un mécanisme central de détection du Fer assurant la régulation de cet élément de haute importance. Ce senseur est un point faible que nous pourrons exploiter dans notre combat contre les infections bactériennes. La protéine Fur, pour « Ferric Uptake Regulator », est un régulateur de transcription métal dépendant qui est impliqué dans vaste réseau de régulation contrôlant principalement l'homéostasie du Fer et l'expression de facteurs de virulence. Le travail présenté dans ce manuscrit complète les études précédentes sur des inhibiteurs de la protéine Fur en utilisant une approche combinée théorique et expérimentale grâce a des expériences de XAS, SAXS et MALLS associé a de la modélisation moléculaire. Nous décrivons pour la première fois la structure de Fur d'E. coli ainsi que la structure d'un tétramère de Fur d'un mutant de P. aeruginosa. Par ailleurs, les profils d'énergie libre des protéines Fur de différentes espèces ont été déterminé, pour des complexes tetramériques ou dans le cas de dimères liés à l'ADN, permettant une compréhension préliminaire de leur mécanistique. Les informations structurales obtenues grâce aux travaux présentés dans ce manuscrit permettront de mieux comprendre les mécanismes d'inhibition des protéines Fur ainsi que fournir de nouvelles opportunités pour le développement de molécules a visée thérapeutique.

Mots clés: Ferric Uptake Regulator ; dynamiques moleculaires ; Profiles d'energie libre ; Structures de diffraction aux rayons X ; characterisation biophysique