Targeting Lsr2/DNA complexation for dysregulation of gene expression in Mycobacterium tuberculosis

Lucile Pinault
The University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation
Theses and Dissertations. 170.
http://utdr.utoledo.edu/theses-dissertations/170
A Thesis

entitled

Targeting Lsr2/DNA Complexation for Dysregulation of Gene Expression in

*Mycobacterium tuberculosis*

by

Lucile Pinault

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Master of Science Degree in Chemistry

______________________________
Dr. Donald R. Ronning, Committee Chair

______________________________
Dr. John J. Bellizzi, Committee Member

______________________________
Dr. Max O. Funk, Committee Member

______________________________
Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2013
An Abstract of

Targeting Lsr2/DNA Complexation for Dysregulation of Gene Expression in *Mycobacterium tuberculosis*

by

Lucile Pinault

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Science Degree in Chemistry

The University of Toledo

May 2013

The identification of new drug targets and mechanisms of action against tuberculosis has become urgent, as the emergence of drug-resistant strains has made many treatments ineffective. Lsr2 is a protein encoded by the *Mycobacterium tuberculosis* genome that acts as a global repressor of transcription by binding AT-rich promoter regions of the mycobacterial DNA. Inhibiting this function could potentially dysregulate a wide array of genes, as Lsr2 binds about 20% of the *Mycobacterium tuberculosis* genome. Among them are key genes involved in mycobacterial virulence, cell wall biosynthesis or antibiotics-inducible genes that confer multi-drug tolerance. The structure of the DNA-binding C-terminal domain has been solved by nuclear magnetic resonance while the dimerization N-terminal domain was solved using X-ray crystallography, but there is no crystal structure of the full-length protein to this day.

The first part of this work describes the development of a high-throughput assay that enabled the identification of Zafirlukast as an inhibitor of the Lsr2/DNA complexation. This ability was shown *in vitro* using a fluorescence polarization assay. Zafirlukast also exhibits growth inhibition of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* on a Kirby Bauer assay, but shows no activity against
Escherichia coli which does not encode the Lsr2 protein. Transcription levels of genes shown to be bound by Lsr2 have been monitored in vivo after applying the drug to a culture of Mycobacterium tuberculosis and the results suggest that Zafirlukast promotes dysregulation of mycobacterial transcription. This drug is already commercially available and is commonly used as a treatment for asthma. As its mechanism of action appears to be distinct from current drugs against tuberculosis, it offers a potential new lead for the treatment of this disease.

The second part of this work describes different attempts made to gain structural knowledge about Lsr2 particularly in the context of an Lsr2/DNA complex. Different fusion proteins consisting of Lsr2 and lysozyme sequences were engineered for this purpose. Oligonucleotides were also designed to try to obtain crystals of the Lsr2/DNA complex.
Acknowledgements

First and foremost I would like to express my complete gratitude to my advisor, Dr. Donald R. Ronning, for all his guidance, kindness and patience throughout my thesis project. Thank you for teaching me so much and being such a great mentor. The experience I gained here has been the most instrumental one in my career development and perspectives. I will always be grateful.

My appreciation extends to my committee members Dr. Max O. Funk and Dr. John J. Bellizzi as well as the whole Chemistry Department of the University of Toledo. I would also like to thank my fellow labmates for passing on their knowledge and being such a friendly team to work with. I spent a wonderful time here with all of you! I want to particularly acknowledge Vidhi for sharing a lot of lab skills when I began working in the lab and Dr. Jimmy Franco for helping me get started on this project.

I must also thank my loving family and friends in France who have been amazingly supportive, especially my best friend Anne and my friend Mathilde.

Finally, I want to thank my friends here for all the laughs and good times that have been my strength. I am particularly thankful for meeting Edma, Mathieu and Jared. Last but not least, a very special thanks goes to my great friend Quentin who shared most of this American adventure with me and was always there for me.
# Table of Contents

Abstract ........................................................................................................................................ iii

Acknowledgements ....................................................................................................................... v

Table of Contents ......................................................................................................................... vi

List of Tables ................................................................................................................................. xi

List of Figures ................................................................................................................................. xii

List of Equations .......................................................................................................................... xiv

List of Abbreviations .................................................................................................................... xv

List of Symbols .............................................................................................................................. xvii

1 Introduction ................................................................................................................................. 1

1.1 Tuberculosis: current status ................................................................................................. 1

1.1.1 Brief introduction to tuberculosis and its infectious process ........................................ 1

1.1.2 Development of treatment regimens .............................................................................. 2

1.1.2.1 The BCG vaccine .................................................................................................... 2

1.1.2.2 First line drugs .................................................................................................... 3

1.1.2.3 The issue of drug resistance and co-infection with HIV .................................... 3

1.2 Roles of Lsr2 in *Mycobacterium tuberculosis* .................................................................. 5

1.2.1 Lsr2: histone-like protein ............................................................................................ 5

1.2.2 Lsr2: global transcriptional regulator .......................................................................... 6

1.2.3 Lsr2: physical barrier against ROIs ............................................................................. 8
1.2.4 Current structural knowledge about Lsr2 ........................................9

1.3 Project goals ..........................................................................................10

2 General Methods: Principles ........................................................................12

2.1 Introduction ..............................................................................................12

2.2 From genes to proteins ............................................................................12

2.2.1 Polymerase chain reaction .................................................................12

2.2.2 Agarose gel electrophoresis .................................................................14

2.2.3 Restriction digest ..................................................................................15

2.2.4 Ligation ..................................................................................................17

2.2.5 Transformation of competent cloning cells: plasmid replication ..........17

2.2.6 Plasmid purification ..............................................................................18

2.2.7 Transformation of competent expression cells .................................19

2.2.8 Induction of protein expression ...........................................................20

2.2.9 Cell harvesting and lysis ......................................................................21

2.3 Protein purification ..................................................................................22

2.3.1 Fast protein liquid chromatography (FPLC) ......................................22

2.3.1.1 Immobilized metal affinity liquid chromatography ..........22

2.3.1.2 Ion exchange chromatography ......................................................23

2.3.2 Protein dialysis ...................................................................................25

2.3.3 SDS-PAGE: characterization of protein purity ..................................26

2.3.4 Determination of protein concentration ..............................................27

2.4 Binding and inhibition studies on the protein of interest ....................29

2.4.1 Fluorescence Polarization assay .........................................................29
Zafirlukast is an Inhibitor of the Lsr2/DNA Complexation and Mycobacterial Growth

4.1 Expression and purification of Lsr2

4.2 Lsr2-DNA complexation

4.3 Zafirlukast inhibits Lsr2-DNA complexation

4.4 FP experiments performed on Zafirlukast derivatives

4.5 Zafirlukast directly binds Lsr2

4.6 Zafirlukast inhibits mycobacterial growth

4.7 Determination of Zafirlukast MIC

4.8 Zafirlukast exhibits bactericidal activity

4.9 Zafirlukast dysregulates gene transcription

5 Attempts at Lsr2 Crystallization

5.1 Synthesis of fusion proteins

5.1.1 Expression and purification of Lysozyme-Lsr2

5.1.2 Cloning of Lyso-Lsr2C

5.1.3 Expression and purification of Lsr2N-T4-Δ54-73-Lsr2C

5.2 DNA design to crystallize the Lsr2-DNA complex

6 Conclusions and Future Work

References
# List of Tables

1.1 Genes bound by Lsr2 and their functions ..............................................................8

3.1 Primers used for amplifying *lsr2* from Mtb H37Rv ........................................39

3.2 Primers used for the quantitative RT-PCR experiments ....................................45

3.3 Primers used for amplifying *lysozyme* from RB69 ............................................47

3.4 Primers used for engineering a plasmid containing *lyso-lsr2C* ......................48

3.5 Primers used for amplifying *T4 lysozyme* .........................................................48

3.6 Primers used for amplifying *T4 lysozyme* and insert “*lsr2* overhangs” ..........49

4.1 Zafirlukast and Zafirlukast derivatives functional groups .................................62
List of Figures

1-1 Lsr2 and H-NS share a similar DNA recognition mechanism ...........................................6
1-2 NMR structure of Lsr2C ........................................................................................................9
1-3 X-ray structure of Lsr2N .....................................................................................................10
2-1 The three major steps of a PCR ........................................................................................13
2-2 Principle of the restriction digest step, example with NcoI and HindIII .........................16
2-3 Principle of the FP binding assay ......................................................................................30
2-4 Conversion of MTT to formazan ......................................................................................33
2-5 Principle of quantitative RT-PCR ....................................................................................36
4-1 Lsr2 nickel affinity chromatogram ..................................................................................53
4-2 Lsr2 cation exchange chromatogram ..............................................................................54
4-3 SDS-PAGE result for Lsr2 after all the purification steps ..............................................55
4-4 Lsr2/DNA complexation monitored by FP ......................................................................56
4-5 Screening of the NCC drugs ............................................................................................57
4-6 Structure of Zafirlukast at physiological pH .................................................................58
4-7 Inhibition of Lsr2/DNA complexation by Zafirlukast shown using FP .........................59
4-8 Comparison between experiments with 10 nM probe and 25 nM probe .......................60
4-9 Model of predicted interactions between Zafirlukast and Lsr2 .......................................61
4-10 Core scaffold of Zafirlukast and its synthesized derivatives ..........................................61
4-11 Comparison of inhibition activity between Zafirlukast and its derivatives .................63
FP tests with higher concentrations of Zd05 .......................................................... 64
Direct binding between Lsr2 and Zafirlukast shown by intensity fading ............ 65
Ion ratio of Zafirlukast/Ritonavir versus the concentration of Lsr2 .................. 66
Kirby-Bauer with M. smegmatis in the presence of Zafirlukast ....................... 67
Kirby-Bauer with Mtb in the presence of Zafirlukast ..................................... 68
Kirby-Bauer with E. coli in the presence of Zafirlukast .................................. 69
Compared effects of kanamycin and kanamycin+Zafirlukast on M. smegmatis .... 70
Kirby-Bauer using combinations of aminoglycosides with Zafirlukast .......... 71
M. smegmatis phenotype with and without Zafirlukast ................................. 72
MTT viability assay for M. smegmatis in presence of Zafirlukast ................. 73
M. smegmatis time course in the presence of Zafirlukast .............................. 74
Killing of M. smegmatis by Zafirlukast .............................................................. 75
Electrophoretic gels and quantitation of the RT-PCR products ........................ 77
Plasmid map after cloning of lysozyme-lsr2 ..................................................... 80
SDS-PAGE result after purification of Lysozyme-Lsr2 .................................... 81
Idea for designing the fused gene lyso-lsr2C ............................................... 82
Schematic representation of Lsr2N-T4-Δ54-73-Lsr2C ................................... 83
Sequence alignment for Lsr2N-T4-Δ54-73-Lsr2C ....................................... 84
SDS-PAGE result of the fusion protein Lsr2N-T4-Δ54-73-Lsr2C after IMAC ... 85
Scheme of the Lsr2/20-mer DNA complexation idea ................................... 86
Scheme of the Lsr2/21-mer DNA complexation idea ................................... 87
Example models of relative orientations of an Lsr2N dimer and DNA ........... 88
List of Equations

2.1 Beer Lambert law.................................................................28
2.2 Polarization equation ..........................................................30
2.3 Binding scheme and dissociation constant equation..................31
3.1 Hill equation........................................................................42
List of Abbreviations

BC .........................before Christ
BCG ..........................bacille Calmette-Guérin
BSA ..........................bovine serum albumin

cDNA ........................complementary DNA
cfu ...........................colony forming unit
CIP ..........................calf intestinal phosphatase

DMSO ........................dimethyl sulfoxide
DNA ..........................deoxyribonucleic acid
dNTP ..........................deoxynucleotide triphosphate

E. coli ......................Escherichia coli

FP ...........................fluorescence polarization
FPLC ..........................fast protein liquid chromatography

HIV ..........................human immunodeficiency virus
H-NS ..........................histone-like nucleoid structuring

IPTG ..........................isopropyl-β-D-thiogalactoside
IMAC ..........................immobilized metal affinity chromatography

LB ..........................Luria Bertani

M. smegmatis ............Mycobacterium smegmatis
MALDI ......................matrix assisted laser desorption/ionization
MDR ..........................multidrug-resistant
MES ..........................2-(N-morpholino)ethanesulfonic acid
MIC ..........................minimum inhibitory concentration
mRNA ........................messenger RNA
MtB ..........................Mycobacterium tuberculosis
MTT ..........................3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

NCC ..........................NIH Clinical Collection
NIH ..........................National Institutes of Health
NMR ................................nuclear magnetic resonance
PCR ................................polymerase chain reaction
PE/PPE .............................proline-glutamate/proline-proline-glutamate
O.D. .................................optical density
ROI .................................reactive oxygen intermediate
RNA ..............................ribonucleic acid
RNI .................................reactive nitrogen intermediate
SDS .................................sodium dodecyl sulfate
SDS-PAGE .......................sodium dodecyl sulfate polyacrylamide gel electrophoresis
TB .................................tuberculosis
TCEP ..............................tris(2-carboxyethyl)phosphine
TDR .................................totally drug-resistant
TEMED .........................tetramethylethylenediamine
Tris .............................tris(hydroxymethyl)aminomethane
UV .................................ultraviolet
WHO .............................World Health Organization
XDR ..............................extensively drug-resistant
List of Symbols

Å..............................angstrom
°C...........................degree Celsius

μM...........................micromolar

aa.............................amino acid
bp............................base pair
Da...........................Dalton
kbp.........................kilo base pair
kDa.........................kilo Dalton
M..............................molar
mAU.........................milli absorbance unit
mM...........................millimolar
mP.........................milli polarization
nm.........................nanometer
nM...........................nanomolar
pH...........................potential of hydrogen
pI............................isoelectric point
rpm........................revolutions per minute
v/v...........................volume/volume
w/v...........................weight/volume
Chapter 1

Introduction

1.1 Tuberculosis: current status

1.1.1 Brief introduction to tuberculosis and its infectious process

Tuberculosis (TB) is a disease caused by exposure to a bacterial pathogen, *Mycobacterium tuberculosis* (Mtb). It was first characterized by the German scientist Dr. Robert Koch in 1882,¹ even though multiple examples that describe the symptoms of the disease can be found in historical texts dating back to BC era.²

Although it is curable, it claims more lives each year than any other treatable disease. About one third of the world population is infected with Mtb.³ There were 8.7 million new cases in 2011 and 1.4 million deaths were reported that year.⁴ India, China and a lot of high-burden countries also located in Asia or Africa account for most of those cases.⁴

The first stage of the infection process of Mtb is transmission through coughing or sneezing from infected patients to healthy subjects. Inhaling as low as 10 bacteria contained in aerosol droplets produced by a contagious patient is enough to develop an infection.⁵ Mtb then travels to the lungs and enters macrophages, which usually serve as a barrier to contain pathogens.⁶ Mild inflammation is caused by Mtb multiplying in the
lungs as they have evolved to survive within macrophages and are not affected by the host immune response. Most bacteria undergo phagocytosis after they are enveloped by a phagosome that resides inside the macrophage. This process typically involves autodestruction of the phagosome due to a hostile environment including acid pH and reactive oxygen and nitrogen intermediates (ROIs and RNIs), which also eliminates the foreign body. However, Mtb has evolved to survive the host immune response by following a specific cellular program that requires coordination of many key genes. The next stage of an Mtb infection consists of the formation of granulomas and often results in a clinically latent infection, which means that the bacteria lie dormant and the patient does not show any clinical symptoms. Granulomas have a central necrotic core partly made of infected macrophages surrounded by activated macrophages and T cells. Those dense layers limit Mtb growth and protect the host from developing an active infection, even though pathogens are not eliminated. The entry into latency represents one of the biggest challenges for the development of treatments effective against those slow-growing bacteria. The final stage of reactivation only occurs in about 10% of the cases, when the host’s immune system is weakened or declining: granulomas are disrupted and airways are directly exposed to free Mtb. Lung tissues are gradually destroyed and cavities appear. At that point the patient presents symptoms of pulmonary disease.

1.1.2 Development of treatment regimens

1.1.2.1 The BCG vaccine

The most commonly known treatment against tuberculosis is the bacille Calmette-Guérin (BCG) vaccine which was developed in 1921 by Albert Calmette and Camille
Guérin.\textsuperscript{12} It is an attenuated pathogen that can potentially induce an immunological response to protect the vaccinee from Mtb. Its use was imposed worldwide in 1950 by the WHO.\textsuperscript{13} It has been administered to 4 billion people ever since. Despite the addition of factors supposed to enhance the immunity produced by BCG, the vaccine has done little to contain the pandemic as it is ineffective once the individual is infected.\textsuperscript{1}

1.1.2.2 First line drugs

First line drugs are antibiotics developed against TB that are core to any treatment program. The anti-TB activity of isoniazid, an inhibitor of the cell wall synthesis,\textsuperscript{14} was first observed in 1951. Ethambutol fulfills the same function through a different mechanism\textsuperscript{15} and was discovered ten years later. The two other first line drugs are pyrazinamide and rifampicin, respectively discovered in 1952 and 1959. One major concern is the fact that the most recent first line drug discovery took place more than 50 years ago.

1.1.2.3 The issues of drug resistance and co-infection with HIV

Drug resistance represents a huge challenge for drug development. It was observed for the first time in 1948 after the isolation of the very first antitubercular agent, streptomycin, when relapses started to occur among patients during the drug trial.\textsuperscript{16} Since then, every new drug administered in monotherapy led to the emergence of new resistant strains. It was soon revealed that the use of multi-drug regimens was a necessary strategy to adopt in the treatment of TB.\textsuperscript{17}
The current typical regimen for treating a drug susceptible Mtb infection has a minimum duration of 6 months: all of the four first line drugs are administered daily for 2 months during the initial phase then isoniazid and rifampicin are given for an additional 4-month period during the continuous phase. The length of the treatment and the necessity of getting access to medication over a long period of time often lead to patient non-compliance which is also to blame for the appearance of resistant strains, especially in developing nations. Over the past two decades, strains classified as multidrug-resistant (MDR) started to complicate TB treatments as they are not sensitive to at least 2 core drugs, isoniazid and rifampicin. Unfortunately, this was followed by the emergence of extensively drug-resistant TB (XDR) that is resistant to all first line drugs as well as second-line drugs usually used to treat MDR-TB, such as fluoroquinolones and aminoglycosides. Most recently, totally drug-resistant (TDR) cases have been reported in India and no available drug with known anti-TB activity was able to cure those patients.

The rising incidence of infections from MDR and XDR strains is not the only issue encountered for treating TB. The HIV pandemic and high rates of Mtb/HIV co-infection have substantially worsened the situation. Weakening of the immune system accelerates the activation of TB and it has been estimated that 90% of HIV-infected patients die within months of contracting TB if they are not properly treated. Besides, HIV treatment and TB treatment are contraindicative. Rifampicin, the most efficient first line drug to date for TB treatment, strongly stimulates cytochrome P450 in the liver, thus increasing metabolism of drugs such as protease and non-nucleoside reverse transcriptase
inhibitors, two major families of drugs used for HIV treatment. Novel routes for treating TB are desperately needed.

1.2 Roles of Lsr2 in *Mycobacterium tuberculosis*

1.2.1 Lsr2: histone-like protein

Lsr2 stands for Leprosy serum reactive clone 2 and was first identified in 1990 in *Mycobacterium leprae* as a potent T-cell antigen. This nucleoid-associated protein can be found in *Mycobacterium tuberculosis* and related actinomycetes. It is a small basic protein (12 kDa) that exists as a homodimer and contains 112 amino-acids (Appendix A). Lsr2 consists of two major domains: the N-terminal dimerization domain (residues 1-65) and the C-terminal DNA-binding domain (residues 66-112). It has been characterized as a functional equivalent of histone–like nucleoid structuring (H-NS) proteins found in *E. coli* and related gammaproteobacteria because of its ability to non-specifically bridge double-stranded DNA, despite the lack of any sequence homology between the two proteins. Genetic complementation experiments have also shown that Lsr2 can complement *hns* knock-outs in *E. coli* and H-NS can complement *lsr2* knock-outs in *M. smegmatis*. H-NS acts as a global silencer of gene transcription by maintaining DNA in a condensed state. Some regions are then harder to access as H-NS plays a shielding role. Similarities with H-NS make Lsr2 a major contributor to nucleoid compaction and bacterial chromatin organization as well as a regulator of gene expression. Additionally, over-expression of H-NS in *E. coli* causes the bacteria to enter an artificial stationary phase where they are viable but unable to grow, similar to the latency stage observed in
Mtb. Analogy between H-NS and Lsr2 suggests a potential role of Lsr2 in the entry of Mtb into a dormant phase, but this has yet to be confirmed.

Structural analysis has provided evidence that H-NS and Lsr2 use a common DNA recognition and binding mechanism through a loop that acts like a hook and contains a conserved Q/RGR motif (Figure 1-1).\textsuperscript{32} The shape of DNA within the minor groove, which is where those two proteins bind, is a determinant of binding specificity: low binding is observed when the minor groove is too wide (GC-rich sequences) or too narrow and the highest affinity is for grooves that exhibit mixed AT-rich sequences.\textsuperscript{32}

Figure 1-1: Lsr2 (residues 80-110 shown) and H-NS (residues 104-126 shown) share a similar DNA recognition mechanism through a loop that acts like a hook (adapted from references 29 and 32). PDB files: 2KNG, 1FQ2, 2L93

1.2.2 Lsr2: global transcriptional regulator

Lsr2 preferentially binds AT-rich regions of GC-rich mycobacterial DNA, generally associated with laterally acquired DNA from phages and transposons and gene promoters.\textsuperscript{33} It therefore acts as a global repressor of transcription.\textsuperscript{34} Chromatin
immunoprecipitation experiments showed that it binds 840 of the 4003 genes identified in Mtb, which represents 21% of the genome, and can silence gene expression through this process.\textsuperscript{29} Past studies performed in \textit{M. smegmatis} (amino acid sequences of Lsr2 in \textit{M. tuberculosis} and \textit{M. smegmatis} are 90% identical) had demonstrated that Lsr2 plays an important role in colony morphology and biofilm formation.\textsuperscript{35} Indeed, \textit{lsr2}-inactivated strains could not produce mycolyl-diaglycerols, thus decreasing the cell wall hydrophobicity and impacting the formation of surface pellicles.\textsuperscript{35} Another study showed sliding motility on solid surfaces of \textit{M. smegmatis} mutants containing an \textit{lsr2} deletion and it was then speculated that Lsr2 could play a regulatory role in mycobacteria, as its inactivation was causing an altered cell wall composition responsible for the observed hypermotile phenotype.\textsuperscript{36} Lsr2 was also found to be involved in regulating the expression of a subset of antibiotics-inducible genes that confer multi-drug tolerance.\textsuperscript{37} Genome-wide mapping confirmed the importance of Lsr2 in gene regulation and allowed the identification of other critical genes bound by Lsr2, for instance involved in energy metabolism, cell wall biosynthesis, DNA replication or stress response (Table 1.1).\textsuperscript{29} It also revealed binding of many key genes in mycobacterial virulence.\textsuperscript{29} The inability to construct Mtb mutants that lack the \textit{lsr2} gene seems to indicate that Lsr2 is essential for viability in \textit{Mycobacterium tuberculosis}.\textsuperscript{37-38}
Table 1.1: Genes bound by Lsr2 and their functions (adapted from reference 29)

<table>
<thead>
<tr>
<th>Functional categories</th>
<th>Number of genes bound</th>
<th>Total number of genes (% bound by Lsr2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virulence, detoxification, adaptation</td>
<td>32</td>
<td>226 (14.2 %)</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>46</td>
<td>238 (19.3 %)</td>
</tr>
<tr>
<td>Information pathways</td>
<td>40</td>
<td>232 (17.2 %)</td>
</tr>
<tr>
<td>Cell wall and cell processes</td>
<td>157</td>
<td>751 (20.9 %)</td>
</tr>
<tr>
<td>Insertion sequences and phages</td>
<td>13</td>
<td>147 (8.8 %)</td>
</tr>
<tr>
<td>PE/PPE (speculated immunological importance)</td>
<td>89</td>
<td>168 (53.0 %)</td>
</tr>
<tr>
<td>Intermediary metabolism and respiration</td>
<td>157</td>
<td>898 (17.5 %)</td>
</tr>
<tr>
<td>Regulatory proteins</td>
<td>42</td>
<td>193 (21.8 %)</td>
</tr>
<tr>
<td>Conserved hypotheticals/unknown</td>
<td>264</td>
<td>1163 (22.7 %)</td>
</tr>
</tbody>
</table>

1.2.3 Lsr2: physical barrier against ROIs

As mentioned previously in the tuberculosis pathogenesis section, reactive oxygen intermediates (ROIs) are generated when the macrophage is infected with a pathogen. This constitutes a defense mechanism from the host that typically kills bacteria by damaging their DNA. The relatively non-specific DNA-binding property of Lsr2 suggested that the protein may act as a physical protection against oxidative damage. Experiments showed that Lsr2 could protect mycobacterial DNA from hydroxyl radical...
degradation through direct DNA-protein binding. Additionally, the <i>M. smegmatis lsr2</i> knockout was more susceptible to H<sub>2</sub>O<sub>2</sub> than the wild type strain.

### 1.2.4 Current structural knowledge about Lsr2

There is no published structure of the full-length protein to date. Lsr2 contains a flexible region at the junction of its domains which is probably the reason why its crystallization has not yet been achieved. However, the C-domain (Lsr2C) and the N-domain (Lsr2N) have been separately solved using two different techniques. The structure of Lsr2C, the DNA-binding domain, was solved by nuclear magnetic resonance (NMR) methods. It consists of two α-helices linked by a long loop (Figure 1-2). NMR titration experiments performed with varying amounts of DNA allowed modeling of its interaction with the protein. This first structure showed that, although Lsr2 and H-NS proteins share a lot of functional similarities, these two proteins are distinct structurally.

![Figure 1-2: NMR structure of Lsr2C (PDB: 2KNG)](image)

29
Crystallization of Lsr2N, the dimerization domain, was eventually achieved and its structure, a compact homodimeric α/β, was solved using X-ray crystallography.\textsuperscript{41} It revealed chains of dimers where each N-terminal region pairs with a neighboring one (Figure 1-3), which suggested a spontaneous oligomerization mechanism of Lsr2 that could explain how the protein achieves DNA condensation.

![Image of X-ray structure of Lsr2N (PDB: 4E1R). Each monomer within each dimer is represented in a different color (green or pink). Other dimers were generated by crystallographic symmetry.](image)

**1.3 Project goals**

As previously mentioned, Mtb is a virulent bacterium that causes a disease increasingly harder to eradicate because of the emergence of more and more drug-resistant strains or co-infection with HIV. Furthermore, most people infected with Mtb have a latent infection, which is not really susceptible to drugs that typically target metabolically active cells. Even though treatment regimens are available and effective for most patients, they are extremely long and stringent. The fact that the last major drug discovery in the field of TB research occurred more than 50 years ago is quite alarming: novel treatments and mechanisms of action are desperately needed.
Because of its important role as a global regulator of gene transcription and its essential role in Mtb viability, Lsr2 appears to be a very attractive target for drug development. Indeed, Mtb virulence requires coordination for expression of key genes, so identifying compounds that can disrupt this order could potentially open the way to global dysregulation of bacterial transcription and hinder its ability to survive within the host.

The first goal of this project was to develop a high-throughput assay allowing screening of a multitude of compounds to identify inhibitors of the Lsr2 DNA-binding activity. The second goal of this project was to try to gain insight into the structure of the full-length protein by attempting its crystallization. Together, these two aims would inform the design of novel compounds potentially leading to new routes of treatment against tuberculosis.
Chapter 2

General Methods: Principles

2.1 Introduction

Our facilities are not equipped to study dangerous organisms, thus Mtb strains are not used to produce the protein of interest. Instead, molecular biology techniques are employed to genetically engineer non-pathogenic E. coli so that they can meet this purpose. This is generally followed by separation steps in order to isolate the protein of interest from the rest of the expressed bacterial proteins. The purified protein can then be utilized in further experiments.

This chapter describes the different techniques used to go from a gene of interest to a protein of interest as well as the purification and characterization procedures. The principles of all the tests and assays employed in this project are also detailed.

2.2 From genes to proteins

2.2.1 Polymerase chain reaction

The polymerase chain reaction (PCR) is a powerful tool that enables amplification of a specific sequence of DNA. The reaction requires the presence of DNA template,
primers, nucleotides (dNTPs) and DNA polymerase. The DNA template contains the target sequence, such as the gene of interest. Primers are small pieces of single-stranded DNA designed to be complementary to the 5’ and 3’ ends of the target sequence. They also contain sequences that are recognized by some specific restriction endonuclease enzymes for cloning purposes that will be detailed in the following steps. The dNTP mix contains deoxynucleotide triphosphates which are single units of the bases A, T, G and C. They are the building-blocks of the new DNA strand. DNA polymerase is the enzyme that assembles the new DNA strand complementary to the DNA template. A buffer solution providing the optimum chemical environment for the polymerase activity and stability is also always added.

A PCR can be divided into three steps: denaturation, annealing and elongation (Figure 2-1).
The denaturation consists in separating the two strands of the double-stranded DNA template by disrupting hydrogen bonds between them at a high temperature, typically 98 °C. Once two separate single-strands are obtained, primers can specifically pair with the ends of the target sequence during the annealing step, usually performed 3 to 5 °C below the predicted melting temperature of the primers. The enzyme then binds the primer-template hybrids and elongates the primers along the DNA template by using complementary nucleotides during the elongation step. This step takes place at the optimal temperature for the activity of the polymerase (68 °C for the Pfx50™ DNA Polymerase and 72 °C for the Phusion® High-Fidelity DNA Polymerase) and its duration depends on the length of the target sequence that needs to be amplified and the enzyme used (1000 bp/min for Pfx50™ and 2000 bp/min for Phusion®). Those three steps constitute one cycle. The cycle is typically repeated 20 to 40 times. Once all the cycles are over, the reaction mixture undergoes the final elongation step at the same temperature as the elongation for 5 minutes and is then held at 4 °C for an indefinite time. To ensure that the desired PCR product was synthesized, the reaction mixture can then be loaded on an agarose gel.

2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a method employed to separate and characterize different DNA fragments. For instance, it is used after a PCR reaction to separate the desired product from the primers and the DNA template. An electric field is applied to the gel and DNA molecules migrate based on their size: larger molecules move slower
than smaller ones. DNA molecules are all negatively charged because of their phosphate backbone and therefore migrate towards the positive electrode.

The gel is made of a polymer matrix, typically agarose, at a concentration varying between 1 and 2% w/v depending on the desired resolution (better separation for 100 bp – 1 kbp fragments when the agarose percentage is higher). Agarose is dissolved in 1X TAE (Tris/Acetate/Ethylene) buffer and SYBR® safe dye is added. The latter is the gel stain that intercalates between the DNA base pairs of the fragments running through the gel, allowing their observation by fluorescence with a blue-light transilluminator.

The samples are loaded into different wells as well as a DNA ladder with a specific set of fragments of known sizes. It is used as a reference that allows evaluation of the size of the “unknown” DNA pieces depending on their relative position on the gel. The gel is visualized on the transilluminator and the DNA products can be recovered by gel extraction. Desired bands are cut out of the gel and treated with a QIAquick® Gel Extraction kit that allows obtaining purified DNA.

### 2.2.3 Restriction digest

The aim of this step is to insert the amplified gene into a plasmid, for further replication. For this purpose, restriction enzymes are used to digest the PCR product and the plasmid in which it has to be inserted. Those enzymes recognize specific palindromic sequences, called cut sites, and cleave DNA within those sequences by hydrolysis of the phosphodiester bond. The enzymes are chosen so that their corresponding recognition sequences do not appear within the target sequence. The primers used in the PCR reaction are engineered to contain the recognition sequences that match the picked
restriction enzymes, which are then present at each end of the PCR product. Plasmids usually contain multiple recognition sites for different restriction enzymes.

The restriction digest is performed on both the PCR product and the chosen plasmid (Figure 2-2).

Figure 2-2: Principle of the restriction digest step, example with NcoI and HindIII

The procedure leads to the formation of “sticky” ends when the cut sites are staggered, which is typically the preferred option as it allows base pairing of the overhangs of the PCR product with the cut plasmid during the ligation step. The reaction mixture generally consists of the amplified target sequence or plasmid, each of the restriction enzymes and their appropriate optimal buffer, as well as bovine serum albumin.
(BSA) to stabilize the enzymes. Calf-intestinal alkaline phosphatase (CIP) is also added for the plasmid digestion to prevent its self-ligation. The mixture is then incubated at 37 °C for 1 hour. The digested plasmid and PCR product then undergo agarose gel electrophoresis and DNA is extracted from the obtained bands.

2.2.4 Ligation

When the purified digested products from the previous step are obtained, they can be pasted together using DNA ligase that catalyzes the formation of a phosphodiester bond between the 5’ phosphate end of a DNA sequence and the 3’ hydroxyl end of another one. The reaction is performed for 5 min at room temperature using the Rapid DNA Ligation Kit from Roche Applied Science and the mixture contains the digested plasmid and PCR product, along with the DNA ligase and its appropriate buffer. If the ligation is successful, the plasmid is then engineered with the target sequence that was amplified during the PCR process and is ready to be incorporated into bacteria.

2.2.5 Transformation of competent cloning cells: plasmid replication

The replication of the new plasmid that contains the gene of interest is achieved by *E. coli* allowed to divide and multiply once they have incorporated it, in order to produce multiple copies. Competent cloning strains are “forced” to take in foreign DNA through the transformation process. They are called competent because of their likelihood to accept DNA from their environment, as their membrane is made permeable by a calcium chloride treatment. Generally, plasmids are engineered with a specific antibiotic resistance gene that allows selection for bacteria that take in the desired plasmid. The
plasmids used in the experiments presented in this thesis are pET-32 vectors, which are ampicillin and carbenicillin resistant (these two drugs can be used interchangeably and are both related to penicillin, they only differ in stability toward heat degradation).

To carry out a transformation, the ligation product is added to competent cloning cells, typically OneShot®TOP10, and the mixture sits on ice for 30 minutes. The cells are then heat-shocked at 42 °C for 30 seconds. This step promotes the uptake of the extragenomic DNA by opening up the discontinuous membrane of bacteria which allows an easier absorption of the plasmid present in solution. Cells then sit on ice again for 2 minutes in order to reconstitute the membrane. Luria-Bertani (LB) growth medium is then added and the bacteria are incubated at 37 °C (optimal temperature for them to grow) for 1 hour. Differentiation between cells that accepted the desired plasmid and those that did not is achieved in the next step, which consists of spreading the liquid culture on a plate made of LB medium, agar and the drug corresponding to the antibiotic resistance gene present in the uptaken plasmid. Thus, only bacteria that contain the plasmid with the gene of interest form colonies. Colonies are visible after an overnight incubation of the plate at 37 °C.

2.2.6 Plasmid purification

Once colonies appear on the plate, it is a fair assumption to think that they contain the plasmid with the gene of interest as they are resistant to the antibiotic present in the plate solid medium. Each colony can then be picked using a sterile pipette tip to inoculate a 5 mL liquid culture (5 mL LB medium as well as the antibiotic to which the chosen bacteria are supposed to be resistant). Cultures are shaken and incubated at 37 °C until
the cell density reaches an O.D. 600nm of 0.6-0.8. Bacteria are then pelleted at the bottom of the tube during a centrifugation step. The QIAquick® Spin Miniprep Kit is used to isolate and purify the desired plasmid, which is in solution and ready for further characterization and experiments at the end of the procedure.

To confirm the presence of the gene of interest in the purified plasmids, a test restriction digest can be performed. A sample of those plasmids undergoes the same process as for a restriction digest and the mixture is then allowed to migrate on a DNA gel. If the right size band (same number of bp as the gene of interest) appears on the gel upon digestion of the plasmid, it is then submitted for DNA sequencing (Eurofins MWG Operon). If the insertion of the target sequence is confirmed, the plasmid will be inserted into expression cells that allow the synthesis of the protein of interest.

2.2.7 Transformation of competent expression cells

The protocol for transforming *E. coli* competent expression cells is the same as for the cloning strain. The only difference is that the expression cells are engineered to express the gene for T7 RNA polymerase, an enzyme necessary for the transcription of the gene of interest present in the pET vectors. The competent cells used for expression of the protein of interest are usually Rosetta™ cells. This strain of *E. coli* also contains an antibiotic resistance gene to chloramphenicol. This time, the liquid culture is spread on a plate that contains both carbenicillin and chloramphenicol and incubated at 37 °C. The colonies grown on the plate are transferred into 5 mL cultures of LB medium and both antibiotics also incubated and shaken at 37 °C. Once cultures are turbid, they can be used to inoculate 100 mL cultures and then 4 L cultures. When the O.D. 600 nm reaches 0.6,
glycerol stocks are prepared: a sample of the culture is taken and glycerol is added to a final concentration of 20% (v/v). These stocks can be stored at -80 °C until needed. Having those glycerol stocks avoids having to go back to the transformation step if more cultures are needed later on.

2.2.8 Induction of protein expression

At this point, the cells are ready to be induced to make the protein of interest. In *E. coli*, the *lac* operon system plays a very important role. Indeed, protein synthesis needs a starting signal as it does not automatically begin. It consists of a promoter starting the translation, an operator and three characteristic genes: *lacZ*, *lacY* and *lacA*, that encode for enzymes involved in the metabolism of lactose. The whole system is regulated by the *lacI* gene encoding for the *lac* repressor. In absence of lactose, the *lac* repressor binds to the *lac* operator. Upon this binding event, the transcription of the *lac* operon system and the subsequent genes is inhibited as RNA polymerase cannot access the promoter anymore. When lactose is added, it allows for the formation of allolactose that binds the *lac* repressor. The resulting conformational change of the repressor prevents it from binding the operator: transcription can then start.

The pET plasmids used in the laboratory are engineered to contain a *lac* operator as well, which offers a double layer of protection (added to the necessary presence of T7 RNA polymerase) against “leaky” expression. Isopropyl-β-D-thiogalactoside (IPTG) is an allolactose analogue commonly used as a starting signal for gene transcription in *E. coli*. When IPTG is added, the complex between the *lac* repressor and the *lac* operator is disrupted on both chromosomal plasmids and pET vectors. The T7 RNA Polymerase can
then be expressed from the chromosomal DNA and starts transcribing the genes inserted between the T7 promoter and the T7 terminator on the pET vector. 1 mM IPTG is typically added to a culture that reached an OD$_{600} \text{nm}$ of 0.6 once its temperature has been allowed to equilibrate to 16 °C. Cells are incubated and shaken at 16 °C for about 16 hours.

### 2.2.9 Cell harvesting and lysis

After bacteria have been allowed to synthesize proteins for a sufficient amount of time, the culture is harvested by centrifugation at 4000 rpm, at 4 °C for 30 minutes, using a swing bucket rotor. The centrifuge is a 5810-R Eppendorf. The obtained pellets are combined and resuspended into a binding buffer chosen upon the type of column used for the first step of protein purification. At that point, bacteria are whole cells in solution that contain the protein of interest: their membrane and cell wall has to be broken down in order to release the bacterial proteins into the solution. Initially, cell wall is enzymatically degraded under the action of lysozyme (3000X stock solution). DNaseI (10000X stock solution) is also added in order to degrade the DNA present in solution. The resuspended cells usually sit on ice for 30 minutes once the enzymes have been added so that they have time to catalyze their respective degradation reactions. The second phase of the lysis procedure is mechanical: the crude lysate is sonicated on a Misonix Sonicator 3000. Sonication utilizes ultrasounds to disrupt the cell membrane. Cell debris and degradation products are then spun down by centrifugation at 12000 rpm for 15 minutes using a fixed-angle rotor. If the protein of interest is soluble, it should remain in solution after
this step. The supernatant is filtered using 0.2 µm syringe filters and is then ready for the first chromatographic step of protein purification.

2.3 Protein purification

2.3.1 Fast protein liquid chromatography (FPLC)

Upon induction of the cells to produce the protein of interest, multiple bacterial proteins are also synthesized. Protein purification consists in isolating the desired protein from the rest of those proteins and the DNA left in solution. The type of chromatography is chosen depending on the characteristics of the protein that needs to be purified such as its size, the presence of any tag or its isoelectric point (pI). The apparatus used for liquid chromatography is an ÄKTA FPLC™ that is compatible with a lot of different columns.

2.3.1.1 Immobilized metal affinity liquid chromatography (IMAC)

When the protein of interest is fused with a poly-histidine tag, IMAC is a method of choice. It is easy to engineer such proteins when pET vectors are used for cloning. Indeed, those plasmids contain codons encoding for 6 consecutive histidine residues that are typically located next to the gene of interest. The protein then has six histidines attached to one of its termini, depending on the restriction enzymes used for cloning. Polyhistidine tags have a very good affinity for nickel (II). The columns used are 5 mL HisTrap™ and their resin is sepharose chelated to Ni^{2+} ions that will coordinate the polyhistidine tag of the protein of interest.

Once the column has been washed and equilibrated by running binding buffer through it, the filtered supernatant obtained after lysis is loaded onto the column using a
syringe. It is made of bacterial proteins and DNA solubilized into the binding buffer (that was used to resuspend the cells in the harvesting step) which contains a low concentration of imidazole (typically 25 mM) to minimize non-specific binding. During the washing step, the same buffer is run through the column to remove impurities. Only the tagged protein should remain attached to the resin. A gradient of high imidazole buffer (typically 250 mM) is then applied to the column and gradually displaces the bound tagged protein during the elution step. Each histidine residue has an imidazole side chain, which explains why an excess of imidazole in solution competes with the tag for binding to Ni
. This forces the protein to come out of the matrix: a peak then appears on the chromatogram as the absorbance at 280 nm is measured the whole time (proteins in solution have an absorbance maximum at this wavelength). Even though this technique allows separation of the protein of interest from a lot of other soluble E. coli proteins, it is often necessary to apply other purification steps to obtain a pure protein.

2.3.1.2 Ion exchange chromatography

Ion exchange chromatography is a purification technique based on the ionic interactions between the protein and the solid support. In cation exchange chromatography, positively charged molecules interact with the negatively charged resin, while for anion exchange chromatography, negatively charged molecules bind to the positively charged resin. The strength of those interactions is dependent on the number of charges on the molecule as well as the pH of the binding buffer. By gradually increasing the salt concentration after the protein has been allowed to bind the resin, the molecules with the weakest ionic interactions start to elute. Stronger interactions require a really
high salt concentration to be disrupted. The isoelectric point (pI) of the protein of interest, which is the pH at which the protein is electro-neutral, is a very important parameter to consider for ion exchange chromatography. If the protein is in a buffer at a higher pH than its pI, it will be negatively charged; similarly, in a buffer at a lower pH than its pI, it will be positively charged.

For the protein studied in this thesis, cation exchange chromatography is used. Performing cation exchange chromatography enables separation of the protein from any remaining nucleic acids and other proteins that display a different overall charge in the buffer used. The column used is a HiTrap™ SP FF. Its stationary phase is made of sepharose cross-linked with negatively charged ligands (sulphopropyl).

Before this purification step can be performed, the protein of interest has to be dialyzed into an appropriate buffer (the principle of dialysis will be detailed in the next subsection). The protein studied in this thesis has a pI of approximately 9, which means that a slightly acidic buffer will give it an overall positive charge and it will be attracted to the negatively charged resin. The binding buffer contains a low concentration (150 mM) of a highly soluble salt, typically NaCl. After the column has been washed, equilibrated and loaded with the solution containing the protein of interest similarly to the steps performed for IMAC, a gradient of salt can be applied (up to 1 M NaCl). As the concentration of NaCl gradually increases, the positive protein exchanges with Na\(^+\) ions and elutes from the column and can be collected. Absorbance at 280 nm is also monitored throughout this process and a peak appears on the chromatogram when the protein is displaced from the column.
2.3.2 Protein dialysis

Dialysis is a very useful procedure that allows separation of proteins from small solutes. Its most frequent application is for changing the buffer conditions of the solution the protein is in. A semi-permeable membrane is used: molecules that have dimensions greater than the pores of the membrane (10000 Da) are retained inside the bag while smaller molecules and ions pass through the pores and are then out of the dialysis bag. For instance, after the IMAC step, when cation exchange chromatography is necessary, the protein has to be transferred into a buffer with a significantly lower pH and imidazole has to be removed: the buffer exchange can be easily and efficiently done using dialysis.

In some instances, when the protein is tagged with an N-terminus polyhistidine tag that needs to be cleaved (for crystallization or assay purposes), aliquots of PreScission Protease are also added to the dialysis bag. pET vectors used in our laboratory are engineered to encode for a PreScission Protease cut site (LeuGluValLeuPheGlnGlyPro) located between the polyhistidine tag and the protein of interest. When the protease is added, it recognizes this specific amino-acids sequence and cleaves between the Gln and Gly residues. The mixture of cleaved tag, protease, untagged protein and tagged protein can be purified using metal affinity chromatography to separate the untagged protein from the rest of the components.

To perform a dialysis, the first step is to cut the desired length of membrane and hydrate it for a few minutes. One end of the bag is then sealed, the protein solution is transferred into the membrane and the other end is sealed as well. The dialysis bag is placed into a large amount of the “new” buffer with the desired conditions at 4 °C and
usually sits overnight. To increase the dialysis efficiency, the buffer can be switched to fresh buffer a few times during this process.

**2.3.3 SDS-PAGE: characterization of protein purity**

SDS-PAGE stands for sodium dodecyl sulfate polyacrylamide gel electrophoresis. It is used to separate proteins or to assess the purity of a protein, the latter being the most common application for this project. This procedure can be performed at any point during protein purification and a gel is typically run after each chromatography step.

Proteins migrate on a gel under an electric field and are separated only according to their molecular weight because of the sodium dodecyl sulfate (SDS) treatment. SDS is a detergent that denatures proteins and gives them an overall negative charge. Proteins present in the sample are consequently all unfolded and negative, so their rate of migration on the gel is not dependent on their original shape or charge and they can all move towards the positive electrode.

The gels are made of acrylamide, SDS, ammonium persulfate and two different Tris buffers with an adjusted pH. Tetramethylethylenediamine (TEMED) is added to initiate polymerization. The concentration of acrylamide in the gel usually varies between 8 and 15%: lower percentages gels are preferred to get a good resolution for larger proteins while higher percentages are necessary to resolve smaller proteins. Samples are prepared by mixing them with loading buffer that notably contains SDS and a blue dye that allows direct visualization of the migration front during electrophoresis. They are then heated up to 95 °C for a few minutes to promote protein denaturation. Each sample is loaded into one of the wells of the gel. A protein marker is also added in another lane.
It contains proteins of known sizes that allow a good estimation of molecular weights for all the proteins present in the different samples.

The gel is run in a Mini-PROTEAN® Tetra Cell filled with 1X Tris/glycine/SDS running buffer at 15 mA for 15 minutes and 30 mA for about 40 minutes. To be able to visualize the different bands corresponding to all the proteins present in solution, the gel has to be stained using a solution of Coomassie brilliant blue. The dye forms a complex with proteins that stabilizes the anionic form of the dye producing the blue color. The gel is then placed in a de-staining solution of 5% acetic acid: after a few hours it becomes colorless but blue bands corresponding to protein/dye complexes are still visible. When purity of a protein needs to be achieved, the desired result is a single band with the expected molecular weight.

2.3.4 Determination of protein concentration

Once the protein has been purified, it can be used for setting up crystallization trays or assays. Being able to determine and adjust the protein concentration for those experiments is a necessity. For this purpose, the absorbance of the protein solution is measured at 260 nm and 280 nm using a Synergy™ H4 Hybrid™ multi-mode microplate reader. Those two wavelengths are chosen because they are respectively absorbance maxima for nucleic acids and proteins. The ratio between the absorbance values $A_{260}/A_{280}$ obtained enables the estimation of the content of the solution. The absorbance rests on the Beer-Lambert law (Equation 2.1).
\[ A = \epsilon lc \]

Equation 2.1: Beer Lambert law

In the Beer-Lambert law, \( \epsilon \) represents the extinction coefficient, \( l \) the optical pathlength and \( c \) the concentration of the sample. The extinction coefficients at 260 nm and 280 nm for proteins (at 1 mg/mL) are 0.57 and 1.00 respectively. For nucleic acid solutions (also at 1 mg/mL), they are 20 and 10 respectively. It means that getting an \( A_{260}/A_{280} \) ratio equal to 0.57 indicates that the protein purity is 100%, while pure nucleic acids would have an \( A_{260}/A_{280} \) ratio of 2.0. Values ranging from 0.5 to 0.7 are acceptable to deem a protein solution pure.

Measurements are performed using a Take3 micro-volume plate and the data is analyzed by the Gen5™ software, which provides \( A_{260}/A_{280} \) values. The molar extinction coefficient of the protein of interest can also be entered in the software that then gives the concentration of protein in the sample. Each concentration measurement only requires 3 \( \mu \)L of sample, which is one of the main advantages of spectrophotometric determination of protein concentration. This technique is also another way of assessing the purity of a protein sample.

The concentration of a protein solution can be adjusted by simple buffer dilution if it is too high. However, when a more concentrated protein sample is needed, Amicon® centrifugal filter units can be used. They consist of a vertical porous membrane placed upon a plastic conical tube and sealed. The protein solution is dispensed into the membrane and the closed tube is centrifuged for a few minutes. The solution remaining inside the membrane still contains the protein as the size of the pores is chosen depending
on the molecular weight of the protein of interest so that it can be retained. Only the buffer and smaller impurities can pass through the membrane. This step can be repeated until the protein solution reaches the desired concentration.

2.4 Binding and inhibition studies on the protein of interest

2.4.1 Fluorescence polarization assays

Fluorescence polarization (FP) is a technique that can be used to analyze molecular binding, and more particularly in this thesis, DNA-protein interactions. A major application of FP in the biopharmaceutical industry is its use for high-throughput screening of compounds that can potentially inhibit interactions between DNA and a specific protein. This aspect of the FP assay was developed in this project.

In an FP experiment, a plane-polarized light is applied to the sample containing molecules each attached to a fluorophore such as fluorescein which will emit light back if it is excited. If those molecules remain stationary between excitation and emission, the emitted light will be in the same plane as the polarized light that was applied in the first place, yielding a high polarization value. However, small molecules tumble rapidly in solution, so the probability of them not rotating is really low. FP assays are based on this observation: if a bigger molecule that binds fluorescently labeled molecules is added to the solution, the complexes will rotate slowly in solution, which increases the chance for the emitted light to be aligned with the initial plane (Figure 2-3).
Figure 2-3: Principle of the FP binding assay

Polarization values are calculated from the fluorescence intensities in the vertical direction ($I_\perp$, perpendicular to the plane of the polarized light) and in the horizontal direction ($I_\parallel$, parallel to the plane of the polarized light) (Equation 2.2).\textsuperscript{46}

$$P = \frac{I_\parallel - I_\perp}{I_\parallel + I_\perp}$$

Equation 2.2: Polarization equation

FP binding assays for DNA-protein interactions involve a fluorescently labeled oligonucleotide. To conduct a DNA-protein binding study, the concentration of fluorescently labeled oligonucleotide (ligand, $L$) is held constant while the amount of protein (receptor, $R$) gradually increases. The FP signal is measured for each different
The concentration of protein and the values are used to generate a binding curve. The dissociation constant $K_d$ can be determined from those data (Equation 2.3).

$$R + L \rightleftharpoons RL$$

$$K_d = \frac{[R][L]}{[RL]}$$

Equation 2.3: Binding scheme and dissociation constant equation

The data fitting and analysis are performed using Prism (GraphPad software). The other situation to consider for FP assays is when a potential inhibitor is added to the mixture of protein and DNA. If adding a molecule to the mixture causes a significant decrease of the FP signal, it is a potential hit as an inhibitor of the interactions between the protein and the DNA probe. High-throughput screening of drug molecules can be developed based on this principle, which will be described later on. The inhibition constant $K_i$ of a specific drug can also be determined using the same software, based on an inhibition curve.

### 2.4.2 Intensity fading MALDI mass spectrometry

Intensity fading is a methodology that uses matrix-assisted laser desorption/ionization (MALDI) to monitor the decrease in intensity of a specific inhibitor upon binding the protein it interacts with. It is useful for characterizing non-covalent interactions between a protein and a drug molecule. A fixed amount of inhibitor is dissolved in a buffer system that does not denature the protein of interest. An internal standard has to be added to the solution. Typically, it is another molecule that does not
bind the protein of interest and is in the same mass range as the drug of interest. The masses should not be too close either so that the two drugs can be differentiated on the spectrum without any confusion. The same amount of solution is then dispensed into several tubes and varying quantities of protein are added. The different mixtures are spotted on a MALDI target plate and analyzed with a Bruker Daltonics UltrafleXtreme mass spectrometer. The peaks corresponding to the inhibitor and the standard are monitored and intensity ratios are calculated. It can be observed that as the concentration of protein increases, the peak corresponding to the inhibitor gradually fades away. This is caused by the formation of the protein/inhibitor complex that decreases the amount of free drug in solution, thus weakening its intensity. The internal standard is not affected by the increase in protein concentration, which is why the intensity of the inhibitor is compared to the intensity of the non-binding drug for each measurement.

2.5 Tests for growth inhibition of bacteria

2.5.1 Kirby-Bauer disk-diffusion assays

Kirby-Bauer in vitro disk-susceptibility tests are an easy way to detect growth inhibition of bacteria due to the presence of a specific drug or amount of drug. For these experiments, a liquid culture of the chosen bacteria is spread on an agar plate that contains an antibiotic if necessary. The desired amounts of antibiotics are dispensed onto several antimicrobial disks that are then placed away from each other on the plate. The plate is incubated at 37 °C until sufficient bacterial growth is observed. When no growth is visible around a disk, it means that the amount of drug present inhibits bacterial
growth. Zones of inhibition can vary in size, characterizing the efficiency of the inhibition.

2.5.2 Cell viability MTT colorimetric assay

The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide or thiazolyl blue tetrazolium bromide (MTT) assay for cell viability is a useful tool for determining the minimum inhibitory concentration (MIC), defined as the lowest concentration of a specific drug necessary to inhibit the growth of an organism in vitro. The assay relies on the conversion of MTT to formazan (Figure 2-4) under the action of reductase enzymes.

![Figure 2-4: Conversion of MTT to formazan](image)

The amount of formazan produced upon MTT exposure is directly proportional to the number of living cells present in a suspension. In dead cells, the enzymes are not functional anymore, thus they cannot convert MTT to formazan. For cell viability and proliferation, absorbance can be monitored in order to estimate the number of living cells in a sample once a standard curve has been established. For MIC determination assays,
the typical procedure is to grow the desired bacteria in LB medium until turbidity is observed and dilute the culture to a final bacterial density of $5 \times 10^5$ colony forming units (cfu)/mL. The diluted culture is then split into several tubes and different amounts of a specific inhibitor are added to each aliquot. The bacteria are incubated in presence of the inhibitor for 24-48 hours and transferred into a 96-well plate. MTT is added to each well. Color changes are generally visible after 20 minutes. When a well contains purple crystals it then means that the concentration of inhibitor was not sufficient to inhibit the growth of all the organisms present in the cell suspension. However, when the solution remains yellow and no crystal is visible, it can be concluded that growth was inhibited. The MIC is defined as the lowest concentration of inhibitor that inhibits growth of all bacteria present in the suspension.

2.5.3 Tests for bactericidal activity of a drug

If a drug is shown to inhibit bacterial growth, this inhibition can be qualified as bacteriostatic, which means that further growth is stopped but the organisms put in presence of the inhibitor are still living, or bactericidal, which means that further growth is stopped and the organisms put in presence of the inhibitor are killed.

For determining what type of inhibition is caused by a particular drug, colony counting upon application of this compound can be performed. Typically, a liquid culture of the desired bacteria is incubated until it yields a bacterial density near $5 \times 10^5$ cfu/mL and split into equal volumes. To minimize the risk of errors, each time point is performed in triplicate. Three aliquots are diluted to an appropriate dilution factor and spread on agar plates. Those will be used as references for “day zero”, when bacteria have not been
exposed to any drug yet. The same concentration of drug is then added to all the other aliquots and they are incubated at 37 °C. The cultures are removed at different time points and diluted to the same dilution factor as the ones for “day zero” prior to plating. Each agar plate is incubated at 37 °C for 1 or 2 days (depending on the bacteria used) and colony count is performed. Colony forming units (cfu)/mL values are obtained by multiplying the number of colonies visible on a plate by the dilution factor and dividing by the original culture volume in mL. A killing curve can then be plotted, and if a decrease of more than 2 log(cfu/mL) is observed, the inhibitor can be qualified as bactericidal.

2.6 Measuring gene expression: reverse transcriptase PCR

Quantitative reverse transcriptase (RT)-PCR is a sensitive method for detecting a specific mRNA and analyzing transcript levels.\textsuperscript{51}

The procedure requires designing of primers that are complementary to the extremities of the gene for which the expression has to be quantified.\textsuperscript{52} The chosen strain of bacteria is grown and induced and total RNA is extracted from the cells and purified using the RNeasy Kit. Complementary DNA (cDNA) is then synthesized from input RNA using a primer as a starting point and a reverse transcriptase enzyme such as SuperScript® III. A regular PCR can then be performed using DNA polymerase and the designed primers mentioned above so that the gene of interest can be amplified (Figure 2-5).
Figure 2-5: Principle of quantitative RT-PCR

The amount of PCR product obtained after agarose gel electrophoresis depends on the level of gene expression in the induced bacteria. The intensity of each band can be quantified using the Quantity One® software. This technique enables the study of changes in gene transcription caused by the addition of an external factor, such as oxidative stress or a drug molecule. It is then recommended to use a gene that is not sensitive to this external factor as an internal standard.
2.7 Protein crystallization strategy

Determining the structure of a protein is instrumental in its characterization. The most common method for this purpose is X-ray diffraction of single protein crystals. Proteins typically form disorganized clusters in solution, and crystallization can only be achieved under some specific conditions (combination of precipitant concentration, buffer conditions, concentration of salts present…) that help the protein to form ordered crystals.

The most commonly used protocol for setting up crystallization trays is the hanging drop vapor diffusion method. Droplets (typically 1 µL) containing the desired concentration of pure protein are dispensed on glass cover slides and the same volume (1 µL) of crystallization solution containing buffer and precipitant is added to each droplet. Generally, each final droplet is a specific combination of protein with a particular solution and all droplets are under different conditions, as the main purpose of setting up a crystallization tray is to screen for the best conditions that allow protein crystallization. Each glass cover slide is inverted and sealed with grease over a reservoir that contains the same solution as the one added into the drop. Once the small system is closed, the drop and the reservoir gradually equilibrate by diffusion and as the concentration of precipitant in the drop increases, it is supposed to slowly promote protein crystallization. Each plate is incubated at 18 °C. A Nikon SMZ1500 microscope is used to look at the different drops and identify the best crystallization conditions.

If crystals are obtained, the solution conditions have to be optimized (detergents and other additives can be added) until single crystals of reasonable size appear. Those
can then be brought to the Advanced Photon Source (APS) at the Argonne National Laboratory for X-ray diffraction.
Chapter 3

Materials and Methods

3.1 Molecular cloning, expression and purification of Lsr2

The *lsr2* gene from Mtb H37Rv was inserted between the NcoI and HindIII restriction sites of a pET32 plasmid. This plasmid encodes a polyhistidine tag near the N-terminus of the fusion protein. The primers used in the polymerase chain reaction to amplify the *lsr2* gene are shown below (Table 3.1).

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>lsr2</em>-ncoI</td>
<td>5’-CAC CAT GGC GAA GAA AGT AAC CGT C-3’</td>
</tr>
<tr>
<td><em>lsr2</em>-hindIII</td>
<td>5’-CAA AAG CTT TCA GGT CGC CGC GTG-3’</td>
</tr>
</tbody>
</table>

Lsr2 was expressed in *E. coli* Rosetta™ using LB media. The culture was incubated at 37 °C until it reached an O.D.\textsubscript{600 nm} of 0.6 and then induced with 1 mM IPTG at 16 °C for 36 hours. First, immobilized metal affinity chromatography was used to purify Lsr2 (Wash buffer: 20 mM Tris pH 7.5, 1 M NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 25 mM imidazole and Elution buffer: 20 mM Tris pH 7.5, 1 M NaCl,
10 % glycerol, 5 mM β-mercaptoethanol, 250 mM imidazole) on a 5 mL HisTrap FF column (GE Healthcare). PreScission Protease was added to the pooled fractions containing Lsr2 during dialysis to remove excess imidazole, the cleaved thioredoxin-polyhistidine tag and protease were removed using immobilized metal affinity chromatography. Lsr2 was further purified using cation exchange chromatography (Wash buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 150 mM NaCl and Elution buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 1 M NaCl) on a 5 mL HiTrap SP FF column (GE Healthcare).

3.2 Identifying potential inhibitors of Lsr2 DNA-binding function

3.2.1 Determination of Lsr2/DNA $K_d$

Fluorescence polarization (FP) assays were performed employing a Biotek Synergy H4 Plate reader and data was analyzed with Gen5 and Prism 4.0. Zafirlukast was purchased from Sigma-Aldrich. The probe was a 5’-fluorescein labeled stem-loop DNA with the sequence 5’-CCTAATTATAACGAAGTTATAATTAGG-3’ from Integrated DNA Technologies at a concentration of 100 µM in deionized water. Prior to conducting FP experiments, the DNA solution was incubated at 95 °C for 15 minutes, cooled on ice, and stored at -20 °C until needed.

The dissociation constant $K_d$ for the Lsr2-DNA complex was determined employing an FP assay performed at 37 °C. Each well of a 384-well microplate contained fluorescently labeled DNA (10 nM) and concentrations of Lsr2 varying from 0 to 200 µM. Buffer solution (20 mM Tris pH 7.5, 50 mM NaCl) was added to reach a final
volume of 25 μL. An excitation wavelength of 485 nm and an emission wavelength of 528 nm were used to measure polarization values.

### 3.2.2 Screening of NIH Clinical Collection

All the tested compounds were from the National Institute of Health Clinical Collection (NCC) purchased from BioFocus. Each well of a black Corning 384-well microplate contained a total volume of 25 μL with the following components in a reaction buffer of 20 mM Tris pH 7.5 and 50 mM NaCl: Lsr2 (13.65 μM), fluorescently labeled DNA probe (10 nM), DMSO (4 % v/v). Controls were performed with those reagents before adding any drug molecule. Each compound tested was added to a concentration of 0.4 mM. Any drug showing a decrease in the polarization value greater than three times the standard deviation was considered a hit.

### 3.2.3 Tests on Zafirlukast derivatives and other potential inhibitors

The same FP assay format as previously described was employed. Each well contained 1 μL of fluorescently labeled DNA (final concentration of 10 nM), 20 μL of Lsr2 (final concentration of 2 μM) and the concentration of each drug was varied by adding a constant volume (1 μL) with increasing concentrations, from 0 to 300 μM (0 to 1.25 mM for the second test performed on Zd05). Additional buffer solution (20 mM Tris pH 7.5, 50 mM NaCl) was added as needed to reach a final volume of 25 μL.
3.3 Studies involving Zafirlukast

3.3.1 Determination of $K_i$

The inhibitory constant $K_i$ was determined using the same format as previously described. Each well contained 1 µL of fluorescently labeled DNA (200 nM), 20 µL of Lsr2 (200 µM) and 1 µL of Zafirlukast with final concentrations ranging from 0 to 500 µM. Buffer was added to adjust each well to a total volume of 25 µL. Data was analyzed using Prism and $K_i$ was calculated employing the Hill equation, where $L_b$ is bound DNA concentration, $L_o$ is the total DNA concentration, and $R_o$ is the total Lsr2 concentration (Equation 3.1).

$$K_i = \frac{[L_b]([C_{50}])K_d}{[L_o][R_o] + [L_b][-R_o] + [L_o] + [L_b] - K_d}$$

Equation 3.1: Hill equation

3.3.2 Kirby-Bauer disk-diffusion assays

*M. smegmatis* ATCC 14468 from a liquid culture was spread on an agar plate containing carbenicillin. Disks soaked with varying amounts of Zafirlukast dissolved in DMSO (49 µg, 98 µg, 196 µg, and 392 µg) were positioned on the surface of the agar layer. The plate was incubated at 37 °C for 48 hours and zones of inhibition were examined. An analogous experiment was performed with *M. tuberculosis* strain mc$^2$6230 (non-pathogenic). A sample from a Middlebrook 7H9 liquid culture containing pantothenic acid and incubated at 37 °C to mid-log was used to inoculate a Middlebrook 7H9 agar plate augmented with pantothenic acid. Wafers impregnated with either 10 µL
of DMSO or 10 µL of DMSO containing 500 µg of Zafirlukast were positioned on the plate, which was incubated at 37 °C until growth was sufficient for the zones of inhibition to be visible. Experiments with *E. coli* in LB medium were performed using a similar protocol but incubated for only 16 hours as these bacteria divide more rapidly.

### 3.3.3 Test of the bactericidal activity of Zafirlukast

*M. smegmatis* in LB medium containing ampicillin (50 µg/mL) was incubated at 37 °C until a bacterial density near 5 x 10^5 cfu/mL was reached. Zafirlukast was added to 2 mL aliquots of the bacterial suspension to a final concentration of 20 µM. The cultures were incubated at 37 °C and centrifuged at 4,000 rpm for 20 minutes after different time points. Each viscous pellet was resuspended in 200 µL of LB media containing 0.2 % Tween 80 v/v and diluted 2000-fold. The diluted sample (250 µL) was spread onto carbenicillin-containing agar plates. Each plate was incubated for 2 days at 37 °C a colony count was performed to determine the number of cfu.

### 3.3.4 Determination of the minimal inhibitory concentration (MIC)

A liquid culture of *M. smegmatis* in LB medium containing ampicillin (50 µg/mL) was grown to reach an O.D._{600 nm} of 0.5 and diluted to a final bacterial density of 5 x 10^5 colony forming units (cfu)/mL. A solution of Zafirlukast was added to 2 mL aliquots of the bacterial suspension to yield final concentrations ranging from 3 to 12 µM. The effect of each concentration of Zafirlukast on bacterial growth was observed in triplicate. After 48 hours of incubation at 37 °C, 0.2 mL from each culture was transferred into a 96-well plate and 10 µL of MTT (10 mg/mL in methanol) was added to each well. Following
incubation at 37 °C for 30 minutes, the MIC was defined as the lowest concentration that did not result in the formation of purple crystals (formazan).

**3.3.5 Semi-quantitative reverse transcriptase PCR**

*M. tuberculosis* strain mc²6230 was put in the presence of Zafirlukast at a final concentration of 12.5 μM and incubated for 16 hours. Total mycobacterial RNA (5 or 10 ng) purified with the RNeasy Kit was used for reverse transcription with SuperScript III. Taq DNA polymerase was used for the subsequent PCR step.

The transcript level of *sigA* was determined in *M. tuberculosis* as an internal standard. Control reactions containing no RNA or reverse transcriptase were also performed. DNA products resulting from the PCR step were run on an agarose gel and band intensity was quantified using the Quantity One software (Bio-Rad). Primers used for these experiments are shown in Table 3.2.
Table 3.2: Primers used for the quantitative RT-PCR experiments

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Name</th>
<th>Sequence</th>
<th>Reaction Step</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>sigA</em></td>
<td>sigA-1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>5’-gaggtcatcagagttcaggt-3’</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td></td>
<td>sigA-nested-F</td>
<td>5’-ccaggacactacgcaccag-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td></td>
<td>sigA-nested-R</td>
<td>5’-gagctcaggtcttgtaagc-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td><em>fbpC</em></td>
<td>fbpC-1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>5’-aggtgtagttcttgccggtg-3’</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td></td>
<td>fbpC-nested-F</td>
<td>5’-tctctgtttacggtcggg-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td></td>
<td>fbpC-nested-R</td>
<td>5’-gatccgcgaacccctgact-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td><em>otsA</em></td>
<td>otsA-1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>5’-aagtagctgccctcaccatt-3’</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td></td>
<td>otsA-nested-F</td>
<td>5’-gtcgatctggagcgtcttcc-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td></td>
<td>otsA-nested-R</td>
<td>5’-cgtggagaatccctcgt-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td><em>sodA</em></td>
<td>sodA-1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>5’-cgtctgtgggtctgtaaccc-3’</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td></td>
<td>sodA-nested-F</td>
<td>5’-gccaaggaagatcactcagc-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td></td>
<td>sodA-nested-R</td>
<td>5’-aaggaactttgtgcgaacac-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td><em>lsr2</em></td>
<td>lsr2-1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>5’-cggtcatcagagttcaggt-3’</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td></td>
<td>lsr2-nested-For</td>
<td>5’-tcactaagaatgccaacag-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
<tr>
<td></td>
<td>lsr2-nested-Rev</td>
<td>5’-acattgtcccgttaacag-3’</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; PCR</td>
</tr>
</tbody>
</table>
3.3.6 Intensity Fading MALDI mass spectrometry assay

Mass spectrometry experiments were performed on a Bruker Daltonics UltrafleXtreme MALDI. Ritonavir was used as an internal standard that does not affect complexation between Lsr2 and DNA. Zafirlukast and ritonavir were dissolved in DMSO and the two solutions were mixed at a ratio to produce ion peaks of similar intensity. The mixture containing both drugs was diluted 10-fold with a 20 mM Tris (pH 8.5) buffer and 2 µL were dispensed into various solutions of Lsr2 (in the same buffer) with concentration ranging from 0 to 60 µM. The different mixtures were incubated for 12 hours at 4 °C. Each sample was mixed using the dried droplet method with a matrix solution (1:2 v/v) of sinapinic acid (10 mg/mL) containing 30 % acetonitrile (v/v) in deionized water and spotted on a MALDI target plate.

3.4 Molecular cloning, expression and purification of fusion proteins

3.4.1 Lysozyme-Lsr2

The lysozyme gene from Enterophage RB69 was placed between two NcoI cut sites of a pET32 plasmid containing the lsr2 gene between the NcoI and the HindIII cut sites. This plasmid encodes a polyhistidine tag near the N-terminus of the fusion protein. The primers used in the polymerase chain reaction to amplify the lysozyme gene are shown below (Table 3.3).
Table 3.3: Primers used for amplifying *lysozyme* from RB69

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lyso-5'</td>
<td>5'-CAC CAT GGT TCG TAA TGA CGA AG-3'</td>
</tr>
<tr>
<td>lyso-3'</td>
<td>5'-AAC CAT GGC TAT ATA CGC TTT CCA AG-3'</td>
</tr>
</tbody>
</table>

Lysozyme-Lsr2 fusion protein was expressed in *E. coli* Rosetta™ using LB media. The culture was grown at 37 °C until reaching an O.D.₆₀₀nm between 0.6 and 0.8 followed by induction with 1 mM IPTG at 16 °C for 36 hours. Lysozyme-Lsr2 was purified using immobilized metal affinity chromatography (Wash buffer: 20 mM Tris pH 7.5, 1 M NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 25 mM imidazole and Elution buffer: 20 mM Tris pH 7.5, 1 M NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 250 mM imidazole) on a 5 mL HisTrap FF column (GE Healthcare). Following tag cleavage by application of PreScission Protease during dialysis to remove excess imidazole, the polyhistidine tag and protease were removed using immobilized metal affinity chromatography. Fractions containing Lysozyme-Lsr2 were pooled and further purified using cation exchange chromatography (Wash buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 150 mM NaCl and Elution buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 1 M NaCl) on a 5 mL HiTrap SP FF column (GE Healthcare).

### 3.4.2 Lyso-Lsr2C

The engineered plasmid produced for the previous fusion protein (Lysozyme-Lsr2) was used as a template for a PCR. The primers used are shown below (Table 3.4)
and consisted of codons corresponding to the Lysozyme C-terminus and nucleotides encoding the first few residues of Lsr2C (starting at residue 65) so that the region encoding Lsr2N could be eliminated.

Table 3.4: Primers used for engineering a plasmid containing lyso-lsr2C

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lyso-lsr2-cterm-cod</td>
<td>5’-CAG GAA CTT GGA AAG CGT ATA TAC GTT CCG GAT CCG GCG GT-3’</td>
</tr>
<tr>
<td>lyso-lsr2-cterm-noncod</td>
<td>5’-ACG GCC GGA TCC GGA ACG TAT ATA CGC TTT CCA AGT TCC TG-3’</td>
</tr>
</tbody>
</table>

3.4.3 Lsr2N-T4-Δ54-73-Lsr2C

The T4 lysozyme gene was placed between the Ndel and the BamHI cut sites of a pET28 plasmid. The primers used in the polymerase chain reaction to amplify the T4 lysozyme gene are shown below (Table 3.5).

Table 3.5: Primers used for amplifying T4 lysozyme

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>t4-ndel</td>
<td>5’-CAC CCA TAT GAA CAT CTT CGA AAT GCT GCG-3’</td>
</tr>
<tr>
<td>t4-bamHI</td>
<td>5’-GGA TCC TCA GTA AGC GTC CCA GGT AC-3’</td>
</tr>
</tbody>
</table>

The following cloning step was to amplify the T4 lysozyme sequence and engineer the PCR product so that its first extremity would match the end of the “lsr2N” gene and the second extremity would correspond to the first bases of “lsr2C”. The T4 lysozyme
was inserted between codons of “lsr2N” encoding residues 1-53 and codons of “lsr2C” encoding residues 74-112. The primers used for this purpose are shown in Table 3.6.

Table 3.6: Primers used for amplifying T4 lysozyme and insert overhangs matching “lsr2N” and “lsr2C”

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-half-lsr2-T4</td>
<td>5’-GAC CTG AAG CAA TGG GTG GCG ATG AAC ATC TTC GAA ATG CTG CGT ATC GAC GAA-3’</td>
</tr>
<tr>
<td>C-half-/Δ54-73-lsr2-T4</td>
<td>5’-CGC GCT CTG CTC GCG GTC GAT CGC GTA AGC GTC CCA GGT ACC GGT A-3’</td>
</tr>
</tbody>
</table>

The fused gene was then synthesized by using both a pET32 vector containing lsr2 and the product from the previous PCR as templates in a new PCR with the primers shown in Table 3.1. The amplified gene was digested and placed between NcoI and HindIII in a pET32 vector. This plasmid encodes a thioredoxin containing a polyhistidine tag near the N-terminus of the fusion protein.

Lsr2N-T4-Δ54-73-Lsr2C fusion protein was expressed in *E. coli* Rosetta™ using LB media. The culture was grown at 37 °C until reaching an O.D.₆₀₀nm between 0.6 and 0.8 followed by induction with 1 mM IPTG at 16 °C for 36 hours. Lsr2N-T4-Δ54-73-Lsr2C was purified using immobilized metal affinity chromatography (Wash buffer: 20 mM Tris pH 7.5, 1 M NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 25 mM imidazole and Elution buffer: 20 mM Tris pH 7.5, 1 M NaCl, 10 % glycerol, 5 mM β-mercaptoethanol, 250 mM imidazole) on a 5 mL HisTrap FF column (GE Healthcare). Following tag cleavage by application of PreScission Protease during dialysis to remove excess imidazole, the thioredoxin-polyhistididine tag and protease were removed using
immobilized metal affinity chromatography. Fractions containing Lsr2N-T4-Δ54-73-Lsr2C were pooled and further purified using cation exchange chromatography (Wash buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 150 mM NaCl and Elution buffer: 20 mM MES pH 6.0, 10 % glycerol, 0.5 mM TCEP, 1 mM EDTA, 1 M NaCl) on a 5 mL HiTrap SP FF column (GE Healthcare).

3.5 Crystallization tray setup

All the attempts at growing protein crystals were made using hanging drop crystallization trays. Droplets (1 µL) containing the desired concentration of Lsr2 or any of the fusion proteins or the Lsr2/DNA complex were dispensed on glass cover slides and the same volume (1 µL) of crystallization solution containing buffer and precipitant was added to each droplet. For initial screening, the Index HT from Hampton Research was used. When any hit was observed, new solutions with varying amount of one or several of the components were made and new crystal trays were set up: this process is referred to as 1D-screening (one varying component) or 2D-screening (two varying components).

3.5.1 Fusion proteins

Before setting up crystallization trays, Lysozyme-Lsr2 was dialyzed against 20 mM Tris (pH 7.5), 150 mM NaCl and concentrated to 14 mg/mL. The crystals obtained were in condition #84 (0.2 M MgCl₂, 0.1 M HEPES at pH 7.5, 25 % w/v PEG 3,350) from the Index HT.

Purified Lsr2N-T4-Δ54-73-Lsr2C was dialyzed against 20 mM Tris (pH 7.5), 150 mM NaCl and concentrated to 6 mg/mL.
3.5.2 Lsr2/DNA complexes

For all the crystal trays set up with an Lsr2/DNA complex, Lsr2 and the oligonucleotide were first premixed at a ratio of 1.1:1. All the DNA molecules received from Integrated DNA Technologies were dissolved in 20 mM Tris (pH 7.5) at a final concentration of 0.8 mM. Annealing was performed by incubating the solutions at 95 °C for 20 minutes and slowly cooling them down until they were at room temperature. Each DNA solution was stored at -20 °C until needed. The oligonucleotide named 20-mer in the results section has the sequence 5’-AAT ATA TGG CGC CAT ATA TT-3’, while the 21-mer is 5’-TAT ATA TGG CGC CAT ATA TAT-3’. The two hits observed for the Lsr2/21-mer complex were in conditions #69 (0.2 M ammonium sulfate, 0.1 M Tris pH 8.5, 25 % w/v PEG 3,350) and #81 (0.2 M ammonium acetate, 0.1 M Tris pH 8.5, 25 % w/v PEG 3,350) from the Index HT. The 12-mer oligonucleotide has the sequence 5’-GCT ATA TAT AGC-3’ and the 13-mer is 5’-GCT ATA TAT AGC T-3’.
Chapter 4

Zafirlukast is an Inhibitor of the Lsr2/DNA Complexation and Mycobacterial Growth

4.1 Expression and purification of Lsr2

Lsr2 was expressed in *E. coli* Rosetta™ using LB media. The bacteria were harvested and lysed into a buffer with a high salt concentration (1 M NaCl) to minimize the amount of DNA binding to the protein. The filtered supernatant obtained after cell lysis was loaded onto a nickel column. Because Lsr2 has a polyhistidine tag on its N-terminus, it binds to the resin and can then be eluted by application of an imidazole gradient. An intense peak absorbing UV light at 280 nm was visible on the chromatogram (Figure 4-1). The fractions corresponding to this peak were pooled and dialyzed into the binding buffer for nickel affinity chromatography in presence of PreScission Protease in order to cleave the polyhistidine tag. The dialyzed protein solution was then loaded onto a nickel affinity column run manually so that the tag and protease could be discarded. The sample collected during the wash step contained the newly untagged Lsr2, while the elution fraction had the tag, PreScission protease (that has a polyhistidine tag that binds to the resin) as well as any remaining tagged Lsr2.
Figure 4-1: Chromatogram of nickel affinity chromatography. The x axis corresponds to the volume of buffer run through the column in mL, the left y axis is the absorbance in milli absorbance units (mAU) and the right y axis is the percentage of imidazole-rich elution buffer run through the column. The absorbance at 280 nm is shown in blue, while the gradient of imidazole-rich elution buffer is shown in red.

Fractions containing untagged Lsr2 were pooled and dialyzed against binding buffer for cation exchange chromatography. As the pI of Lsr2 is 10.7 and the pH of this buffer is 6.0, the protein is positively charged in the solution, which allows it to bind the negatively charged resin of the column. An NaCl gradient was applied and gradually displaced Lsr2 from the column. An intense peak was again observed on the chromatogram, corresponding to the eluted protein (Figure 4-2).
Figure 4-2: Chromatogram of cation exchange chromatography. The x axis corresponds to the volume of buffer run through the column in mL, the left y axis is the absorbance in milli absorbance units (mAU) and the right y axis is the percentage of NaCl-rich elution buffer run through the column. The absorbance at 280 nm is shown in blue, while the gradient of imidazole-rich elution buffer is shown in red.

Fractions containing Lsr2 were pooled and protein was concentrated down to the desired concentration depending on the experiments it needed to be used for. A sample was run on SDS-PAGE to attest to its purity (Figure 4-3).
4.2 Lsr2-DNA complexation

To get a better understanding of Lsr2-DNA complexation and identify compounds that alter Lsr2 DNA-binding function, a fluorescence polarization assay was performed using a fluorescein-labeled AT-rich DNA probe previously shown to bind Lsr2. The concentration of DNA that was judged appropriate for binding studies was determined by a quick FP triplicate test for several probe concentrations. The lowest concentrations of DNA that gave consistent polarization values were ranging from 5 to 10 nM. A concentration of 10 nM labeled hairpin DNA was then used for all experiments. Binding experiments where the concentration of probe was held constant while the concentration of Lsr2 increased were performed. The polarization values showed a consistent increase that reflected complexation of Lsr2 with the DNA probe. After data analysis and fitting,
a $K_d$ of 0.60 μM ± 0.06 ($R^2=0.97$) was obtained for the Lsr2-DNA complex, which corresponds well with previously determined values (Figure 4-4).  

A high-throughput FP assay was then developed in order to perform library screening for identifying compounds that inhibit Lsr2/DNA complexation. Lsr2 was preincubated with the DNA probe. All the wells containing the Lsr2/DNA complex were provided with different drug molecules from the NCC (one compound for each well). The polarization value was measured for every mixture of Lsr2, DNA, and drug (Figure 4-5). Seven compounds were identified as hits based on an FP signal decrease (upon disruption of the Lsr2-DNA complex) greater than three times the standard deviation from the mean polarization measurement.
Figure 4-5: Screening of the NCC drugs. Each data point corresponds to one of the compounds of the collection. Any compound showing a decrease in the FP signal that was greater than three times the standard deviation was considered a hit.

Of the potential leads observed, six are known DNA intercalators or groove binders so they were not studied further, since the goal of this study was to identify a compound that specifically inhibits Lsr2 function by binding the protein and not the DNA probe. The only remaining compound was Zafirlukast, a drug orally administered as tablets and commonly used in patients suffering from asthma.

4.3 Zafirlukast inhibits Lsr2-DNA complexation

Zafirlukast is commercially available under the name Accolate (AstraZeneca) and used as an oral prophylactic treatment for asthma (Figure 4-6). Leukotrienes are
substances responsible for some of the symptoms of asthma such as bronchoconstriction or enhanced secretion of mucus\textsuperscript{54}. Zafirlukast acts as a leukotriene receptor antagonist that blocks the actions of leukotrienes in the pulmonary system. Clinical trials have shown that this drug is safe and generally well tolerated.\textsuperscript{55}

![Figure 4-6: Structure of Zafirlukast at physiological pH](image)

The same FP assay format as described above was used to determine the inhibitory constant ($K_i$) for Zafirlukast (Figure 4-7). DNA and Lsr2 concentrations were kept constant while the concentration of Zafirlukast was varied in order to perform a dose-dependence study. A gradual decrease of FP signal as more Zafirlukast was added could be observed, confirming the disruption of the preincubated complex between Lsr2 and the probe. The dose-dependent inhibition of Lsr2/DNA complexation by Zafirlukast exhibited an $IC_{50}$ of 64.55 $\mu$M ($R^2 = 0.989$) and a calculated $K_i$ value of 13.9 $\mu$M.
Our assumption was that the inhibition of complexation was caused by direct interactions between Zafirlukast and Lsr2 and that the drug did not interact with DNA. Other FP experiments were run under the exact same conditions except for the concentration of probe that was raised to 25 nM. The inhibition curve obtained was almost superimposable with the one shown above (Figure 4-8). If the drug was interacting with labeled DNA, higher concentrations of Zafirlukast would have been necessary to reach the same level of inhibition as in the previous experiment. This information supported our hypothesis, which was later confirmed by intensity fading experiments.
4.4 FP experiments performed on Zafirlukast derivatives

Based on the FP data obtained for Zafirlukast, our collaborator Dr. Jimmy Franco built a model using the published NMR structure of Lsr2C\textsuperscript{29} (DNA-binding domain) to look at the possible interactions between the drug and the protein. AutoDOCK software was employed and the prediction suggested that Zafirlukast binds at the same position as bacterial DNA on Lsr2, which supports our findings, as potential interactions of Zafirlukast with DNA-binding residues Ser80, Arg84, Arg88 and Arg97 are observed (Figure 4-9).\textsuperscript{29}
Dr. Franco also performed docking studies with Lsr2C by the intermediate of the same software then designed and synthesized Zafirlukast “derivatives” that showed potential interactions with Lsr2. Those are based on the same scaffold as Zafirlukast but display different substituting groups (Figure 4-10, Table 4.1).

Figure 4-9: Model of predicted interactions between Zafirlukast and Lsr2. PDB file used for Lsr2C: 2KNG

Figure 4-10: Core scaffold of Zafirlukast and its synthesized derivatives
Table 4.1: Zafirlukast and Zafirlukast derivatives functional groups

<table>
<thead>
<tr>
<th>Name</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zafirlukast</td>
<td>[Structure]</td>
<td>-</td>
<td>![Structure]</td>
</tr>
<tr>
<td>Zd01</td>
<td>![Structure]</td>
<td>$\text{CH}_3$</td>
<td>-</td>
</tr>
<tr>
<td>Zd02</td>
<td>![Structure]</td>
<td>$\text{CH}_3$</td>
<td>$\text{OH}$</td>
</tr>
<tr>
<td>Zd03</td>
<td>![Structure]</td>
<td>$\text{CH}_3$</td>
<td>![Structure]</td>
</tr>
<tr>
<td>Zd04</td>
<td>![Structure]</td>
<td>$\text{CH}_3$</td>
<td>![Structure]</td>
</tr>
<tr>
<td>Zd05</td>
<td>![Structure]</td>
<td>$\text{H}$</td>
<td>![Structure]</td>
</tr>
<tr>
<td>Zd06</td>
<td>![Structure]</td>
<td>$\text{H}$</td>
<td>![Structure]</td>
</tr>
</tbody>
</table>

All those Zafirlukast derivatives were tested in FP experiments using the same protocol as for inhibition studies with Zafirlukast. Raw data from those studies showed that none of the new compounds exhibited any inhibition activity comparable to the one obtained when using Zafirlukast (Figure 4-11). The only compound that seemed to actually have some effect on Lsr2-DNA complexation was Zd05.
Compound Zd05 was used for further FP inhibition experiments. The first tests shown above seemed to indicate some weak inhibition compared with Zafirlukast, but the concentration of Zd05 was only ranging from 0 to 300 µM. The FP signal for the Lsr2-DNA complex in presence of higher concentrations of Zd05 (up to 1.25 mM) was measured and compared with the decrease obtained when Zafirlukast is used (Figure 4-12). Even when the concentration of Zd05 was 1.25 mM, the inhibition was not as good as when Zafirlukast is employed at much lower concentrations. As none of those designed compounds exhibited any inhibition activity that looked more efficient than Zafirlukast, they were not studied further.
Docking studies also led toward six compounds that seemed to potentially have good interactions with Lsr2 in the model. These molecules all had scaffolds very different from Zafirlukast and its derivatives. They were all tested in FP inhibition studies with the same protocol as before, but none of them seemed to exhibit activity against the complexation of Lsr2 with labeled DNA.

**4.5 Zafirlukast directly binds Lsr2**

An intensity fading MALDI mass spectrometry assay was used to provide evidence that Zafirlukast interacts specifically with Lsr2. Ritonavir was used as an internal standard as it does not have any effect on Lsr2 complexation with DNA. Increasing concentrations of Lsr2 were applied to solutions containing fixed quantities of
Ritonavir and Zafirlukast. The intensities of the drugs ion peaks were monitored employing MALDI experiments (Figure 4-13). The concentration of Lsr2 in each sample increases from top to bottom promoting a decrease in the relative intensity of the Zafirlukast + Na$^+$ ion peak (m/z = 598.5) relative to the Ritonavir + Na$^+$ ion peak (m/z = 743.9).

![Intensity fading experiment showing direct binding between Lsr2 and Zafirlukast](image)

Figure 4-13: Intensity fading experiment showing direct binding between Lsr2 and Zafirlukast

Relative ion peaks intensities were used to calculate the ratio of free Zafirlukast (Z) to Ritonavir (R) in each experiment (Figure 4-14). The decrease observed can be
directly attributed to a direct binding between Zafirlukast and Lsr2, which reduces the amount of free Zafirlukast as the concentration of protein is increased.

Figure 4-14: Plot of the ion ratio of Zafirlukast (Z) over Ritonavir (R) versus the concentration of Lsr2

4.6 Zafirlukast inhibits mycobacterial growth

To attest to the ability of Zafirlukast to enter mycobacterial cells, inhibit Lsr2 function and affect growth, a Kirby-Bauer assay was first performed using *Mycobacterium smegmatis* (Figure 4-15). Different amounts of Zafirlukast ranging from 49 to 392 μg were administered since it was unknown to what extend the drug would inhibit growth. Following incubation of the plate, zones of inhibition whose size increased in a dose-dependent manner were clearly visible. The experiment confirmed that the drug is capable of penetrating the cell wall and inhibiting mycobacterial growth in *M. smegmatis*. 
Figure 4-15: Kirby-Bauer growth inhibition studies with *M. smegmatis* in the presence of Zafirlukast. The amount of Zafirlukast used on each disk (dissolved in DMSO) is indicated.

After the activity of Zafirlukast against *M. smegmatis* was demonstrated, the drug had to be tested in Mtb. An attenuated strain (mc\(^2\)6230) was grown by our collaborators at Wayne State University and put in the presence of two disks: the first one impregnated with only DMSO and the second one with Zafirlukast dissolved in DMSO (Figure 4-16). Since it is typical for anti-mycobacterial compounds to have better efficacy against *M. smegmatis* than Mtb, it was expected that Mtb would not respond as strongly to Zafirlukast. Taking this information into account, a Zafirlukast dose of 500 \(\mu\)g was used for the Mtb mc\(^2\)6230 Kirby-Bauer growth assay. Even if higher concentrations of Zafirlukast are needed in Mtb to reach the same level of growth inhibition as with *M. smegmatis*, the fact that the drug was also able to enter Mtb cells was very encouraging as
it is often an obstacle in drug development for TB treatment. Those data combined together showed that Zafirlukast can inhibit growth of mycobacteria.

![Image of Kirby-Bauer assay](image)

Figure 4-16: Kirby-Bauer growth inhibition study with Mtb in the presence of Zafirlukast

A control Kirby-Bauer assay employing *E. coli* was also performed. Indeed, *E. coli* does not encode Lsr2, thus it should not be affected by the presence of Zafirlukast if its inhibitory action is specific to Lsr2. When exposing a laboratory strain of *E. coli* to the same amount of compound as in experiments with mycobacteria, no growth inhibition was exhibited (Figure 4-17). Ampicillin (Amp) and kanamycin (Kan) are well-known antibiotics that inhibit growth of *E. coli* and were used as comparative standards in the same Kirby-Bauer experiment. Clear zones of inhibition could be observed around disks containing those compounds.
These data clearly show that *E. coli*, which instead of Lsr2 uses the H-NS protein to modulate genome structure, is not affected by Zafirlukast. In contrast, the effect on growth of two different species of mycobacteria suggests that the inhibitory activity of Zafirlukast may be specific to organisms that encode Lsr2.

Kirby-Bauer assays involving drug combinations of commonly used antibiotics with Zafirlukast were also performed to look for any synergistic effects that would cause enhanced growth inhibition of mycobacteria. *M. smegmatis* cells were employed for those tests. Combinations of Zafirlukast with vancomycin, tetracycline, ampicillin, ethambutol, isoniazid and kanamycin were initially tested. In those assays, one plate had disks containing only one drug per disk, while the second plate had disks impregnated with the different combinations, so that the zones of inhibition caused by the antibiotics could be easily compared with the ones caused by Zafirlukast/antibiotic combinations for
each experiment. Zones of inhibition observed for combinations were about the size of the zone of inhibition caused by the antibiotic added to the zone of inhibition caused by Zafirlukast, so they did not seem to indicate any particular synergistic effect. However, an interesting result was noticed for kanamycin, as regrowth of mycobacteria was visible within the zone of inhibition caused by kanamycin only, whereas the zone around the disk that contained both the antibiotic and Zafirlukast remained clear (Figure 4-18).

Figure 4-18: Compared effects of kanamycin and kanamycin + Zafirlukast on M. smegmatis after 3 days of incubation at 37 °C

Kanamycin belongs to the aminoglycoside family of antibiotics. Other aminoglycosides (G418, amikacin, neomycin, gentamycin, paromomycin) were also tested in combination with Zafirlukast. Unfortunately, the regrowth result could not be reproduced and there was no synergistic effect observed for the use of those antibiotics with Zafirlukast (Figure 4-19).
Figure 4-19: Kirby-Bauer assays using combinations of aminoglycosides with Zafirlukast. Those different plates were all incubated at 37 °C for 3 days.

As observed zones of inhibition obtained with the different drug combinations did not seem to indicate any interesting lead, those experiments were not further pursued.

4.7 Determination of Zafirlukast MIC\(^{53}\)

Different attempts were made to determine the minimum inhibitory concentration (MIC) to further characterize the anti-mycobacterial activity of Zafirlukast. Concentrations varying from 0.5 to 100 μM were added to cultures of \(M. \text{smegmatis}\). Bacterial growth was assessed by measuring the O.D.\(_{600 \text{nm}}\). At Zafirlukast concentrations of 10 μM and above, the culture media clarified after three days and viscous debris accumulated at the bottom of the culture tube (Figure 4-20). No growth was exhibited on
agar plates inoculated with the culture supernatant, suggesting a lack of viable planktonic
*M. smegmatis* bacteria.

Figure 4-20: *M. smegmatis* phenotype after 3 days of incubation with and without Zafirlukast

MTT viability assays on *M. smegmatis* grown in the presence of Zafirlukast showed an MIC of 6 μM. As previously explained, the appearance of purple crystals in the plate wells indicates the presence of living cells that still contain functioning reductase enzymes, allowing the conversion of yellow MTT to insoluble purple formazan. The MIC was determined as the lowest concentration of Zafirlukast that prevented the formation of those crystals, thus inhibited growth of mycobacteria (Figure 4-21).
Figure 4-21: MTT viability assay for *M. smegmatis* in presence of Zafirlukast. Concentrations of Zafirlukast are indicated above each lane of triplicate experiments.

4.8 Zafirlukast exhibits bactericidal activity

To characterize the activity of Zafirlukast as bactericidal or bacteriostatic, a plating assay was performed using material from an *M. smegmatis* culture incubated with 20 μM Zafirlukast. The number of colonies on the plates was recorded at different time points after the addition of Zafirlukast. This number decreased after only 24 hours of incubation with clarification of the liquid media after 3 days of incubation with the drug. As previously observed, cellular detritus accumulated at the bottom of the culture tube. Since this is a common phenomenon for mycobacteria, it was necessary to determine if this material contained any living *M. smegmatis*.

To address the question of the viability of bacteria present in the debris, the pelleted detritus were resuspended in new culture media containing 0.2 % v/v Tween 80 to minimize bacterial agglutination. The resuspended sample from each time point was diluted and plated on agar plates containing carbenicillin (Figure 4-22).
Figure 4-22: *M. smegmatis* plated on agar plates after different times of incubation in the presence of Zafirlukast. A plate of *M. smegmatis* before any drug is added on day 0 is shown as a reference.

The experiment was run in triplicate, so each time point was analyzed based on three different cultures incubated at the same time under the same conditions. Following addition of Zafirlukast, a time course over a seven-day period showed a significant decrease in the number of colony forming units (Figure 4-23).
After seven and ten days, zero colony forming units were observed from each of
the three different cultures. Although the insoluble material contains viable *M. smegmatis*
in the days immediately following application of the drug, these numbers are lower than
expected if Zafirlukast were simply bacteriostatic. Ultimately, the lack of viable bacteria
after seven days suggests that the activity of Zafirlukast is bactericidal for *M. smegmatis*.

### 4.9 Zafirlukast dysregulates gene transcription\textsuperscript{53}

To gain insight into the mechanism of growth inhibition by Zafirlukast, semi-
quantitative reverse transcriptase PCR (RT-PCR) was used to examine the effects on
transcription of five Mtb genes previously shown to be bound by Lsr2\textsuperscript{29} upon application
of the drug (Figure 4-24). The experiments were performed by our collaborator at
Myongji University in the Republic of Korea. The two first genes, *otsA* and *fbpC*, were chosen because they play important roles in the biosynthetic pathway of trehalose and the production of the mycobacterial outer membrane, respectively.\(^{57}\) The third gene, *sodA*, encodes the superoxide dismutase which is involved in virulence.\(^{58}\) It has also been shown that Lsr2 binds its own gene, *lsr2*, so the mRNA levels of that gene were also quantified.\(^{29}\) The last gene, *sigA*, was examined as a standard since it is a housekeeping gene that is constitutively expressed, which means its transcription level is not significantly affected by external stress. Two cultures of Mtb were exposed to Zafirlukast and RNA was extracted from each of them. Two different amounts of RNA, 5 and 10 ng, were used for reverse transcription followed by amplification of the targeted genes. The intensity of the resulting bands was quantified.
Figure 4-24: Electrophoretic gels and quantitation of the reverse transcriptase PCR products. Each major column indicates an experiment either with or without application of Zafirlukast. Specific genes are labeled and shown are results using either 5 or 10 ng of RNA. The intensities relative to the sigA band intensities are shown in the bottom row.

These results concomitantly show that application of Zafirlukast promotes the increase of gene expression in the four mycobacterial genes known to be bound by Lsr2. The largest increase observed was in the mRNA levels of *lsr2*, which exhibits an 8.8 fold increase in *lsr2* expression. Transcription of the *otsA* and *fbpC* genes displayed 2.3 and 1.3 fold increases, respectively, while *sodA* exhibited a 4-fold increase. The *sigA* mRNA levels decreased slightly, which is consistent with previous data. All of these changes are concomitant with the hypothesis that Zafirlukast inhibits Lsr2 activity which causes dysregulation in gene expression.
Chapter 5

Attempts at Lsr2 Crystallization

The structure of full-length Lsr2 has not been published yet. As previously mentioned, so far only an NMR structure of Lsr2C\textsuperscript{29} and an X-ray structure of Lsr2N\textsuperscript{41} are available. The presence of a highly flexible region (residues 54-73) between the oligomerization domain and the DNA-binding domain of Lsr2 is probably to blame for the lack of order necessary to crystallize the full-length protein. Although the available NMR structural study offered a lot of information about the complexation of Lsr2 with DNA, there are still unresolved questions concerning this binding. Indeed, a protein sequence alignment shows conservation of residues 56-65, and six of them are positively charged arginine residues, which could play a role in binding the negatively charged DNA backbone. This conserved region has not been solved so far since the X-ray structure of Lsr2N covered residues 4-59 and the NMR structure was solved between residues 66-112.\textsuperscript{29, 41} A bigger picture of the Lsr2-DNA complex showing all their different interactions would be helpful to get a better understanding of the binding process. Crystallization of the full protein on its own or bound to DNA pieces similar to the ones used in FP assays (no fluorescent label) or bound to Zafirlukast had been previously attempted but remained unsuccessful. This chapter focuses on other attempts
made to obtain an X-ray structure of Lsr2, particularly in the context of an Lsr2/DNA complex.

5.1 Synthesis of fusion proteins

The first idea to promote crystallization of Lsr2 was to fuse it with a protein that crystallizes very easily, an appealing technique that has been successfully employed to solve a lot of protein structures. Indeed, in that case crystal packing can potentially be mediated by the interactions between residues of the added protein sequence.

5.1.1 Expression and purification of Lysozyme-Lsr2

The gene encoding Lysozyme from Enterobacteria phage RB69 was initially used to design a fusion protein that had the lysozyme fused at the N-terminus of Lsr2. The idea for cloning was to insert the lysozyme gene into the pET vector that already contained lsr2. As lsr2 was inserted between NcoI and HindIII cut sites in a pET32 plasmid, the vector was cut using the NcoI restriction enzyme. The gene encoding for the lysozyme was PCR amplified and engineered with NcoI cut sites at each extremity. It was then digested with NcoI and ligated into cut pET32 containing lsr2 (Figure 5-1). Typical subsequent cloning steps were then performed until purified plasmids were obtained and sent out for sequencing, which confirmed that the desired fused gene was obtained.
Expression cells (Rosetta™) were transformed with the plasmid, grown in larger cultures and protein expression was induced. After cell harvesting and lysis, the fusion protein named Lysozyme-Lsr2 was purified using IMAC (nickel) and cation exchange chromatography (Figure 5-2). The polyhistidine tag attached to the fusion protein was cleaved by PreScission protease treatment.
Figure 5-2: SDS-PAGE result after purification of the fusion protein Lysozyme-Lsr2

Purified protein was dialyzed and concentrated, and an Index Screen was set up. Very small crystals could be observed in one of the crystallization solutions. One dimensional and 2D screenings were performed for optimization of the crystallization conditions and bigger crystals could be obtained. Those were brought to the APS in order to collect data for the Lysozyme-Lsr2 fusion protein but unfortunately, diffractions patterns revealed that the crystals were in fact crystals of MgCl₂, present in the crystallization solution.

5.1.2 Cloning of Lyso-Lsr2C

The second designed fusion protein consisted of Lysozyme RB69 fused with only the DNA-binding domain of Lsr2 (Lsr2C). For this cloning, the plasmid engineered to express Lysozyme-Lsr2 was used as a starting point. The purpose of this cloning was to eliminate lsr2 codons encoding Lsr2N and obtain a fused gene containing the lysozyme gene and codons encoding Lsr2C (Figure 5-3).
Primers designed for this purpose had codons corresponding to the Lysozyme C-terminus and nucleotides encoding the first few residues of Lsr2C (starting at residue 65). A test digestion was performed after the PCR and a band appeared between 500 bp and 1 kbp. As the lysozyme gene is 473 bp and the “lsr2C” gene is 144 bp, the fused gene lyso-lsr2C should be approximately 600 bp long. The new plasmid obtained after PCR was then sequenced. Cloning was successful. Expression cells (Rosetta™) were transformed with the plasmid, grown in larger cultures and protein expression was induced. Unfortunately, protein expression was never achieved: no band could ever be observed on SDS-PAGE gels and no peak absorbing at 280 nm was visible on any Nickel affinity chromatogram.

Figure 5-3: Idea for designing the fused gene lyso-lsr2C based on the plasmid containing the lysozyme-lsr2 gene
5.1.3 Expression and purification of Lsr2N-T4-Δ54-73-Lsr2C

The gene encoding T4 Lysozyme was used to design a fusion protein that had the lysozyme fused between Lsr2N and Lsr2C. As previously mentioned, Lsr2 contains a high flexibility region located between the two domains (residues 54-73). The idea for designing a fusion protein that incorporates the T4 Lysozyme was to insert the lysozyme between the two domains Lsr2N and Lsr2C, but with a deletion of the high flexibility region. The protein was named Lsr2N-T4-Δ54-73-Lsr2C (Figure 5-4).

Figure 5-4: Schematic representation of Lsr2N-T4-Δ54-73-Lsr2C

For cloning, the T4 lysozyme gene was first amplified and engineered with ends matching the end of the “lsr2N” sequence and the beginning of the “lsr2C” sequence (starting at the codon encoding residue 54 or 74 depending on the desired version of the fused gene). The whole fused genes were then amplified and inserted between NcoI and HindIII cut sites in a pET32 plasmid. Sequencing confirmed that cloning was successful for the gene encoding Lsr2N-T4-Δ54-73-Lsr2C (Figure 5-5).
Expression cells (Rosetta™) were transformed with the plasmid containing the gene encoding Lsr2N-T4-Δ54-73-Lsr2C, grown in larger cultures and protein expression was induced. After cell harvesting and lysis, the fusion protein was purified using nickel affinity chromatography (Figure 5-6).
The fractions containing Lsr2N-T4-Δ54-73-Lsr2C were pooled and further purified using cation exchange chromatography. Thioredoxin-polyhistidine tag attached to the fusion protein was cleaved by PreScission protease treatment. Purified protein was dialyzed and concentrated, and an Index Screen was set up. No crystals have been observed so far.

5.2 DNA design to crystallize the Lsr2-DNA complex

The other avenue explored for attempting to obtain a structure of Lsr2 was to crystallize the Lsr2-DNA complex. The presence of DNA molecules could allow the dimer to be locked into place and the high flexibility of the middle region of the protein would then not be such an issue.
The first idea was to design a DNA molecule that could be bound at two sites by an Lsr2 dimer (Figure 5-7). A palindromic 20-mer was designed for this purpose. Two AT-rich regions are separated by a GC-rich region so that two DNA-binding domains are allowed to bind the same DNA molecule, potentially creating a rigid form of the dimer that could promote crystallization.

![Diagram of the Lsr2/20-mer DNA complexation idea for promoting crystallization](image)

Figure 5-7: Scheme of the Lsr2/20-mer DNA complexation idea for promoting crystallization

The 20-mer was annealed in order to obtain a double stranded DNA molecule and preincubated with purified Lsr2. A 96-well hanging drop Index screen was then set up. Some very small crystals were observed and one dimensional screenings were set up (varying amounts of one of the components of the buffer condition that promoted crystallization) but no larger crystals or diffraction-quality crystals could be obtained.

A new oligonucleotide containing 21 bases was designed. The sequence was similar to the one of the 20-mer but this time, a thymine overhang was added at the 3’ end. The idea was that it could potentially base pair with a neighboring thymine overhang.
from another 21-mer molecule and then form chains of complexes that could form a higher organization structure, leading to crystal formation (Figure 5-8).

The double stranded 21-mer was preincubated with purified Lsr2. A 96-well hanging drop Index screen was then set up. This time better-quality crystals were observed with two conditions and one dimensional screenings were set up. Several new conditions produced single crystals and those were brought to the APS at the Argonne National Laboratory and shot with X-rays. Unfortunately, the crystals obtained turned out to be salts.

After Summers et al. published an X-ray structure of the oligomerization domain of Lsr2 in 2012, new oligonucleotides were designed as it allowed for an idea of the distance between the two DNA-binding domains of a dimer. Testing several orientations of B-form DNA with Lsr2N in PyMOL, it was found that a fragment between 8 and 12 bp should allow the formation of a rigid Lsr2/DNA complex (Figure 5-9).
It was also decided that instead of separating the two AT-rich regions with a GC-rich region, the oligonucleotide would be mostly made of A and T bases, capped with GC-rich regions at each extremity. Two versions of oligonucleotides with similar sequences were designed: a palindromic 12-mer with blunt ends and a 13-mer that had the same sequence with an additional thymine overhang, for the reason previously mentioned. The double stranded DNA molecules were preincubated with purified Lsr2. Each complex was screened in a 96-well hanging drop with solutions from the Index screen. No crystals could be observed so far.
Chapter 6

Conclusions and Future Work

Lsr2 is key to regulating the transcription of a vast portion of the *Mycobacterium tuberculosis* genome. This essential function, along with the fact that it is specifically found in mycobacteria, makes it a very attractive drug target. The primary purpose of this study was to identify small molecules that inhibit the DNA-binding activity of Lsr2. Developing a sensitive FP assay that monitors Lsr2 binding to a DNA probe provided evidence that Zafirlukast prevents the complexation between Lsr2 and DNA. To our knowledge, Zafirlukast is the first reported inhibitor of the DNA-binding function of Lsr2. Intensity Fading data supported the hypothesis that this inhibition occurs through direct interactions between the drug and the protein. It was also shown using Kirby-Bauer disk diffusion assays that Zafirlukast is able to penetrate the mycobacterial cell wall and clearly inhibit the growth of *M. smegmatis* and Mtb in a dose-dependent manner. Other compounds identified through docking predictions and Zafirlukast derivatives did not exhibit any better or equivalent inhibitory activity against the Lsr2/DNA complexation.

Visible growth inhibition of Mtb and *M. smegmatis* upon administration of Zafirlukast was a really interesting aspect of this project. Since mycobacteria tend to form drug-insensitive clusters or biofilms even in the presence of bactericidal compounds such
as isoniazid, it was necessary to determine the number of colony forming units in the viscous bacterial detritus observable in cultures incubated in the presence of Zafirlukast.\textsuperscript{62} Bacterial death of \textit{M. smegmatis} incubated with the drug was recorded after different time points and the killing curve obtained implies that this inhibitor exhibits bactericidal activity.

The RT-PCR assay revealed that incubating a culture of \textit{Mt} in the presence of Zafirlukast had a measureable effect on transcription levels of the tested mycobacterial genes that are known to be bound by Lsr2 under normal conditions. A significant increase in gene expression was observed upon application of the drug inhibitor for all of the tested genes, especially \textit{lsr2}. This result was expected since Lsr2 is a global repressor of gene expression. It has been shown that Lsr2 expression levels increase slightly in response to heat shock and addition of iron, but that change is minor compared with the almost 9-fold change observed when Zafirlukast is applied. This difference suggests that the response triggered by the presence of Zafirlukast uses a mechanism different from the antibiotic-induced stress caused by drugs from typical treatment regimens. Therefore, it is reasonable to propose that Zafirlukast is acting through a novel molecular mechanism.\textsuperscript{63}

The first considered path by which Zafirlukast inhibition of Lsr2 could lead to growth inhibition of mycobacteria is by overexpressing the Lsr2 protein. This could lead to saturation of the mycobacterial genomic DNA with Lsr2, which would strongly inhibit expression of genes important for growth. However, this possibility is not consistent with the RT-PCR results that show a global increase in transcription levels of the tested genes.

Another mechanism of action could result from toxicity due to the overexpression of some genes upon inhibition of Lsr2 activity. Indeed, some proteins are beneficial to the
bacteria at moderate levels but are detrimental to viability at higher concentrations due to their enzymatic activity or the overproduction of toxic metabolites. This has already been observed in Mtb. For instance, the production of ChiZ, a cell wall hydrolase involved in cell division, increases dramatically upon exposure to DNA damaging agents and promotes the survival of the bacterium under those conditions.\(^{64}\) However, if ChiZ levels are too high under non-stressing conditions, cell division is compromised and the overall viability of the bacteria is decreased, possibly because overexpressing this protein promotes degradation of cell wall components.\(^{64-65}\)

Toxicity due to metabolite accumulation in mycobacteria was also recently shown through the example of GlgE.\(^{66}\) Bacteria in which the gene encoding GlgE has been deleted see an accumulation of maltose 1-phosphate that promotes the increased expression of enzymes necessary for production of the toxic metabolite, thereby producing a positive feedback loop that promotes rapid self-poisoning of Mtb.

The main purpose of this project was to identify inhibitors of Lsr2 DNA-binding activity for the development of novel compounds utilizable in TB treatment. Although the exact mechanism of action by which this inhibition occurs is yet to be discovered, the presented data suggest inhibition of Lsr2 DNA-binding activity by Zafirlukast and a physiological response by mycobacteria to this inhibition. Zafirlukast appears to be specific to Lsr2. As previously mentioned, multiple obstacles can hamper Mtb drug development. The first one is ensuring that the drug can enter mycobacterial cells.\(^{67}\) This does not seem to be a barrier for Zafirlukast as suggested by the growth study: the drug can clearly penetrate the mycobacterial cell wall. The second major obstacle is making certain that the desired drug can get to the site of infection. Zafirlukast is currently used
as a treatment for asthma, which ensures that the molecule is able to access the primary site of Mtb infection: the lungs. Additionally, Zafirlukast is already commercially available and safe for daily use. This drug could have an immediate impact on TB treatment, especially if used in addition to other typical therapies. The discovery of its anti-tubercular activity paves the way for the development of new classes of compounds affecting the same target but with improved efficacy.

Unfortunately, no additional structural knowledge could be gained from fusion proteins or complexes of Lsr2 and DNA since attempts at growing crystals remained unsuccessful. The next idea to try to promote crystallization is to design a fusion protein consisting of only DNA-binding Lsr2C and the T4 lysozyme, which could also be premixed with the designed oligonucleotides mentioned in this thesis before setting up the hanging drop trays. Moreover, further FP assays using different sizes of labeled hairpin DNA will be employed to determine the optimal size of oligonucleotide that exhibits the tightest binding to an Lsr2 dimer. Those results will be used to design new DNA molecules for crystallization purposes.

One of the other leads for future work on Lsr2 inhibition is to purchase or synthesize Zafirlukast analogs and study changes in inhibitory activity upon modifying its different functional groups or shortening the molecule. This would be particularly interesting for evaluating the importance of the sulfonamide bond and its substituent groups, since our current thought is that it mimics DNA phosphate backbone, which could possibly be how Zafirlukast disrupts the Lsr2/DNA complex. Another idea is to design Zafirlukast-like molecules that would be capable of binding both DNA-binding domains of Lsr2, based on the inherent 2-fold symmetry of the protein.
Finally, it would be interesting to look for possible mutations within the mycobacterial genome of colonies resistant to Zafirlukast. Preliminary experiments have been performed through sequencing of flanking regions (+/- 200 bp) of the *lsr2* gene in *M. smegmatis* after two days of incubation in the presence of the drug, but only silent mutations could be observed. Eventually, global analysis of mycobacterial transcription levels in response to Zafirlukast application will be required to more clearly assess the mechanism of growth inhibition.
References


disease: the official journal of the International Union against Tuberculosis and Lung Disease 1999, 3 (10 Suppl 2), S231-79.


57. (a) Belisle, J. T.; Vissa, V. D.; Sievert, T.; Takayama, K.; Brennan, P. J.; Besra, G. S., Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* 1997, 276 (5317), 1420-2; (b) De Smet, K. A.; Weston, A.; Brown, I. N.; Young, D. B.; Robertson, B. D., Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology* 2000, 146 (Pt 1), 199-208; (c) Sassetti, C. M.; Boyd, D. H.; Rubin, E. J.,


63. (a) Betts, J. C.; Lukey, P. T.; Robb, L. C.; McAdam, R. A.; Duncan, K., Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by


Appendix A

lsr2 and Lsr2 sequences

Gene sequence

M. tuberculosis H37Rv|Rv3597c
lsr2: 339 bp

1 - atg gcg aag aaa gta acc gtc acc ttg gtc
31 - gac gat ttc gac ggt tgc ggc gcc gac
61 - gaa aac gtc gaa ttc ggg ctt gac ggg gtc
91 - acc tat gag atc gac ctt tcc act aag aat
121 - gcc acc aaa ctc gtc gac cgt aag cca
151 - tgg gtc ggc ggc cgt cgc gtc gtt ggg
181 - cgc cgg gcc cgc cgt ctc gga tcc ggc cgt
211 - gga cgt ggc cgc atc gac cgc gag cag aca
241 - gcg cgc atc cgc gaa tgg gct cgt cgt aac
271 - ggg cac aat gtc tgc aag cga ggc cgg atc
301 - cgg gcc gac gtc atc gac gca tac cac cgg
331 - gcg acc tga

Protein sequence

M. tuberculosis H37Rv|Rv3597c
Lsr2: 112 aa

MAKKVTVTLVDDFDGSGAADETVEFLDGVTVYIEIDLSTKNATKLRGDLKQWVA AGRRVGGRRRGRSGSGRGGRGAIDREQSAIAIREWARRNGHNVSTRGRIPADVIDAYHAAT