Structural Characterization of *Mycobacterium tuberculosis* RNA Polymerase Binding Protein A (RbpA) and Its Interactions with Sigma Factors

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Alessio Bortoluzzi

Department of Biochemistry

University of Leicester

July 2013

Abstract

Structural Characterization of *Mycobacterium tuberculosis* RNA Polymerase Binding Protein A (RbpA) and Its Interactions with Sigma Factors.

Alessio Bortoluzzi, July 2013

The RNA polymerase binding protein A (RbpA) is a 13 kDa protein, encoded by the gene *Rv2050,* that was shown to be essential for the growth and survival of the important human pathogen *Mycobacterium tuberculosis.* Although is not clear yet why RbpA is essential in *M. tuberculosis*, significant progress has been made in the characterization of the protein. For instance, it was shown that RbpA binds to the β-subunit of the RNA polymerase (RNAP) and activates transcription. Interestingly, it was reported that RbpA can enhance the transcription activity of the RNAP containing the primary σ-subunit σ^A but does not have any detectable effect if the RNAP is associated with the alternative σ-subunit σ^F. Moreover, it was also shown that RbpA might influence the response of *M. tuberculosis* to the current frontline anti-tuberculosis drug rifampicin.

The research project described in this thesis contributes to the ongoing efforts to characterize RbpA by providing the structure of the protein and identifying the principle σ-subunit σ^A, and the principle-like σ-subunit σ^B, as interaction partners. The solution structure of RbpA reveals the presence of a central structured region and highly dynamic N- and C- termini. Both termini are involved in the formation of a tight complex with the σ-subunit but only the C-terminal region appears to be essential for this interaction. The finding that RbpA also binds to the RNAP σ-subunit suggests new possibilities for the mechanism of action used by RbpA to activate transcription. Furthermore, preliminary data obtained using a Δ*Rv2050* conditional mutant strain of *M. tuberculosis* suggest that the interaction with the σ-subunit is essential for the functionality of RbpA.

Acknowledgements

Being here for four years of my life as a PhD student was a very positive experience. It is not an easy task to acknowledge all the people that have contributed in making this possible, but I will try. Firstly, I would to thank my supervisors Dr. Helen O'Hare and Prof. Mark Carr for their support throughout my PhD, from the first days in the lab to the search of a postdoctoral position. Sincerely, thank you very much for all your help. I would also like to thank my committee members Dr. Galina Mukamolova, Dr. Richard Bayliss and Dr. Christine Wells for their scientific inputs. Thanks to Prof. Daniela Ghisotti and her team to kindly provide us with the *ΔRv2050* conditional mutant strain of *M. tuberculosis* and thanks to the university of Leicester for funding my PhD. Many thanks to all the colleagues from lab 1\12 and lab 213b. It is nice going to the lab each morning knowing that you will find people able to help you with your experiments but also able to help you to have a laugh from time to time. In particular I would like to thank Andrew, Abdul, Shaun, Barbara and Lorna to proofread chapters and sections of my thesis.

I would also like to acknowledge all the people that support me in these four years outside the lab-life. I am not sure if I would have left for this adventure in Leicester with the same enthusiasm and serenity without having a solid and supportive family. Thank you Mum, Dad and Sister! Many thanks to Pigi, Dalla, Simone, Paolo, Dario, Toto, Andrea and all the friends who made me feel always welcomed in Biella each time I was coming back, even after four years. To my housemates Tomasz and Laimonas to help me in seeing Leicester as my new home. And finally, thank you Nadia. With you at just few hours of train from Leicester all the problems of these four years have been divided by two and all the joys have multiplied by two.

Contents

Abbreviations

DNA bases:

Adenine	A	Guanine	G
Cytosine	C	Thymine	T

Single and three letter codes for amino acids:

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

Abbreviations:

AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
CD	Circular dichroism
dH_2O	De-ionised water
DNA	Deoxyribonucleic acid
DTT	1,4,-dithiothreitol
E	RNAP-core
EDTA	Ethylenediaminetetraacetic acid
FeBABE	Fe(III) (*s*)-1-(*p*-bromoacetamidobenzyl) ethylenediamine tetra-acetic acid
GST	Glutathione S-transferase
HIV	Human immunodeficiency virus
HSQC	Heteronuclear single quantum coherence
IC_{50}	Half inhibitory concentration
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-1-thio-β-D-galactoside
K_D	Equilibrium dissociation constant
kDa	kilo Dalton
LB	Luria-Bertani broth
MDR-TB	Multidrug-resistant TB
MIC	Minimum inhibitory concentration
mRNA	Messenger RNA
MW	Molecular weight
Ni-NTA	Nickel nitrilotriacetic acid

NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NTPs	Nucleoside triphosphates
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PI	Pristinamycin I
ppm	Parts per million
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNAP	RNA polymerase
SEC	Size exclusion chromatography
SDS	Sodium dodecyl sulphate
TB	Tuberculosis
TEV	Tobacco etch virus
TOCSY	Total correlation spectroscopy
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
TROSY	Transverse relaxation-optimized spectroscopy
v/v	Volume per volume
w/v	Weight per volume
WHO	World Health Organisation
XDR-TB	Extensively drug-resistant TB

Chapter 1

Introduction

1.1 Human tuberculosis: the current situation.

Tuberculosis (TB) is an infectious disease that has plagued humanity for thousands of years. Molecular evidence of TB infection has been found in an Egyptian mummy (1), proving that TB has existed in humankind for at least 5000 years. Several TB epidemics have been recorded in the story of humanity, but it was probably during the industrial revolution and the formation of the first big urban centres, which favoured the spreading of infectious diseases, that TB became one of the most feared diseases (2). By the 19th century TB was the leading cause of death in most European countries (3). In the 20th century the discovery of the first anti-TB drugs resulted in the introduction of an effective multidrug chemotherapy for TB that led to a significant reduction in the incident cases in the most developed countries (4). However, on the global scale the TB burden continued to rise and in 1993 the World Health Organization (WHO) declared TB a global emergency. Significant efforts have been undertaken worldwide to contain the spreading of the disease and since 2006 the absolute number of new TB cases has constantly decreased (5). However, despite this encouraging data the global TB burden remains enormous. According to the WHO report "Global Tuberculosis Report, 2012" (5) in 2011 1.4 million people died from TB and there were an estimated 8.7 million new cases of TB. The geographic distribution of the TB burden is uneven (figure 1.1) and often correlates with a reduced socio-economic status (5).

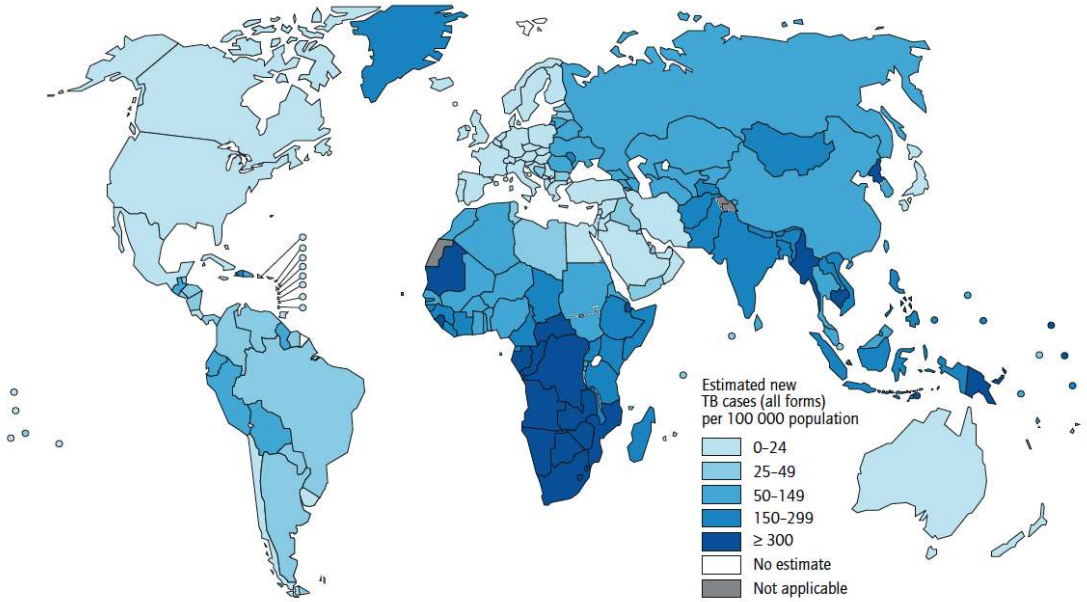


Figure 1.1. Estimated TB incident rates, 2011. Representation of the geographic distribution of the estimated new TB cases for the year 2011. Image modified from the WHO report "Global Tuberculosis Report, 2012" (5).

The five countries with the highest number of TB incident cases in 2011 were India (2.0–2.5 million), China (0.9–1.1 million), South Africa (0.4–0.6 million), Indonesia (0.4–0.5 million) and Pakistan (0.3–0.5 million) (5). However, TB affects also economically favoured countries. For example, in the UK 8,963 cases of TB were reported in 2011 confirming a stable incidence rate since 2005. The majority of the cases were notified from urban centres, with London accounting for 39% of all the cases reported in the UK (figure 1.2) (6).

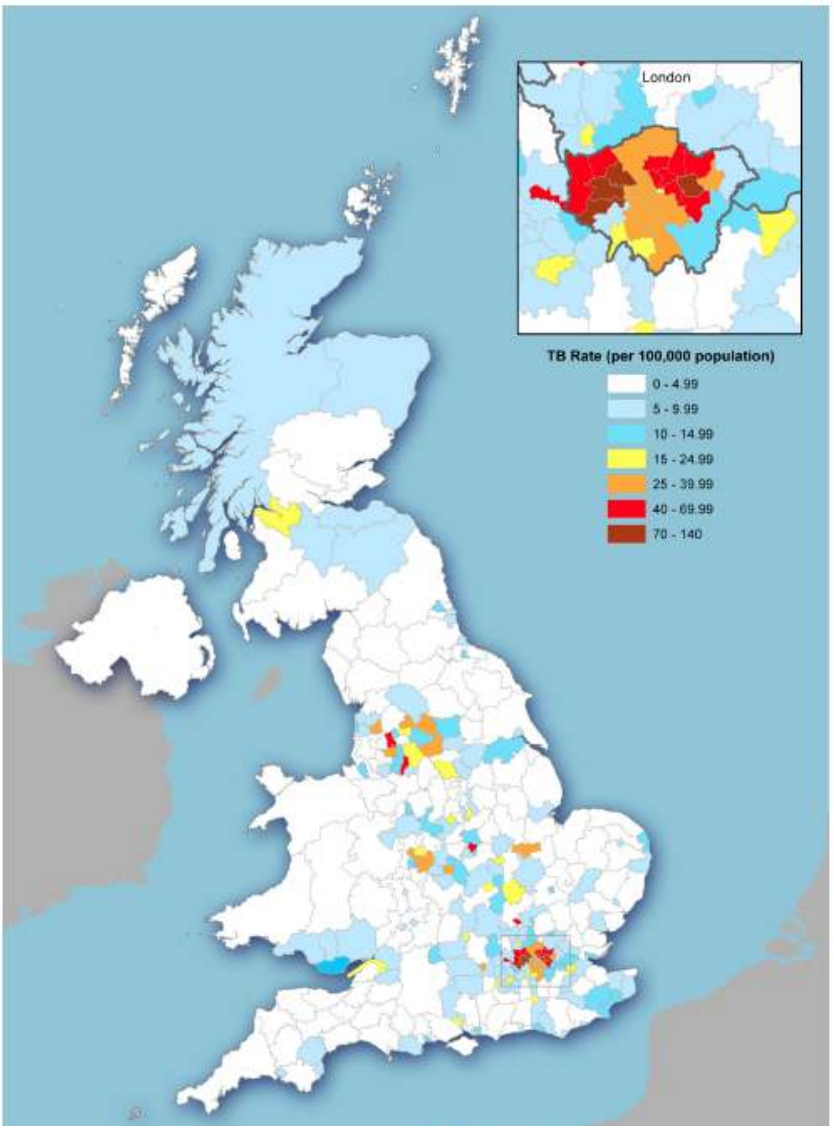


Figure 1.2. Three years (2009-2011) average TB case rates in the UK. Image modified from the Health Protection Agency (HPA) report "Tuberculosis in the UK: 2012 report" (6).

A factor that significantly affects the rate and spreading of TB is the Human Immunodeficiency Virus (HIV) epidemic (2). Both TB and HIV undermine the efficiency of the immune system of the host and potentiate each other (2). It has been estimated that 5–10% of patients infected with *Mycobacterium tuberculosis,* the main etiological agent of TB in humans, are likely to develop active TB during their lives, but

this risk is increased to 50% in the case of HIV co-infection (2). Sub-Saharan Africa is particularly plagued by the synergy between HIV and TB; in the 2011, 79% of worldwide TB patients co-infected with HIV were reported in Africa (5).

The data reported about the worldwide TB burden demonstrates how *M. tuberculosis* has not yet released its grip on humankind and underline the need of a continued global effort to eradicate TB.

1.2 The etiology of TB.

Until the late 19th century the cause of TB was unknown. Different theories were formulated trying to explain the cause of TB, including that it was a contagious disease, heritable, or due to nutritional factors. Evidence that TB was due to an infectious agent was produced in the second half of the 19th century (7). The nature of the agent remained unknown until 1882 when Robert Koch held his famous presentation "The etiology of tuberculosis" where he demonstrated that characteristic bacilli were always present in TB infections (8). The bacterium identified by Robert Koch was *M. tuberculosis* (8), but TB can be caused by the infection of the host with any bacterium member of the *M. tuberculosis* complex, which components include: *M. tuberculosis, M. africanum, M. canettii, M. microti, M. bovis, M. caprae* and *M. pinnipedii*. *M. tuberculosis* is the most frequent and important agent causing TB in human (7). It is an aerobic, rod-shaped, non-spore forming bacterium with a particularly slow duplication time (between 12 and 24 hours in optimum conditions) (7).

The cell wall of members of the genus *Mycobacterium* shows peculiar features. It has an unusually high level of lipid content compared to most prokaryotic organisms and its main structural element is a cross linked network of peptidoglycan. The cell wall results in a highly impermeable layer that is thought to play a pivotal role in the intrinsic

resistance of *M. tuberculosis* to common antibiotics and lytic enzymes (9). The peculiar composition of the cell wall makes the staining of *M. tuberculosis* with commonly employed staining procedures, such as Gram's staining, difficult and sometimes misleading. *M. tuberculosis* is classified as an "acid-fast" bacterium, indeed once stained, typically with arylmethane dyes such as carbol fuchsin, it is able to retain the dye even after several cycles of washes with acid-alcohol solutions. The "acid-fastness" of *M. tuberculosis* is the key feature on which several mycobacteria staining methods (*e.g.* the Ziehl-Neelsen technique) are based (7).

M. tuberculosis is most commonly transmitted by patients with pulmonary TB. Coughing, sneezing or talking can produce aerosolized droplets containing bacilli. The droplets can remain in the atmosphere for several hours (10) and if inhaled by healthy patients have the potential to cause TB and start a new cycle of infection. The infection process occurs with high efficiency, indeed approximately one third of the human population is infected with *M. tuberculosis* (11). However, it has been estimated that only 10% of patients infected with *M. tuberculosis* will develop active TB in their life (12). In fact, when infectious bacilli reach the lungs they are recognized by the immune system of the host which in most of the cases is able to contain the infection inside a complex multicellular structure termed a granuloma and avoid further dissemination of TB (7). *M. tuberculosis* is a highly adapted intracellular pathogen that is able to interfere with and resist the common antimicrobial strategies employed by the immune system. Thus, complete eradication of *M. tuberculosis* by the immune system of the host is extremely rare, if not absent, but as long the bacilli remain confined in the granuloma the disease is asymptomatic and not transmissible (13). However, in case of failure of the containment the infection spreads to other loci within the body and, if untreated, in 70% of the cases this will lead to death of the patient (14).

1.3 Available therapies for TB and their limitations.

Prior to the development of antibiotic treatments, TB patients were subjected to a range of treatments that often resulted into little or no improvements. In the first decades of the 20th century it was common practice for TB patients to be healed in sanatoriums, where the therapy was based on absolute rest of body and mind, good nutrition and exposure to high altitude and fresh air (15). The first step toward the development of an effective therapy for TB was the discovery in 1944 of streptomycin and para-amino salt of salicylic acid. During the years 1950-60s several other molecules active against TB were discovered such as: isoniazid, pyrazinamide, ethambutol and rifampicin, that are still the base of the multi drugs chemotherapy currently employed for TB (table1.1) (16).

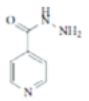
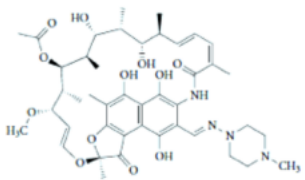
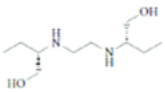
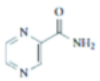
Antibiotic	Chemical structure	Mechanism and target	Mutation associated with resistance
Isoniazid		Inhibits mycolic acid synthesis; primary target is InhA and secondary targets are KasA ar DfrA	*KatG* (required for drug activation); *inhA* (promoter mutations); and others
Rifampicin		Inhibits transcription; RNA polymerase β-subunit	*rpoB*
Ethambutol		Inhibits arabinogalactan synthesis; possibly EmbB	*embB*
Pyrazinamide		Inhibits trans-translation; ribosomal protein S1 (RpsA) (17)	*pncA* (required for drug activation); *rpsA* (17)

Table 1.1. First line drugs used for the treatment of TB. Modified from reference (18).

The current guidelines for the treatment of patients with drug-susceptible TB recommend a six month regimen of: isoniazid, rifampicin, pyrazinamide and ethambutol for the first two months (intensive phase) and rifampicin and isoniazid for the next four months (continuation phase) (19). The long duration of the TB therapy increases the risk of mismanagement of the treatment, especially in countries where there is not easy access to healthcare facilities and patients are more likely to fail to complete the course. This results in a poor outcome of the therapy and in an increased risk of emergence resistant strains of *M. tuberculosis*. In the 2011 there were an estimated 310,000 cases of multidrug-resistant (MDR) TB (resistance to at least rifampicin and isoniazid) (5). In cases of infection with MDR-TB the duration of the therapy is extended to twenty months (20) and involves the use of second line drugs which, in general, are more expensive and have a higher risk of develop side effects. Unfortunately, there exist *M. tuberculosis* strains that have a higher level of drug resistance than MDR, they are defined as extensively resistant (XDR: resistance to rifampicin, isoniazid, fluoroquinolone and at least one of three injectable second-line drugs such amikacin, kanamycin or capreomycin) and in case of infection with XDR-TB the outcome of the TB therapy is further compromised (5). The first-line TB drugs currently employed have been essentially the same since the introduction of a specific multi-drug TB chemotherapy and resistance mechanisms for each drug are known (table 1.1). The discovery of new active molecules against *M. tuberculosis* that allow a reduction in the length of the therapy and to overcome the problem of drug resistance would greatly improve TB therapy.

Another important strategy adopted to control TB is vaccination that is performed with the use of an attenuated strain of *Mycobacterium bovis* known as bacille Calmette-Guèrin (BCG). It is generally accepted that the vaccination based on BCG is effective

for extra-pulmonary TB in young people. However, in case of adult pulmonary TB the protection given by the vaccine is extremely variable: from 0 to 80% (21).

1.4 Latent TB requires important changes in the metabolism of *M. tuberculosis*.

One of the reasons that has greatly contributed to the TB pandemic is the ability of *M. tuberculosis* to survive, even for decades, inside the host leading to an asymptomatic and not infective status of the disease defined as "latency". Patients with a latent TB infection represent the major reservoir of *M. tuberculosis* bacilli (22) and the latent status can be easily reversed to an active and transmissible infection. *M. tuberculosis* survives during the phase of latency by a complex series of interactions with the immune system of the host and by entering into a so-called "dormant state" (23) which is characterized by a non-replicating persistence (24). In the dormant state *M. tuberculosis* metabolism undergoes important changes, protein synthesis is reduced and the available resources are used to maintain the cell wall and genome integrity (25). Unfortunately, if the immunosurveillance of the host is compromised, for example in case of HIV co-infection, poor nutrition and aging, *M. tuberculosis* can recover from the dormant state and develop an active infection (10). Patients with an active TB infection in the lungs (pulmonary-tuberculosis is the most frequent form of TB) produce aerosolized droplet containing infectious bacilli that have the potential to start a new cycle of infection in a healthily patient.

The changes of *M. tuberculosis* metabolism at different stages of the infection and the interferences with the immune system of the host are critical aspects to complete successfully an infection cycle and are expected to require a tight gene expression control (26).

1.4.1 The DNA dependent RNA polymerase.

Gene expression control in bacteria occurs primarily during transcription initiation (27). In *M. tuberculosis,* as in all the prokaryotic organisms, the DNA dependent RNA polymerase (RNAP) is responsible for all the transcription activity (28), thus playing a pivotal role in gene expression control. The bacterial RNAP is formed by a highly conserved core of five subunits ($\alpha_2\beta\beta'\omega$) (28-33), which is competent for transcription elongation, and by a sixth σ-subunit (or σ-factor) that is essential for promoter recognition and transcription initiation. The complex formed by the RNAP core and the σ-subunit is usually defined as RNAP holoenzyme. The overall structure of the complex is reminiscent of a "crab-claw" like structure, in which the β and β' subunits are the two pincers and form the main RNAP channel through which the template DNA can reach the catalytic site of the complex (figure 1.3).

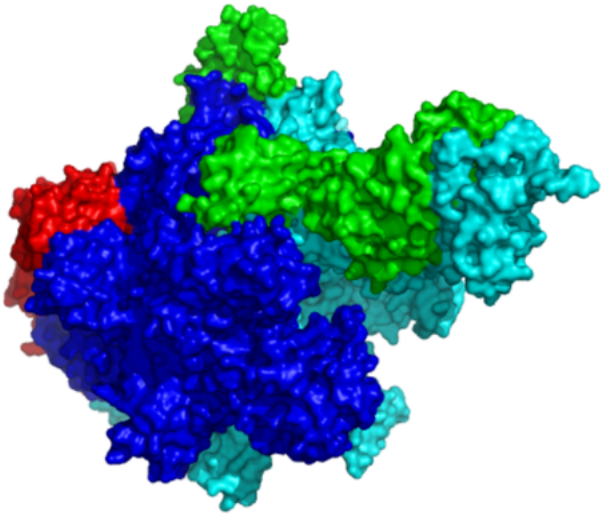


Figure 1.3. The *Thermus thermophilus* RNAP holoenzyme structure. A surface representation of the *T. thermophilus* RNAP holoenzyme (PDB: 1IW7) with the subunits part of the complex highlighted with different colours: α in red, β in blue, β' in cyan and σ in green. The ω subunit is not visible.

The two identical α-subunits have a role in binding certain DNA promoters, but the main function is as a scaffold in assembling the large β and β' subunits which form the active site of the complex (34,35). The ω-subunit has no active role in transcription, but seems to function as chaperone for the folding of the β'-subunit (36).

The process that leads to transcription involves a series of steps that have been recently reviewed (37) and can been summarized as follows. The RNAP holoenzyme (R) recognizes and binds the DNA promoter (P) forming the closed complex (RP_c). The formation of the RP_c triggers a series of changes in both DNA and RNAP that ultimately leads to DNA melting for approximately 13 base pairs around the transcription site (+1) and to establish the open complex (RP_o) formation. The RP_o is competent for transcription and synthesizes short abortive RNA chains that are released from the complex. When a longer RNA chain, typically 11 nucleotides, is formed, the contacts between RNAP and the DNA promoter are disrupted and the RP_o stability is increased. Further elongation of the nascent RNA chain triggers the release of the σ-subunit and the resulting RNAP-core DNA complex continues transcription with high efficiency.

1.4.1.1 The RNAP is the target of the front line anti-TB drug rifampicin.

Rifamycins are an important class of antibiotics targeting the bacterial RNAP which were isolated for the first time in 1959 (38). Within ten years rifampicin, probably the most known compound that belongs to rifamycins, was introduced into the TB therapy reducing the treatment time from 19 to 9 months. Rifampicin can easily diffuse through living cells, making it particularly effective against intracellular pathogens like *M. tuberculosis* (39). Unfortunately, the rate of spontaneous mutations that leads to rifampicin resistance in *M. tuberculosis* is relatively high: 2.25×10^{-9} mutations per bacterium per generation (40). For this reason rifampicin is used almost exclusively in

combination with other drugs and its usage, in the United States, is restricted to TB therapy and to cases of clinical emergencies (41). Mutations that confer rifampicin resistance occur almost exclusively in the *rpoB* gene (encoding for the β-subunit of the RNAP) (42). Indeed, rifampicin antimicrobial activity is due to inhibition of transcription activity by its binding to the β-subunit of the bacterial RNAP. The rifampicin binding site is far from the catalytic site, the closest approach is at about 12 Å, and the transcription inhibition action is thought to be due to steric hindrance of the nascent RNA chain (41) (figure 1.4).

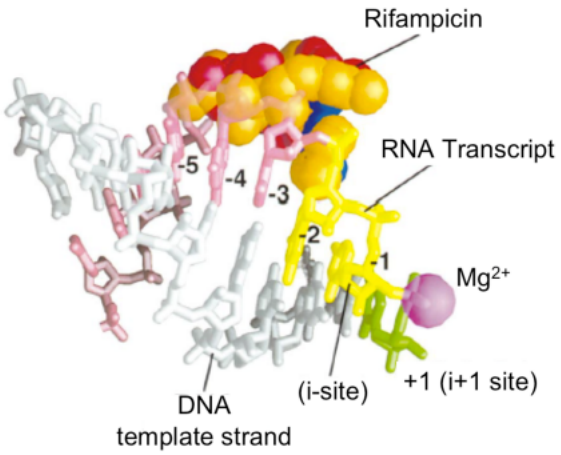


Figure 1.4. Rifampicin blocks transcription by steric hindrance. Model of the elongation complex showing the 9 bp RNA/DNA hybrid (from +1 to -8) and the Mg^{2+} ion (magenta sphere) in the RNAP active site. The RNAP and the rest of the nucleic acids are omitted for clarity. The incoming nucleotide in position +1 is coloured in green. The RNA transcript that can be accommodated in presence of rifampicin (-1 and -2 positions) is coloured in yellow and the RNA transcript that cannot be accommodated in presence of rifampicin (from -3 to -8 positions) is coloured in pink. The template DNA strand is coloured in gray. Rifampicin is shown in a CPK representation (carbon atoms are coloured in orange, oxygen in red and nitrogen in blue) (41).

Several biochemical and genetic studies support a steric-block model for the mechanism of action of rifampicin (for a comprehensive review see reference (43)), including the solution of the structure of *Thermus aquaticus* RNAP core bound to rifampicin (41). However, a simple steric-block model cannot easily account for some of the evidence previously reported. For example, a number of mutations in *rpoB* conferring rifampicin resistance map far from the inhibitor binding site (44) and there were observed differences in the rifampicin sensitivity in function of the type of σ-factor contained in the RNAP (45). The lack of the σ-subunit in the structure of the *T. aquaticus* RNAP core in complex with rifampicin (41) could hide some aspects of the rifampicin mechanism of action. In 2005, Artsimovitch *et al.* solved the structures of the *T. thermophilus* RNAP holoenzyme in complex with rifabutin and rifapentine (rifampicin, rifabutin and rifapentine are the only rifamycins currently employed in therapy) (46). Interestingly, the structures reported lacked the catalytic Mg^{2+} ion, in contrast to most of the RNAP structures previously described, and showed interactions between the inhibitors and the σ-subunit. Thus, Artsimovitch *et al.* (46) proposed a new mechanism of action for rifamycins, which should work in addition to the steric-block model, where the binding of the inhibitor would trigger an allosteric signal that is transmitted through both the β and σ subunits to the active site and culminate in the release of Mg^{2+} ion. However, in a recent work, Feklistov *et al* (47) showed that rifamycins do not affect the binding affinity of Mg^{2+} to the RNAP questioning the allosteric model. In summary, although the steric hindrance caused by rifampicin on the nascent RNA chain seems to be the key event that leads to transcription inhibition there could be other molecular events induced by rifampicin that have not yet been fully characterized.

1.4.2 The role of the σ-subunit in gene expression control and classification and overview of the *M. tuberculosis* σ-factors.

The first step in transcription is the recognition and binding of the gene promoter by the RNAP. In bacteria this step is mainly accomplished by the σ-subunit, which is able to bind specific DNA sequences conferring promoter specificity to the RNAP holoenzyme. All bacterial genomes encode one primary σ-subunit that is responsible for the transcription of housekeeping genes and is essential for the growth of the organism (26). However, the total number of σ-subunits encoded can vary from one in *Mycoplasma pneumoniae* (48) to 65 in *Streptomyces coelicolor* A3(2) (49). Each σ-subunit has different promoter specificity so the use of different σ-subunits, which is often linked to a specific stimulus such as stress response, represents an efficient mechanism of gene expression control. In *M. tuberculosis* the possibility of switching between different σ-subunits is exploited to express particular regulons to sustain multiple stages during host infection (*e.g.*: adhesion, invasion, intracellular replication and dissemination to other sites) (50).

The σ-factors have been divided in two families: sigma 70 (σ^{70}) and sigma 54 (σ^{54}). The classification is based on the similarity to the 70 kDa primary σ-factor or to the 54 kDa nitrogen-regulator σ-factor from *Escherichia coli* (51). All bacterial genomes encode at least one σ-factor that belongs to the σ^{70} family. Otherwise, members of the σ^{54} family are more rare and are not found in mycobacteria (52). The σ-factors belonging to the σ^{70} family can be further divided in four phylogenetically related groups (53). Group 1 contains essential primary σ-factors, with a high sequence homology to the σ^{70} of *E. coli,* which are responsible for the transcription of housekeeping genes. The σ-factors in group 2 are closely related to primary σ-factors, but are dispensable for growth. Group 3 σ-factors share a lower sequence homology

with the primary σ-factor and are normally activated in response to precise stimulus. Finally, proteins in group 4 are the most diverged from the primary σ-factor and often are responsive to extracytoplasmic signals (54).

Sequence alignment of members that belong to the σ^{70} family revealed the presence of four highly conserved regions (53) that can be further dived in sub-regions. Regions 2 and 4 are conserved in all the σ^{70} family members and are involved in the binding of the -10 and -35 promoter elements, respectively (55,56). Region 3 is less conserved and recognizes the "extended -10" promoter element (57). Only σ-factors from the first and second group contain region 1, which harbours the negatively charged subregion 1.1 that has been observed to inhibit the binding of free σ-factor to the DNA (58). Structural analysis of the RNAP holoenzyme (56,57,59,60) has shown that the primary σ-factors are organized in three compact domains: σ_2, σ_3 and σ_4 which have conserved folds and include the conserved regions 2,3 and 4, respectively. The conformation of the σ-subunit is extended, with the domains spread along one face of the RNAP core.

In *E. coli* several σ-factors compete for a limited pool of RNAP-core. The primary σ-factor has a higher affinity for the binding with the RNAP-core compared to alternative σ-factors. This guarantees the expression of the housekeeping genes essential for the viability of the organism (61,62). However, bacteria must be able to switch between different σ-factors in order to adapt their gene expression profile to specific stimuli. This is often achieved by controlling the expression of the alternative σ-factors and the expression of anti-σ-factors, proteins that bind and sequester specific σ-factor resulting in drastic change in the equilibrium dynamics for the binding with the RNAP core (63). The fine-tuning of the σ-factor choice can be further regulated by the expression of anti-anti-σ-factors, proteins that bind and inactivate the anti-σ-factors.

M. tuberculosis encodes 13 distinct σ-subunits (σ^A-σ^M) (64). σ^A is believed to be the primary σ-subunit because disruption of *sigA*, the gene encoding for σ^A, has not been possible in *M. tuberculosis* (26). Furthermore, the σ^A orthologue in *M. smegmatis* was shown to be essential, in fact its inactivation leads to non-viable *M. smegmatis* strains unless an extra copy of the gene is provided (65). σ^A seems to have a role also in the virulence of *M. tuberculosis.* Indeed, overexpression of σ^A enhances the growth of *M. tuberculosis* in human macrophage and in mice lungs (66). Moreover, a point mutation in σ^A (R515H) was shown to attenuate the virulence of a *M. bovis* strain in a guinea pig infection model (67), but the same mutation failed to show attenuation in a different animal model (Australian brushtail possum) (68). σ^B is a primary-like σ-subunit that shares a high degree of sequence homology with the last 300 residues of σ^A. *M. tuberculosis* and *M. smegmatis* strains with *sigB* gene disrupted are viable showing that σ^B is not essential (65,69). However, *M. tuberculosis* strains harbouring *sigB* mutations are more sensitive to surface stress (SDS-mediated), heat shock, oxidative stress, exposure to vancomycin and hypoxia (52,69).

The remaining *M. tuberculosis* σ-factors (σ^C-σ^M) are alternative σ-subunits (26). All the four phylogenetically related σ-factor groups are represented in *M. tuberculosis* and their composition is summarized in figure 1.5.

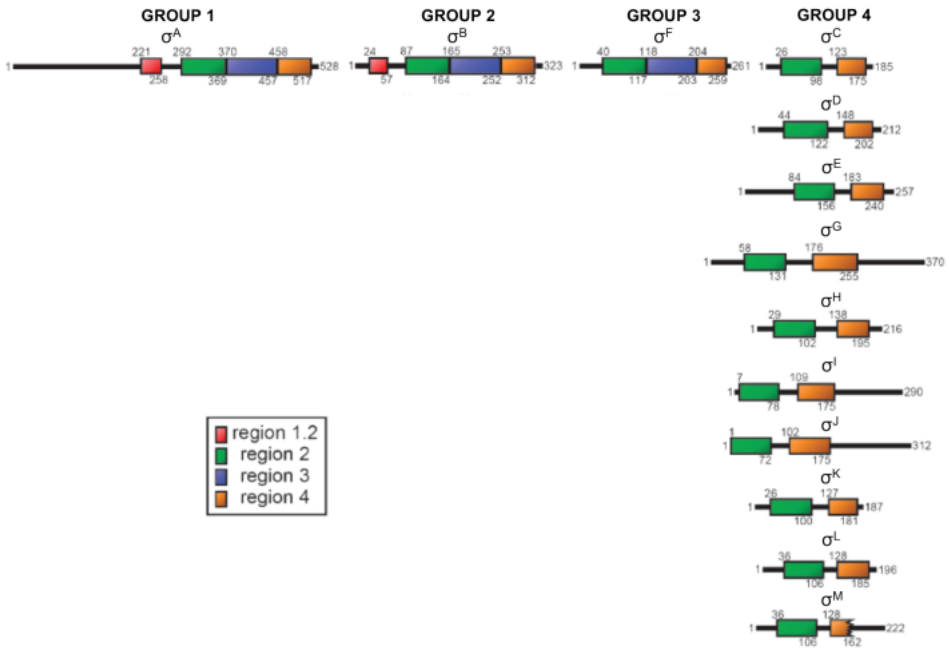


Figure 1.5. ***M. tuberculosis*** **σ-factors classification.** Adapted from reference (52).

During exponential growth all the 13 *M. tuberculosis* σ-factors are expressed (70,71). The expression of σ^A is constant and the *sigA* messenger RNA (mRNA) has a very long half-life (~40 minutes) (72), for these reasons σ^A is often chosen as internal control during mRNA quantification (73). Despite the remarkable homology between σ^A and σ^B sequences the expression of σ^B is clearly increased in response to several stress stimuli as heat shock, low aeration and exposure to hydrogen peroxide (70,72) and there is no overlap of the regulons activated by σ^A and σ^B (74,75). The σ^A and σ^B sequences and loci are conserved in all mycobacteria. Between the *M. tuberculosis* alternative σ-factors only σ^E is conserved across the whole genus (76) and σ^C is conserved only in pathogenic mycobacteria (77). The conservation of the remaining σ-factors is relatively variable within different species (76). Interestingly, *M. tuberculosis* has the highest ratio

of alternative σ-subunits compared to genome size of any obligate pathogen (52), possibly reflecting the complex stages of infection that require tight regulation.

The possibility to switch between 13 different σ-subunits to reprogram transcription is for *M. tuberculosis* an effective strategy to control gene expression, but certainly is not the only one. Indeed, following the completion of the *M. tuberculosis* genome sequence were identified a total of 190 potential gene expression regulators (78). The expression of a particular gene is the result of the interaction of different players in a complex and dynamic network that, for *M. tuberculosis*, is still not completely understood.

1.4.3 Transcription regulator proteins that do not bind DNA

There are several known mechanisms used by bacteria to control the transcription activity of the RNAP including: non-protein ligands, the folded bacterial chromosome structure and transcription factors (27,79). The best characterized mechanisms often involve direct interaction with the DNA. However, there is a class of transcription modulators that act interacting directly with the RNAP but not DNA and recent studies have elucidated aspects of their mechanism of action. Gre-factors are small proteins (~20 kDa), evolutionarily conserved, including in *M. tuberculosis*, that activate transcription rescuing arrested RNAP (80). During transcription elongation certain DNA sequences, generally enriched in the content of Thymine, can slow or even stop the RNAP activity (81). The arrest of transcription induces RNAP to proceed backward leading to disengagement of the nascent RNA chain from the catalytic site making further advance in transcription elongation impossible (82). The blocked RNAP can resume transcription by cleaving the newly synthesized RNA chain and generating a new 3' end in the catalytic site. This process is favoured by Gre-factors that have been shown to stimulate the endonucleolytic activity of the RNAP (83). Gre factors, and their

homologues, are composed of two domains: a coiled-coil N-terminal domain and a globular C-terminal domain (80). According to the currently accepted model the C-terminal domain binds on the surface of the RNAP and the N-terminal domain reaches the catalytic site of the holoenzyme through the RNAP secondary channel where it activates the endonucleolytic reaction (84-86). The RNAP secondary channel is solvent exposed providing an entry for the nucleoside triphosphates (NTPs) in transcription elongation and accommodates the nascent RNA chain during backtracking (33,87). An increasing number of evidence show that the secondary channel is also the route used by a class of small proteins, which share structural similarities, to directly interact with the RNAP and influence the transcription activity (figure 1.6).

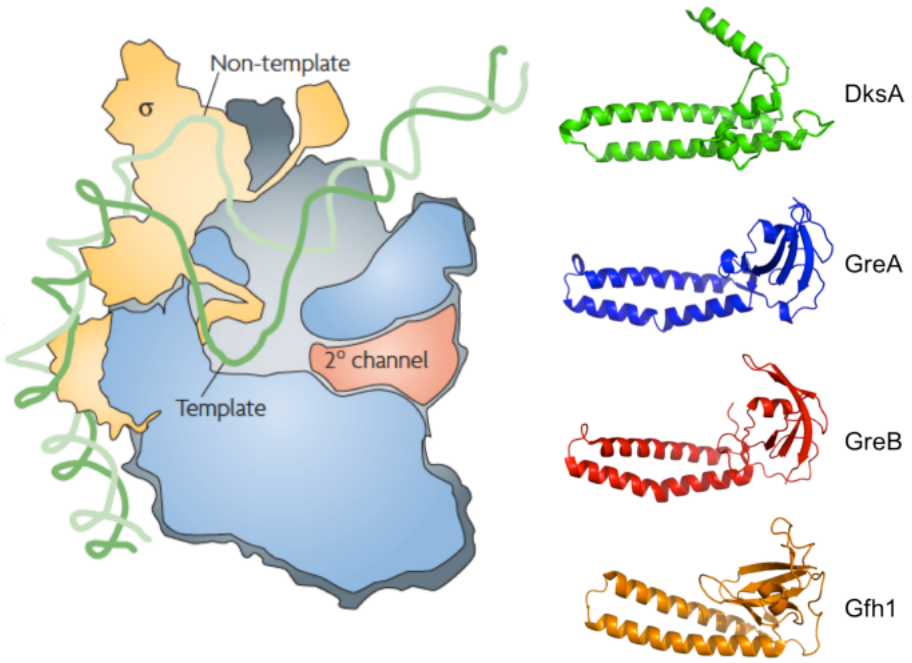


Figure 1.6. Transcription regulators that bind to the RNAP secondary channel. Cartoon representation of the RNAP holoenzyme-DNA complex. The DNA template and non-template filaments are shown in green and light green respectively. The σ-subunit is shown in light orange, the secondary channel in red and the rest of the RNAP in blue and gray. The structures of proteins that bind the secondary RNAP channel are also shown: DksA (PDB: 1TJL), GreA (PDB: 1GRJ), GreB (PDB: 2P4V) and Gfh1 (PDB: 2EUL). Figure modified from reference (79).

DksA is another transcription regulator protein that acts without direct binding with the DNA and in *E. coli* it has been shown to play a pivotal role in the stringent response (88). The stringent response is a general mechanism activated in bacteria in response to nutrient starvation characterized by an increase in the synthesis of hyperphosphorylated guanine nucleotides ppGpp and pppGpp, ((p)ppGpp) (89). High levels of (p)ppGpp lead to specific transcription inhibition of ribosomal RNA (rRNA) and some transfer RNA (tRNA) that ultimately will result in low ribosome biogenesis and decreased growth rate

(90). However, the ribosome synthesis inhibition induced by (p)ppGpp is virtually abolished by DksA inactivation (88). The molecular details of transcription inhibition induced by (p)ppGpp and DksA remain unknown. However, the structure similarities observed between DksA and the Gre-factors (figure 1.6) and evidence that the N-terminal coiled-coil domain of DksA binds the RNAP polymerase secondary channel (91) suggest that DksA and Gre-factor mechanisms of action could share important similarities.

Despite the high conservation of the stringent response mechanism, in several bacteria, including *M. tuberculosis,* there are no obvious DksA homologues. However, it has been found that a *M. tuberculosis* protein, named CarD, can functionally replace DksA in *E. coli* (92). CarD binds directly to the β-subunit of the RNAP, is essential for the stringent response in *M. tuberculosis* and for the control of rRNA synthesis also during normal growth conditions (92). Despite the lack of sequence homology DksA and CarD seem to have overlapping functions, but there are also significant differences between the two proteins. For example, CarD is essential for *M. tuberculosis* viability and binds to the RNAP outside the secondary channel, suggesting that *M. tuberculosis* has evolved a different mechanism to regulate transcription during the stringent response.

A further interesting example of a transcriptional regulator protein that does not interact with DNA is Crl. In *E. coli* Crl binds to σ^S and leads to an increased affinity for σ^S in the binding with the RNAP core (93). The details of the Crl mechanism of action are not known yet, but its binding with σ^S stimulates the formation of RNAP holoenzyme containing σ^S ($E\sigma^S$) resulting in an enhanced $E\sigma^S$ dependent transcription activity (94,95). Otherwise, in absence of Crl, the higher concentration and higher affinity of the primary σ-factor would limit or block the formation of the $E\sigma^S$ holoenzyme. No homologues of *E. coli* Crl have been identified in *M. tuberculosis.*

1.5 The *M. tuberculosis* RNAP Binding Protein A, RbpA.

A novel small RNAP binding protein RbpA was identified in *S. coelicolor* and is highly conserved in the actinobacteria (96). RbpA of *M. tuberculosis* contains 111 amino acids, is encoded by *Rv2050,* and is thought to play a role in the gene expression control network stimulating the transcription activity of the RNAP (97). RbpA expression is essential for the growth of *M. tuberculosis* (98,99) and is upregulated in several stress conditions: starvation, heat shock, oxidative stress, hypoxia, in mouse macrophages and in the presence of rifampicin and vancomycin (25,100-102). Interestingly, RbpA and its homologues in *S. coelicolor* and *M. smegmatis* have been shown to reduce the RNAP inhibition caused by rifampicin (97,101,102). Despite the recent progress achieved in the characterization of RbpA it is not clear yet which is the essential function that it underpins in *M. tuberculosis* or the precise molecular mechanism through which RbpA activates transcription and rescues the RNAP from the rifampicin inhibitor effect.

1.5.1 Mapping of the RbpA binding site on the RNAP

RbpA was identified for the first time as an RNAP binding protein in 2001 during a study on the oxidative stress response in *S. coelicolor* (96). The interaction between RNAP and RbpA was found to occur also in *M. smegmatis* and *M. tuberculosis* and a series of cross-linking and proteolysis experiments revealed that the RbpA binding site was located on the β-subunit (97,102). Precise mapping of the binding sites of RbpA and of its *M. smegmatis* homologue were attempted (97,103) adopting different strategies. RbpA from *M. smegmatis* was cross-linked with the β-subunit of the RNAP and the resulting RbpA:β-subunit adduct and the β-subunit were subjected to in-gel trypsin digestion, followed by mass spectroscopy analysis of the peptides obtained. The binding site of *M. smegmatis* RbpA on the β-subunit was then found by analyzing the

mass spectroscopy spectrum of the RbpA:β-subunit adduct for the absence of peaks present in the spectrum collected for the β-subunit (103). Otherwise, the *M. tuberculosis* RbpA binding site was mapped by proteolysis experiments. The single Cys residue of RbpA was covalently tethered to the chemical protease Fe(III) (*s*)-1-(*p*-bromoacetamidobenzyl) ethylenediamine tetra-acetic acid (FeBABE) (97). The resulting RbpA-FeBABE was incubated with the RNAP, only β-subunit cleavage products were detected and their sizes were determined based on the migration of the relative bands during SDS-PAGE analysis. The sizes found were used to calculate the position of cleavage sites on the β-subunit, which were attributed to the residues 523 ±10, 578 ±26 and 659 ±8. These residues are thought to represent the binding site of *M. tuberculosis* RbpA on the β-subunit (97). However, despite the fact that the β-subunit and RbpA in *M. tuberculosis* and in *M. smegmatis* are virtually identical (sequence identity > 90%), the putative binding sites found are different (figure 1.7).

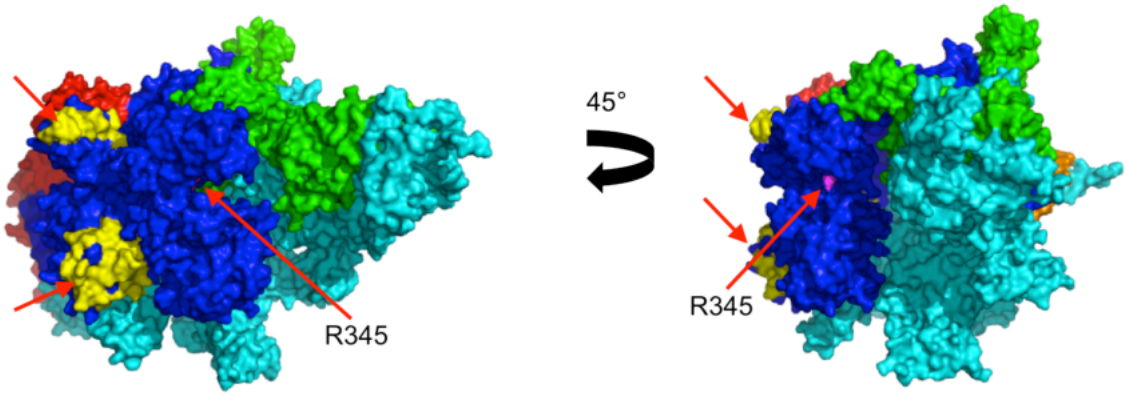


Figure 1.7. The RbpA binding site on the RNAP holoenzyme. The binding sites found for RbpA (97) and for the RbpA homologue in *M. smegmatis* (103) have been mapped on the *T. thermophilus* RNAP holoenzyme structure (PDB: 1IW7) and highlighted with different colours: yellow and magenta, respectively. For clarity the position of the binding sites of RbpA and of *M. smegmatis* RbpA (R345) are indicated by red arrows. The different subunits forming the RNAP holoenzyme have been coloured as follow: α in red, β in blue, β' in cyan, σ in green and ω in orange.

In *M. smegmatis* only one binding site was found for RbpA on the β-subunit R381 (*T. thermophilus* R345), located in the RNAP cleft close to the rifampicin binding site (103). Otherwise, for *M. tuberculosis* RbpA have been identified two binding sites on two protrusions of the β-subunit on the outer side of the RNAP. These binding sites are symmetric and consistent with a dimeric organization of RbpA (97), which is in agreement with the dimeric organization found by size exclusion chromatography (SEC) for the RbpA homologue in *S. coelicolor*. However, it should be noticed that the RbpA homologue in *M. smegmatis* has been found to be monomeric based on the elution volume detected during SEC analysis (103).

1.5.2 RbpA affects rifampicin susceptibility.

Early evidence that RbpA can affect the susceptibility to rifampicin was reported in 2006 studying the RbpA homologue in *S. coelicolor* (101). Newell *et al* demonstrated that *ΔrbpA* mutant strains of *S. coelicolor* had a higher sensitivity to rifampicin, up to ~15 fold (101). The RbpA homologue in *M. smegmatis* was also tested to investigate its role in rifampicin resistance. Dey *et al.* showed that *M. smegmatis* strains overexpressing the *M. smegmatis* RbpA homologue had a MIC for rifampicin of ~4 fold higher (102). Further studies, *in vitro* transcription assays, have demonstrated that RbpA and its homologues in *M. smegmatis* and in *S. coelicolor* can rescue the RNAP from the inhibitory effect of rifampicin (102). In details, a range of rifampicin concentrations were tested, from 1 to 100 μM, showing that transcription activity could be detected, at each inhibitor concentrations assessed, only when the *M. smegmatis* RbpA was added at the reaction mixture (97,101,102).

There is disagreement in the literature about the mechanism used by RbpA to rescue RNAP from rifampicin. Dey *et al.* propose that RbpA binding on the RNAP directly trigger the release of rifampicin (103). However, Hu *et al* argued that RbpA binding to RNAP does not affect the IC_{50} of rifampicin. Hence, they hypothesized that the stimulation of transcription induced by RbpA could account also for the higher transcription activity observed in the presence of rifampicin (97).

1.5.3 RbpA is part of the gene expression control network of *M. tuberculosis.*

The role of RbpA in the gene expression control network was hypothesized following the observation that the RbpA homologue in *S. coelicolor* can increase the RNAP transcription activity from a ribosomal promoter (101). This early evidence was confirmed in a recent work (97) where it was shown that RbpA activates transcription

also from non-ribosomal promoter and that RbpA activity is σ-factor specific. Hu *et al.*, performed a set of single-round transcription assays showing that RbpA increases the $E\sigma^A$ transcription activity of ~2-fold, but not of $E\sigma^F$ (97). Because no direct binding was observed at the σ-subunit Hu and colleagues concluded that the binding of RbpA on the β-subunit could induce a structural modification of the RNAP core increasing the affinity for σ^A and stabilizing the transcriptional competent open complex (97). Moreover, no influence of RbpA was observed on the rate of transcription elongation using promoter-less scaffold DNA, thus it was concluded that the function of RbpA is restricted to transcription initiation (97).

Rv2050 (the gene encoding for RbpA in *M. tuberculosis*) was predicted to be an essential gene for optimal growth of *M. tuberculosis* in a saturation mutagenesis study (98). This early observation was confirmed recently generating an *Rv2050* conditional mutant strain of *M. tuberculosis* where the expression of *Rv2050* is under the control of the antibiotic pristinamycin I (PI) (99). Forti *et al.* (99), showed that both in liquid and solid media the addition of PI is essential for the growth of the conditional mutant strain, concluding that *Rv2050* is an essential gene for *M. tuberculosis* viability. The literature data available suggests that RbpA function is linked to transcription regulation, but this does not explain why *Rv2050* expression is essential in *M. tuberculosis* (98,99). Hu *et al.*, formulated an interesting hypothesis about the RbpA essentiality in *M. tuberculosis* following the observation that σ^A and σ^F of *M. tuberculosis* were required at similar concentrations for *in vitro* transcription (97). They proposed that the relatively low affinity between σ^A and RNAP could lead to a reliance on RbpA to increase the σ^A competitiveness for the binding with the RNAP core. Thus, the transcription of housekeeping genes, particularly in times of stress when alternative σ-factors are expressed, would require the expression of RbpA (97). Interestingly, the RbpA

homologue in *S. coelicolor* is non-essential, although the gene-deletion mutant has a severe growth defect (101). This could potentially be due to differences in the biology between the organisms or to the presence in *S. coelicolor* of the *C54.03c* gene, the product of which is 29% identical to RbpA and could potentially fulfill a similar role in strains lacking RbpA.

1.6 Aims of the project

The essentiality of RbpA in *M. tuberculosis* (98,99) combined with the knowledge that it is conserved only in actinobacteria led to the proposal of RbpA as a potential target for anti-TB drug (97,100). This observation and the evidence that RbpA has a role in regulation of gene expression and in rifampicin resistance, make this protein an important subject for study (97,101-103).

When the research project described in this thesis began the available literature data showed that the RbpA binding site was located at the β-subunit of the RNAP, but unpublished yeast two-hybrid screens and protein fragment complementation assays in *E. coli* demonstrated that RbpA is a binding partner of σ^B (personal communication). The interaction between RbpA and the σ-subunit opened new scenarios in RbpA mechanism of action in transcription activation and in the RNAP-RbpA interaction mode. The aim of this project was to contribute to the ongoing effort to characterize RbpA performing a structural and functional study. The main focuses were to solve the structure of RbpA using nuclear magnetic resonance (NMR) techniques, to confirm and characterize the interaction between RbpA and the σ-subunit and to investigate the RbpA function using an *Rv2050* conditional mutant strain in *M. tuberculosis*.

Chapter 2

Expression, purification and structural characterization of *M. tuberculosis* RbpA

2.1 Introduction

When the research project described in this thesis began no structural information were available for RbpA or for its homologues. The knowledge of the spatial organization of a protein is potentially highly informative, it could be very important in the understanding of the protein function and mechanism of action. Moreover, atomic resolution structures can be used for drug design or to rationally project further experiments. With the aim to contribute at the characterization of RbpA it was decided to solve its structure by NMR.

In this chapter it will be described the process that led to generate structural information on RbpA. The first step was to optimize protocols that allowed the purification of RbpA. During this process it was noticed that RbpA had a strongly concentration dependent oligomeric organization. Bioinformatics analysis and preliminary NMR experiments led to hypothesize that the C-terminal region of RbpA could be responsible for the peculiar oligomeric organization of the protein. Truncation of the last 32 residues of RbpA abolished oligomerization of the protein and improved the quality of the NMR data collected. Therefore, it was chosen to solve the structure of the first 79 residues of RbpA ($RbpA_{1-79}$) that revealed the presence of a structured central domain with a conserved hydrophobic surface that may be a potential protein interaction site. Otherwise, the N- and C- termini of $RbpA_{1-79}$ are both highly dynamic and unstructured. The atomic coordinates of the central structured domain of $RbpA_{1-79}$ were used to perform a structural homologues search, several structures were found harbouring a

similar fold but all had a relatively low homology score and none of them can be consider a genuine structural homologue of $RbpA_{1-79}$.

2.2 Materials and methods.

2.2.1 Sequence alignment and secondary structure prediction of RbpA.

Evolutionary conserved regions of RbpA were identified analyzing a sequence alignment performed as follows. RbpA homologue sequences were identified with the BLAST algorithm (104) using the non-redundant protein sequences database and the default settings except for the maximum number of aligned sequences to display that was increased from 100 to 500. The process led to the identification of 416 sequences. The alignment was performed using the software COBALT (105) and with sequences from the *Actinobacteria* phylum. When multiple sequences were identified from the same genus only the sequence with the highest similarity score was selected. The resulting alignment was used to assign at each RbpA residue a conservation score, calculated using the AMAS method (106) implemented in Jalview (107). Secondary structure prediction of RbpA was performed using the Jpred 3 package (108).

2.2.2 Preparation of expression vectors.

The gene encoding RbpA (*Rv2050*) and the ORF encoding for its truncated version $RbpA_{1-79}$ were amplified by polymerase chain reaction (PCR) from *M. tuberculosis* H37Rv genomic DNA and cloned by a ligase independent method (using the in-house service Protex – University of Leicester) into an expression vector derived from pET-43.1a(+) (Novagen) encoding a cleavable His_6-tag at the N-terminus. All the constructs obtained were subjected to DNA sequencing at the University of Leicester Protein and Nuclear Acid Chemistry Laboratory (PNACL).

2.2.3 Protein analysis by SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

SDS-PAGE analyses were performed in order to control the status of the proteins at different stages of the purification process. Typically, samples were prepared mixing 20 μl of protein sample with 10 μl of 0.2 mM 1,4,-dithiothreitol (DTT) and 10 μL of 4x NuPAGE® LDS sample buffer (Invitrogen). Samples were heated at 70 °C for 10 minutes and loaded on precast NuPAGE® Bis-Tris gels with a 4-12% acrylamide gradient (Invitrogen). At least one lane in each gel was loaded with 3 μl of Novex® sharp protein standard (Invitrogen) that contain a mixture of 12 protein bands in the range of 3.5-260 kDa or with 3 μl of protein marker, broad range (New England Biolabs) that contain a mixture of 8 protein bands in the range of 7-175 kDa. The electrophoresis run was performed in 2-(*N*-morpholino)ethanesulfonic acid-SDS (MES-SDS) running buffer (Invitrogen) with a constant voltage of 200 V for 35 minutes. The gel was stained with Instant*Blue*™ (Expedeon) for 1 hour with gentle rocking then destained in deionised water (dH_2O).

2.2.4 Expression of unlabelled and ^{15}N,^{13}C,$^{15}N/^{13}C$ labelled RbpA and $RbpA_{1-79}$.

E. coli BL21(DE3) cells, transformed with the relevant expression vector, were grown in Luria-Bertani media (LB, appendix) containing 100 μg/ml ampicillin at 37 °C with shaking to an absorbance at 600 nm of ~0.7. Protein expression was then induced by the addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and the cell growth continued at 30 °C for 4 hours. Cells were harvested by centrifugation (8000 *g* for 20 minutes at 4 °C) and the pellet was stored at -20 °C.

To uniformly label RbpA with ^{15}N, ^{13}C or $^{15}N/^{13}C$, *E. coli* BL21(DE3) cells transformed with the relevant expression vector were grown in modified Spizizen minimal media (109) (appendix) containing 100 μg/ml ampicillin, where the sole source of nitrogen

and carbon were 1 g/l of ^{15}N-$(NH_4)_2SO_4$ and 2 g/l of ^{13}C-D-(+)-glucose as appropriate. The same expression protocol was applied to uniformly label $RbpA_{1-79}$, except for $^{15}N/^{13}C$ labelled $RbpA_{1-79}$, which was prepared using non-isotopically labelled aromatic amino acids (His, Tyr, Trp and Phe at 50 mg/l) in the minimal media. The expression of the labelled proteins was performed as described for unlabelled.

2.2.5 Expression of $^{15}N/^{13}C/^{2}H$ labelled RbpA.

E. coli BL21(DE3) cells transformed with the relevant expression vector were subjected to double selection of high-level expression colonies as described in Sivashanmugam *et al.* (110). The colony that gave the best protein expression was used to inoculate 50 ml of 2YT media (appendix) in D_2O (containing 100 μg/ml of ampicillin) and the culture was grown at 37 °C with shaking for 20 hours. The cells were harvested by centrifugation (8000 *g* for 20 minutes at 4 °C), resuspended in optimized high-cell-density-IPTG-induction minimal media (appendix) (110) and the culture growth was continued at 37 °C until the absorbance at 600 nm was increased by at least one unit. The protein expression was then induced by adding 0.25 mM IPTG and the culture growth was continued for 5 hours at 30 °C. Cells were harvested by centrifugation (8000 *g* for 20 minutes at 4 °C) and the pellet was stored at -20 °C.

2.2.6 Purification of RbpA and $RbpA_{1-79}$.

Cell pellet was resuspended in a buffer containing 50 mM Tris-HCl pH 8.0, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) and 0.1 mg/ml lysozyme from chicken egg white (Sigma). Cells were sonicated in ice using Misonix sonicator 3000 with power set on 7 for 10 minutes with 20 seconds on-off intermittency. The resulting cell lysate was

centrifuged (15,000 *g* for 30 minutes at 4 °C). The supernatant was then loaded on a Ni-Sepharose 5 ml column (GE Healthcare) pre-equilibrated with a buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 30 mM imidazole. The same buffer was used to wash the column (40 ml) and to elute the His_6-tagged protein with a linear imidazole gradient (30-500 mM, 50 ml). Fractions containing the His_6-tagged protein were subjected to SEC using a Superdex 75 16/60 (GE Healthcare) column and a buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 2 mM DTT. The same buffer was used to calibrate the SEC column using the low molecular weights calibration kit (GE Healthcare). The concentration of the purified protein was calculated measuring the absorbance at 280 nm.

2.2.7 His_6-tag cleavage

The His_6-tag was removed from the His_6-tagged proteins by enzymatic cleavage using His_6-TEV (tobacco etch virus) protease (provided by PROTEX, University of Leicester). The reaction was conducted at 4 °C, over-night and using 2 units of His_6-TEV protease for 100 μg of His_6-tagged protein. Successful His_6-tag cleavage resulted in complete removal of the tag leaving a single Ser residue. His_6-TEV protease and the His_6-tag were removed from the reaction mixture using a 5 ml Ni-NTA column (QIAGEN).

2.2.8 Analytical size exclusion chromatography (SEC)

The oligomeric organisation of RbpA was assessed by analytical SEC using a Superdex 75 10/300 GL column (GE Healthcare) and a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA and NaN_3 0.02% w/v. The SEC column was previously calibrated using the low molecular weights calibration kit (GE Healthcare).

2.2.9 NMR spectroscopy

NMR spectra were acquired from 0.35 ml samples of RbpA (~0.2 mM) and $RbpA_{1-79}$ (~0.35 mM) dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM Tris(2-carboxyethyl) phosphine (TCEP) with 10% D_2O-90% H_2O or 100% D_2O as appropriate. All NMR data were recorded at 35 °C using either 600 MHz Bruker Avance/DRX systems, or an 800 MHz Bruker Avance II spectrometer fitted with cryogenically-cooled probe heads.

The 2D and 3D spectra recorded to obtain sequence specific assignments for RbpA were: ^{1}H-^{1}H TOCSY (111) with mixing times of 35 ms and 55 ms, ^{1}H-^{1}H NOESY (112) with an NOE mixing time of 80 ms, ^{15}N/^{1}H TROSY (113), ^{15}N/^{1}H NOESY-TROSY with an NOE mixing time of 50 ms (114), ^{13}C/^{1}H HC(C)H-TOCSY and (H)CCH-TOCSY (115) with mixing times of 18 ms, ^{15}N/^{13}C/^{1}H HBHA(CO)NH (116), TROSY-HNCO (117), TROSY-HNCA (117), TROSY-HN(CO)CA (118), CBCA(CO)NH (119), and TROSY-HNCACB (120). A uniformly ^{15}N/^{13}C/^{2}H labelled sample of RbpA was used to acquire the ^{15}N/^{1}H TROSY and ^{15}N/^{13}C/^{1}H TROSY-HNCACB spectra. Typical acquisition times in the 2D experiments were: 60 ms (^{15}N) and 19 ms (^{1}H) in F_1 and 65-72 ms (^{1}H) in F_2. Typical acquisition times in the 3D experiments were: 4-9 ms (^{13}C), 9-18 ms (^{15}N) in F_1 and F_2 and 65-83 ms in F_3 (^{1}H).

The 2D and 3D spectra recorded to obtain sequence specific assignments and ^{1}H-^{1}H distance constraints for $RbpA_{1-79}$ were: ^{1}H-^{1}H TOCSY (111) with mixing times of 35 and 55 ms, ^{15}N/^{1}H HSQC (113), TOCSY-HSQC (121) with a mixing time of 60 ms, ^{15}N/^{13}C/^{1}H HNCACB (120) , HNCO (117), HNCACB (120) ^{1}H-^{1}H NOESY (112) and ^{13}C/^{1}H NOESY-HSQC (122), ^{15}N/^{1}H NOESY-HSQC (114) with NOE mixing times of

200 ms. An $^{15}N/^{13}C$ labelled sample of $RbpA_{1-79}$ with non-isotopically labelled aromatic amino acids (His, Tyr, Trp and Phe) was used to acquire the $^{15}N/^{13}C/^{1}H$ HNCACB, HNCO and $^{13}C/^{1}H$ NOESY-HSQC spectra. Typical acquisition times in the 2D experiments were: 60 ms (^{15}N), 35-39 ms (^{1}H) in F_1 and 80-120 ms (^{1}H) in F_2. Typical acquisition times in the 3D experiments were: 9-12 ms (^{13}C), 19-34 ms (^{15}N) and 15-18 ms (^{1}H) in F_1 and F_2 and 70-100 ms (^{1}H) in F_3.

The WATERGATE method (123) was used to suppress the water signal when required. The NMR data collected were processed using the program Topspin (Bruker BioSpin Ltd.) with linear prediction used to extend the effective acquisition time up to twofold in F_1 and F_2. The spectra were analysed with the Sparky package (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

The chemical shifts referencing was performed following the guidelines described by Wishart *et al.*(124) which have been implemented in the program Topspin. Briefly, the primary ^{1}H chemical shift standard is 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). The ^{13}C and ^{15}N chemical shifts are defined in terms of the frequency ratios $^{13}C/^{1}H$ = 0.251449530 and $^{15}N/^{1}H$=0.101329118, where the ^{1}H frequency is that of DSS in water.

2.2.10 Sequence-specific assignment.

Sequence-specific backbone assignments were obtained for RbpA from the identification of intra- and inter-residue connectivities in a series of double and triple resonance spectra: $^{15}N/^{1}H$ TROSY (113), NOESY-TROSY (114), $^{15}N/^{13}C/^{1}H$ HBHA(CO)NH (116), TROSY-HNCO (117), TROSY-HNCA (117), TROSY-HN(CO)CA (118), CBCA(CO)NH (119) and TROSY-HNCACB (120). Assignments were then extended to the side chains using correlations observed in: ^{1}H-^{1}H TOCSY (111), NOESY (112), $^{13}C/^{1}H$ HC(C)H-TOCSY, (H)CCH-TOCSY (115) and NOESY-

TROSY (114). Initially, tentative assignments for $RbpA_{1-79}$ were obtained by transposing the assignment from RbpA onto the ^{1}H-^{1}H TOCSY (111), ^{1}H-^{1}H NOESY (112), $^{15}N/^{1}H$ HSQC (113) and TOCSY-HSQC (121) spectra. The inter-residue amide nitrogen/proton to Cα and Cβ connectivities observed in the HNCACB (120) and the NH to NH and CH to CH NOEs identified in the ^{15}N- and ^{13}C-edited NOESY spectra (114,122) were used to confirm the assignment.

2.2.11 Far UV circular dichroism (CD) spectroscopy.

CD spectra were acquired from RbpA or $RbpA_{1-79}$ samples (0.25 mg/ml) dissolved in a buffer containing 5 mM K_2HPO_4 pH 6.5 and 50 mM Na_2SO_4 at a temperature of 20 °C using a Chirascan™ CD spectrometer. Each spectrum represents the average of 4 accumulations recorded from 190 to 260 nm with a bandwidth of 1 nm, a sampling time of 0.5 second per step and using a cuvette with a path length of 0.5 mm. CD spectra were deconvoluted using the CD spectra deconvolution software Version 2.1 (CDNN 2.1) written by Dr. Gerald Böhm.

2.2.12 Secondary structure determination of $RbpA_{1-79}$.

The primary sequence of $RbpA_{1-79}$ and the chemical shifts of the Cα, Cβ, CO, Hα, NH and N atoms were used to determine the secondary structure of $RbpA_{1-79}$ using the Torsion Angle Likeliness Obtained from Shift and Sequence Similarity + (TALOS +) (125).

2.2.13 Structure calculation

The program Cyana (126) was used to calculate the family of converged structures of $RbpA_{1-79}$ in a two-stage process. Initially, the combined automated NOE assignment and structure determination protocol (CANDID) was used to automatically assign NOE cross-peaks and to generate a preliminary family of converged structures. The inputs used for this first step were: ^{15}N, ^{13}C, and ^{1}H resonance assignments, a set of 72 backbone torsion angle constraints, determined by program TALOS+ (125), and three lists of manually picked NOE cross-peaks that were identified in a 2D NOESY spectrum (399 peaks) and in 3D ^{15}N- and a ^{13}C-edited NOESY spectra (881 and 1049 peaks, respectively). The CANDID calculation was carried out using the standard parameters of Cyana but with a chemical shift tolerance set at 0.03 ppm for ^{1}H and at 0.4 ppm for ^{15}N and ^{13}C. In the second stage, the final family of converged $RbpA_{1-79}$ structures was produced through several cycles of simulated annealing combined with redundant dihedral angle constraints (REDAC) (127) resulting in 98 structures with no distance violations greater than 0.5 Å and no dihedral angle violations greater than 5°. The structures obtained were then refined by Dr. Fred Muskett in a generalised Born solvent model (128) using the AMBER 10 package (129) as described previously (130), but with one cycle of restrained molecular dynamics simulated annealing. The 35 structures with the lowest AMBER energy were selected and analyzed using MOLMOL (131) and CING (132)

2.2.14 Comparative analysis between RbpA and its structurally homologous proteins.

The coordinates of the structure of $RbpA_{1-79}$ structured central region (residues 26-66) were used as input in the software DaliLite V. 3 (133) to search for RbpA structural homologues between all the structures deposited in the PDB. The software performs a structural homology search using the distance-matrix alignment (DALI) algorithm (134). The algorithm transforms 3D protein structures into 2D matrices that contain all the pairwise distances between Cα atoms and structural homologues are found identifying matrices with similar Cα-Cα distances pattern.

2.3 Results and discussion.

2.3.1 Secondary structure prediction and sequence alignment for RbpA.

Most of the RbpA sequence is predicted to lack secondary structure except for 3 β-strands in the centre of the sequence (β_1 residues 28-33, β_2 39-45 and β_3 62-64) and 2 α-helices at the C-terminus (α_1 residues 81-87 and α_2 91-108) (figure 2.1 panel A).

A helical wheel projection shows that the predicted α-helix between residues 91-108 has a marked amphipathic character (figure 2.1 panel B). The RbpA homologue of *M. smegmatis* and *S. coelicolor* are the only RbpA homologues for which there are published studies (96,101-103). Interestingly, secondary structure prediction performed for these homologues (results not shown) reveals that, similarly to RbpA, an α-helix is likely to be formed between residues 91-108 of RbpA homologue in *M. smegmatis* and between residues 92-107 of RbpA homologue in *S. coelicolor*. In both proteins the amphipathic character of the α-helix is maintained (figure 2.1 panel B).

The alignment of RbpA homologous sequences across the *Actinobacteria* phylum reveals a cluster of residues (79-89) at the C-terminus highly conserved, notably 9 out of 11 residues of this cluster are invariant across all the sequences analyzed. The first 20 residues of RbpA are also relatively conserved, otherwise the rest of the sequence does not show evident conservation patches (figure 2.1 panel C). A set of aligned sequences, representative of the multi sequence alignment performed, is shown in figure 2.1 panel D.

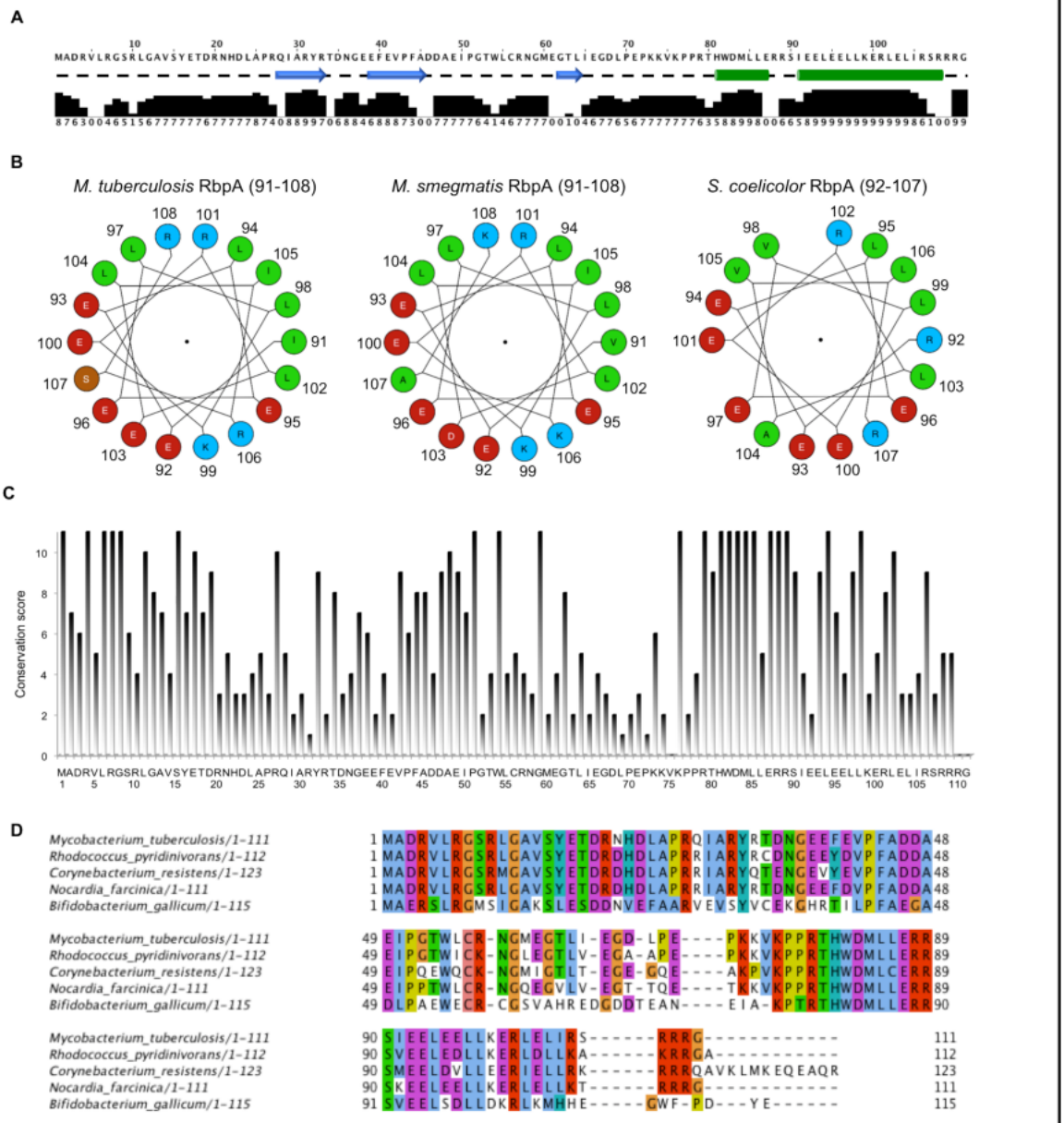


Figure 2.1. Bioinformatics analysis of RbpA. A. The sequence of RbpA and its secondary structure prediction, performed using JPred 3 (108), with the relative prediction confidence underneath (0-9). Cyan arrows represent β-strands and green cylinders α-helices, a dashed line indicates lack of secondary structure. **B.** The helical wheel projections of: *M. tuberculosis* RbpA residues 91-108 (IEELEELLKERLELIRSR), *M. smegmatis* RbpA residues 91-108 (VEELEELLKERLDLIKAK) and *S. coelicolor* RbpA residues 92-107

(REELEEVLEERLAVLR). The following colour code is used to highlight differently charged residues: red for acidic residues, blue for basic residues, green for non-polar residues and orange for polar, uncharged residues. **C.** The histogram shows the conservation score (calculated using the AMAS method (106) implemented in Jalview (107)) assigned to RbpA residues. A score of 11 indicates that the relative residue is invariant across all the sequences analyzed and a score of 10 indicates that the properties of the relative residue are maintained. **D.** Shows a set of aligned sequences that is representative of the multi sequence alignment reported in panel C. The Clustal x Colour Scheme implemented in Jalview (107) was used to highlight conserved features across the alignment.

2.3.2 Expression and purification of His_6-RbpA.

Expression of His_6-RbpA was induced from *E. coli* BL21(DE3) cells transformed with the relevant expression vector adding IPTG and the culture growth was continued for 4 hours at 30 °C. Culture samples collected pre-induction and 4 hours post-induction were lysed (section 2.2.6) and subjected to SDS-PAGE analysis. The resulting gel (figure 2.2) shows that an overexpressed band is produced following induction. This band is partially soluble and migrates in a position consistent with the calculated molecular weight of His_6-RbpA.

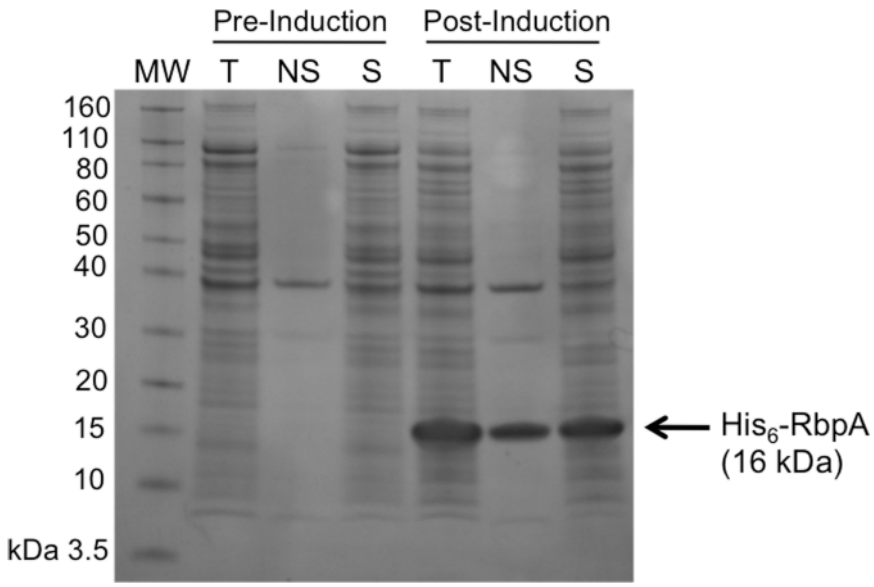


Figure 2.2. His_6-RbpA expression. SDS-PAGE analysis of culture samples collected pre-induction and 4 hours post-induction of His_6-RbpA expression. Lane MW is molecular weights markers, lanes T the total cell lysate, lanes NS the non-soluble fraction and lanes S soluble fractions.

The protocol adopted for the purification of His_6-RbpA involved two main steps: immobilized-metal ion affinity chromatography (IMAC) and SEC. In the first step the

soluble fraction resulting from the lysis of the cell culture expressing His_6-RbpA was loaded on a Ni-Sepharose column (GE Healthcare), subsequently His_6-RbpA was eluted with a linear imidazole gradient. SDS-PAGE analysis of samples collected during different stages of the chromatography reveals that the fractions under the main elution peak (elution volume ~104-108 ml) contain a high quantity of relatively pure His_6-RbpA (figure 2.3).

Typically, fractions containing pure His_6-RbpA were pooled together and subjected to a TEV cleavage reaction over-night to remove the His_6-tag. The second step of purification was performed the following day loading the reaction mixture on a SEC column (Superdex 75 16/60, GE Healthcare). In the resulting chromatogram were detected two main peaks with elution volumes of 46 ml (peak A) and 67 ml (peak B) that correspond to apparent molecular weights of >75 kDa (outside the range of resolution of the column, which is 6.5-75 kDa) and ~30 kDa, respectively (figure 2.4 panel A).

RbpA and its homologue in *S. coelicolor* were proposed to be homodimers (97,101), but not the RbpA homologue in *M. smegmatis* that was shown to be monomeric (103).The elution profile obtained during the SEC analysis of RbpA (figure 2.4 panel A) is consistent with a dimeric organization of the protein. In fact, the theoretical molecular weight calculated for RbpA is 13 kDa and most of RbpA is eluted at 67 ml (figure 2.4 panel A) that corresponds to an apparent molecular weight of ~30 kDa. However, approximately 30% of the RbpA sample loaded on the SEC column gave rise to a distinct oligomeric species with a higher apparent molecular weight (> 75 kDa). Interestingly, this high molecular weight oligomeric species is generated by His_6-RbpA and only in minor part (< 10%) by RbpA (figure 2.4 lane 46). Indeed, SDS-PAGE analysis of the fractions collected during the SEC run shows that the His_6-tag cleavage

occurred only for ~70% of the protein (figure 2.4 panel B, lane L). The protein that retained the His_6-tag is eluted under peak A (apparent molecular weight > 75 kDa), suggesting that is probably aggregated (figure 2.4 panel A). In order to investigate if the His_6-tag influences the oligomeric organization of RbpA SEC analyses were performed also on His_6-RbpA samples. The elution profiles obtained (results not shown) were characterized by the presence of two distinct peaks with elution volumes comparable to the ones found for peaks A and B of the chromatogram reported in figure 2.4 panel A, leading to the conclusion that the His_6-tag does not affect the oligomeric status of RbpA. Thus, it was hypothesized that a fraction of His_6-RbpA purified by IMAC underwent aggregation. This would cause a reduction of the His_6-tag cleavage reaction efficiency. In fact, protein aggregation could limit the access of the TEV protease at the cleavage site in this fraction of the RbpA sample.

Aggregate RbpA was discarded, only the fraction of the sample eluted inside the range of resolution of the column was kept and used for further experiments, resulting in a purification yield of ~12 mg for each litre of initial *E. coli* culture. Comparable yields were obtained for the purification of ^{15}N, ^{13}C, $^{15}N/^{13}C$ or $^{15}N/^{13}C/^{2}H$ labelled RbpA.

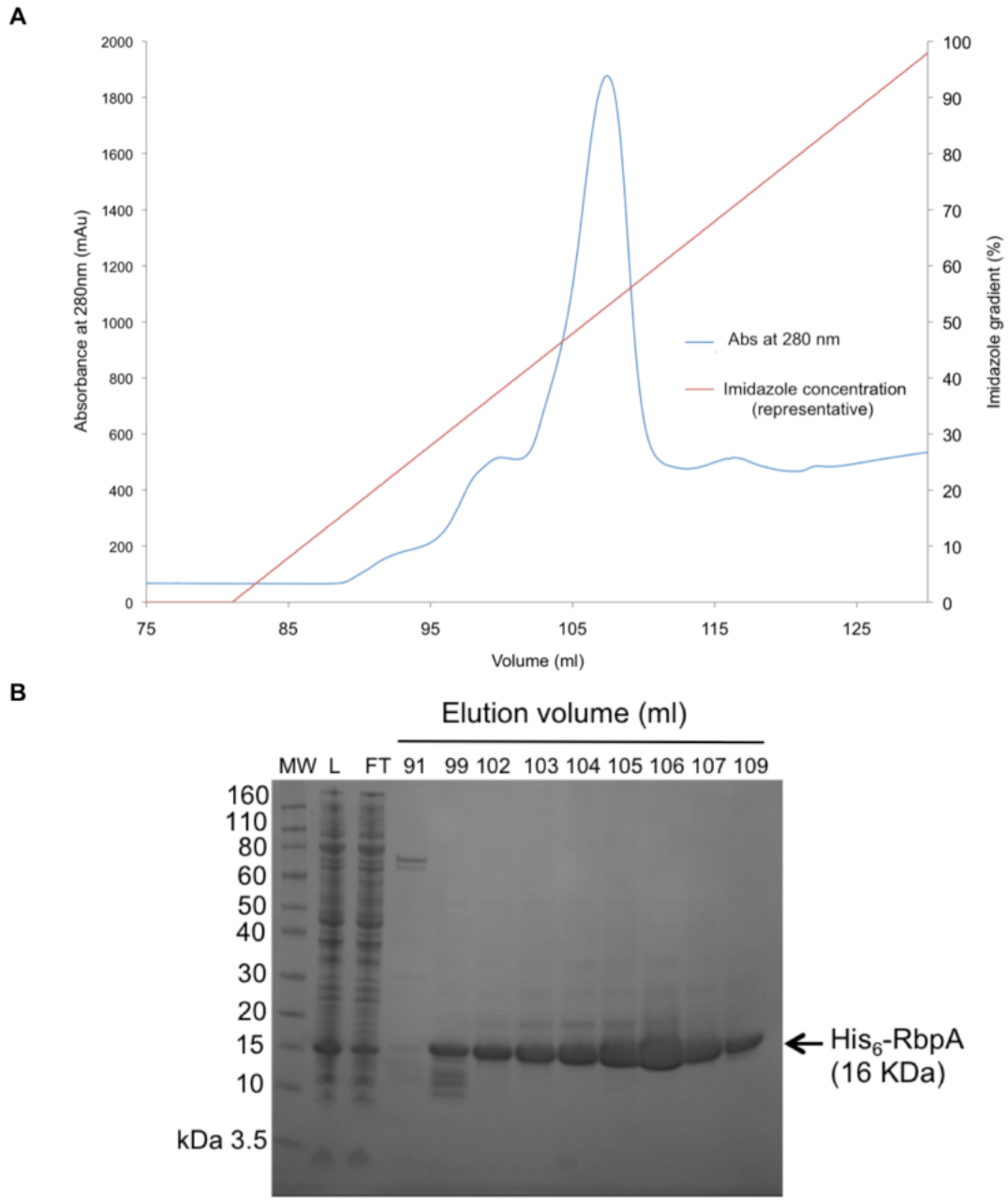


Figure 2.3. His_6-RbpA purification by IMAC. A. The typical elution profile obtained during the purification of His_6-RbpA by IMAC. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process is shown. Lane MW contains the molecular weight markers, lane L the column load, lane FT the column flow-through and lanes 91-109 are labelled according to the elution volume (ml).

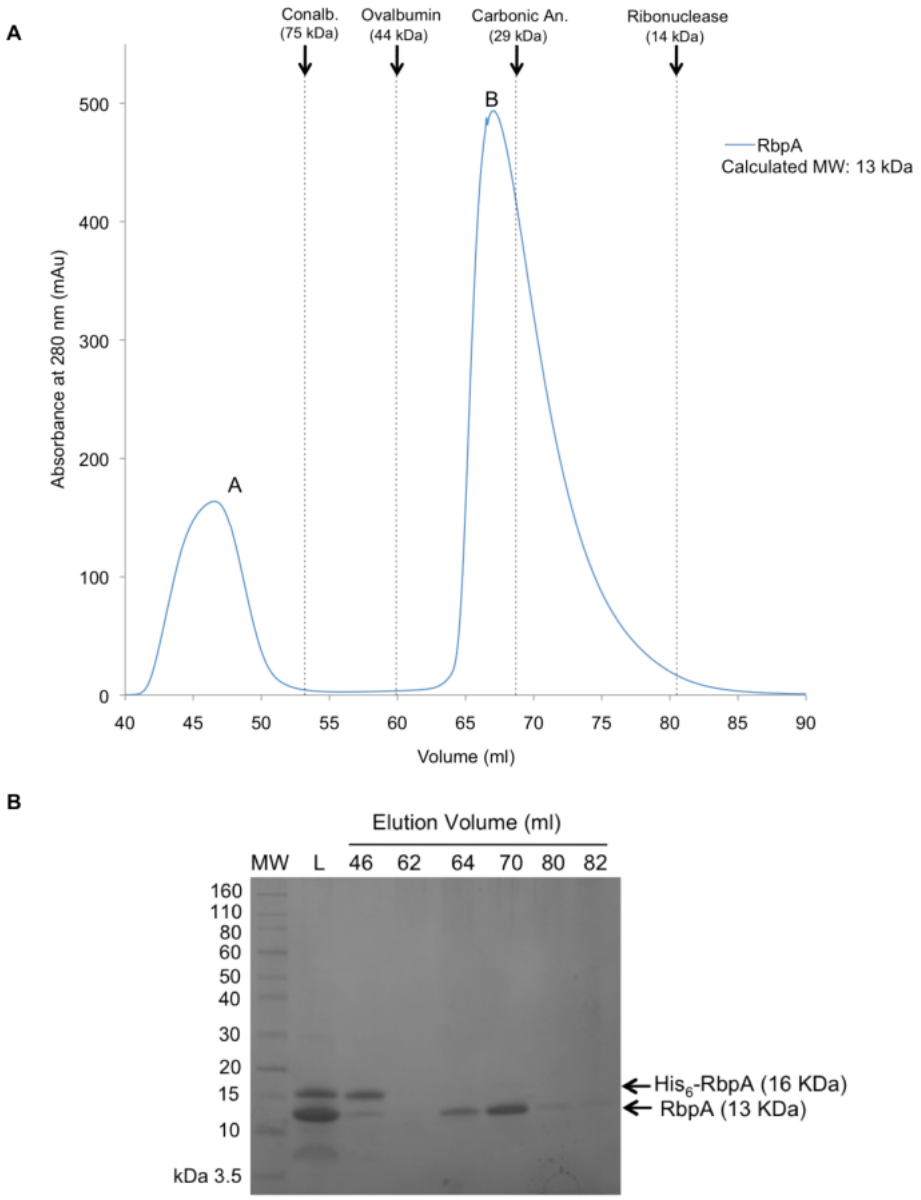


Figure 2.4. RbpA purification by SEC chromatography. A. A typical elution profile obtained during the purification of RbpA by SEC (Superdex 75 16/60). In the chromatogram two main peaks are detected and are labelled with the letters A and B and the elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process is shown. Lane MW is the molecular weight markers, lane L the column load and lanes 46-82 are labelled according to the elution volume (ml).

2.3.3 Characterization of the RbpA oligomeric organization by analytical SEC.

From the comparison of the SEC chromatograms recorded for different RbpA purifications it was noticed that the elution volumes obtained were affected by the concentration of the sample analyzed. In order to investigate the relationship between protein concentration and the oligomeric status of RbpA a series of samples with a concentration range from 3 to 200 µM were analysed by analytical SEC (figure 2.5).

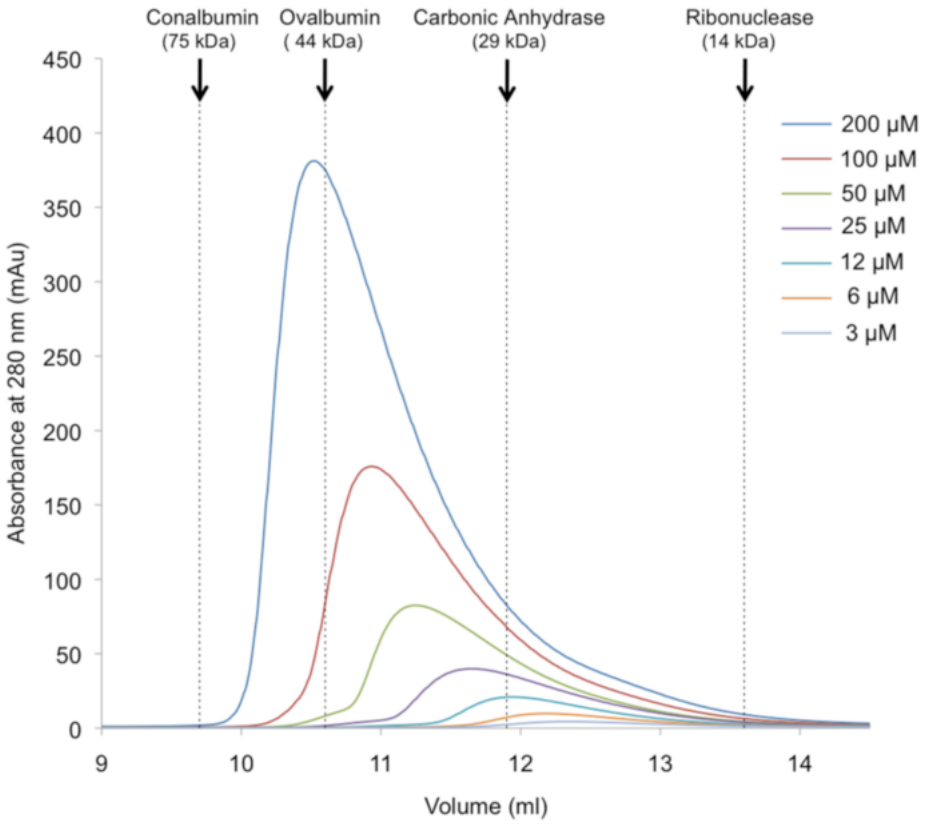


Figure 2.5. Overlay of the SEC elution profiles of RbpA samples loaded at different concentrations. Samples of RbpA were loaded at different concentrations onto an analytical SEC column (Superdex 75 10/300). The elution volumes of protein molecular weight standards used to calibrate the column are indicated by dashed lines.

The resulting chromatograms reveal asymmetric peaks with an elution volume strongly dependent on the concentration of the sample. These observations suggest that there is a dynamic and concentration dependent equilibrium between different oligomeric species of RbpA.

There are conflicting data on the likely oligomerization state of RbpA and its homologues. The retention during SEC was the reason that RbpA (97) and its homologue from *S. coelicolor* (101) were proposed to be homodimers, whereas the same method led Dey *et al*. (103), to propose that the homologue from *M. smegmatis* is monomeric (103). Because of the high amino acid identity these homologues share with RbpA (55% of sequence identity for *S. coelicolor* RbpA and 92% for *M. smegmatis* RbpA), it is unlikely that the oligomeric organization is not maintained in different organisms. Rather, the finding that RbpA can self-oligomerize in a concentration dependent manner (figure 2.5) could account for the inconsistent data found in the literature (97,101,103).

2.3.4 Sequence-specific assignment of RbpA.

The procedure described previously (section 2.3.2) allowed the production of RbpA samples suitable for NMR analyses. Figure 2.6 shows a 1D ^{1}H NMR spectrum recorded for RbpA, which displays some of the key features of a folded protein like a significant dispersion of the signals from backbone NH groups (6.5-9.5 ppm) and several high field shifted methyl resonances (-0.3-0.5 ppm).

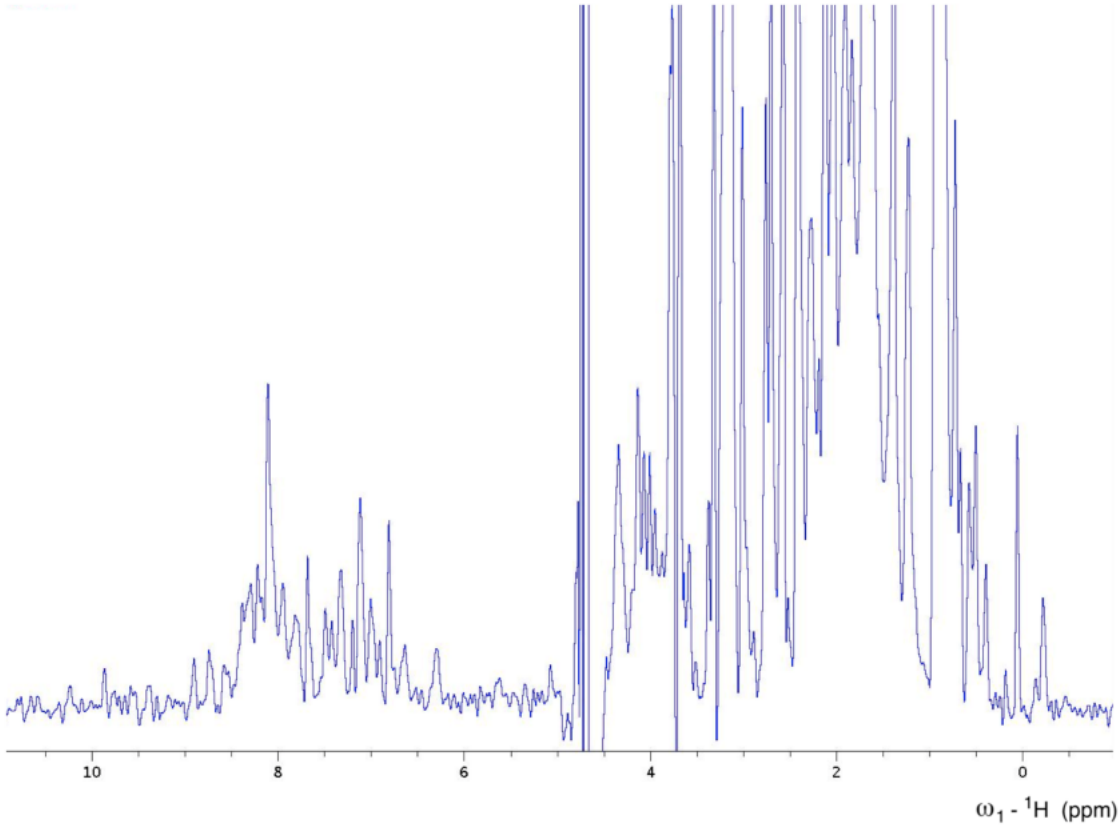


Figure 2.6. 1D ^{1}H-NMR spectrum of RbpA. The spectrum shown results from four scans and was acquired at 35 °C from an RbpA sample (0.2 mM) dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O.

Isotopically labelled samples of RbpA were prepared and used to record a set of 2D and 3D NMR spectra required to make the backbone assignments to the protein. In figure 2.7 is shown an ^{15}N/^{1}H HSQC spectrum recorded on ^{15}N RbpA, it shows relatively wide dispersion of the signals except for the centre of the spectrum where a number of peaks are overlapped. The spectrum was expected to show 104 backbone NH peaks because RbpA contains 112 residues, including the extra Ser residue retained at the N-terminus after the His_6-tag cleavage, but the 7 Pro residues and the Ser residue at the N-terminus

will not be detected. However, only 87 of the peaks detected in the $^{15}N/^{1}H$ HSQC spectrum were found to belong to backbone NH groups. Thus, it was concluded that 17 backbone NH groups did not generate clearly detectable NMR signals.

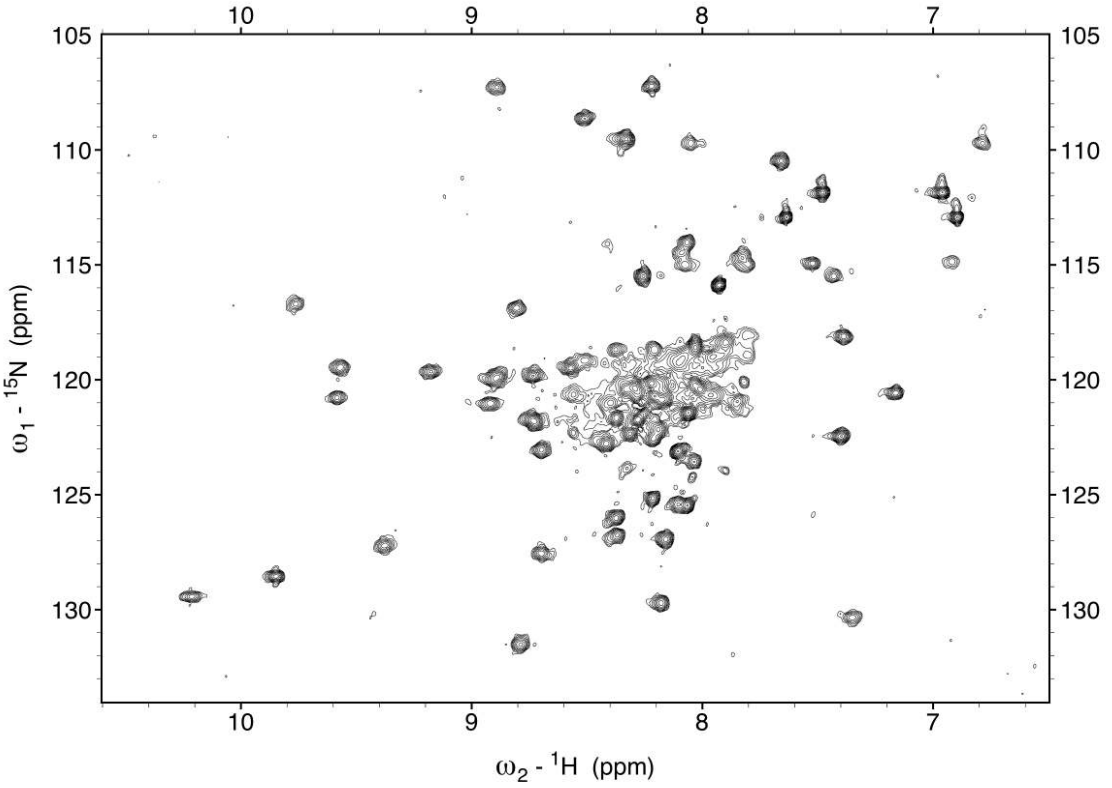


Figure 2.7. 2D $^{15}N/^{1}H$ HSQC spectrum of RbpA. The spectrum reported was acquired at 35 °C from an ^{15}N RbpA sample (0.2 mM) dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O.

The relatively high concentration of RbpA used in NMR experiments, typically ~200 μM, induces homoligomerization of the protein (figure 2.5) that ultimately causes a reduction in the signal/noise ratio and in resolution (due to peak broadening) of the

NMR spectra collected. These problems were partially overcome by preparing deuterated samples of RbpA that improved significantly the quality of the NMR data recorded. In figure 2.8 are shown NMR spectra recorded on $^{15}N/^{13}C$ RbpA (panel A) or on $^{15}N/^{13}C/^{2}H$ RbpA (panel B). The comparison of these two spectra highlight how deuteration of the protein led to a significant improvement of the signal/noise ratio of the NMR data collected.

Near complete assignments of backbone (N, NH, Hα, Cα, Hβ and Cβ) resonances were made for residues 1-76 except: Met_1 (N and NH), Arg_7 (Hα, Hβ2 and Hβ3), Gly_8 (N and NH), Arg_{10} (N and NH), Glu_{61} (Hα, Hβ2 and Hβ3), Glu_{66} (Hα), Glu_{71} (Hα, Hβ2 and Hβ3) and Lys_{73} (Hα). However, about half of the residues within the region 77-111 do not generate observable NMR signals. This prevented unequivocal assignment of the last 35 residues of RbpA. The lack of backbone NMR signals for this region implies that it samples a number of discrete structural states on a second to millisecond timescale.

Assignments were obtained also for the side chain resonances of residues 1-76 except: Arg_4 (Nε and Hε), Arg_7, Arg_{10} (Hγ2, Hγ3, Hε and Nε), His_{22}, Arg_{27} (Hε and Nε), Asn_{36} (Hδ21, Hδ22 and Nη2), Arg_{57} (Hε and Nε), Trp_{54} (Hζ3 and Hη2), Glu_{61}, Glu_{66}, Glu_{71} and Val_{75} (Cγ1, Cγ2, Hγ1 and Hγ2). For residues 1-76, the extent of the assignment is 94.1% of all the aliphatic ^{13}C resonances, 88.1% and 91.5% of all the ^{15}N and ^{1}H resonances, respectively.

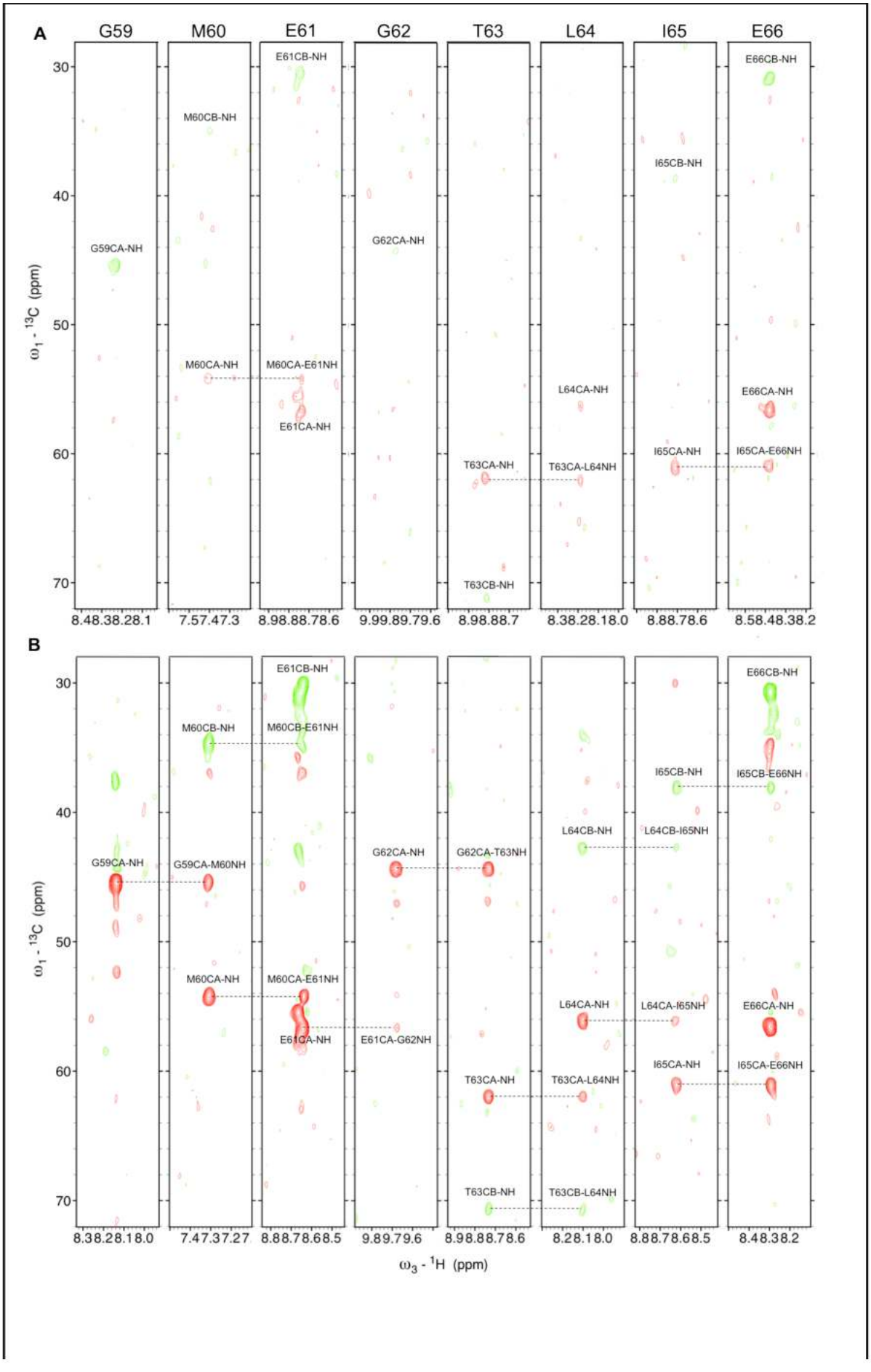

A
G59
M60
E61
G62
T63
L64
I65
E66
E61CB-NH
M60CB-NH
G59CA-NH
G62CA-NH
I65CB-NH
E66CB-NH
M60CA-NH
M60CA-E61NH
E61CA-NH
L64CA-NH
E66CA-NH
T63CA-NH
T63CA-L64NH
I65CA-NH
I65CA-E66NH
T63CB-NH
B
M60CB-E61NH
G59CA-M60NH
G62CA-T63NH
I65CB-E66NH
L64CB-NH
L64CB-I65NH
E61CA-G62NH
L64CA-I65NH
T63CB-L64NH
ω1 - 13C (ppm)
ω3 - 1H (ppm)

Figure 2.8. ^{15}N strips for the residues 59-66 from the CBCANH and the TROSY-HNCACB spectra recorded for RbpA. **A.** Shows ^{15}N strips for the residues 59-66 from the CBCANH spectrum acquired at 35 °C from an ^{15}N/^{13}C RbpA (0.2 mM) sample dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O. Dashed lines indicate sequential connections of Cα and Cβ. **B.** Shows ^{15}N strips for the residues 59-66 from the TROSY-HNCACB spectrum acquired in the same conditions described in panel A on an ^{15}N/^{13}C/^{2}H RbpA (0.2 mM) sample.

2.3.5 The role of the C-terminal region (residues 80-111) of RbpA in oligomerization.

Conformational exchange occurring with a frequency comparable with the NMR time-scale will lead to signal broadening and to reduction in the signal to noise ratio. The poor quality of the NMR data collected for RbpA residues 77-111, which prevented their assignment, could be due to conformational exchange occurring in this region of the protein. It was hypothesized that the hydrophobic side of the amphipathic α-helix, which is likely to be formed between residues 91-108 (figure 2.1 panel B), could interact with the equivalent face on a second RbpA molecule leading to homo-oligomerization and ultimately to conformational exchange of the C-terminal region. Indeed, the elongated shape of the peaks detected during SEC analyses of RbpA (peak B in 2.4 and figure 2.5) suggests that there might be a constant interchange between different oligomerization species. In order to investigate the role of the C-terminal region in oligomerization, a shorter version of RbpA was expressed and purified,

termed $RbpA_{1-79}$, which consists of residues 1-79 and therefore lacks the predicted amphipathic helix.

$RbpA_{1-79}$ was cloned in to a vector that allowed the expression of a recombinant version of the protein in fusion with His_6-tag at its N-terminus. His_6-$RbpA_{1-79}$ was expressed and purified following the procedure adopted for the full-length version of the protein (section 2.3.2). Following the IMAC purification step and the His_6-tag cleavage, $RbpA_{1-79}$ was loaded on a SEC column in order to increase the purity of the protein preparation and to investigate the oligomeric organization of $RbpA_{1-79}$. The resulting chromatogram (figure 2.9 panel A) shows a single peak with an elution volume of 81 ml that corresponds to an apparent molecular weight of ~13 kDa, which is consistent with a monomeric organization of the protein ($RbpA_{1-79}$ calculated molecular weight is 9 kDa). Interestingly, the His_6-tag was completely removed from $RbpA_{1-79}$ during the TEV cleavage reaction. This is evident from the comparison between lanes A and L of the gel reported in figure 2.9 panel B which contain samples of the protein before and after the His_6-tag cleavage reaction, respectively. This is in contrast with the partial His_6-tag cleavage observed for RbpA (figure 2.4 lane L) suggesting that the homogenous monomeric status of $RbpA_{1-79}$ facilitates the access of TEV protease at its cleavage site.

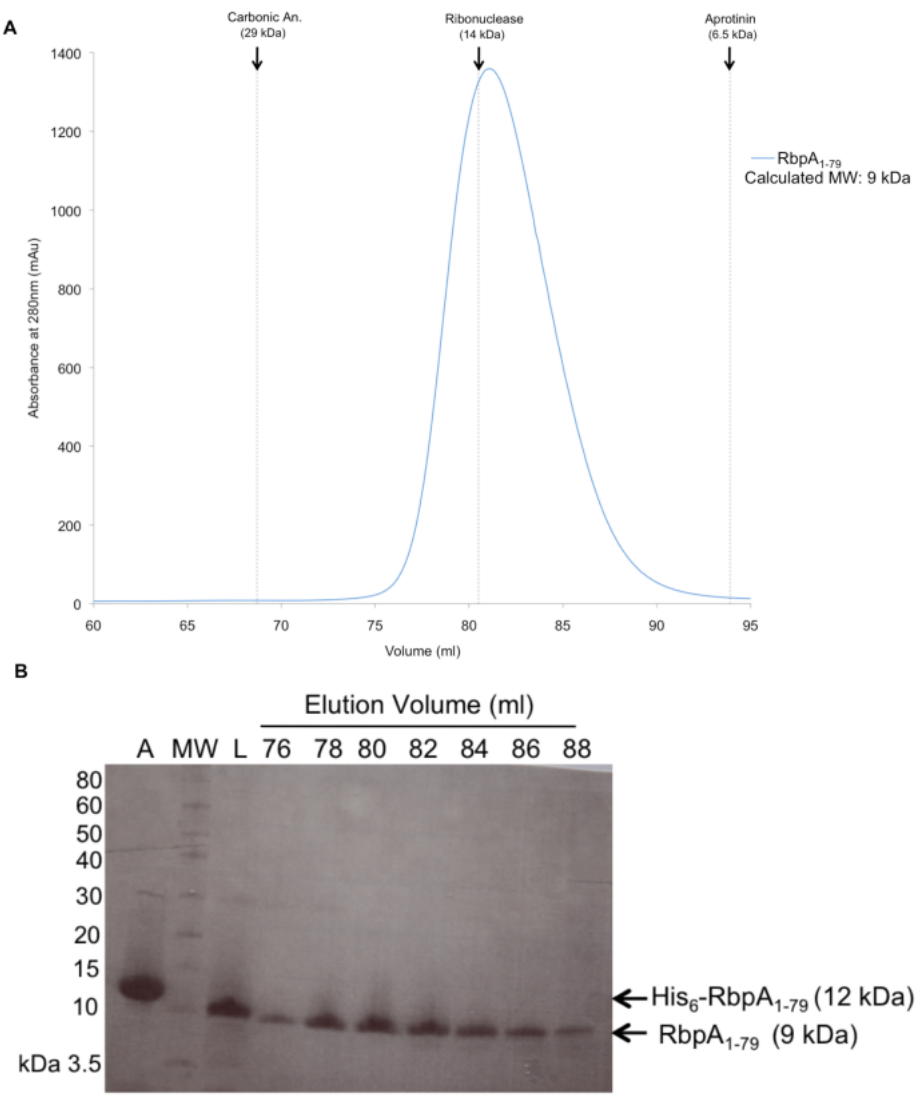


Figure 2.9. $RbpA_{1-79}$ purification by SEC chromatography. A. A typical elution profile obtained during the purification of $RbpA_{1-79}$ by SEC (Superdex 75 16/60). In the chromatogram the elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process is reported. In the gel lane A is His_6-$RbpA_{1-79}$ before TEV cleavage, MW the molecular weight markers, lane L the column load and lanes 76-88 are labelled according to the elution volume (ml).

An overlay of the SEC chromatograms obtained for RbpA and $RbpA_{1-79}$ clearly shows how the C-terminal region of RbpA affects the elution profile of the protein (figure 2.10). The peak generated by $RbpA_{1-79}$ does not show the characteristic elongated shape found for RbpA and has a higher elution volume than RbpA. It should be noted that $RbpA_{1-79}$ is eluted with an apparent molecular weight of ~13 kDa, which is consistent with a monomeric status of the protein, despite the relatively high concentration of the $RbpA_{1-79}$ sample (~300 μM) loaded on the SEC column (figure 2.10 red trace). Otherwise, RbpA full length at a concentration of ~80 μM, which is ~3.7 fold lower than the concentration used for $RbpA_{1-79}$, shows signs of homoligomerization (figure 2.10 blue trace). The calculated molecular weight difference between RbpA and $RbpA_{1-79}$ is 4 kDa, but the apparent molecular weights found during SEC for the two proteins differs by 17 kDa. Taken together these observations suggest that the C-terminal region might play an important role in the homoligomerization of RbpA.

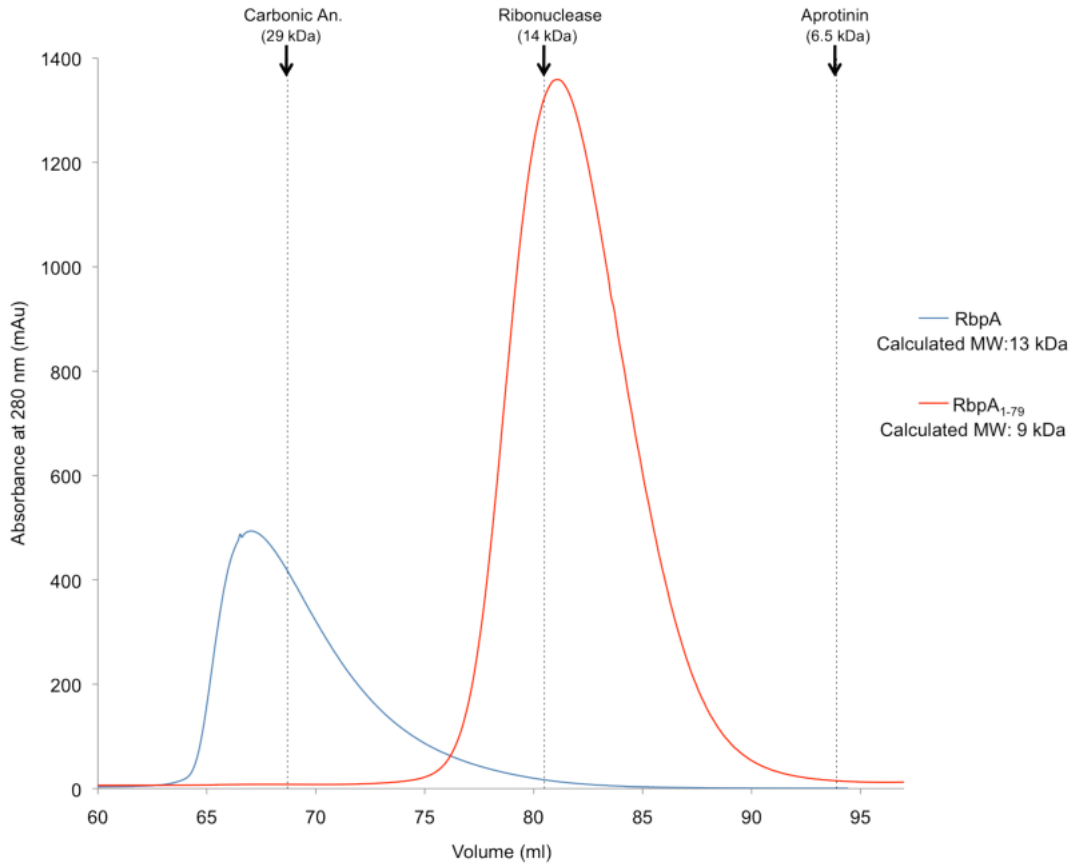


Figure 2.10. The role of the C-terminal region of RbpA in oligomerization. Overlay of the SEC (Superdex 75 16/60) elution profiles of RbpA (~ 80 μM) and $RbpA_{1-79}$ (~300 μM). The elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines.

2.3.6 Far UV circular dichroism (CD) spectroscopy of RbpA and $RbpA_{1-79}$.

The bioinformatics analysis performed for RbpA predicted that the C-terminal region of the protein (residues 80-111) is organized in two α-helices: α_1 residues 81-87 and α_2 91-108 (figure 2.1 panel A). The latter, is an amphipathic α-helix (figure 2.1 panel B) and it was hypothesized that could be responsible for the concentration dependent oligomerization observed for RbpA (section 2.3.5). The SEC analysis performed on $RbpA_{1-79}$ supported this hypothesis (figure 2.10).

In this section, in order to confirm that the C-terminal region of RbpA has a helical secondary structure organization, CD spectra were recorded for both RbpA and $RbpA_{1-79}$ and compared with each other. Figure 2.11 panel A shows an overlay of the CD spectra recorded for RbpA (black trace) and $RbpA_{1-79}$ (red trace). A difference between the two traces is the deflection between ~214-222 nm observed for RbpA. This suggests that the truncation of the C-terminal region led to a significant change in the secondary structure content of the protein. The secondary structure contents of RbpA and $RbpA_{1-79}$ were estimated by deconvoluting the relative CD spectra (table 2.1).

The change in secondary structure content between the two versions of the protein was highlighted by subtracting the $RbpA_{1-79}$ CD spectrum (panel A, red trace) from the spectrum recorded for RbpA (panel A, black trace). The resulting trace (figure 2.11 panel B), assuming that the structure of the first 79 residues of RbpA is independent from the C-terminal region of the protein, would represent the contribution of the C-terminal region to the CD spectrum of RbpA. The CD trace obtained (figure 2.11 panel B) shows features characteristic of α-helical proteins: a positive peak at 193 nm and negatives peaks at 208 and 222 nm (135). This observation and the bioinformatics analysis performed in section 2.3.1 suggest that residues 80-111 have high content of α-helical secondary structure (figure 2.1 panel A).

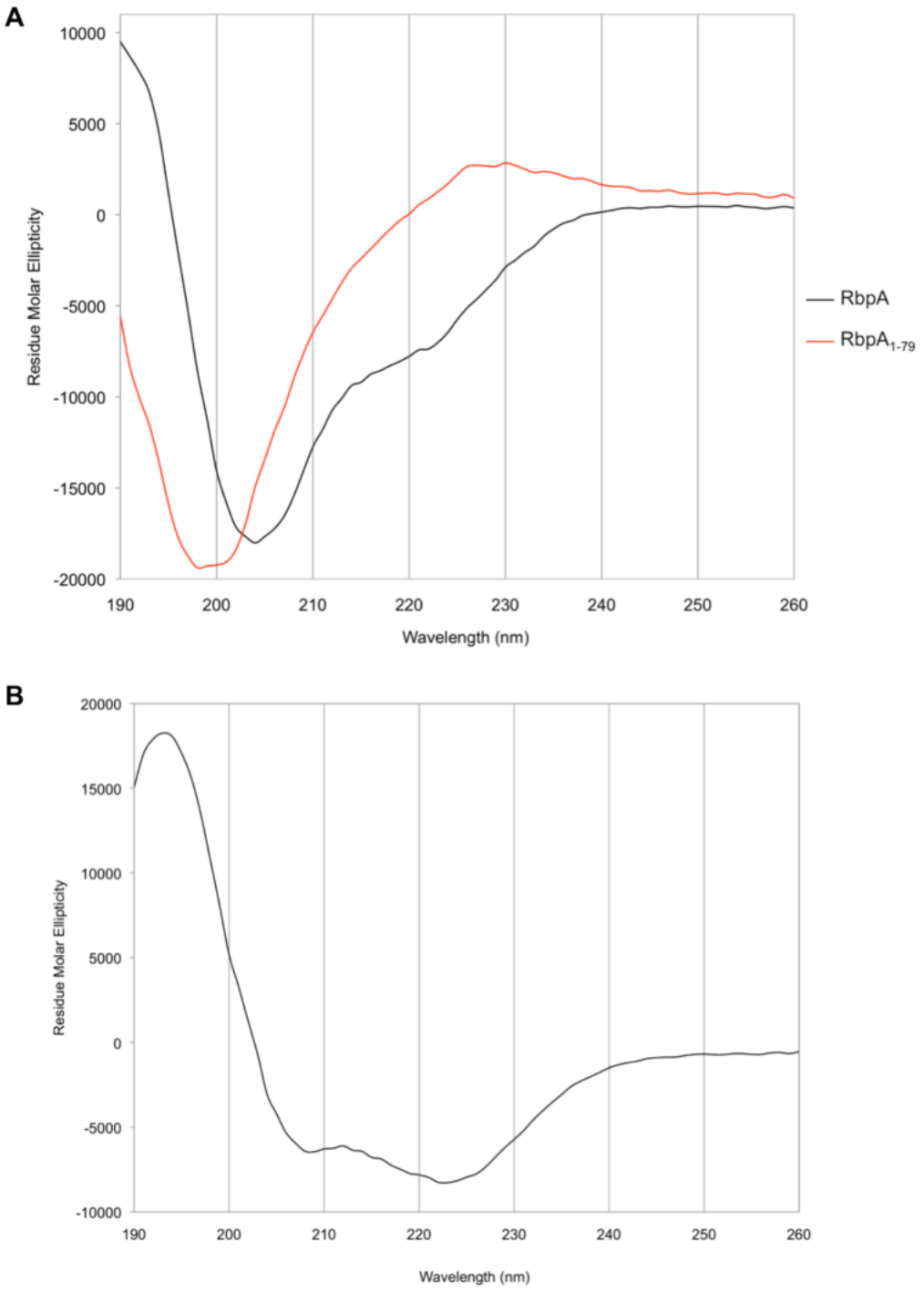


Figure 2.11 Far UV CD spectroscopy of RbpA and RbpA$_{1\text{-}79}$. A. The CD spectra obtained for RbpA (black trace) and RbpA$_{1\text{-}79}$ (red trace) samples (0.25 mg/ml) at 20 °C. **B.** Shows the trace resulting from the subtraction of the CD spectrum recorded for RbpA$_{1\text{-}79}$ (panel A, red trace) from the RbpA spectrum (panel A, black trace).

	RbpA	$RbpA_{1-79}$
Helix	28.6%	14.2%
Sheet	20.0%	36.3%
Turn	18.8%	22.9%
Random-coil	28.3%	43.4%
Total sum	95.7%	116.8%

Table 2.1. Deconvolution of the CD spectra of RbpA and $RbpA_{1-79}$. The deconvolution of the RbpA and $RbpA_{1-79}$ CD spectra, performed with the deconvolution software CDNN 2.1, provided an estimate of the secondary structure content of the two versions of the protein.

2.3.7 Sequence-specific assignment of $RbpA_{1-79}$.

A 1D ^{1}H NMR spectrum was recorded and analyzed for $RbpA_{1-79}$ (figure 2.12). The spectrum displays some of the key features of a folded protein like a significant dispersion of the signals from backbone NH groups (6.5-9.5 ppm) and several high field shifted methyl resonances (-0.3-0.5 ppm). Thus, it was concluded that $RbpA_{1-79}$ was produced in a folded status.

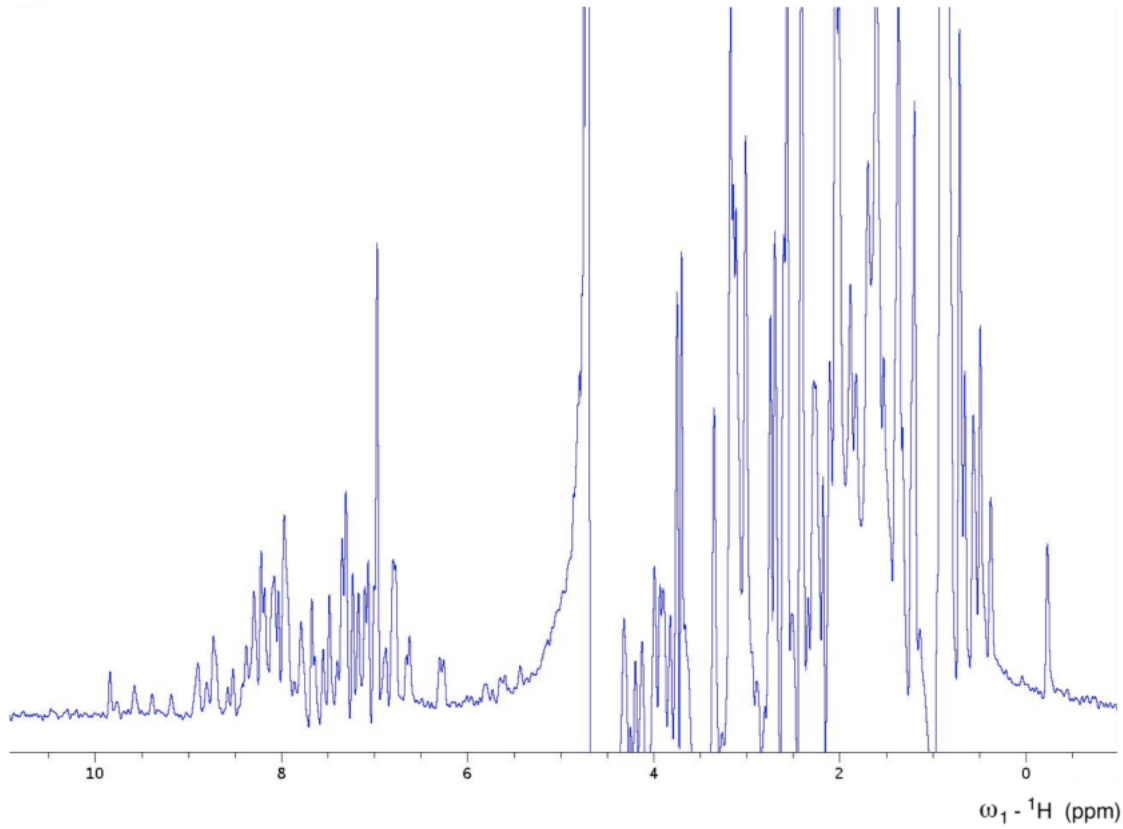


Figure 2.12. 1D ^{1}H-NMR spectrum of $RbpA_{1-79}$. The spectrum shown results from four scans and was acquired at 35 °C from an $RbpA_{1-79}$ sample (0.2 mM) dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O.

Isotopically labelled $RbpA_{1-79}$ samples were prepared in order to continue the NMR analysis of the protein. In figure 2.13 is reported an overlay of the $^{15}N/^{1}H$ HSQC spectra recorded on ^{15}N $RbpA_{1-79}$ (shown in red) and on ^{15}N RbpA (shown in black). The overlay highlights only minor changes (figure 2.13), indicating that the structure of the first 79 residues of RbpA is maintained in the shorter $RbpA_{1-79}$ construct.

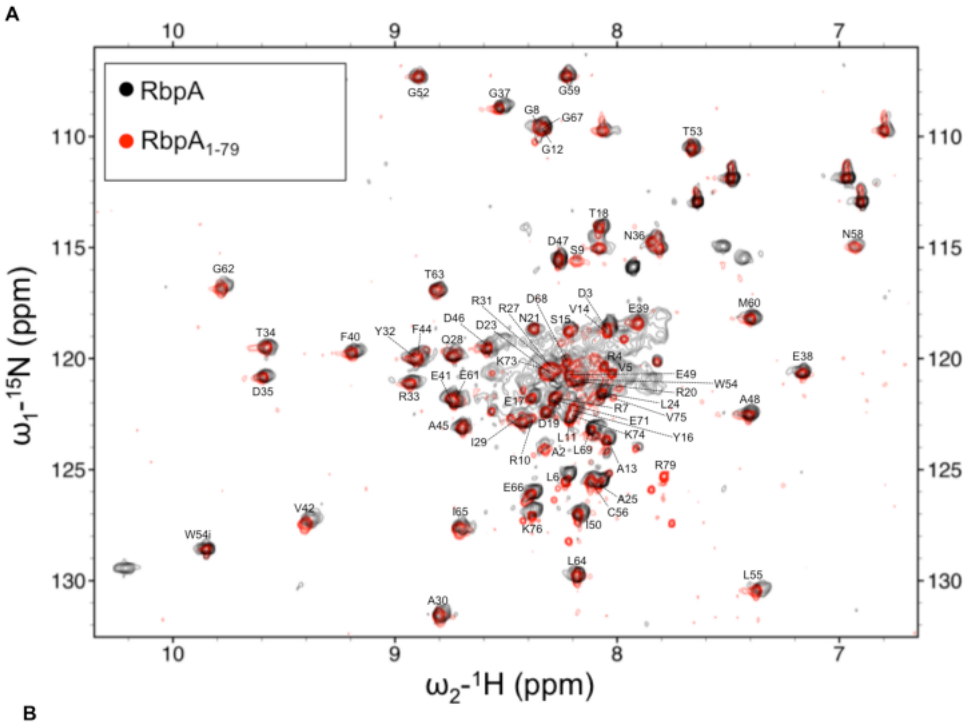


Figure 2.13. Assignment of the backbone amide groups of $RbpA_{1-79}$. A. Overlay of $^{15}N/^{1}H$ HSQC spectra recorded on RbpA and $RbpA_{1-79}$. The spectra reported were acquired at 35 °C from ^{15}N RbpA (0.2 mM), shown in black, and ^{15}N $RbpA_{1-79}$ (0.35 mM), shown in red, samples dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O. The assignments of the backbone amide groups of $RbpA_{1-79}$ are reported in the picture. An "i" indicates peaks generated by indole groups. **B.** The amino acid sequence of $RbpA_{1-79}$. Residues for which the backbone amide group was assigned are highlighted in blue.

There are 8 NH unassigned backbone resonances in the $^{15}N/^{1}H$ HSQC spectrum recorded on $RbpA_{1-79}$ (figure 2.13 panel A, red spectrum). However, only three amide groups of $RbpA_{1-79}$ result unassigned (figure2.13 panel B). The extra resonances might arise from conformational exchange or from minor proteolysis occurring during NMR experiments.

A set of 2D and 3D NMR experiments were recorded and analyzed for $RbpA_{1-79}$ leading to essentially complete assignment. For example, assignments of backbone (N, NH, Hα, Cα, Hβ and Cβ) resonances were made for all residues except: Met_1 (N, NH, Hα, Hβ2, Hβ3 and Cβ), Gly_8 (Hα2 and Hα3), Tyr_{16} (Cα and Cβ), His_{22}, Tyr_{32} (Cα and Cβ), Phe_{40} (Cα and Cβ), Phe_{44} (Cα and Cβ), Arg_{57} (N and NH), Pro_{70} (Hα, Hβ2 and Hβ3), Pro_{77} and Pro_{78} (Hα, Hβ2 and Hβ3). Complete assignments were obtained also for the side chain resonances except: Met_1, Arg_4 (Hγ2, Hγ3, Nε, Hε, Cγ and Cδ), Leu_6 (Cγ, Cδ1 and Cδ2), Arg_7 (Nε, Hε, Cγ and Cδ), Gly_9 (Hα2 and Hα3), Arg_{10}, Leu_{11} (Hγ, Hδ1, Cγ, Cδ1 and Cδ2), Arg_{20} (Nε, Hε, and Cδ), His_{23}, Leu_{24} (Cδ1), Arg_{27} (Nε, Hε and Cδ), Phe_{40} (Hζ), Arg_{57} (Nε, Hε and Cδ), Glu_{61} (Cγ, Hγ2 and Hγ3), Glu_{66} (Cγ), Leu_{69} (Cγ and Cδ1), Pro_{70}, Glu_{71} (Cγ), Lys_{74} (Hδ2, Hδ3, Hε2, Hε3, Cγ, Cδ and Cε), Lys_{76} (Hδ2, Hδ3, Cγ, Cδ and Cε), Pro_{77}, Pro_{78} and Arg_{79} (Nε, Hε, Cγ and Cδ). The extent of the assignment is 78.5% of all the aliphatic ^{13}C resonances, 86.4% and 91% of all the ^{15}N and ^{1}H resonances, respectively. A representative set of ^{15}N strips from HNCACB and ^{15}N-edited NOESY spectra used for the assignment of $RbpA_{1-79}$ resonances is reported in figure 2.14.

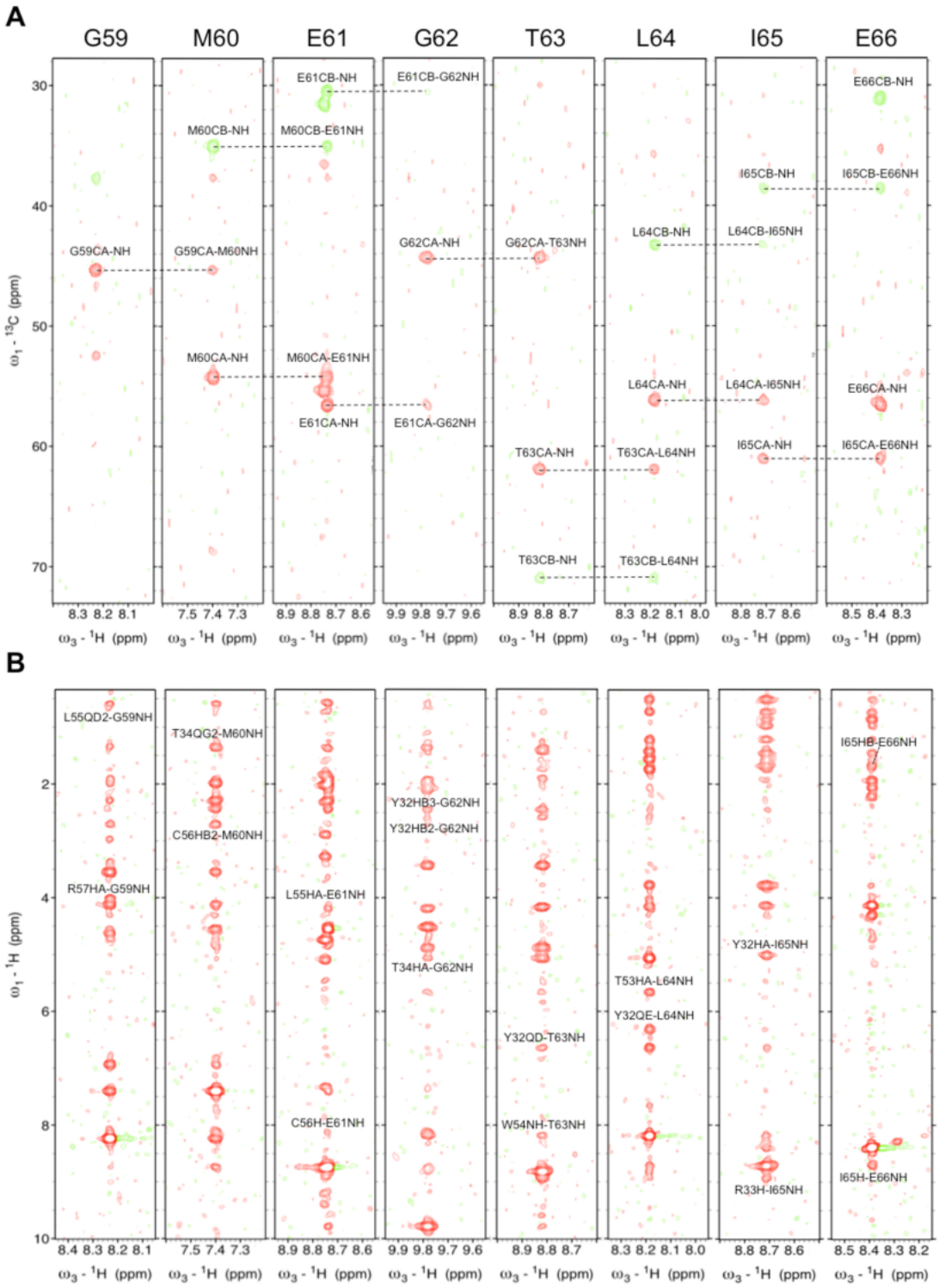


Figure 2.14. ^{15}N strips for the residues 59-66 from the HNCACB and ^{15}N-edited NOESY spectra recorded for RbpA$_{1-79}$. A. Shows ^{15}N strips for the residues 59-66 from the HNCACB spectrum acquired at 35 °C from ^{15}N/^{13}C RbpA$_{1-79}$ (0.35 mM)

sample dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF, 0.5 mM TCEP and 10% D_2O v/v. Dashed lines indicate sequential connections of Cα and Cβ. **B.** Shows ^{15}N strips for the $RbpA_{1-79}$ residues 59-66 from the ^{15}N-edited NOESY spectrum acquired in the same conditions described in panel A and with a mixing time of 200 ms. Representative examples of NOE assignments are reported for each strip.

2.3.8 Secondary structure determination of $RbpA_{1-79}$ using TALOS+.

There is a strong correlation between protein secondary structure and NMR chemical shifts and it has been shown that backbone chemical shifts can be used to determine the secondary structure of a protein (136-141). The TALOS+ software (125) determines the secondary structure of a query protein based on its sequence and chemical shifts homology with the entries of a database containing 200 proteins for which complete or nearly complete resonance assignments and high-resolution structures are available (125). TALOS+ compares strings of three sequential residues from the query sequence with all the amino acids triplets contained in the database. If the top 10 triplets with the highest sequence and chemical shifts homology have similar Φ and Ψ backbone angles their average can be used with high confidence to estimate the backbone torsion angles of the query triplet. The availability of comprehensive backbone resonances assignment for $RbpA_{1-79}$ allowed determination of the secondary structure of the protein using the software TALOS+. The elements of regular secondary structure found for $RbpA_{1-79}$ are four β-strands: β_1 residues 28-34, β_2 39-44, β_3 54-55 and β_4 63-65 (figure 2.15).

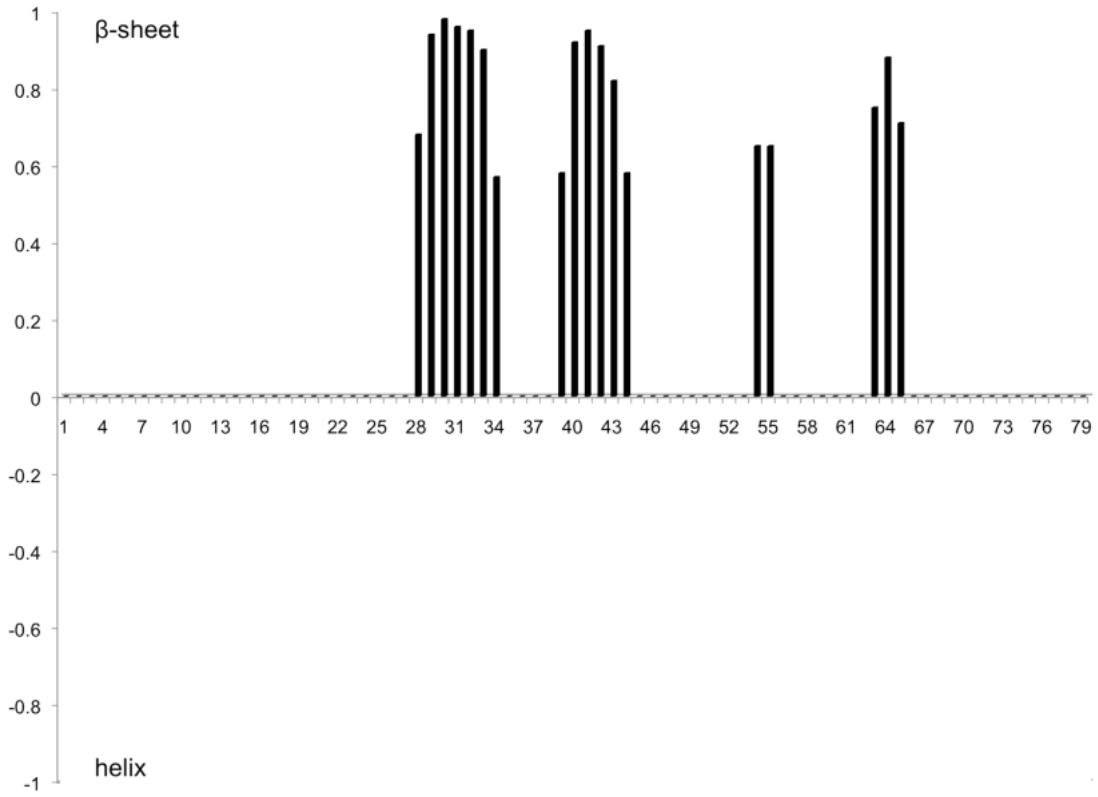


Figure 2.15. TALOS+ secondary structure determination of RbpA$_{1\text{-}79}$. The height of the bars in the histogram reflects the probability (from 0 to ±1) of the relative RbpA$_{1\text{-}79}$ residue to be located in β-sheet (positive values) or helix (negative values).

2.3.9 Solution structure calculation of RbpA$_{1\text{-}79}$.

There was a significant improvement in the quality of the NMR data obtained from the RbpA$_{1\text{-}79}$ as compared to full length RbpA, which is evident from both the increased signal-to-noise ratio and decreased line-widths observed for RbpA$_{1\text{-}79}$ (figure 2.16). Furthermore, comparison of the $^{15}N/^{1}H$ HSQC spectra recorded on RbpA and on RbpA$_{1\text{-}79}$ (figure 2.13) suggested that the fold of RbpA residues 1-79 is independent from the C-terminal region of the protein. Therefore, RbpA$_{1\text{-}79}$ was chosen for further structural characterization.

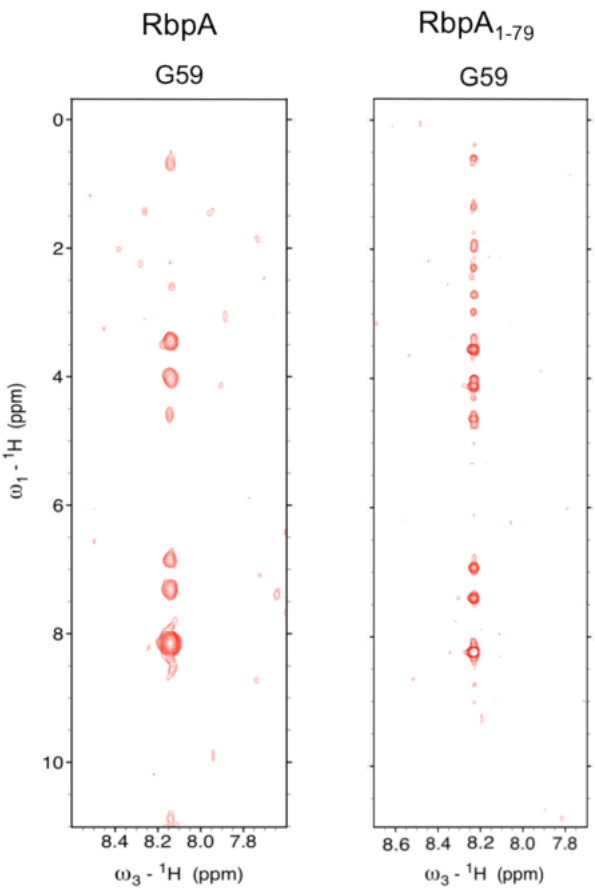


Figure 2.16. ^{15}N strips for the residue Gly$_{59}$ from ^{15}N-edited NOESY spectra recorded for RbpA and RbpA$_{1-79}$. The ^{15}N-edited NOESY spectra were acquired at 35 °C from ^{15}N/^{13}C RbpA (0.2 mM) and ^{15}N/^{13}C RbpA$_{1-79}$ (0.35 mM) samples with mixing times of 80 and 200 ms respectively. The samples were dissolved in a buffer containing 25 mM K_2HPO_4 pH 6.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O.

The CANDID protocol (126) was used to automatically assign the NOE cross peaks identified in 2D ^{1}H-^{1}H NOESY and 3D ^{15}N- and ^{13}C- edited NOESY spectra. Unique assignments were obtained for 87% of the NOE cross peaks identified. The structure of RbpA$_{1-79}$ was calculated with the program Cyana (126) using the constraints listed in table 2.2. Subsequently, the final family of converged RbpA$_{1-79}$ structures was produced

through several cycles of simulated annealing combined with redundant dihedral angle constraints (REDAC) (127) resulting in 98 satisfactorily converged structures out of 100 initial random structures. These had no distance or Van der Waals violations greater than 0.5 Å, no dihedral angle violations greater than 5° and had an average value for the Cyana target function of 1.17 $Å^2$. The structures were refined using the AMBER 10 package (129) and the 35 structures with the lowest AMBER energy were selected to be deposited in the PDB (accession code 2M4V). The final family of converged structures has no distance constraint violations greater than 0.37 Å and no dihedral angle violations.

Superposition of the final family of converged structures (figure 2.17 panel A) reveals the presence of a well-defined central domain (residues 26-66), whereas the N- and C-terminal regions of the protein are both disordered and highly flexible. The structure of the central domain was determined to high precision, as is clearly evident from the overlay shown in figure 2.17 panel B, and is reflected in low RMSD values to the mean structure for both the backbone and all the heavy atoms of 0.32 ±0.11 Å and 1.08 ±0.16 Å, respectively (a summary of the NMR constraints and structural statistics is reported in table 2.2). The central domain is primarily composed of four distinct β-strands: β_1 (residues 27-33), β_2 (39-45), β_3 (53-55) and β_4 (61-65), which fold to form two antiparallel β-sheets linked by turns and loops (figure 2.17 panel C). The two β-sheets lie perpendicular to each other forming a β-sandwich like structure, which is stabilized by a cluster of aromatic and non-polar residues (Tyr_{32}, Val_{42}, Phe_{44} and Trp_{54}).

NMR constraints	
No. of upper distance limits	1276
Short range |i-j|≤1	618
Medium range 1≤|i-j|≤5	165
Long range |i-j|≥5	493
No. of backbone torsion angle constraints	72
Violations	
Maximum distance violation	0.37 Å
Maximum dihedral angle violation	0°
Energies	
Mean AMBER energy	-3696.46 kcal mol^{-1}
Mean NOE energy	11.51 kcal mol^{-1}
Deviation for idealized geometry	
Bond lengths	0.0108 ± 1.09 x 10^{-4} Å
Bond angles	2.139 ± 0.036°
r.m.s.d. from mean structure (residues 26-66)	
Backbone	0.32 ± 0.11 Å
Heavy atoms	1.08 ± 0.16 Å
Ramachandran plot (residues 26-66)	
Residues in most favoured regions	97.6 %
Residues in additionally allowed regions	2.4 %
Residues in generously allowed regions	0 %
Residues in disallowed regions	0 %

Table 2.2. **NMR constraints and structural statistics for the $RbpA_{1-79}$ structure.** Summary of the NMR constraints used for the calculation of the $RbpA_{1-79}$ structure and of the structural statistics for the final family of 35 $RbpA_{1-79}$ converged structures.

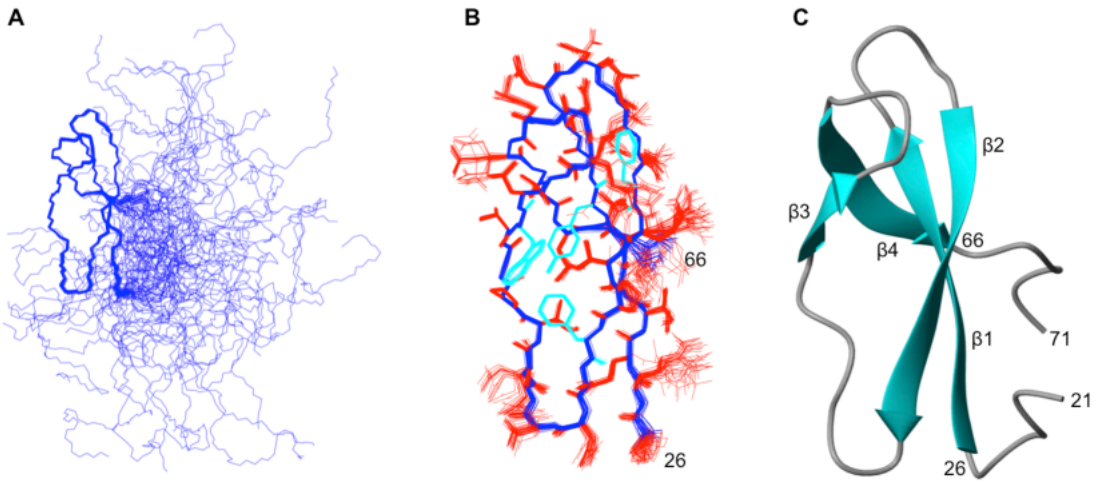


Figure 2.17. Solution structure of RbpA$_{1\text{-}79}$. A. A best-fit superposition (residues 26-66) of the final family of 35 converged structures obtained for RbpA$_{1\text{-}79}$. **B.** A best-fit superposition of the structured core (residues 26-66) of the final family of converged structures of RbpA$_{1\text{-}79}$. The backbone is shown in blue, the side chains are coloured in red except for the aromatic side chains that are coloured in cyan. The orientation is identical to that shown in panel A. **C.** A ribbon representation of the RbpA$_{1\text{-}79}$ structure, from residue 21 to residue 71, with an orientation identical to that shown in panel A and B.

Analysis of the surface of the central domain (figure 2.18) revealed the presence of a cluster of conserved residues (Arg$_{27}$, Val$_{42}$, Phe$_{44}$, Ala$_{45}$, Asp$_{47}$, Ala$_{48}$, Glu$_{49}$ and Trp$_{54}$) that form a surface patch of 742.2 Å^2 (calculated using MOLMOL (131)) with a hydrophobic centre surrounded by charged residues. This patch is the most conserved region on the surface of the central domain (figure 2.18 panel B) and may form a potential protein interaction site.

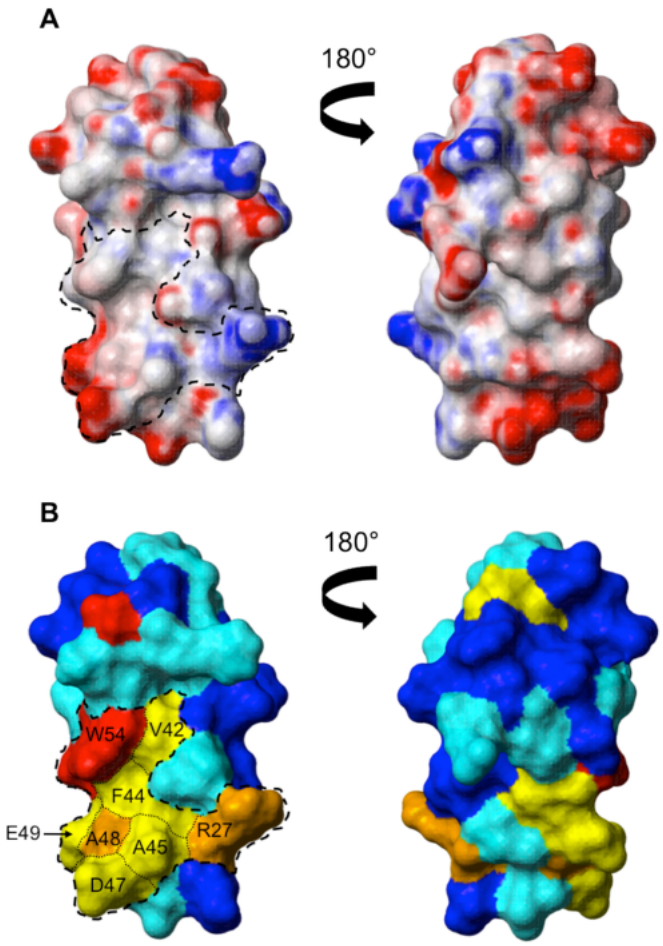


Figure 2.18. Surface view of the structured core of RbpA (residues 26-66).

A. Contact surface views of the central domain coloured according to electrostatic potential. Neutral areas are shown in white and areas of significant charge are shown in red (negative) and blue (positive). **B.** The same views are coloured according to amino acid conservation amongst homologues with the least conserved residues in blue, scaling to completely conserved residues in orange and invariant residues in red (conservation scores were taken from figure 2.1 panel C, blue 0-3, cyan 4-6, yellow 7-9, orange 10 and red 11 (107)).

2.3.10 RbpA structural homologues search.

RbpA sequence is specific to the *Actinobacteria* phylum and no structural information was available for any of its homologues when the research project described in this thesis was undertaken. However, because protein structures can be maintained through evolution despite poor sequence conservation, it was decided to investigate if the structure found for RbpA shared similarities with other known protein folds performing a structural homologues search. Comparison of the RbpA central domain structure (residues 26-66) with the known folds in the PDB using Dali (133) led to the identification of fourteen structures (table 2.3) that share structural homology (z-score > 2). However, the structural homology of each with RbpA central domain is relatively low (z-score < 3.5) and no specific function has been assigned to this fold. The highest scoring structural homologue (PDB: 3UXQ chain 6) is a ribosomal protein that has an RMSD to the RbpA central domain of 2.8 Å over 41 Cα-s (figure 2.19) but there is no sequence homology (2% amino acid identity).

Results	PDB ID	Z-Score*	RMSD	LALI**	NRES***	% ID****	DESCRIPTION
1	3uxq-6	3.5	2.8	41	53	2	Ribosomal protein
2	3ir9-a	3	2.8	40	162	10	Peptide chain release factor subunit 1
3	1woo-a	2.6	2.0	38	362	5	Aminomethyltransferase
4	1ltl-e	2.6	2.2	35	242	14	DNA replication initiator
5	2lcd-a	2.4	2.2	37	118	8	RBBP1 tudor domain
6	2qqr-a	2.4	2.5	39	118	10	JMJC tudor domain
7	1vs6-1	2.4	2.8	40	54	3	Ribosomal protein
8	1pj6-a	2.3	2.1	38	829	11	N,N-Dimethylglycine oxidase
9	3oe3-c	2.3	3.2	39	88	15	Periplasmic lysozyme inhibitor
10	1vlo-a	2.2	2.1	39	364	5	Aminomethyltransferase
11	2ky9-a	2.2	2.5	39	132	8	Uncharacterized protein from *B. subtilis*
12	3f6z-b	2.2	3.4	37	93	16	Periplasmic lysozyme inhibitor
13	2cuw-a	2.1	2.9	37	83	19	Formylglycinamide Ribonucleotide Amidotransferase
14	4exw-a	2.1	2.6	32	131	19	Dna damage response B protein

Table 2.3. Structural homologues search for the central structured domain (residues 26-66) of $RbpA_{1-79}$ performed with Dali (133).

* the Dali Z-score measures the quality of the structural alignment based on the distance between Cα of matched residues (unmatched residues do not contribute to the overall score). Alignments with a Z-score < 2 are considered spurious and are not reported.

** number of residues aligned in the structure superposition. *** number of residues in matched structure. **** sequence identity of aligned residues.

Figure 2.19. Overlay of the RbpA$_{1\text{-}79}$ and the ribosomal protein L33 structures. Overlay between the RbpA$_{1\text{-}79}$ structured central domain (residues 26-66) shown in red and the structure of the *T. thermophilus* ribosomal protein L33 (PDB: 3UXQ-chain 6) shown in blue.

2.4 Conclusions

The data presented in this chapter contribute to the ongoing effort to characterize RbpA shedding light on its oligomeric status and on its structural organization.

The finding that the elution profile of RbpA during SEC analyses is strongly dependent on the concentration of the protein (figure 2.5) could account for the disagreement found in the literature about the likely oligomeric organization of RbpA. In fact, RbpA and its homologues were found to be monomeric or dimeric on the basis of SEC experiments (96,97,102). Whether oligomerization of RbpA occurs *in vivo* would then depend on its concentration, as well as the local environment and competition with other (non-self) binding partners.

The C-terminal region of RbpA (residues 79-111) is clearly essential for oligomerization (figure 2.10). It was noticed that an amphipathic α-helix is likely to be formed in this region of the protein (between residues 91-108, figure 2.1 panel B) and it was proposed that its hydrophobic side could interact with the equivalent face on a second RbpA molecule leading to oligomerization. An amphipathic α-helix is likely to be formed also at the C-terminal of *M. smegmatis* and *S. coelicolor* RbpA homologues (figure 2.1 panel B), hence the oligomerization mechanism proposed could be conserved also in these organisms.

The involvement of the C-terminus of RbpA in oligomerization could account for the poor quality of the NMR data collected for this region of the protein: some signals are weak, some are very broad and some are missing. The elongated shape of the peaks detected in SEC analyses of RbpA (figure 2.5) suggests that there is a dynamic equilibrium between different oligomeric species of the protein, this could cause conformational exchange of the protein region involved in oligomerization and ultimately would affect the quality of the NMR data collected.

When the research project described in this thesis began there was no structural information available for RbpA or its homologues. The results described in this chapter show that RbpA consists of three regions: residues 1-25 are flexible random coil, residues 26-66 form a stably folded central domain, which is connected by an unstructured linker to an helical region, residues 80-111. Comparison of the $^{15}N/^{1}H$ HSQC spectra collected on RbpA and $RbpA_{1-79}$ reveals only minor changes (figure 2.13) despite the different oligomeric organization of the two proteins (figure 2.5). Because chemical shifts are extremely sensitive to their chemical environment, the similarity between spectra suggests that residues 1-79 are not involved in the oligomerization of RbpA and that their fold is independent from the C-terminal region. Interestingly, the N- and C- termini, which are the most conserved regions of RbpA (figure 2.1 panel C), are both unstructured. In contrast to the central region of RbpA (residues 26-66) that is structured and does not show clusters of highly conserved residues. However, the analysis of the RbpA surface reveals that the 3D arrangement of residues 26-66 forms a conserved patch (figure 2.18) with a hydrophobic centre surrounded by charged residues. The patch identified could be a potential site for RbpA binding partners.

Chapter 3

Characterization of the interaction between RbpA and the RNAP σ-subunit.

3.1 Introduction.

RbpA was identified as an RNAP binding protein during a study on the disulphide stress response of *S. coelicolor* (96). The interaction between RbpA and the RNAP was then confirmed also in *M. smegmatis* and in *M. tuberculosis* and preliminary attempts have been made to map the RbpA binding site on the RNAP, which was found on the β-subunit (97,102,103). No other subunits of the RNAP were reported to be involved in the interaction (97,103). However, a yeast two-hybrid screen followed by a protein fragment complementation assay, performed before the research project described in this thesis began, suggested that RbpA could interact also with the RNAP σ^B-subunit (142). The interaction between RbpA and σ^B had not been reported previously and clearly its confirmation and characterization would contribute to broaden the knowledge on the function of RbpA.

The data reported in this chapter demonstrate, by *in vitro* experiments, that RbpA can bind to both σ^B and σ^A (the primary sigma factor of *M. tuberculosis*) and provide important insights in the interaction mode between RbpA and the σ-subunit. In the first part of the chapter are described the protocols adopted for the purification of σ^A and σ^B and the SEC experiments performed to demonstrate that RbpA can form a tight complex with both σ^A and σ^B. Furthermore, it is shown that the first 228 residues of σ^B ($\sigma^B_{1\text{-}228}$) are sufficient for the interaction with RbpA and that the RbpA•$\sigma^B_{1\text{-}228}$ complex occurs with a 1:1 stoichiometry. In the second part of the chapter is described how a set of NMR experiments combined with SEC interaction assays, using truncated versions of RbpA, led to the identification of those regions of RbpA that are involved in the

binding with the σ-subunit. It was concluded that the structured central domain of RbpA is not involved in the formation of the complex with the σ-subunit but both ends of the protein are affected by the interaction, although only the C-terminus is essential.

3.2 Materials and methods.

3.2.1 Preparation of expression vectors.

The genes encoding σ^A (*Rv2703*), σ^B (*Rv2710*) and the ORFs encoding for residues 1-228 of σ^B (σ^B_{1-228}), residues 1-92 of RbpA ($RbpA_{1-92}$) and residues 24-111 of RbpA ($RbpA_{24-111}$) were cloned into an expression vector derived from pET-43.1a(+) (Novagen) encoding a cleavable His_6-tag at the N-terminus. The ORFs encoding for σ^B_{1-228} and for residues 228-323 of σ^B ($\sigma^B_{228-323}$) were cloned into an expression vector derived from pET-43.1a(+) (Novagen) encoding a cleavable GST-tag at the N-terminus. All the vectors were prepared following the procedure described in section 2.2.2 and were subjected to DNA sequencing at the University of Leicester Protein and Nuclear Acid Chemistry Laboratory (PNACL).

3.2.2 Expression of His_6-σ^A, His_6-σ^B, His_6-σ^B_{1-228}, GST-σ^B_{1-228} and GST-$\sigma^B_{228-323}$.

E. coli BL21(DE3) cells, transformed with the relevant expression vector, were grown in LB media (appendix) containing 100 μg/ml ampicillin at 37 °C with shaking to an absorbance at 600 nm of ~0.7. Protein expression was then induced by the addition of 0.05 mM IPTG and the cell growth continued at 16 °C over-night. Cells were harvested by centrifugation (8000 *g* for 20 minutes at 4 °C) and the pellet was stored at -20 °C.

3.2.3 Purification of His_6-σ^A, His_6-σ^B and His_6-σ^B_{1-228}.

Cell pellet was resuspended in a buffer containing 50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 1 mM AEBSF, 0.1 mg/ml lysozyme, 0.1 mg/ml deoxyribonuclease I and 5 mM $MgCl_2$. Cells were sonicated on ice using Misonix sonicator 3000 with power set on 7 for 10 minutes with 20 seconds on-off intermittency. The resulting cell

lysate was centrifuged (15,000 *g* for 30 minutes at 4 °C). The supernatant was then loaded on a 5 ml Ni-NTA column (QIAGEN) pre-equilibrated with a buffer containing 50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol and 20 mM imidazole. The His_6-tagged protein was eluted in the same buffer with a linear imidazole gradient (20-500 mM). Fractions containing the His_6-tagged protein were subjected to SEC using a Superdex 75 16/60 (GE Healthcare) column and a buffer containing 50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 0.1 mM EDTA and 0.5 mM DTT. The SEC column was calibrated as described previously (section 2.2.6). The concentration of the purified protein was calculated measuring the absorbance at 280 nm.

3.2.4 SEC interaction assays between RbpA and the σ-subunit.

The interaction between RbpA and the sigma factors was assessed by SEC using a Superdex 75 10/300 GL column (GE Healthcare) or a Superdex 200 10/300 GL column (GE Healthcare) and a buffer containing 50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 0.1 mM EDTA and 0.5 mM DTT. The same buffer was used to calibrate the Superdex 75 10/300 GL column using the low molecular weights calibration kit (GE Healthcare). The column Superdex 200 10/300 GL was previously calibrated in a buffer containing 25 mM Tris-HCl pH 8 and 150 mM NaCl using the low and high molecular weights calibration kits (GE Healthcare).

3.2.5 The structural model of *M. tuberculosis* σ^B.

In order to design the expression vectors for the purification of σ^B domains a structural model of the *M. tuberculosis* σ^B was calculated using the Protein Homology\analogY Recognition Engine (Phyre) (143). The software Phyre generates a structural model for a query protein sequence based on known structures of homologous proteins. The

structural model of *M. tuberculosis* σ^B was calculated on the basis of the structure of the principal σ-subunit of *T. thermophilus* (PDB 1IW7, 45% of sequence identity).

3.2.6 Purification of GST-σ^B_{1-228} and GST-$\sigma^B_{228-323}$.

Cell pellet was resuspended in a buffer containing 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM AEBSF, 0.1 mg/ml lysozyme, 0.1 mg/ml deoxyribonuclease I and 5 mM $MgCl_2$. Cells were sonicated on ice using Misonix sonicator 3000 with power set on 7 for 10 minutes with 20 seconds on-off intermittency. The resulting cell lysate was centrifuged (15,000 *g* for 30 minutes at 4 °C). The supernatant was then loaded on a 5 ml GSTrap HP column (GE Healthcare) pre-equilibrated with a buffer containing 25 mM Tris-HCl pH 7.4 and 150 mM NaCl. The GST-tagged protein was eluted using a buffer containing 25 mM Tris-HCl pH 8, 150 mM NaCl and 10 mM reduced glutathione. GST-σ^B_{1-228} was further purified by SEC using a Superdex 75 16/60 column (GE Healthcare) and a buffer containing 25 mM Tris-HCl pH 7.4 and 150 mM NaCl. The SEC column was calibrated as described previously (section 2.2.6). The concentration of the purified protein was calculated measuring the absorbance at 280 nm.

3.2.7 Pull-down assays of RbpA binding to GST-σ^B_{1-228} and GST-$\sigma^B_{228-323}$.

In a typical pull-down assay 1 ml of GST-tagged protein at a concentration of 20 μM in a buffer containing 25 mM Tris-HCl pH 7.4 and 150 mM NaCl (binding buffer) was loaded on a gravity column containing 1 ml of glutathione-agarose resin (Sigma-Aldrich). The resin was previously equilibrated in binding buffer. Following loading, the column was washed with 5 ml of binding buffer. Afterwards, 1 ml of untagged RbpA at concentration of 40 μM was loaded on the column. The unbound RbpA was

removed by washing the resin with 10 ml of binding buffer. Finally, the GST-tagged protein was eluted in 3 ml of a buffer containing 25 mM Tris-HCl pH 8, 150 mM NaCl and 10 mM reduced glutathione.

3.2.8 NMR spectroscopy.

In order to identify changes in the positions of RbpA NMR signals resulting from binding to His_6-σ^B_{1-228}, NMR spectra were acquired from 0.35 ml samples of $^{15}N/^{13}C/^{2}H$ RbpA (0.19 mM) and of $^{15}N/^{13}C/^{2}H$ RbpA (0.1 mM) bound to unlabelled His_6-σ^B_{1-228} (1:1 molar ratio) dissolved in a buffer containing 25 mM K_2HPO_4 pH 7.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O. The 2D and 3D spectra recorded were: $^{15}N/^{1}H$ TROSY and $^{15}N/^{13}C/^{1}H$ TROSY-HNCO. Acquisition times in the 2D experiment were: 26 ms in F_1 and 65 ms in F_2. Acquisition times in the 3D experiment were: 12 ms in F_1 (^{13}C), 10 ms F_2 (^{15}N) and 65 ms in F_3 (^{1}H). The 2D NMR spectra were acquired typically for about 0.5 hours and the 3D spectra for about 65 hours. The spectra were processed and analyzed as described in section 2.2.9.

3.2.9 NMR minimal shift assay.

Changes in the positions of RbpA NMR signals resulting from binding to His_6-σ^B_{1-228} were analysed using the minimal shift approach (144-146). The combined HN, N and CO chemical shift differences ($\Delta\delta$) between the peaks detected in HNCO spectra recorded on free $^{15}N/^{13}C/^{2}H$ RbpA and on $^{15}N/^{13}C/^{2}H$ RbpA bound to unlabelled His_6-σ^B_{1-228} were calculated using the following equation: $\Delta\delta=\sqrt{((\Delta\delta_{HN})^2 + (\Delta\delta_N*0.2)^2 + (\Delta\delta_{CO}*0.35)^2)}$ where $\Delta\delta_{HN}$, $\Delta\delta_N$, and $\Delta\delta_{CO}$ correspond to the differences in ^{1}H, ^{15}N and ^{13}C chemical shifts between pairs of HNCO peaks and 0.2 and 0.35 are scaling factors

required to account for differences in the range of amide proton, amide nitrogen and carbonyl chemical shifts. For each peak detected in the HNCO spectrum recorded on free $^{15}N/^{13}C/^{2}H$ RbpA the minimal shift induced by the binding with His_6-σ^{B}_{1-228} was taken as the lowest calculated combined shift value ($\Delta\delta$).

3.3 Results

3.3.1 Expression and purification of His_6-σ^B.

Expression of His_6-σ^B was induced from *E. coli* BL21(DE3) cells transformed with the relevant expression vector adding IPTG and the culture growth was continued at 16 °C over-night. Culture samples collected ~20 hours post-induction were lysed (section 3.2.3) and subjected to SDS-PAGE analysis. The resulting gel (figure 3.1) shows an overexpressed band that migrates in a position consistent with the calculated molecular weight of His_6-σ^B (39 kDa) and that is highly soluble.

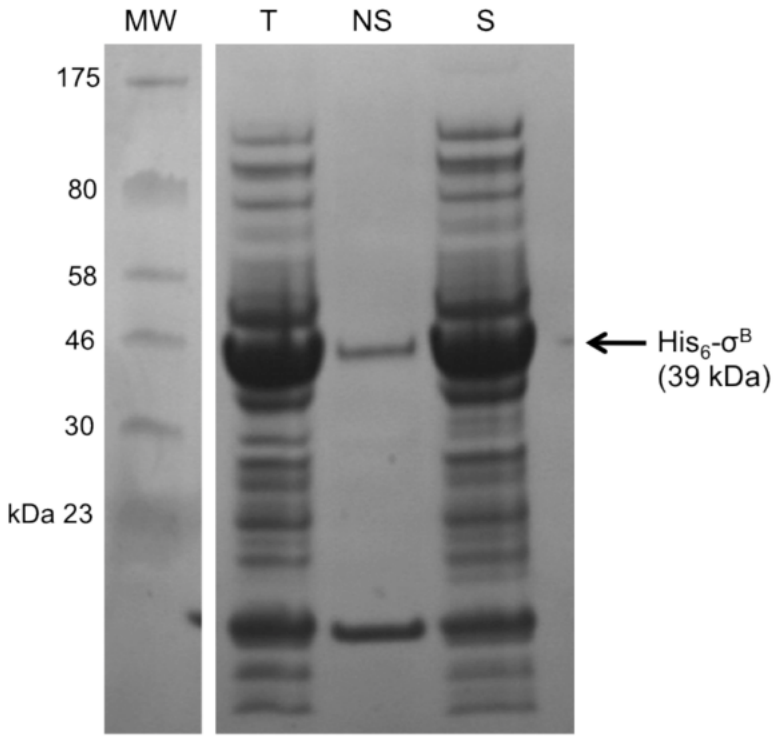


Figure 3.1. His_6-σ^B expression. SDS-PAGE analysis of culture samples collected ~20 hours post-induction of His_6-σ^B expression. Lane MW is molecular weights markers, lane T the total cell lysate, lane NS the non-soluble fraction and lane S soluble fraction.

The protocol adopted for the purification of His_6-σ^B involved two main steps: IMAC and SEC. In the first step the soluble fraction resulting from the lysis of the cell culture expressing His_6-σ^B was loaded on a Ni-NTA column (QIAGEN) and subsequently His_6-σ^B was eluted with a linear imidazole gradient. In the resulting chromatogram (figure 3.2 panel A) is it possible to distinguish two peaks, which are labelled with the letters "A" (elution volume 186 ml) and "B" (elution volume 194 ml). The SDS-PAGE analysis of the fractions collected at different stages of the chromatography (figure 3.2 panel B) reveals that His_6-σ^B was eluted under peak B and the fractions under peak A contained contaminant proteins and were discarded. The fractions under peak B were pooled together and loaded on a SEC column (Superdex 75 16/60, GE Healthcare). The typical elution profile obtained during the SEC of His_6-σ^B (figure 3.3 panel A) shows one peak eluted outside the range of resolution of the column (peak A, elution volume 45 ml) and a second peak that corresponds to an apparent molecular weight of ~40 kDa (peak B, elution volume 63 ml). The SDS-PAGE analysis of the fractions collected during the chromatography reveals that His_6-σ^B was eluted under both peaks (figure 3.3 panel B). However, the elution volume found for His_6-σ^B eluted under peak A suggests that this fraction of the protein might be aggregated. Furthermore, the fractions under peak A also contain several other contaminant proteins and for these reasons were discarded. Otherwise, His_6-σ^B eluted under peak B is relatively pure and its apparent molecular weight is of ~40 kDa, which is consistent with a monomeric oligomeric status of the protein (the calculated molecular weight of His_6-σ^B is of 39 kDa). Thus, only His_6-σ^B eluted under peak B was kept for further experiments. In order to increase the fraction of His_6-σ^B eluted under peak B several optimization attempts were undertaken (*e.g.* reduction of the concentration of the sample loaded on the column and increase of the ionic strength of the buffer used during SEC). However, the amount of

monomeric His_6-σ^B purified remained essentially unchanged. The protocol described led to a purification yield of ~2 mg of His_6-σ^B for each litre of initial *E. coli* culture.

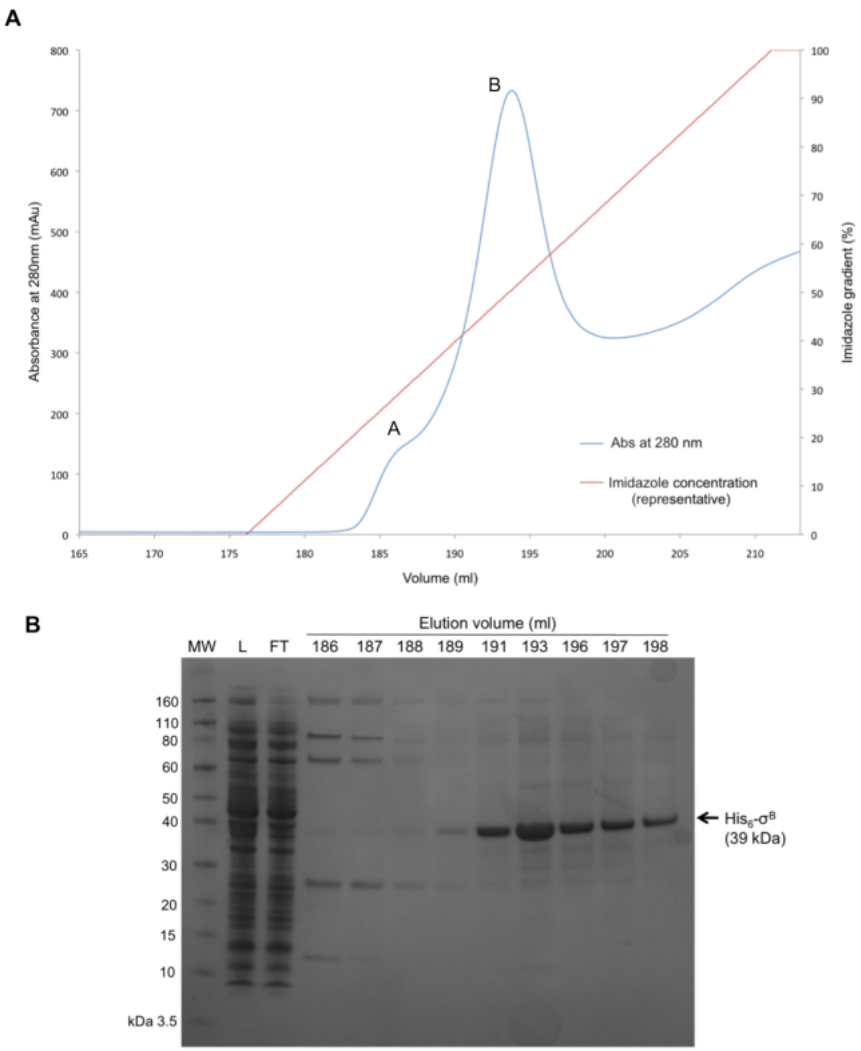


Figure 3.2. His_6-σ^B purification by IMAC. A. The typical elution profile obtained during the purification of His_6-σ^B by IMAC. Two peaks were detected during the elution of His_6-σ^B and in the chromatogram are labelled with the letters A and B. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is the molecular weight markers, lane L the column load, lane FT the column flow-through and lanes 186-198 are labelled according to the elution volume (ml).

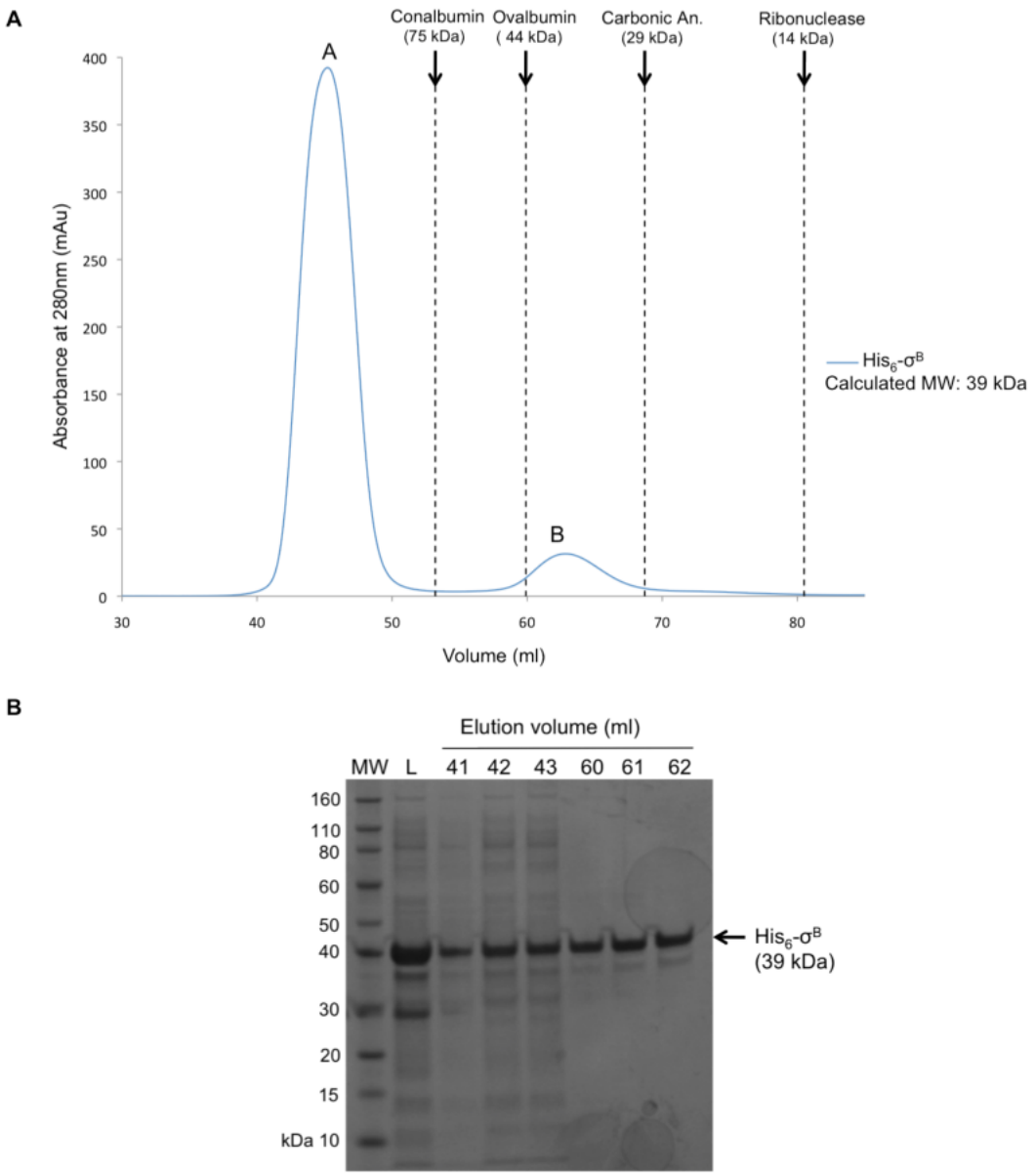


Figure 3.3. His_6-σ^B purification by SEC. A. The typical elution profile obtained during the purification of His_6-σ^B by SEC (Superdex 75 16/60). In the chromatogram two main elution peaks were detected and are labelled with the letters A and B. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is the molecular weight markers, lane L the column load and lanes 41-62 are labelled according to the elution volume (ml).

3.3.2 SEC interaction assay between RbpA and His_6-σ^B.

A yeast two-hybrid screen followed by a protein fragment complementation assay, performed before the research project described in this thesis began, suggested that RbpA could interact with the RNAP σ^B-subunit (142). Given the availability of highly purified RbpA (section 2.3.2) and His_6-σ^B (section 3.3.1) it was possible to test the interaction between the two proteins *in vitro* by SEC interaction assays. RbpA and His_6-σ^B samples were loaded on an analytical SEC column (Superdex 75 10/300 GL) individually or mixed together with a molar ratio of 1:1. Figure 3.4 shows an overlay of the resulting chromatograms and the SDS-PAGE page analysis of the fractions collected during the different runs. RbpA and His_6-σ^B analyzed individually give rise to elution peaks at 11.7 ml and 10.2 ml, respectively. Interestingly, when the two proteins are mixed together only one main elution peak is detected with an elution volume of 9.4 ml (figure 3.4 panel A) suggesting that RbpA and His_6-σ^B are co-eluted. The co-elution of the two proteins is confirmed by the SDS-PAGE analysis of the fractions collected during the different runs (figure 3.4 panel B). The co-elution of RbpA and His_6-σ^B in SEC demonstrates that the two proteins interact with each other and form a tight complex. Furthermore, because no major contaminants were present in the preparations of both RbpA and His_6-σ^B (figure 3.4 panel B, L lanes), it is possible to conclude that the formation of the RbpA•σ^B complex is by direct interaction and does not involve additional proteins.

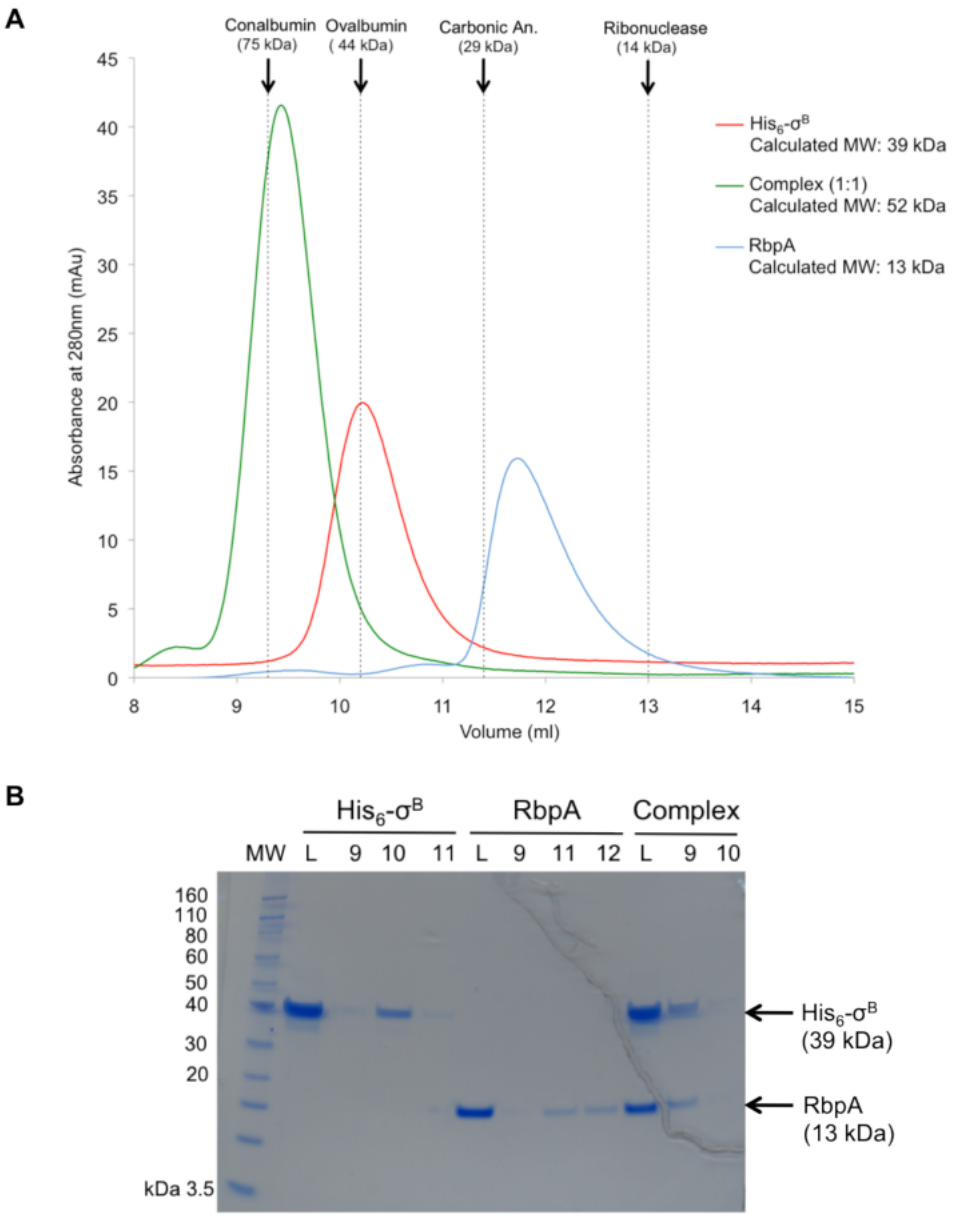


Figure 3.4. RbpA binds and forms a tight complex with His_6-σ^B. A. An overlay of the elution profiles from a SEC column (Superdex 75 10\300 GL) of RbpA (10 μM) or His_6-σ^B (10 μM) individually and the complex formed between them (1:1 molar ratio). The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the different SEC runs. Lane MW is molecular weight markers, lanes L the column load and lanes 9-12 labelled according to the elution volume (ml).

3.3.3 Expression and purification of His_6-σ^A.

The expression and purification of His_6-σ^A were performed following the protocols adopted for His_6-σ^B (section 3.3.1). The gel shown in figure 3.5 shows the result of the SDS-PAGE analysis of culture samples collected pre-induction and ~20 hours post-induction of His_6-σ^A expression. Following induction an overexpression band is produced, it is soluble but migrates in a position slightly higher than that expected for the calculated molecular weight of His_6-σ^A (60 kDa). Similar anomalous migration of *M. tuberculosis* σ^A during SDS-PAGE analysis was reported previously (147).

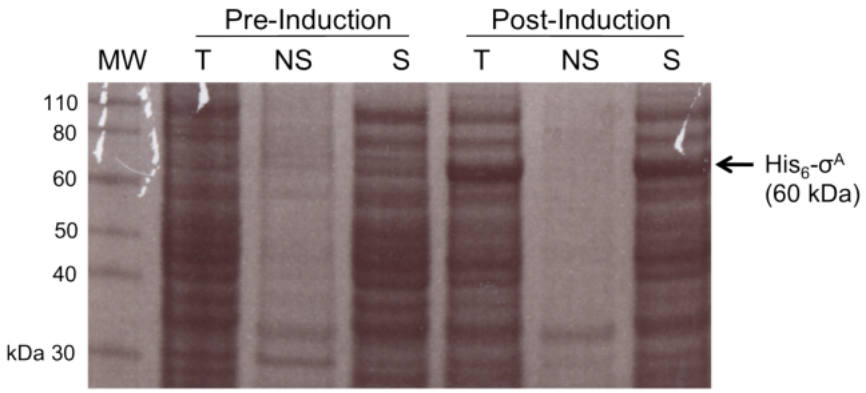


Figure 3.5. His_6-σ^A expression. SDS-PAGE analysis of culture samples collected pre-induction and ~20 hours post-induction of His_6-σ^A expression. Lane MW is molecular weights markers, lanes T the total cell lysate, lanes NS the non-soluble fraction and lanes S soluble fractions.

In figure 3.6 panel A, which shows the elution profile obtained during the first step of the purification of His_6-σ^A (IMAC), a single elution peak was detected (elution volume 193 ml). SDS-PAGE analysis of the fractions collected during different stages of the chromatography reveals that His_6-σ^A was eluted under this peak (figure 3.6 panel B). However, a number of other bands are clearly detectable in the same fractions where His_6-σ^A was eluted. In order to increase the purity of the protein preparation the

fractions containing His_6-σ^A were pooled together and loaded on a SEC column (Superdex 75 16/60, GE Healthcare). The typical elution profile obtained during the SEC of His_6-σ^A and the SDS-PAGE analysis of the fractions collected during the chromatography are shown in figure 3.7 panel A and panel B, respectively. At least three elution peaks are distinguishable in the chromatogram shown in figure 3.7 panel A. Peaks A and B are almost completely overlapped and both contain His_6-σ^A (figure 3 panel B), otherwise peak C does not contain His_6-σ^A (results not shown). The fractions collected under peak A (elution volume 47-51 ml) are contaminated with high molecular weight proteins (~160 kDa) and were discarded. Only the fractions eluted under peak B and with minor low molecular weight contamination (elution volume 52-54 ml) were kept and used in further experiments. His_6-σ^A is eluted close to the upper-limit of resolution of the column used and is not possible assess the oligomeric status of the protein. The protocol described led to a purification yield of ~2 mg of His_6-σ^A for each litre of initial *E. coli* culture.

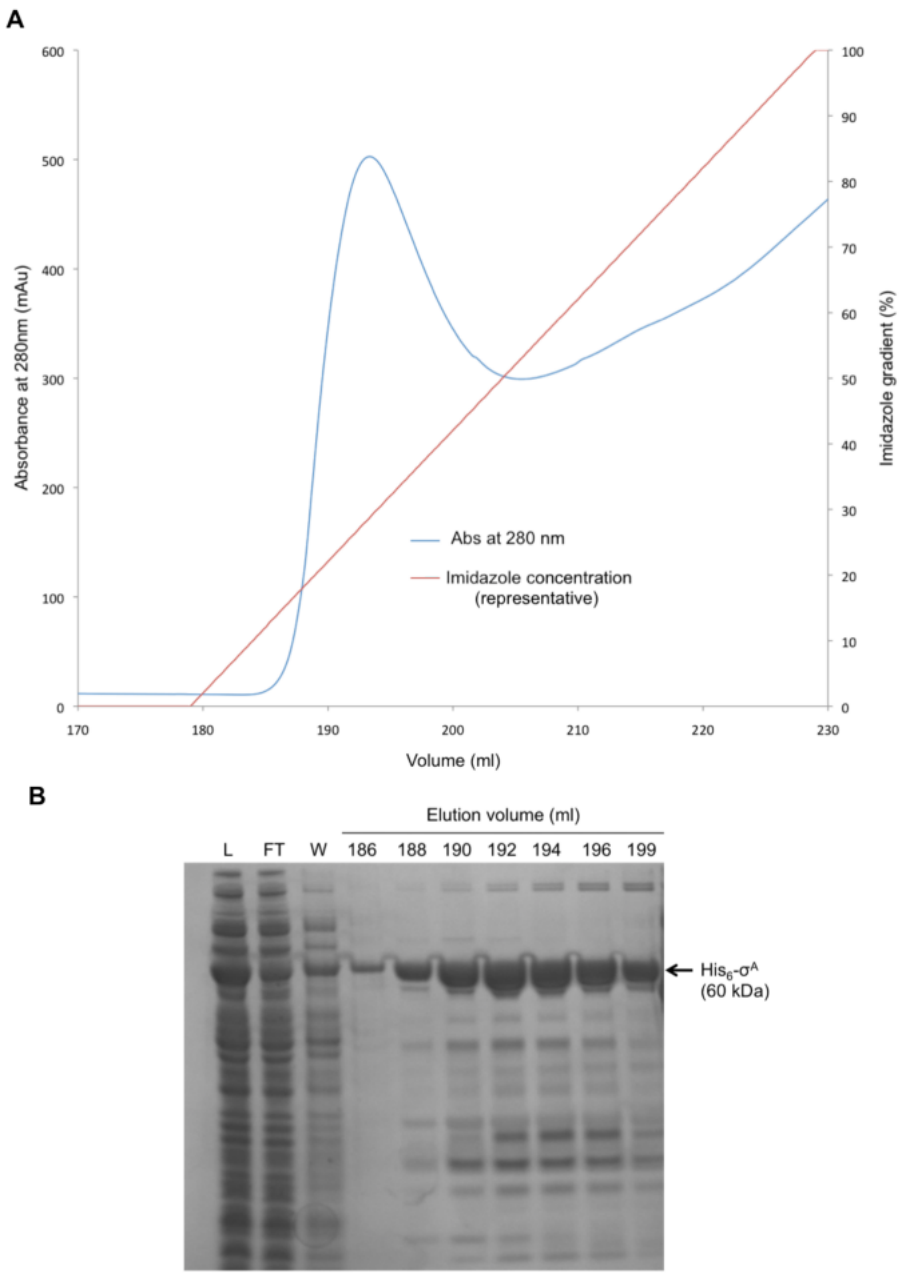


Figure 3.6. His_6-σ^A purification by IMAC. A. The typical elution profile obtained during the purification of His_6-σ^A by IMAC. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane L is the column load, lane FT the column flow-through and lanes 186-199 are labelled according to the elution volume (ml).

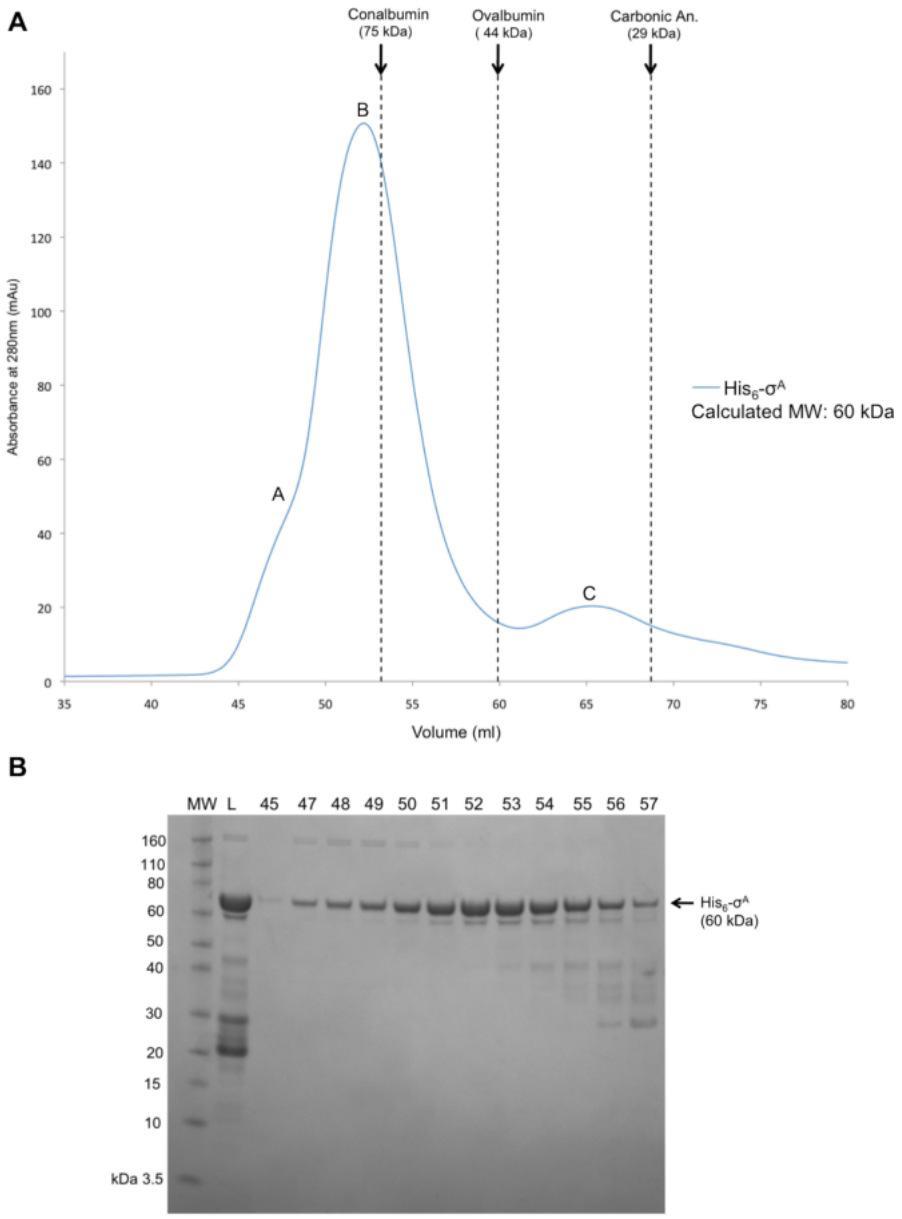


Figure 3.7. His_6-σ^A purification by SEC. A. The typical elution profile obtained during the purification of His_6-σ^A by SEC (Superdex 75 16/60). In the chromatogram are distinguishable three elution peaks that are labelled with the letters A, B and C. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is the molecular weight markers, lane L the column load and lanes 45-57 are labelled according to the elution volume (ml).

3.3.4 SEC interaction assay between RbpA and His_6-σ^A.

σ^A is the primary sigma factor of *M. tuberculosis*, it is responsible for the transcription of housekeeping genes and is essential for the growth of the bacteria. Based on the high sequence homology with σ^A and σ^B (63% of amino acid identity) it was hypothesized that σ^A could interact with RbpA. To test this hypothesis RbpA and His_6-σ^A samples were loaded on an analytical SEC column (Superdex 200 10/300 GL) individually or mixed together with a molar ratio of 1:1. In figure 3.8 are shown an overlay of the resulting chromatograms and the SDS-PAGE analysis of the fractions collected during the different runs. When RbpA and His_6-σ^A are mixed together only one main elution peak is detected, which is eluted earlier than the peaks detected when the two proteins were analyzed individually (figure 3.8 panel A) suggesting that RbpA and His_6-σ^A are co-eluted. The co-elution of the two proteins is confirmed by the SDS-PAGE analysis of the fractions collected during the different runs (figure 3.8 panel B). This experiment shows that RbpA binds and forms a stable complex also with the primary sigma factor of *M. tuberculosis* (σ^A).

His_6-σ^A was eluted at 12.3 ml (figure 3.8 panel A, red trace) that corresponds to an apparent molecular weight of ~220 kDa. Given the calculated His_6-σ^A molecular weight of 60 kDa the elution position of His_6-σ^A during SEC suggests that the protein has an oligomeric status between trimer and tetramer. However, the functional oligomeric organization of σ-factors is thought to be monomeric (56,57,59,60). The anomalous apparent molecular weight found for *M. tuberculosis* σ^A could be due to an extended structure of the protein or the no-physiological conditions in which the SEC experiment was performed.

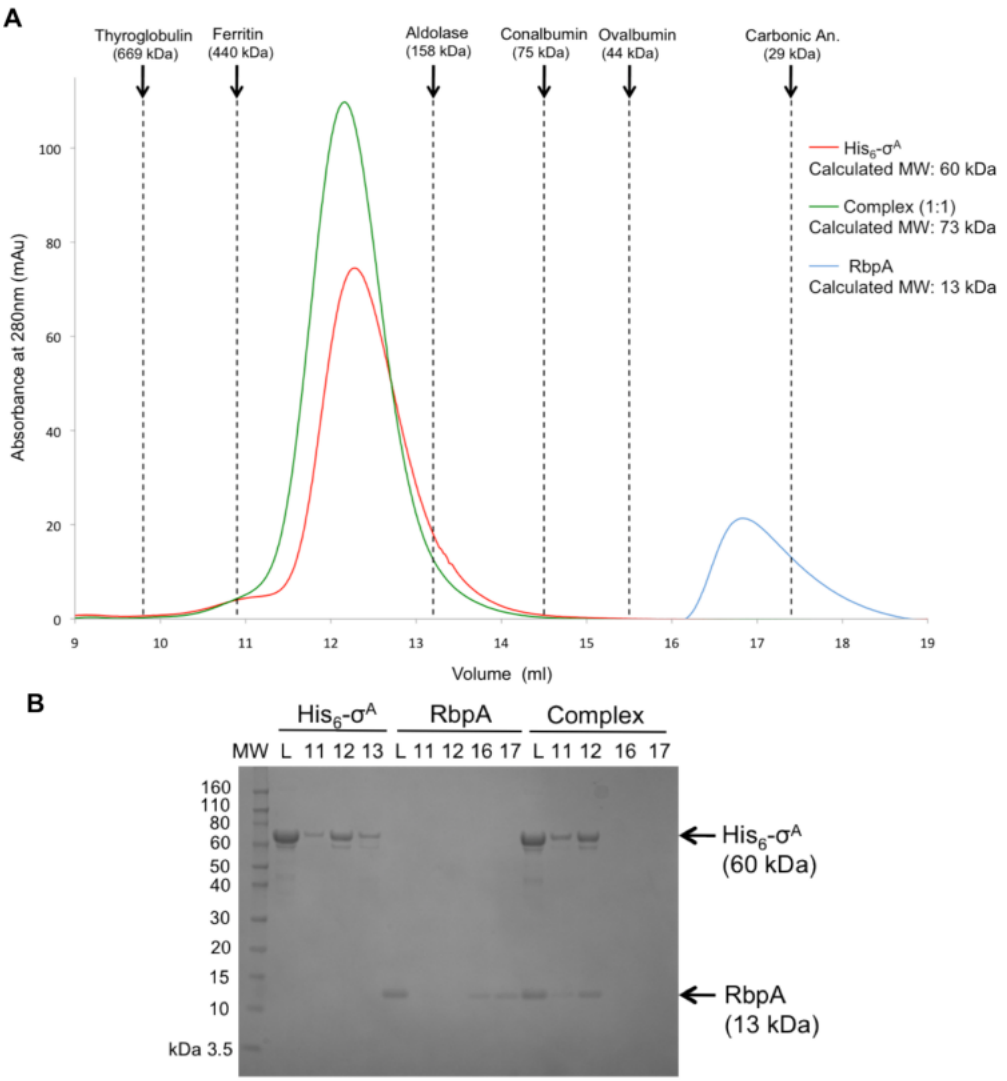


Figure 3.8. RbpA binds and forms a tight complex with His_6-σ^A. A. An overlay of the elution profiles from a SEC column (Superdex 200 10\300) of RbpA (12 μM) or His_6-σ^A (12 μM) individually and the complex formed between them (1:1 molar ratio). The elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** SDS-PAGE analysis of the fractions collected during the different SEC runs. Lane MW is molecular weight markers, lanes L the column load and lanes 11-17 labelled according to the elution volume (ml).

3.3.5 Design of expression vectors for the purification of truncated versions of σ^B.

To perform an initial mapping of the regions of σ^B responsible for the interaction with RbpA it was decided to purify two truncated versions of σ^B: $\sigma^B_{1\text{-}228}$ (residues 1-228) that comprises the conserved regions 1.2-3.1 and $\sigma^B_{228\text{-}323}$ (residues 228-323) that comprises the conserved region 4. The site of the truncation was chosen on the basis of multiple sequence alignment and using the *M. tuberculosis* σ^B structural model shown in figure 3.9. The model was calculated using the software Phyre (143) on the basis of the of the structure of the principal σ-subunit of *T. thermophilus* (PDB 1IW7, 45% sequence identity). The resulting model structure (figure 3.9) comprises σ^B residues 21-322 (93.5% of the whole σ^B sequence) and was determined with an accuracy score > 90%, which presumes an RMSD to the native structure of 2-4 Å.

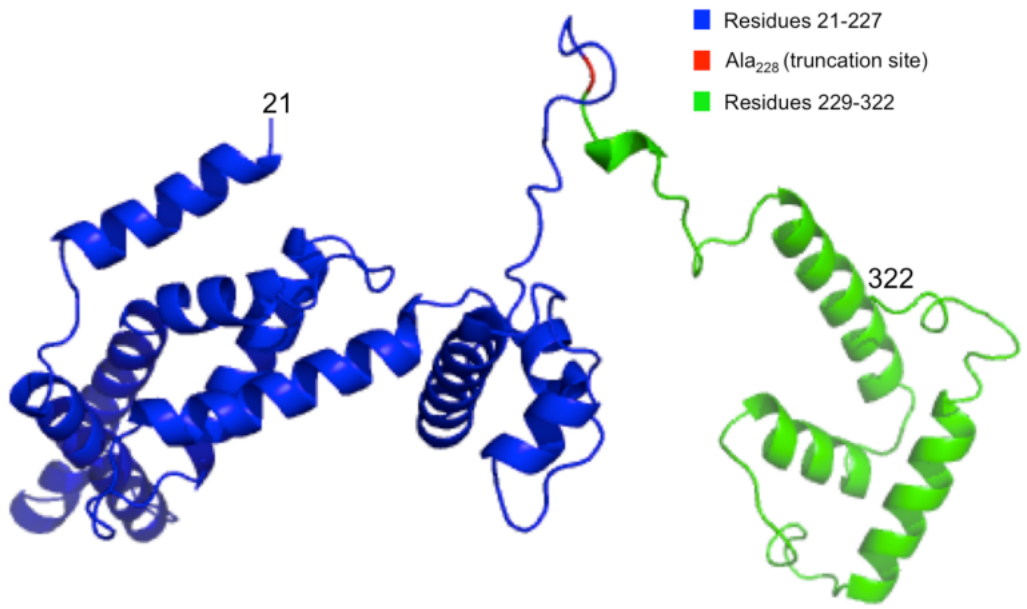


Figure 3.9. Structural model of the *M. tuberculosis* σ^B. The structural model of the *M. tuberculosis* σ^B here reported was calculated using the software Phyre (143) on the basis of the structure of the principal σ-subunit of *T. thermophilus* (PDB 1IW7, 45% sequence identity). The regions of the model that encompass the σ^B_{1-228} and $\sigma^B_{228-323}$ domains are shown in blue and green, respectively. The truncation site (Ala_{228}) is highlighted in red.

3.3.6 Expression of GST-σ^B_{1-228} and GST-$\sigma^B_{228-323}$.

The expression of GST-σ^B_{1-228} and GST-$\sigma^B_{228-323}$ was performed following the procedure described for His_6-σ^B (section 3.3.1). The gels reported in figure 3.10 show that for both constructs an overexpression band is produced following induction, it is soluble and migrates in a position consistent with the calculated molecular weight of GST-σ^B_{1-228} (51 kDa) and GST-$\sigma^B_{228-323}$ (36 kDa).

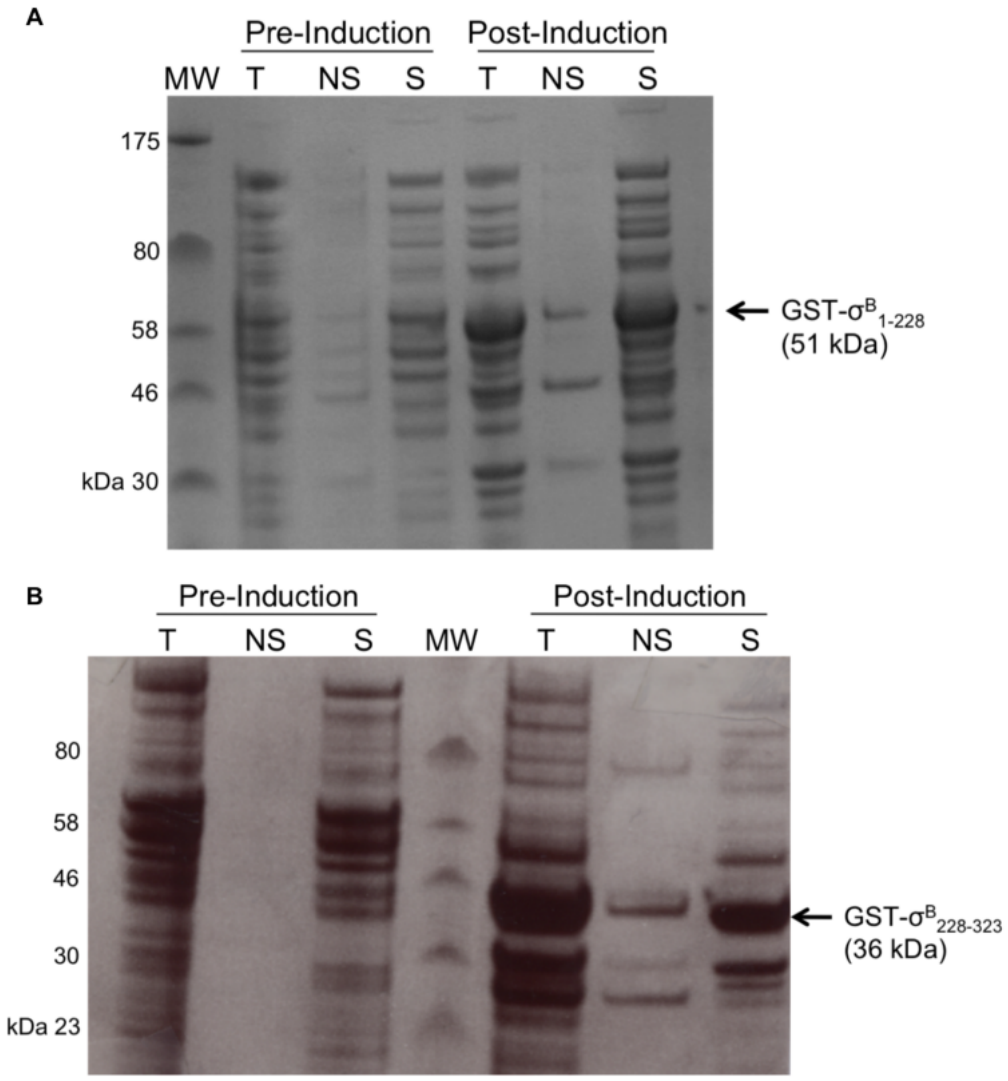


Figure 3.10. Expression of GST-σ^{B}_{1-228} and GST-$\sigma^{B}_{228-323}$. SDS-PAGE analysis of culture samples collected pre-induction and ~20 hours post-induction of GST-σ^{B}_{1-228} (**A**) and GST-$\sigma^{B}_{228-323}$ (**B**) expression. Lanes MW are molecular weights markers, lanes T the total cell lysate, lanes NS the non-soluble fraction and lanes S soluble fractions.

3.3.7 Purification of GST-σ^B_{1-228}.

The soluble fraction resulting from the lysis of the cell culture expressing GST-σ^B_{1-228} was loaded on a GSTrap HP column (GE Healthcare) and subsequently GST-σ^B_{1-228} was eluted with a buffer containing 10 mM reduced glutathione. Figure 3.11 shows the typical elution profile obtained (panel A) and the SDS-PAGE analysis of the fractions collected at different stages of the chromatography (panel B). The SDS-PAGE analysis revealed that all the fractions analyzed contain GST-σ^B_{1-228} (figure 3.11 panel B). However, a number of other bands were detected in the same fractions where GST-σ^B_{1-228} was eluted. In order to increase the purity of the protein preparation a second purification step was performed loading the fractions containing GST-σ^B_{1-228} on a SEC column (Superdex 75 16/60, GE Healthcare). The typical elution profile obtained during SEC for GST-σ^B_{1-228} is shown in figure 3.12 panel A. It shows one peak eluted outside the range of resolution of the column (peak A, elution volume 45 ml) and a second peak that corresponds to an apparent molecular weight of ~40 kDa (peak B, elution volume 60 ml). The SDS-PAGE analysis of samples from the fractions collected during the chromatography (figure 3.12 panel B) reveals that GST-σ^B_{1-228} was eluted under peak A, whereas peak B contains low molecular weight bands (< 30 kDa). Only the fractions eluted under peak A were kept and used in further experiments. Because GST-σ^B_{1-228} was eluted outside the range of resolution of the column (6.5-75 kDa) the oligomeric organization of the protein remains unknown. However, the SEC run increased the purity of the GST-σ^B_{1-228} preparation, as can be appreciated comparing lane L with lanes 44-48 of the gel shown in figure 3.12 panel B, although some lower molecular weight contaminants remained.

The protocol described led to a purification yield of ~18 mg of GST-σ^B_{1-228} for each litre of initial *E. coli* culture.

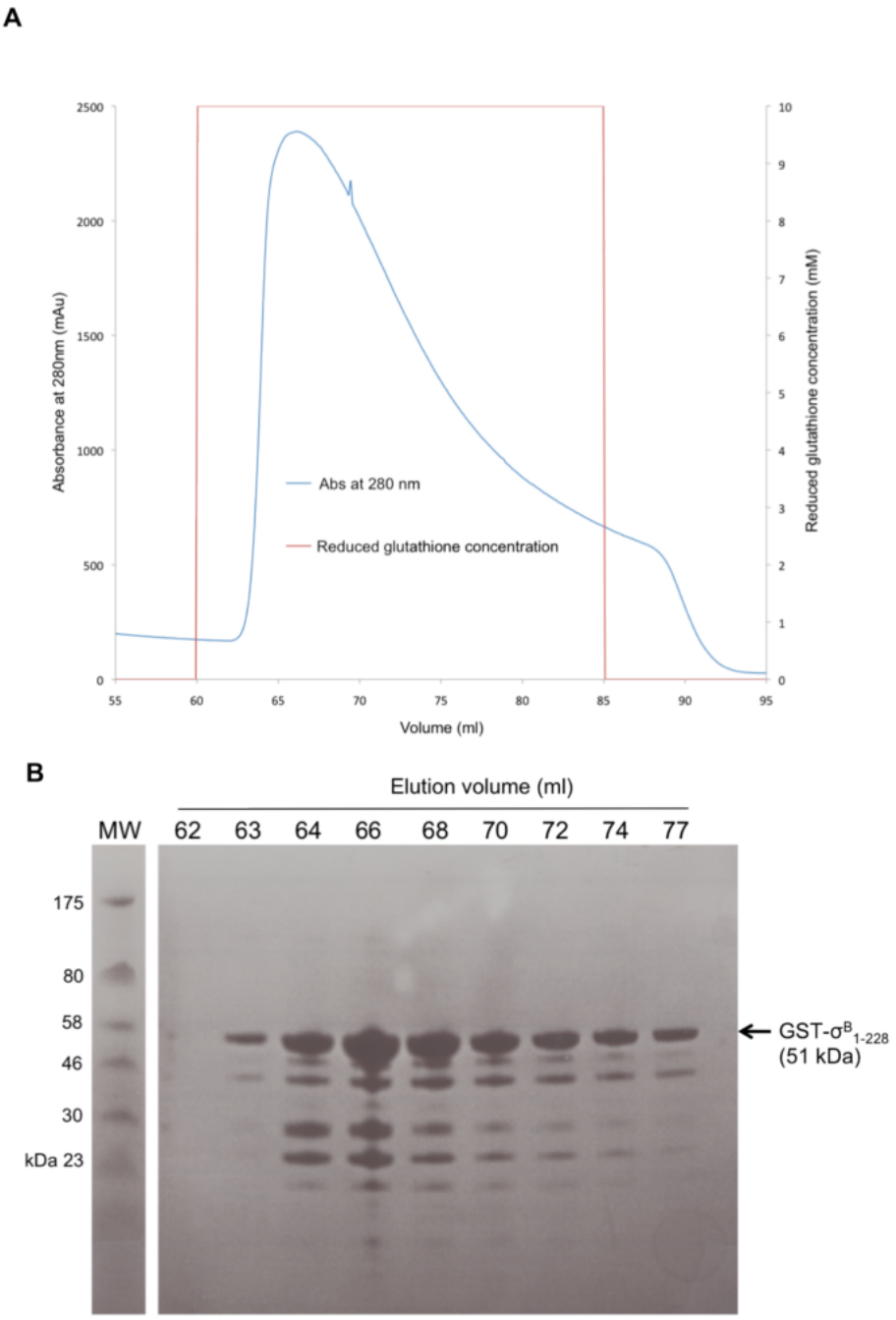


Figure 3.11. GST-$\sigma^{B}_{1\text{-}228}$ purification by affinity chromatography. A. The typical elution profile obtained during the purification of GST-$\sigma^{B}_{1\text{-}228}$ by affinity chromatography. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is molecular weight markers and lanes 62-77 are labelled according to the elution volume (ml).

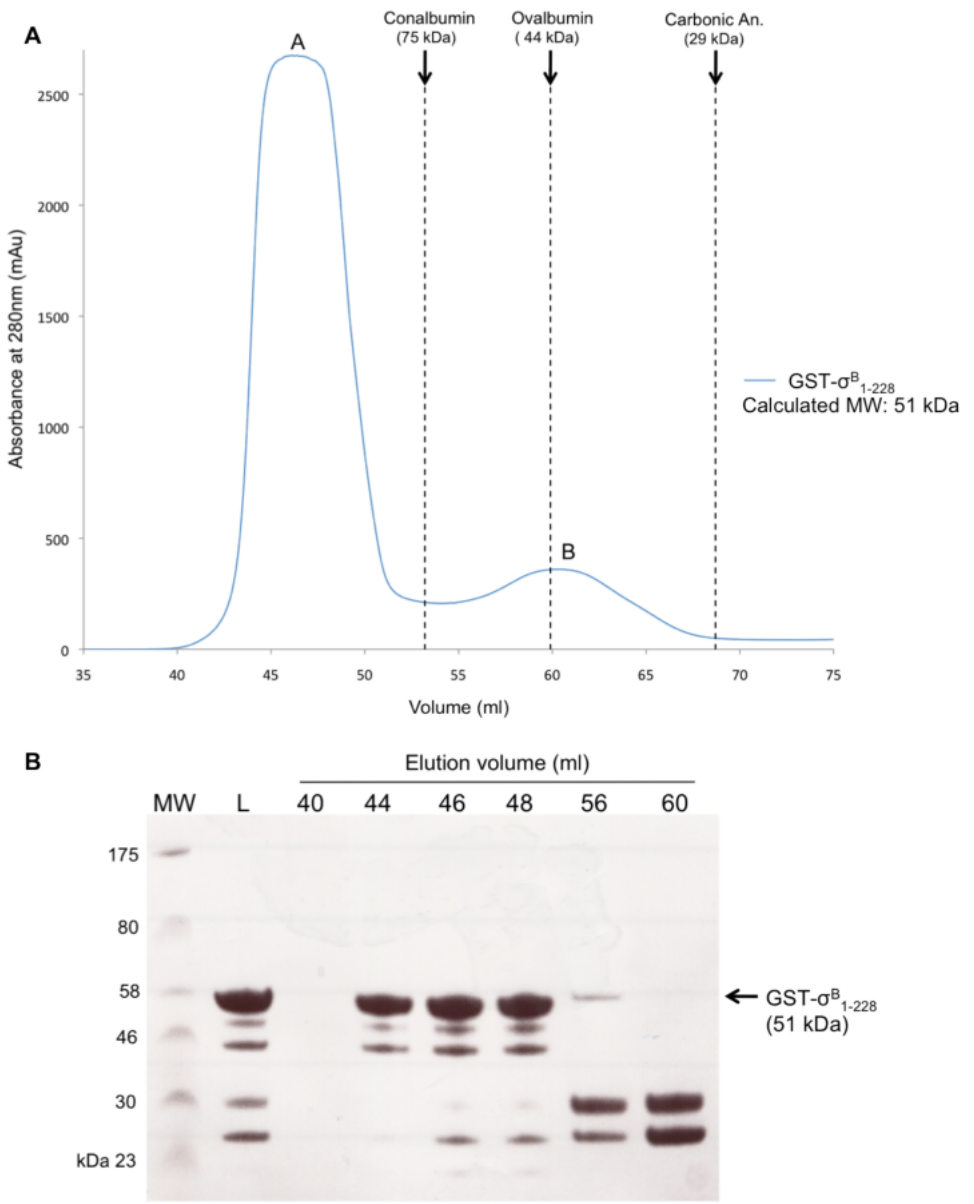


Figure 3.12. GST-$\sigma^{B}_{1\text{-}228}$ purification by SEC. A. The typical elution profile obtained during the purification of GST-$\sigma^{B}_{1\text{-}228}$ by SEC (Superdex 75 16/60). In the chromatogram two main elution peaks are distinguishable and are labelled with the letters "A" and "B". The elution volumes of the protein molecular weight standards used to calibrate the column are indicated by dashed lines. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is the molecular weight markers, lane L the column load and lanes 40-60 are labelled according to the elution volume (ml).

3.3.8 Pull-down assay of RbpA binding to GST-σ^B_{1-228}.

The interaction between RbpA and σ^B_{1-228} was assessed by a pull-down assay and the SDS-PAGE analysis of samples collected at different stages of the experiment is shown in figure 3.13. GST-σ^B_{1-228} was loaded on a column containing glutathione-agarose resin and the unbound protein was washed with 5 ml of binding buffer. In the column flow-through (figure 3.13 lane FT) and in the first fraction collected during the washing step (figure 3.13 lane 1) there is a band that corresponds to GST-σ^B_{1-228}, indicating that not all the GST tagged protein bound to the resin. Afterwards, untagged RbpA was added to the column and a further washing step was performed using 10 ml of binding buffer. Finally, GST-σ^B_{1-228} bound to the resin was eluted in 3 ml of buffer containing 10 mM of reduced glutathione. In the first fraction collected during the second washing step (figure 3.13 lane 3) an intense band corresponding to RbpA was detected, indicating that most of the untagged protein was not retained. However, in the first and second fractions collected during the elution step (figure 3.13 lanes E_1 and E_2) a faint band that corresponds to RbpA is detectable. The same band is not visible in the final fraction collected during the washing (figure 3.13 lane 4), suggesting that RbpA was specifically bound to GST-σ^B_{1-228}.

The presence of a faint band corresponding to RbpA in E_1 and E_2 lanes (figure 3.13) suggests that the first 228 residues of σ^B could be sufficient to bind with RbpA. However, the pull-down assay described in this section cannot be considered a definitive evidence of the interaction between σ^B_{1-228} and RbpA. Indeed, is not clear why only a minor part of RbpA (< 5%) seems to interact with GST-σ^B_{1-228}. Furthermore, no controls were run to check if untagged RbpA interacts with the column used in the pull-down assay and several contaminant bands are detectable in the

preparation of GST-σ^{B}_{1-228} (figure 3.13, lane L of GST-σ^{B}_{1-228}) that could potentially interfere with the interaction assay.

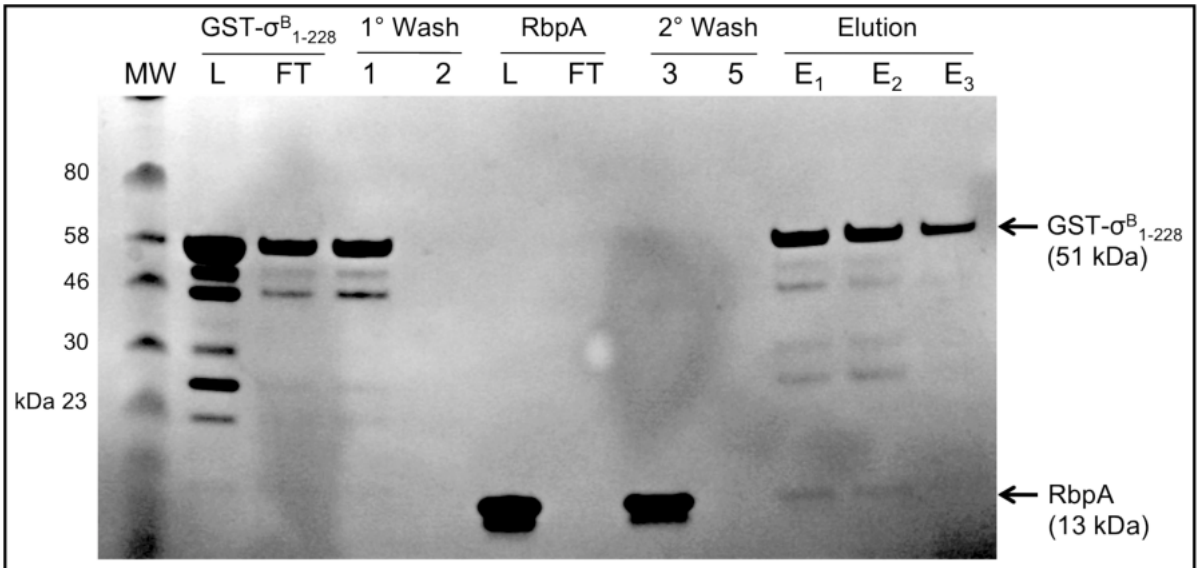


Figure 3.13. Pull-down assay of RbpA binding to GST-σ^{B}_{1-228}. The gel resulting from the SDS-PAGE analysis of samples collected at different stages of the pull-down assay. Lane MW is the molecular weight markers, lanes L the column load, lanes FT the column flow-through, lanes 1 and 2 the first and last fractions collected during the first wash step, lanes 3 and 4 the first and last fractions collected during the second wash step and lanes E_1-E_3 the fractions collected during the elution step.

3.3.9 Purification of GST-$\sigma^{B}_{228-323}$.

The purification of GST-$\sigma^{B}_{228-323}$ was performed following the protocol adopted for GST-σ^{B}_{1-228} (section 3.3.7). In figure 3.14 are shown the typical elution profile obtained during the purification of GST-$\sigma^{B}_{228-323}$ by affinity chromatography (panel A) and the SDS-PAGE analysis of the fractions collected at different stages of the purification (panel B). The SDS-PAGE analysis of a sample collected from the column flow-through (figure 3.13 lane FT) reveals that not all the GST-tagged protein was retained into the column. However, most of the protein was bound to the glutathione-agarose resin as is evident from the large amount of GST-$\sigma^{B}_{228-323}$ (figure 3.14 panel B) contained in the fractions collected under the main elution peak (figure 3.14 panel A). Fractions containing GST-$\sigma^{B}_{228-323}$ were pooled together and dialyzed in a buffer containing 25 mM Tris-HCl pH 7.4 and 150 mM NaCl. A 20 μM GST-$\sigma^{B}_{228-323}$ sample was analyzed by SDS-PAGE (figure 3.14 panel C) showing that the protein preparation does not contain major contaminants, thus no further purification steps were performed. The protocol described led to a purification yield of ~25 mg of GST-$\sigma^{B}_{228-323}$ for each litre of initial *E. coli* culture.

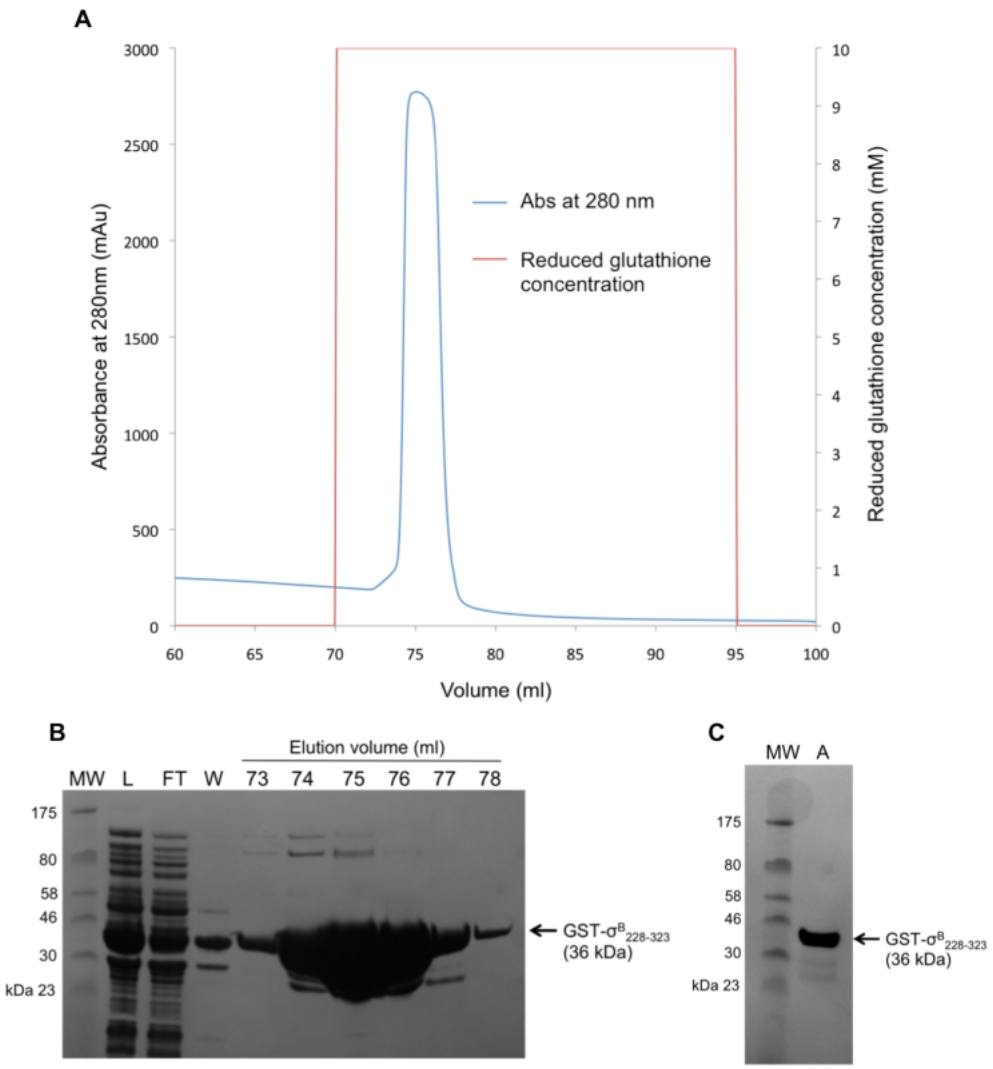


Figure 3.14. GST-$\sigma^{B}_{228-323}$ purification by affinity chromatography. A. The typical elution profile obtained during the purification of GST-$\sigma^{B}_{228-323}$ by affinity chromatography. **B.** The gel resulting from the SDS-PAGE analysis of the fractions collected during the purification process. Lane MW is molecular weight markers, lane L the column load, lane FT the column flow-through, lane W the column wash and lanes 73-78 are labelled according to the elution volume (ml). **C.** The gel resulting from the SDS-PAGE analysis of a 20 μM GST-$\sigma^{B}_{228-323}$ sample (lane A). Lane MW is molecular weight markers.

3.3.10 Pull-down assay of RbpA binding to GST-$\sigma^{B}_{228\text{-}323}$.

The interaction between RbpA and GST-$\sigma^{B}_{228\text{-}323}$ was assessed by a pull-down assay following the procedure described in section 3.3.8. Figure 3.15 shows the SDS-PAGE analysis of samples collected at different stages of the experiment. It was not possible to detect RbpA in any of the elution fractions (E_1-E_3) suggesting that residues 228-323 of σ^{B} are not sufficient to form a stable interaction with RbpA.

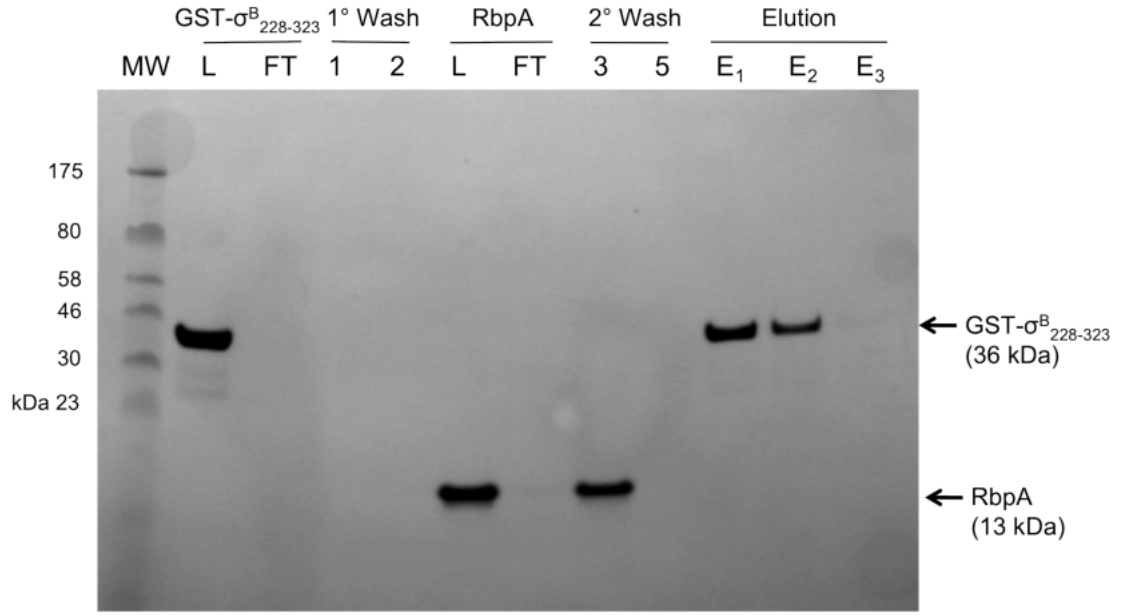


Figure 3.15. Pull-down assay of RbpA binding to GST-$\sigma^{B}_{228\text{-}323}$. The gel resulting from the SDS-PAGE analysis of samples collected at different stages of the pull-down assay. Lane MW is the molecular weight markers, lanes L the column load, lanes FT the column flow-through, lanes 1 and 2 the first and last fractions collected during the first wash step, lanes 3 and 4 the first and last fractions collected during the second wash step and lanes E_1-E_3 the fractions collected during the elution step.

3.3.11 SEC interaction assay between His_6-RbpA and His_6-σ^B_{1-228}.

The set of pull-down assays performed in section 3.3.8 suggested that the first 228 residues of σ^B are sufficient for the binding with RbpA. To confirm this initial hypothesis it was decided to test the interaction between RbpA and σ^B_{1-228} by a SEC interaction assay.

The ORF encoding for the σ^B residues 1-228 was cloned into an expression vector that allowed the over expression of σ^B_{1-228} in fusion with a cleavable His_6-tag at the N-terminus (section 3.2.1). The resulting construct, His_6-σ^B_{1-228}, was expressed and purified following the procedure described for His_6-σ^B full length (section 3.3.1). The purification process allowed the purification of ~3 mg of His_6-σ^B_{1-228}, in a monomeric form, for each litre of initial *E. coli* culture (results not shown).

The SEC interaction assay between His_6-RbpA and His_6-σ^B_{1-228} was performed following the procedure described in section 3.3.2. His_6-RbpA and His_6-σ^B_{1-228} samples were loaded on an analytical SEC column (Superdex 75 10/300 GL) individually or mixed together with a molar ratio of 1:1. When the two proteins were mixed together only one main elution peak was detected, which is eluted earlier than the peaks detected when the two proteins were analyzed individually (figure 3.16 panel A). This indicates that His_6-RbpA and His_6-σ^B_{1-228} were co-eluted and it was confirmed by the SDS-PAGE analysis of the fractions collected during the different runs (figure 3.16 panel B). The experiment described confirms that residues 1-228 of σ^B are sufficient to bind and form a stable complex with RbpA.

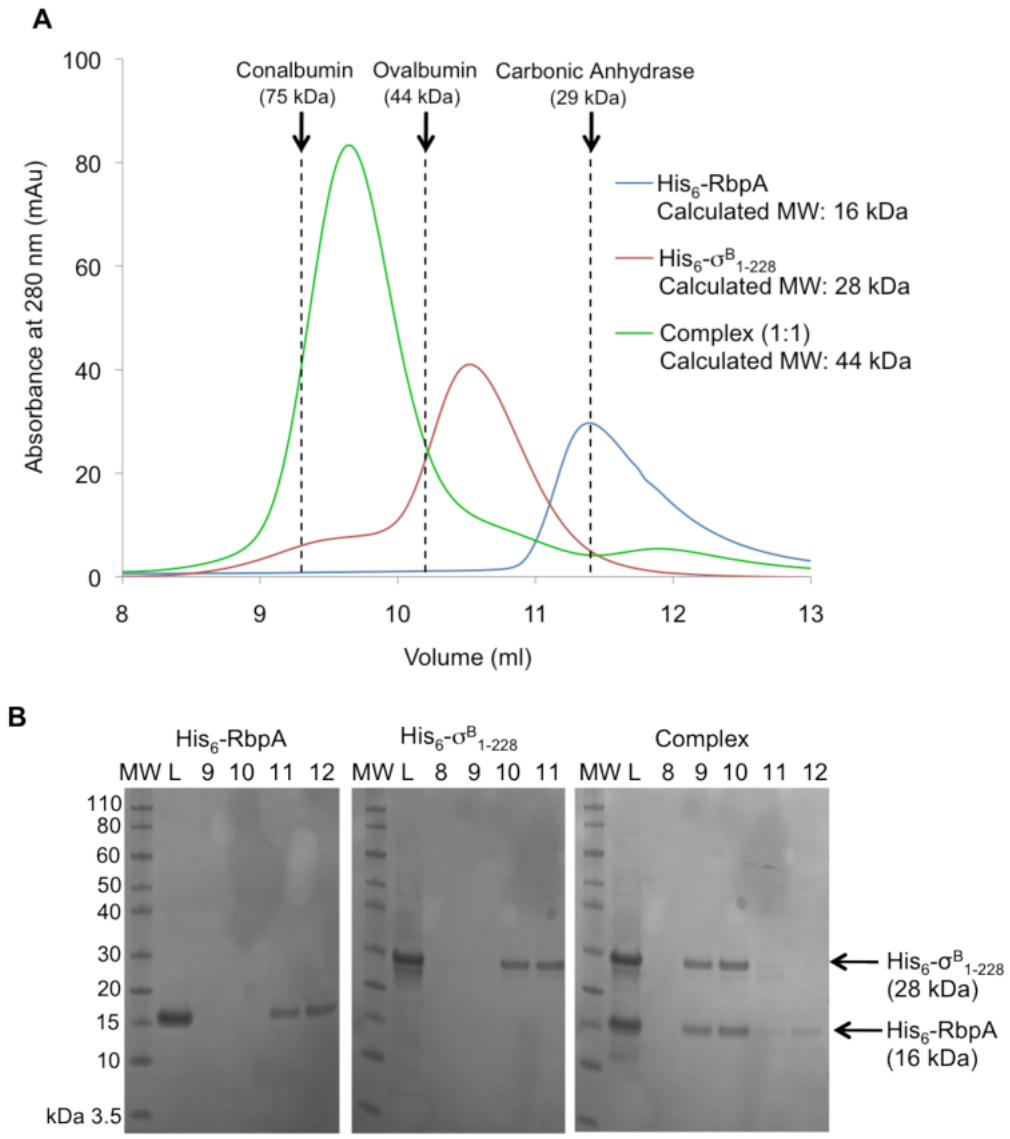


Figure 3.16. RbpA binds and forms a tight complex with His_6-σ^B_{1-228}. A. An overlay of the elution profiles from a SEC column (Superdex 75 10\300 GL) of RbpA (15 μM) or His_6-σ^B_{1-228} (15 μM) individually and the complex formed between them (1:1 molar ratio). The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the different SEC runs were analysed by SDS-PAGE. Lanes MW are molecular weight markers, lanes L the column load and lanes 8-12 labelled according to the elution volume (ml).

3.3.12 The stoichiometry of the RbpA•σ^{B}_{1-228} complex.

In order to investigate the stoichiometry of the complex formed between RbpA and σ^{B}_{1-228} the SEC interaction assay described in the section 3.3.11 was repeated using different molar ratios of RbpA:σ^{B}_{1-228}: 2:1 and 1:2 (figures 3.17 and 3.18, respectively). In each case the His_6-RbpA•His_6-σ^{B}_{1-228} complex was formed, as indicated by the presence of a peak eluted at ~9.7 ml (peaks A in panels A of figures 3.17 and 3.18), and the protein loaded in excess was eluted as expected for the free protein (peaks B in panels A of figures 3.17 and 3.18). The SDS-PAGE analyses of the fractions collected during the different SEC runs confirm that both His_6-RbpA and His_6-σ^{B}_{1-228} are found under the peaks A and that only the proteins loaded in excess were detected under peaks B (panels B of figures 3.17 and 3.18). The set of SEC experiments described suggests that RbpA binds to σ^{B}_{1-228} with an equal molar ratio. The apparent molecular weight found for the His_6-RbpA•His_6-σ^{B}_{1-228} complex during SEC analyses (figures 3.16, 3.17 and 3.18) was of ~62 kDa that is higher than the expected molecular weight (44 kDa) for the complex occurring with a 1:1 stoichiometry. However, a His_6-RbpA•His_6-σ^{B}_{1-228} complex occurring with a 2:2 stoichiometry would have a calculated molecular weight of 88 kDa and in SEC experiments should be eluted in the void volume because it would be outside the range of resolution of the column (6.5-75 kDa). The peak corresponding to the His_6-RbpA•His_6-σ^{B}_{1-228} complex was resolved during the SEC analysis (figure 3.16, 3.17 and 3.18) suggesting that the His_6-RbpA•His_6-σ^{B}_{1-228} complex occurs with a 1:1 stoichiometry.

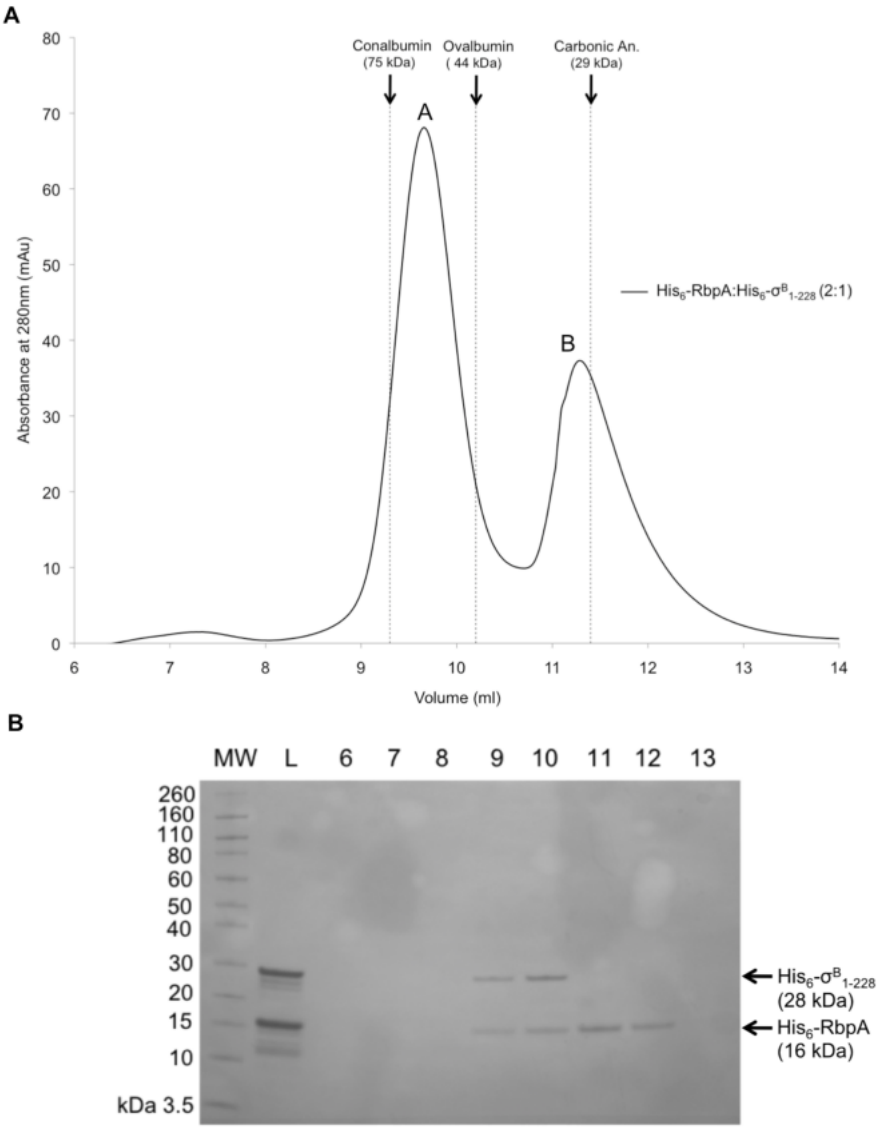


Figure 3.17. SEC analysis of a sample containing His_6-RbpA and His_6-σ^B_{1-228} combined in a 2:1 molar ratio. A. The elution profile from a SEC column (Superdex 75 10\300 GL) of a sample containing His_6-RbpA and His_6-σ^B_{1-228} combined in a 2:1 molar ratio. The two main elution peaks detected in the chromatogram are labelled with the letters A and B. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the SEC run were analysed by SDS-PAGE. Lane MW is molecular weight markers, lane L the column load and lanes 6-13 labelled according to the elution volume (ml).

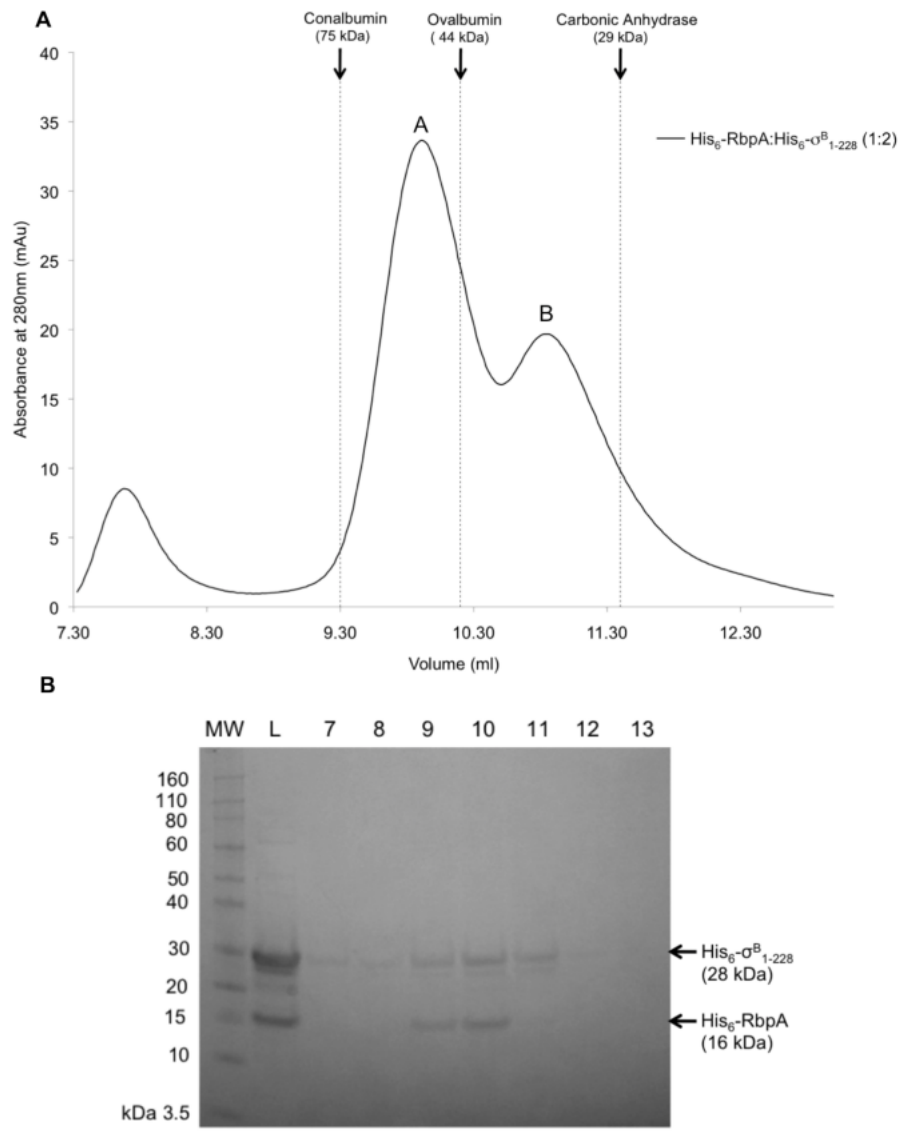


Figure 3.18. SEC analysis of a sample containing His_6-RbpA and His_6-σ^B_{1-228} combined in a 1:2 molar ratio. A. The elution profile from a SEC column (Superdex 75 10\300 GL) of a sample containing RbpA and His_6-σ^B_{1-228} combined in a 2:1 molar ratio. The two main elution peaks detected in the chromatogram are labelled with the letters A and B. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the SEC run were analysed by SDS-PAGE. Lane MW is molecular weight markers, lane L the column load and lanes 7-13 labelled according to the elution volume (ml).

3.3.13 Insights on the interaction between RbpA and σ^B.

The chemical shifts detected during NMR experiments are extremely sensitive to the chemical environment that surrounds each residue of the protein analyzed. Thus, on interaction between a protein and its binding partner the chemical shifts of the residues involved in the interaction are expected to be affected. To better understand the binding mode of RbpA with the σ-subunit it was decided to take advantage of the RbpA NMR assignments to investigate which residues of RbpA are involved in the interaction with $\sigma^B_{1\text{-}228}$.

To this purpose a sample of triply labelled $^{15}N/^{13}C/^{2}H$ RbpA was mixed in an equimolar ratio with unlabelled $His_6\text{-}\sigma^B_{1\text{-}228}$ and the resulting complex was purified by SEC. TROSY spectra were acquired on the complex sample and on a triply labelled sample of $^{15}N/^{13}C/^{2}H$ RbpA in the free form. The differences observed between the two spectra are highlighted in the TROSY spectra overlay shown in figure 3.19 panel A. In total, at least 19 RbpA backbone amide groups are clearly affected by the binding with $His_6\text{-}\sigma^B_{1\text{-}228}$ and ten of these are unassigned resonances. All the unassigned RbpA backbone amide groups – except Met_1, Gly_8 and Arg_{10} – lie in the last 35 residues of the protein (figure 3.19 panel B). Therefore, the relatively high number of perturbations observed for unassigned peaks suggests that the C-terminus of RbpA is affected by the binding with $His_6\text{-}\sigma^B_{1\text{-}228}$. Furthermore, the only resonance that was assigned between the residues 77-111, the indole group of Trp_{82}, is clearly affected by the binding with $His_6\text{-}\sigma^B_{1\text{-}228}$ (figure 3.19 panel A).

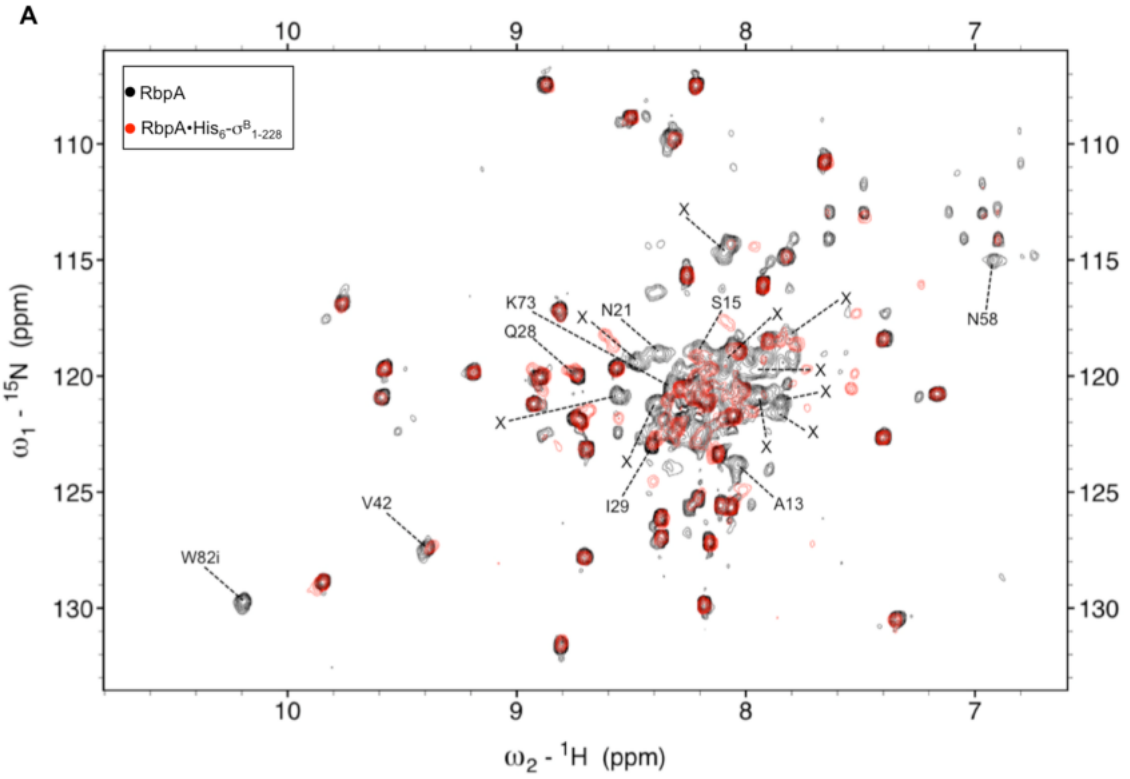


Figure 3.19. RbpA chemical shift changes induced by the binding of His_6-σ^B_{1-228}. A. Overlay of the TROSY spectra acquired at 35 °C on $^{15}N/^{13}C/^{2}H$ RbpA (190 μM) (shown in black) and on $^{15}N/^{13}C/^{2}H$ RbpA•His_6-σ^B_{1-228} complex (100 μM) (shown in red) samples dissolved in a buffer containing 25 mM K_2HPO_4 pH 7.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O. Peaks that have been clearly shifted on formation of the $^{15}N/^{13}C/^{2}H$ RbpA•His_6-σ^B_{1-228} complex have been highlighted. An "i" indicates peaks generated by indole groups and an "X" indicates unassigned peaks that are affected by His_6-σ^B_{1-228} binding. **B.** The amino acid sequence of RbpA. Residues for which the backbone amide group of RbpA in the free form was assigned are highlighted in blue.

Further analysis of the RbpA chemical shift changes induced by the binding with His_6-σ^B_{1-228} was performed by calculating the NMR minimal chemical shift (section 3.2.9). To this purpose TROSY-HNCO spectra were acquired on $^{15}N/^{13}C/^{2}H$ RbpA in complex with His_6-σ^B_{1-228} and on $^{15}N/^{13}C/^{2}H$ RbpA in the free form. An overlay of selected regions from the two spectra is shown in figure 3.20 panel A. Clear changes were detected in the RbpA TROSY-HNCO spectrum following the formation of the complex with His_6-σ^B_{1-228}. The histogram shown in figure 3.20 panel B summarizes the results of the minimal shift assay performed. It shows a clear cluster of residues at the N-terminus of RbpA, between residue 11 and residue 20, that are affected by the binding with His_6-σ^B_{1-228}. Otherwise, the rest of the sequence displays only minor shifts and no other clusters of residues affected by the His_6-σ^B_{1-228} binding were identified.

Overall, the analysis of the RbpA chemical shift changes induced by binding of His_6-σ^B_{1-228} suggests that both the N- and the C- termini of RbpA are involved in the binding with the σ-subunit. The central ordered domain of RbpA (residues 26-66) does not appear to be affected by the formation of the complex (figure 3.20 panel B). The relatively high number of unassigned resonances affected by the binding with His_6-σ^B_{1-228} (figure 3.19 panel A) suggests that the C-terminal region of RbpA is involved in the formation of the complex with the σ-subunit. Unfortunately, due to conformational exchange occurring at the C-terminal region of RbpA (section 2.3.5), none of the residues 77-111 were assigned preventing precise mapping of the residues involved in the binding with His_6-σ^B_{1-228}. A significant number of assigned (indicated with red circles in the histogram shown in figure 3.20 panel B) and unassigned resonances detected in the HNCO spectrum collected on $^{15}N/^{13}C/^{2}H$ RbpA in the free form were not found in the spectra collected on the $^{15}N/^{13}C/^{2}H$ RbpA•His_6-σ^B_{1-228} complex. This can clearly been seen for some residues in the TROSY overlay shown in figure 3.19 panel A

(*e.g.* the resonance generated by the indole group of Trp_{82}). Taken together these observations suggest that after the binding with His_6-σ^B_{1-228} a number of the $^{15}N/^{13}C/^{2}H$ RbpA resonances are not longer detectable. The disappearance of these signals could be due to the formation of the RbpA•His_6-σ^B_{1-228} complex in intermediate exchange on the chemical shift timescale. Such exchange rate would cause broadening of the peaks that belong to residues involved in the interaction, which could make them hard to detect.

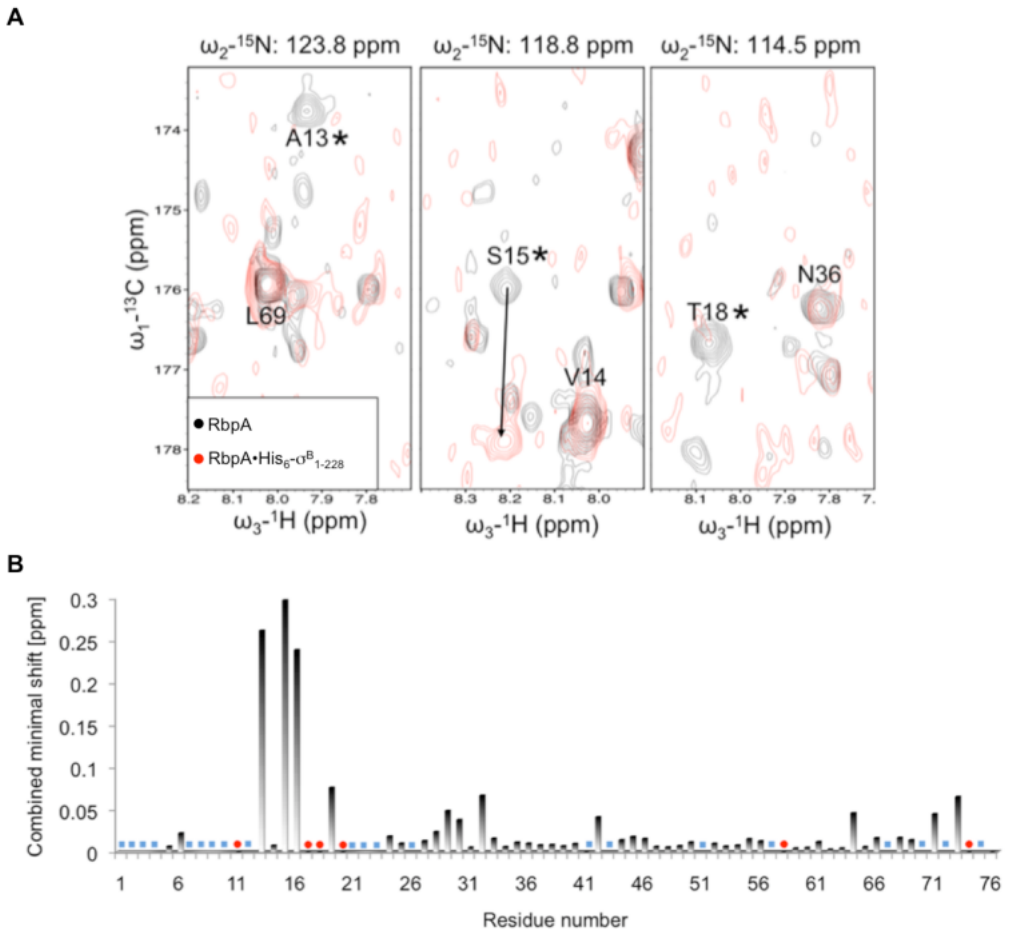


Figure 3.20. NMR minimal shift assay. A. An overlay of selected regions from TROSY-HNCO spectra acquired at 35 °C on $^{15}N/^{13}C/^{2}H$ RbpA (190 μM) (shown in black) and on $^{15}N/^{13}C/^{2}H$ RbpA•His_6-σ^B_{1-228} complex (100 μM) (shown in red) samples dissolved in a buffer containing 25 mM K_2HPO_4 pH 7.5, 100 mM KCl, 0.5 mM DTT, 0.5 mM EDTA, 0.02% w/v NaN_3, 0.2 mM AEBSF and 0.5 mM TCEP with 10% D_2O-90% H_2O. Asterisks (*) indicate peaks that are affected by His_6-σ^B_{1-228} binding and an arrow shows the shift for the Ser_{15} peak. **B.** Combined minimal shift changes observed comparing the TROSY-HNCO spectra described. Assigned resonances for RbpA that cannot be detected in the complex are marked with a red circle (•), whereas RbpA residues for which full assignment in the TROSY-HNCO spectrum is missing are indicated with a blue square (■).

3.3.14 SEC interaction assays between a set of truncated versions of RbpA and the σ-subunit.

The analysis of the RbpA chemical shifts changes induced by His_6-σ^B_{1-228} binding, described in the previous section, suggested that both the N- and the C- termini of RbpA are involved in the binding with the σ-subunit. To confirm this hypothesis $RbpA_{1-79}$ and two other truncated versions of RbpA: $RbpA_{1-92}$ (residues 1-92) and $RbpA_{24-111}$ (residues 24-111) were tested for the interaction with the σ-subunit by SEC assays. $RbpA_{1-79}$ is the construct used to solve the structure of the protein, which lacks the whole C-terminal region. The truncation site at Glu_{92} was chosen to remove the amphipathic C-terminal helix but retain residues 79-89 that are almost invariant amongst RbpA homologues (figure 2.1 panel C). Furthermore, $RbpA_{1-92}$ includes Trp_{82}, which is thought to be involved in His_6-σ^B_{1-228} binding (figure 3.19 panel A). Otherwise, Leu_{24} was chosen as a truncation site to remove the entire flexible N-terminal region including residues 11-20 that are affected by the binding with His_6-σ^B_{1-228} (figure 3.20 panel B). The truncated versions $RbpA_{1-92}$ and $RbpA_{24-111}$ were cloned, expressed and purified following the procedures described for RbpA full length. All the SEC interaction assays between the different truncated versions of RbpA and the σ-subunit were performed as described in section 3.3.2.

3.3.14.1 SEC interaction assays between RbpA$_{1-79}$ and His$_6$-σ$^B_{1-228}$ and His$_6$-σ^A.

RbpA$_{1-79}$ and His$_6$-σ$^B_{1-228}$ samples were loaded on an analytical SEC column (Superdex 75 10/300 GL) individually or mixed together with a molar ratio of 1:1. Figure 3.21 shows an overlay of the resulting chromatograms and the SDS-PAGE page analysis of the fractions collected during the different runs. RbpA$_{1-79}$ and His$_6$-σ$^B_{1-228}$ mixed together are eluted in the same positions observed when the two proteins were analyzed individually (figure 3.21 panel A). This indicates that RbpA$_{1-79}$ and His$_6$-σ$^B_{1-228}$ do not co-elute during SEC and it is confirmed by the SDS-PAGE analysis of the fractions collected during the different runs (figure 3.21 panel B). It should be noticed that His$_6$-σ$^B_{1-228}$ generates two elution peaks (figure 3.21 panel A, red trace): the first peak (elution volume ~ 7.8 ml), whereas the second (elution volume ~ 10.7 ml) is resolved during the SEC run. The former peak probably stems from aggregation of a fraction of the His$_6$-σ$^B_{1-228}$ sample.

The same experimental approach was maintained to test the interaction between RbpA$_{1-79}$ and His$_6$-σ^A but using a different analytical SEC column: Superdex 200 10\300. The overlay of the chromatograms recorded for the different SEC runs (figure 3.22 panel A) and the relative SDS-PAGE analysis (figure 3.22 panel B) show that RbpA$_{1-79}$ and His$_6$-σ^A are not co-eluted when mixed together.

The data described in this section demonstrate that the residues 79-111 of RbpA are essential to form a stable complex with both His$_6$-σ$^B_{1-228}$ and His$_6$-σ^A suggesting that the C-terminal region of RbpA plays an important role in the formation of the complex with the σ-subunit.

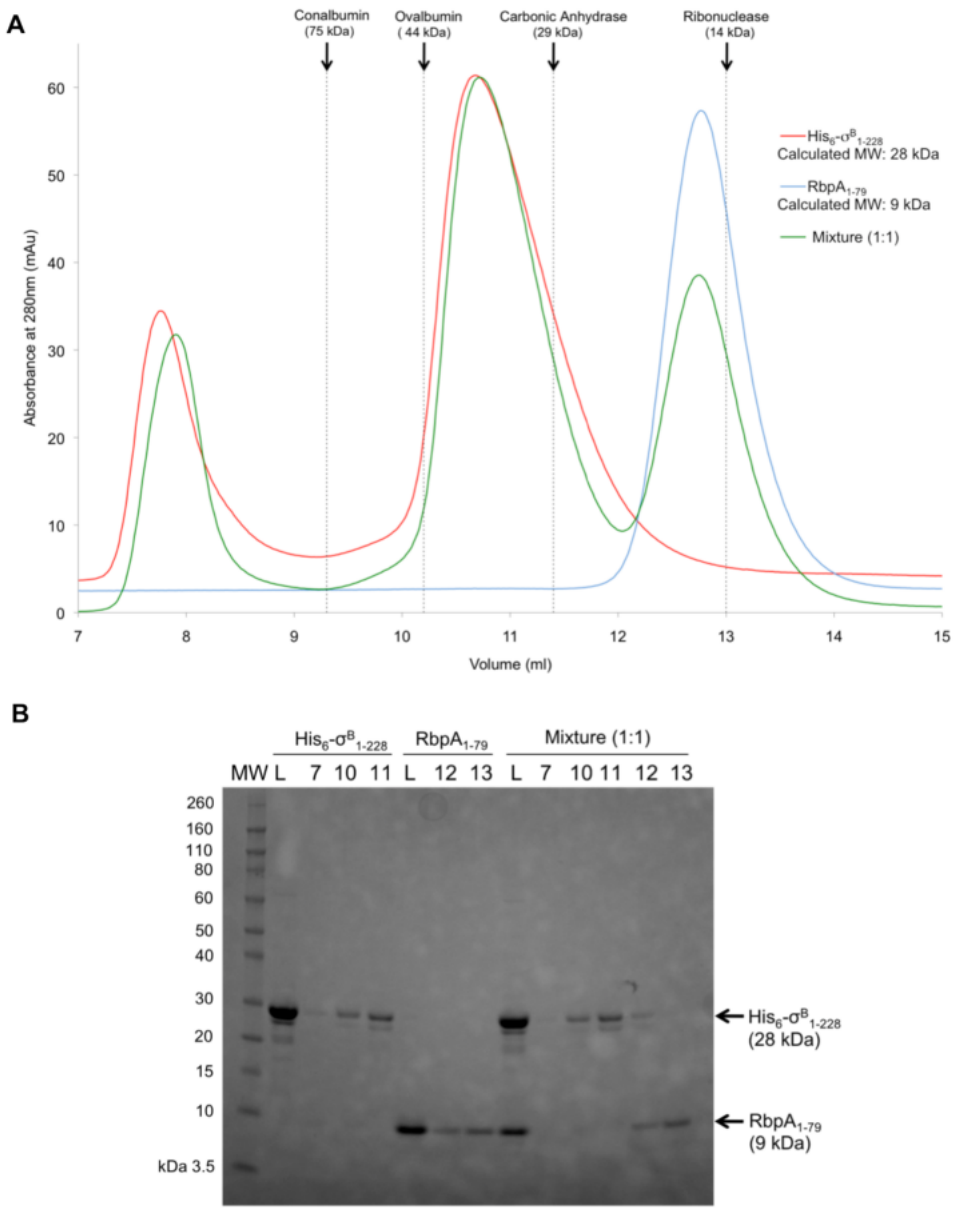


Figure 3.21. $RbpA_{1-79}$ does not bind to His_6-σ^B_{1-228}. A. An overlay of the elution profiles from a SEC column (Superdex 75 10\300 GL) of $RbpA_{1-79}$ (60 μM) or His_6-σ^B_{1-228} (30 μM) individually and mixed together with a molar ratio of 1:1. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the different SEC runs were analysed by SDS-PAGE. Lane MW is molecular weight markers, lanes L the column load and lanes 7-13 labelled according to the elution volume (ml).

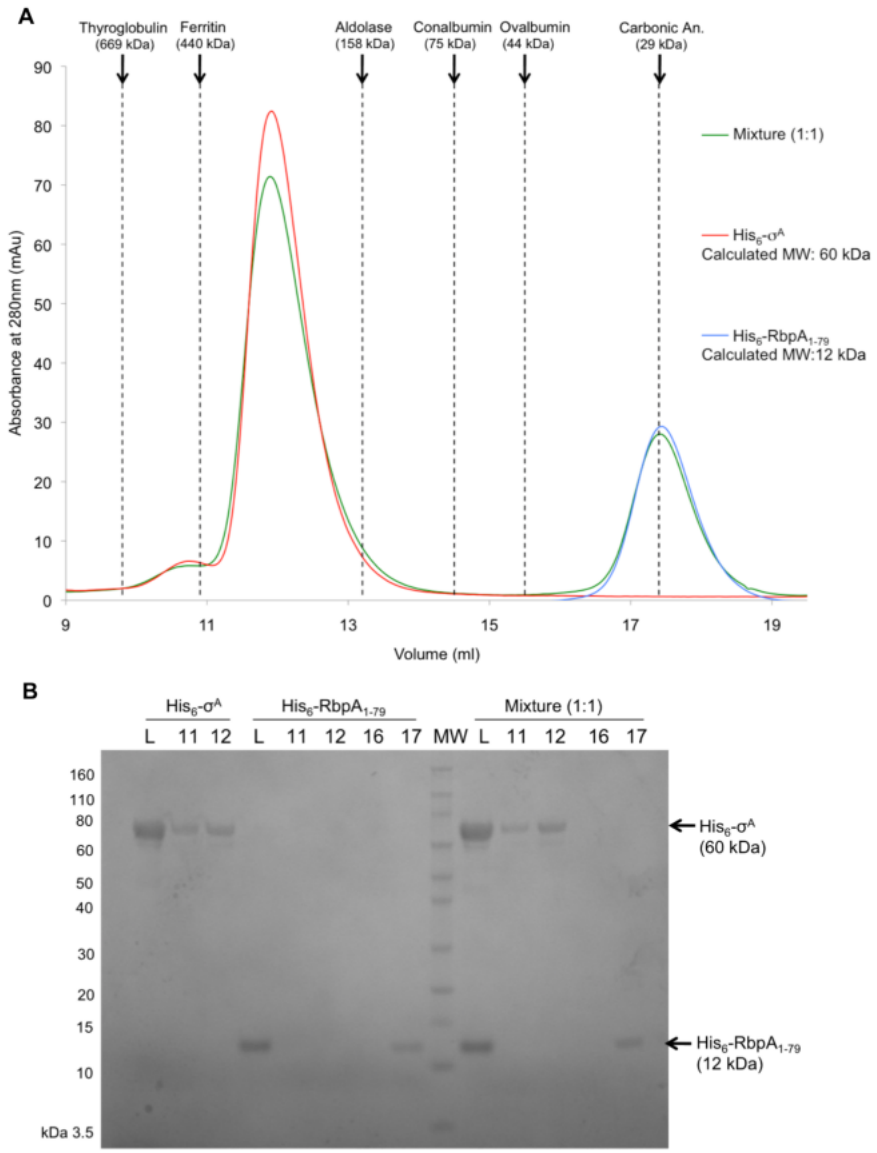


Figure 3.22. His_6-$RbpA_{1-79}$ does not bind to His_6-σ^A. A. An overlay of the elution profiles from a SEC column (Superdex 200 10\300) of His_6-$RbpA_{1-79}$ (20 μM) or His_6-σ^A (20 μM) individually and mixed together with a molar ratio of 1:1. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the different SEC runs were analysed by SDS-PAGE. Lane MW is molecular weight markers, lanes L the column load and lanes 11-17 labelled according to the elution volume (ml).

3.3.14.2 SEC interaction assays between His_6-$RbpA_{1-92}$ and His_6-σ^B_{1-228}.

His_6-$RbpA_{1-92}$ and His_6-σ^B_{1-228} samples were loaded on an analytical SEC column (Superdex 75 10/300 GL) individually or mixed together with a molar ratio of 1:1. In figure 3.23 are shown an overlay of the resulting chromatograms and the SDS-PAGE analysis of the fractions collected during the different runs. His_6-$RbpA_{1-92}$ is eluted at 12.5 ml (figure 3.23 panel A, blue trace) that correspond to an apparent molecular weight of ~18 kDa and is consistent with a monomeric status of the protein (the calculated molecular weight of His_6-$RbpA_{1-92}$ is of 13 kDa). When His_6-$RbpA_{1-92}$ and His_6-σ^B_{1-228} are mixed together and analyzed by SEC they are eluted in the same positions observed when the two proteins were analyzed individually (figure 3.23 panel A). This indicates that $RbpA_{1-92}$ and His_6-σ^B_{1-228} do not co-elute during SEC and it is confirmed by the SDS-PAGE analysis of the fractions collected during the different runs (figure 3.23 panel B).

The experiment described in this section shows that RbpA residues 93-111 are essential to form a stable complex with the σ-subunit. Additionally, the finding that $RbpA_{1-92}$ is monomeric supports the hypothesis that the amphipathic α-helix between residues 91-108 is responsible for the oligomerization of RbpA (section 2.3.5).

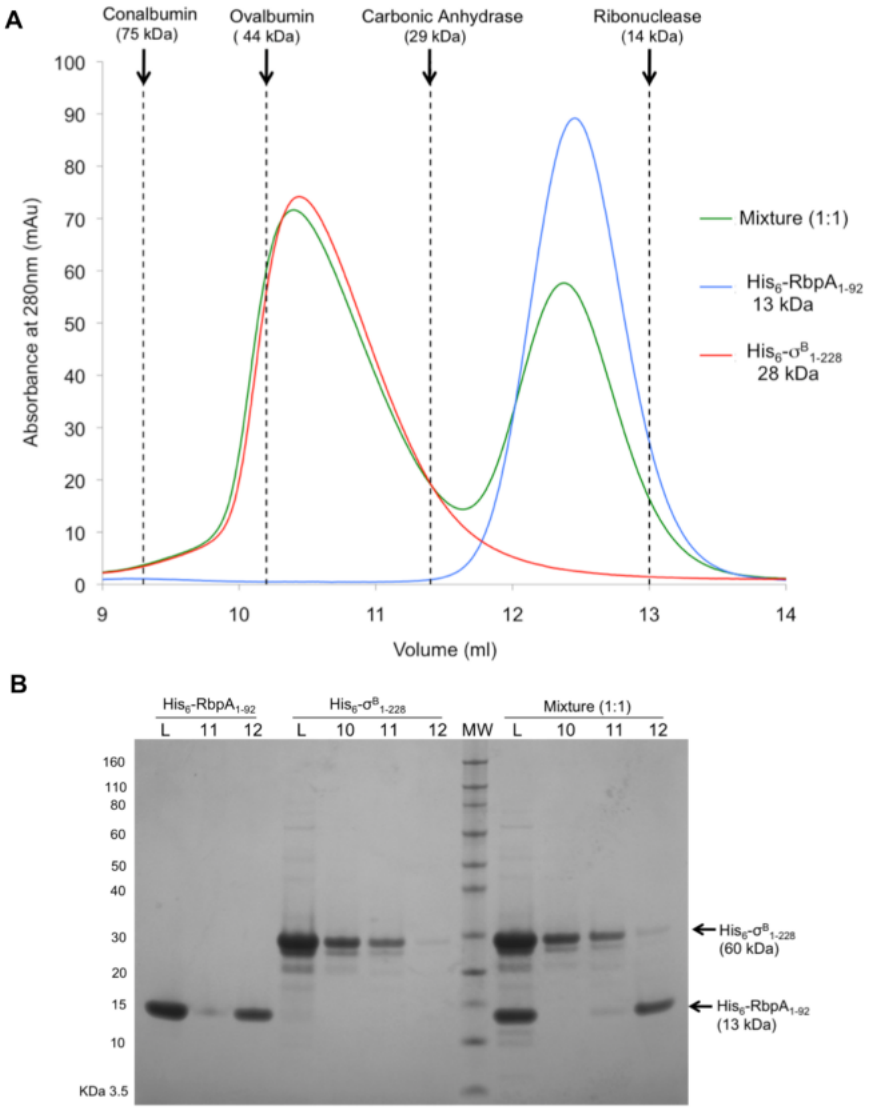


Figure 3.23. His_6-$RbpA_{1-92}$ does not bind to His_6-σ^B_{1-228}. A. An overlay of the elution profiles from a SEC column (Superdex 75 10\300 GL) of His_6-$RbpA_{1-92}$ (35 μM) or His_6-σ^B_{1-228} (25 μM) individually and mixed together with a molar ratio of 1:1. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the different SEC runs were analysed by SDS-PAGE. Lane MW is molecular weight markers, lanes L the column load and lanes 10-12 labelled according to the elution volume (ml).

3.3.14.3 SEC interaction assays between His_6-$RbpA_{24-111}$ and His_6-σ^B_{1-228}.

His_6-$RbpA_{24-111}$ and His_6-σ^B_{1-228} samples were loaded on an analytical SEC column (Superdex 75 10/300 GL) individually or mixed together with a molar ratio of 2:1. The two proteins mixed together are co-eluted during the SEC run, as indicated by the presence of a peak eluted earlier than the peaks detected when His_6-$RbpA_{24-111}$ and His_6-σ^B_{1-228} were analyzed individually, and His_6-$RbpA_{24-111}$ in excess is eluted as expected for the free protein (figure 3.24 panel A). The SDS-PAGE analysis of the fractions collected during the different SEC runs confirms that His_6-$RbpA_{24-111}$ and His_6-σ^B_{1-228} were co-eluted (figure 3.24 panel B) suggesting that the first 23 residues of RbpA are not required for the interaction with His_6-σ^B_{1-228}. This was unexpected given the cluster of chemical shifts perturbations observed between residue 11 and residue 20 in the NMR minimal shift assay (figure 3.20 panel B). However, the N-terminal region might play other roles in the formation of the complex, for example it could increase the affinity of the binding between RbpA and σ-subunit to further stabilize the complex.

The chromatogram obtained for His_6-$RbpA_{24-111}$ (figure 3.24 panel A, blue trace) displays a strongly asymmetric peak at 11.7 ml that corresponds to an apparent molecular weight of 25 kDa, which is close to the double of the calculated molecular weight of His_6-$RbpA_{24-111}$ (13 kDa). These features are reminiscent of the typical elution profile obtained for RbpA full length (figure 2.5) suggesting that a concentration dependent oligomerization might be expected also for $RbpA_{24-111}$.

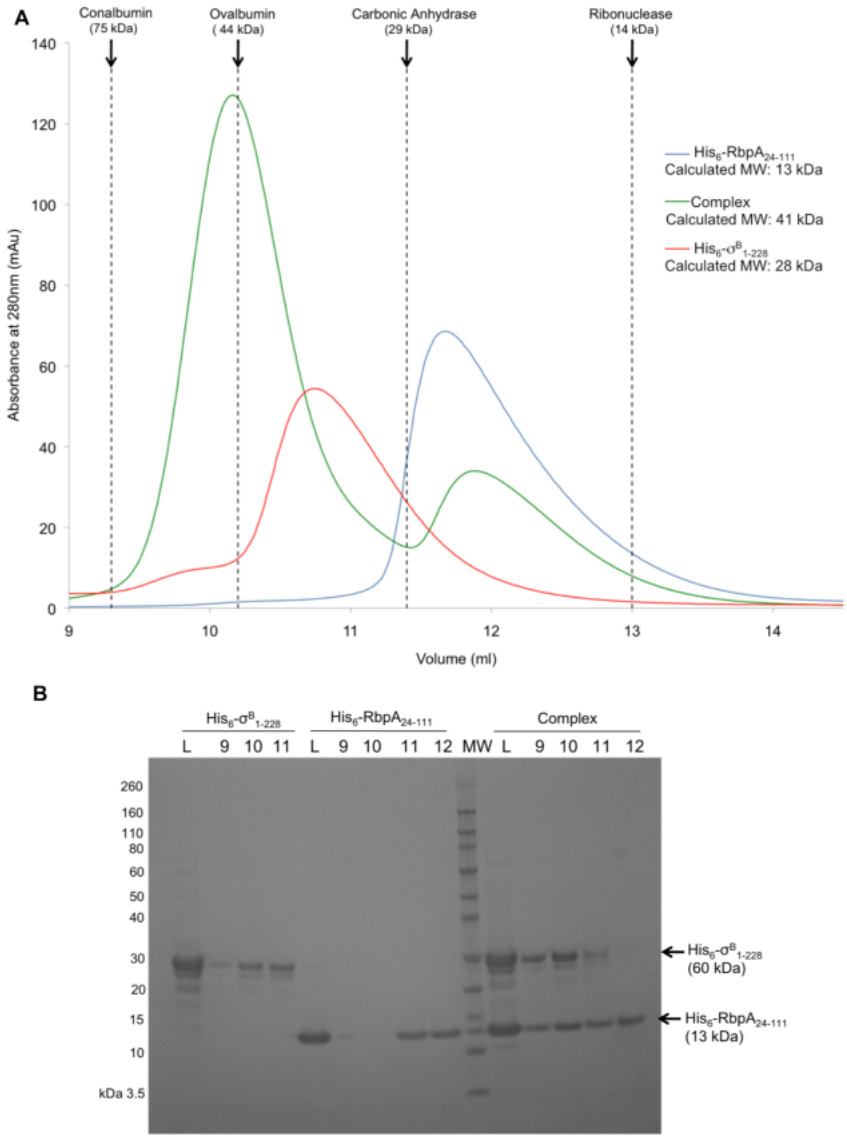


Figure 3.24. His_6-$RbpA_{24\text{-}111}$ binds and forms a tight complex with His_6-$\sigma^B_{1\text{-}228}$. A. An overlay of the elution profiles from a SEC column (Superdex 75 10\300 GL) of His_6-$RbpA_{24\text{-}111}$ (50 μM) or His_6-$\sigma^B_{1\text{-}228}$ (25 μM) individually and mixed together with a molar ratio of 2:1. The elution volumes of the protein molecular weight standards used to calibrate the column are indicated above the picture by dashed lines. **B.** Samples from the fractions collected during the different SEC runs were analysed by SDS-PAGE. Lane MW is molecular weight markers, lanes L the column load and lanes 9-12 labelled according to the elution volume (ml).

3.4 Conclusions

The set of results presented in this chapter demonstrate that RbpA is a binding partner of the RNAP σ-subunit. This interaction was not reported before and opens new scenarios in the RbpA mechanism of action to activate transcription. Hu *et al.* (97), showed that RbpA increases the $E\sigma^A$ transcription activity of ~2-fold, but does not activate $E\sigma^F$. Because the β-subunit was the only RNAP subunit reported to interact with RbpA the specificity of RbpA toward σ^A was justified proposing an allosteric mechanism of action (97). Hu *et al*. (97), suggested that the binding of RbpA on the β-subunit could induce a structural modification of the RNAP core increasing the affinity for σ^A and stabilizing the transcriptional competent open complex. However, the finding that RbpA can bind both β-subunit (97,102) and σ^A (section 3.3.4) provides a simpler mechanism by which RbpA could stabilize the holo-RNAP containing σ^A. The data reported in section 3.3.2 show that RbpA can bind σ^B as well, implying that RbpA might also stimulate the $E\sigma^B$ transcription activity. None of the remaining *M. tuberculosis* σ-subunits (σ^C-σ^M) were tested for the interaction with RbpA during this research project. However, between the *M. tuberculosis* alternative σ factors, σ^F shares the highest sequence homology with both σ^A and σ^B, and σ^F-driven transcription is not activated by RbpA (97). This would suggests that RbpA binding is restricted to σ^A and σ^B and its activity restricted to transcription driven by primary or primary-like σ factors. Attempts to map precisely the RbpA binding site for the σ-subunit were undertaken comparing the NMR spectra recorded on RbpA in complex with His_6-σ^B_{1-228} and on RbpA in the free form (section 3.3.13). The results obtained show that the central folded domain of RbpA is not affected by the formation of the complex with the σ-subunit but both the N- and C- termini are clearly involved. The structure of $RbpA_{1-79}$ (figure 2.17 panel A) shows that the N- and C- termini are on the same side of the structured core

suggesting that also in RbpA full length the C-terminal region could be found in close proximity to the N-terminal region. Therefore, both ends could take part in the interaction with the σ-subunit. To support this observation a set of SEC interaction assays were performed using different truncated versions of RbpA (figure 3.25). Taken together, these experiments show that the whole C-terminal region of RbpA is essential to generate a stable complex with the σ-subunit but not the N-terminus.

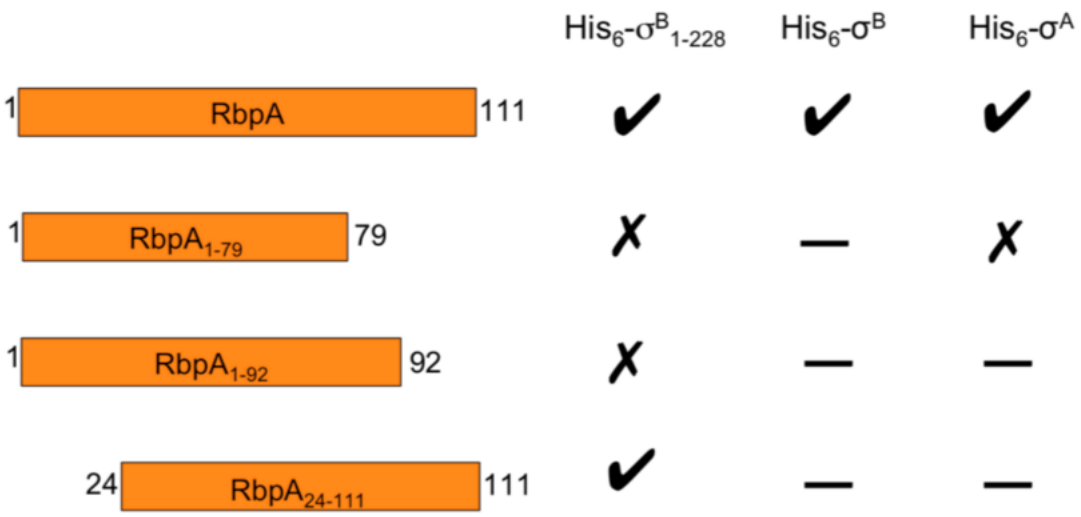


Figure 3.25. Summary of the SEC experiments performed to investigate binding between RbpA and the σ-factors. A tick (✔) indicates that the proteins formed a complex, a cross (✗) indicates that no interaction was detected and a dash (−) indicates that the combination was not tested.

It was hypothesized (section 2.3.5) that the poor quality, including missing signals, of the NMR data collected for the RbpA residues 77-111 was due to conformational exchange occurring in this region of the protein. The SEC experiments performed demonstrate that the C-terminal region of RbpA is essential for the interaction with the

σ-subunit (figure 3.25). It is not unusual for conformational averaging due to protein flexibility to occur at high affinity protein-protein interaction sites, indeed there are many documented examples (145,148-151).

The SEC experiments shown in section 3.3.12 indicate that the complex between RbpA and His_6-σ^B_{1-228} is likely to have a 1:1 stoichiometry. Together with the established 1:1 stoichiometry of σ-subunit binding to RNAP, this suggests that a single RbpA binds per RNAP. However, *in vivo* oligomerization of RbpA cannot be ruled out, indeed it could be relevant for other functions of the protein. For example, protection from non-specific interactions, as reported for the anti-σ-factor AsiA (152).

Previous studies have found that RbpA principally binds to the β-subunit of the RNAP (97,102,103). However, the evidence reported in this chapter show that RbpA binds also the σ-subunit. As shown in figure 1.7, which depicts the structure of the RNAP holoenzyme of *T. thermophilus* (PDB: 1IW7), in the RNAP complex the β and σ subunits are found in close proximity. Thus, it is reasonable to hypothesize that RbpA could bind simultaneously to the β and σ subunits. In support to this hypothesis it should be noticed that the long and dynamic ends of RbpA, through which RbpA binds the σ-subunits (section 3.3.13), could potentially allow the protein to span relatively long distances. Assuming a stretched conformation, the N-terminal region of RbpA (residues 1-25) can reach a length of 90 Å (153). Whilst, an extended conformation of the two α-helices predicted to be formed between residues 81-108 (section 2.3.1) would allow at the C-terminal region to span a length of at least 37 Å (154). However, given the structural flexibility of the termini of RbpA, and the fact that σ-subunit interaction occurs through these termini, it is not possible to model the RbpA structure on the RNAP to propose which, if any, of the previously identified binding sites on the β-

subunit (97,103) might be compatible with RbpA binding to the β and σ subunits simultaneously.

Chapter 4

Complementation of an *Rv2050* conditional mutant strain in *M. tuberculosis.*

4.1 Introduction.

The gene encoding RbpA (*Rv2050*) was predicted to be essential for the growth of *M. tuberculosis* by high density mutagenesis (98). This early observation was confirmed recently by Professor Daniela Ghisotti and her team constructing a *M. tuberculosis* strain where the expression of *Rv2050* is under the control of the antibiotic pristinamycin I (PI) (99) (*M. tuberculosis* Δ*Rv2050*). This strain exhibits a conditional lethal phenotype when PI is withdrawn (99).

Professor Daniela Ghisotti kindly provided us with the *M. tuberculosis* Δ*Rv2050* strain. The data reported in this chapter confirm that the conditional mutant strain shows growth inhibition when PI is withdrawn. Furthermore, it will be shown that ectopic expression of RbpA leads to complementation of the conditional mutant strain. However, complementation failed using the constructs $RbpA_{1-79}$ and $RbpA_{24-111}$ suggesting that both the N- and C- terminus are important for the function of the protein.

4.2 Materials and methods.

4.2.1 *Rv2050* conditional mutant strain in *M. tuberculosis* (*M. tuberculosis ΔRv2050*).

An *Rv2050* conditional mutant strain of *M. tuberculosis* H37Rv was obtained by Forti *et al*. (99) using a Himar1 derived transposon. In this strain the gene encoding for RbpA (*Rv2050*) is under the control of the *Streptomyces* pristinamycin (PI) inducible *ptr* promoter (*pptr*). The transposon inserts randomly in TA sites and was designed to harbour the gene conferring resistance to hygromycin, *pptr* and the gene encoding for the *pptr* suppressor (*pip*). The transposon was inserted in the thermo-sensitive mycobacteriophage phAE87, which was used for *M. tuberculosis* H37Rv transduction. Clones resistant to hygromycin were screened for PI-growth dependence and subsequently sequenced. Only the clones in which a single insertion was detected were retained. In one of these clones *pptr* was found 45 bases upstream at the transcription initiation site of *Rv2050* making the expression of RbpA PI-dependent. Figure 4.1 shows a schematic representation of the *Rv2050* genetic context in the *M. tuberculosis ΔRv2050* conditional mutant strain.

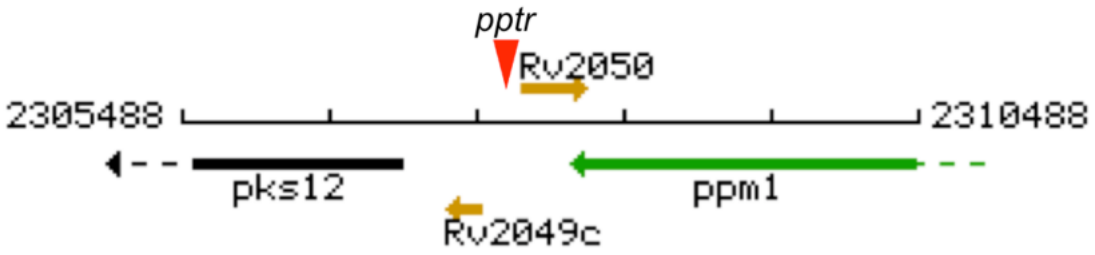


Figure 4.1. Schematic representation of the *Rv2050* genetic context in the *M. tuberculosis ΔRv2050* conditional mutant strain. The red triangle represents the insertion point of *pptr* (45 bases upstream at the transcription initiation site of *Rv2050*).

4.2.2 Culture of the *M. tuberculosis* Δ*Rv2050* strain.

A glycerol stock of the *M. tuberculosis* Δ*Rv2050* strain was used to inoculate 5 ml of Middlebrook 7H9 supplemented with 10% ADC (appendix), hygromycin (100 µg/ml), Tween80 (0.05% v/v) and PI (0.5 µg/ml). Typically, the culture was incubated at 37 °C with gentle shaking for 10 days or until sufficient bacterial growth was detected (absorbance at 580 nm > 0.2). The cells were harvested by centrifugation (1500 *g* for 20 minutes) and the resulting pellet was washed with Middlebrook 7H9 supplemented with 10% ADC, Tween80 and hygromycin. The cells were harvested by centrifugation and washed for a second time in order to remove residual PI. The resulting culture was used to inoculate 10 ml of Middlebrook 7H9 supplemented with 10% ADC, hygromycin, Tween80 with or without PI at an optical density of ~0.05. The cultures were incubated at 37 °C for 8 days and their optical density was measured regularly. For comparison, the same procedure was performed in parallel with a *M. tuberculosis* H37Rv glycerol stock but without addition of hygromycin or PI to the media used.

4.2.3 Complementation of the *M. tuberculosis* Δ*Rv2050* strain.

The coding region of the gene *Rv2050* with the 304 base pair promoter region was cloned into plasmid pMV306 to obtain the plasmid pMV306-RbpA. pMV306 is an integrative plasmid in which the expression cassette of pMV361 is replaced by a multiple cloning site (155). Two truncated versions, pMV306-$RbpA_{1-79}$ and pMV306-$RbpA_{24-111}$, were constructed in an identical manner. The three constructs were kindly provided by Dr. Barbara Rieck. Each complementing plasmid and, as a control, pMV306, was electroporated into the conditional mutant. The electroporation was performed as follows. Typically, 100 ml of a *M. tuberculosis* Δ*Rv2050* culture to an absorbance at 580 nm of ~0.7 was harvested by centrifugation (2500 *g* for 15 minutes) and the resulting pellet was washed in a solution containing 10% v/v glycerol. The washing step was repeated for four times reducing gradually the total volume of the bacterial culture to 1 ml. The resulting bacterial culture was divided in 100 μl aliquots and each aliquot was electroporated using 1 μg of plasmid. Colonies were obtained on Middlebrook 7H10 agar plates supplemented with 10% ADC, kanamycin (40 mg/ml), hygromycin (100 μg/ml) and PI (0.5 μg/ml). A few clones for each transformation were grown in Middlebrook 7H9 medium supplemented with ADC, Tween80, kanamycin, hygromycin and PI for 10 days and then serially diluted and spotted in parallel onto 7H10 agar plates with and without PI. Photographs were taken after 24 days of incubation.

4.3 Results and discussion.

4.3.1 The *M. tuberculosis* Δ*Rv2050* strain exhibits growth inhibition when PI is withdrawn.

Cultures of the *M. tuberculosis* Δ*Rv2050* strain with or without PI and cultures of *M. tuberculosis* H37Rv were grown at 37 °C for 8 days and their optical density was measured regularly (figure 4.2). It is clear that when PI is withdrawn from the media, *M. tuberculosis* growth is inhibited (figure 4.2, red trace). Otherwise, supplementing the media with PI (figure 4.2, green trace) results in a growth rate comparable to the one recorded for *M. tuberculosis* H37Rv (figure 4.2, blue trace). Growth inhibition is detectable only after four days of incubation at 37 °C suggesting that in the first days the conditional mutant strain can grow in media without PI. A similar phenomenon was also observed in the work of Forti *et al.* (99) and could be due to residues of the inducer (PI) or of RbpA. The half-life of RbpA is not known, but a slow turnover of the protein could justify the growth of the conditional mutant strain in media without PI in the first four days. Furthermore, small amounts of PI could be retained inside the cells despite the washing steps performed before the inoculum of the cultures (section 4.2.2) and could be sufficient to induce RbpA expression. However, residues of PI or RbpA inside the cells will be diluted during the bacterial growth and after a number of cell divisions their concentration will not be sufficient to sustain the growth of the *M. tuberculosis* Δ*Rv2050* strain.

The results reported in this section are comparable to the data published by Forti *et al.* (99) confirming that the *M. tuberculosis* Δ*Rv2050* strain exhibits growth inhibition when PI is withdrawn.

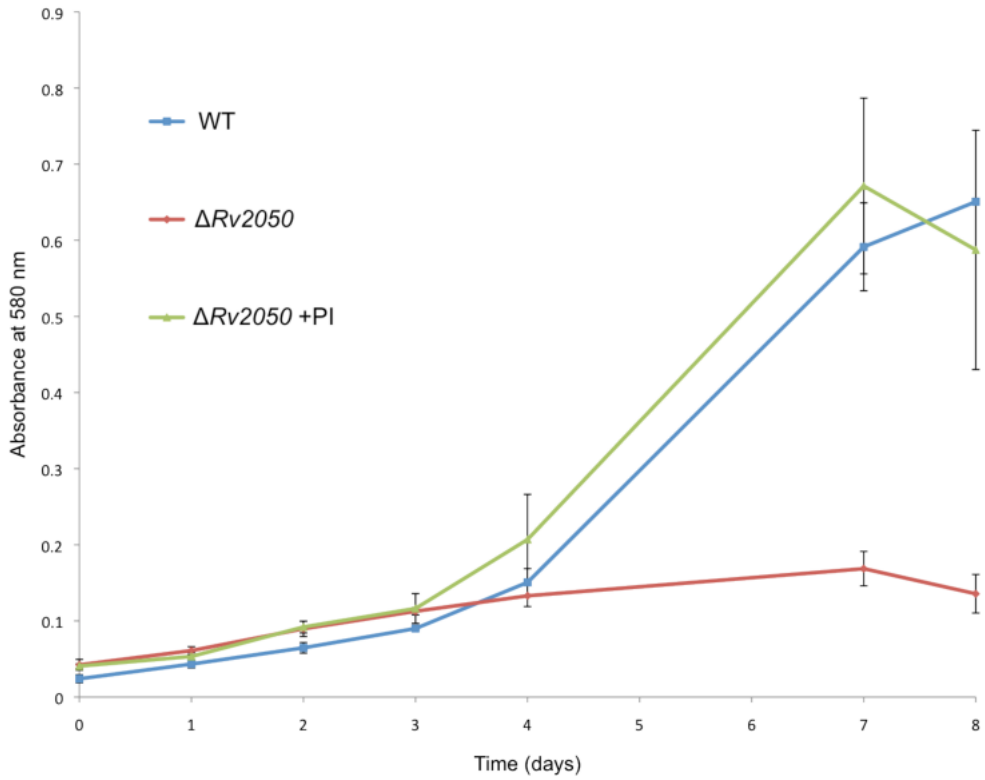


Figure 4.2. PI is essential for the optimal growth of the *M. tuberculosis* ΔRv2050 strain. Growth curves of: *M. tuberculosis* H37Rv (WT), *M. tuberculosis* Δ*Rv2050* strain with PI (Δ*Rv2050* +PI*)* or without PI (Δ*Rv2050*). Each point in the graph is the average of the absorbance values recorded for three equivalent cultures. The error bars represent the standard deviation to the mean.

4.3.2 Complementation of the *M. tuberculosis* Δ*Rv2050* strain.

To confirm that the growth inhibition of the conditional mutant strain induced by PI withdrawal is due to the lack of RbpA expression, and to investigate the domains of RbpA, the mutant strain was complemented with RbpA, $RbpA_{1-79}$ or $RbpA_{24-111}$ expressed from the plasmid pMV306. The resulting strains were grown in liquid media, serially diluted, and plated on agar with or without PI (figure 4.3). As expected growth

was detected for all cultures when the media was supplemented with PI showing that all strains were viable. However, only the strain transformed with pMV306-RbpA was able to grow on media without PI. The results obtained showed that the growth inhibition of the *M. tuberculosis* Δ*Rv2050* strain is due to lack of RbpA expression and suggests that both the N- and the C- terminal regions of RbpA are important for the function of the protein.

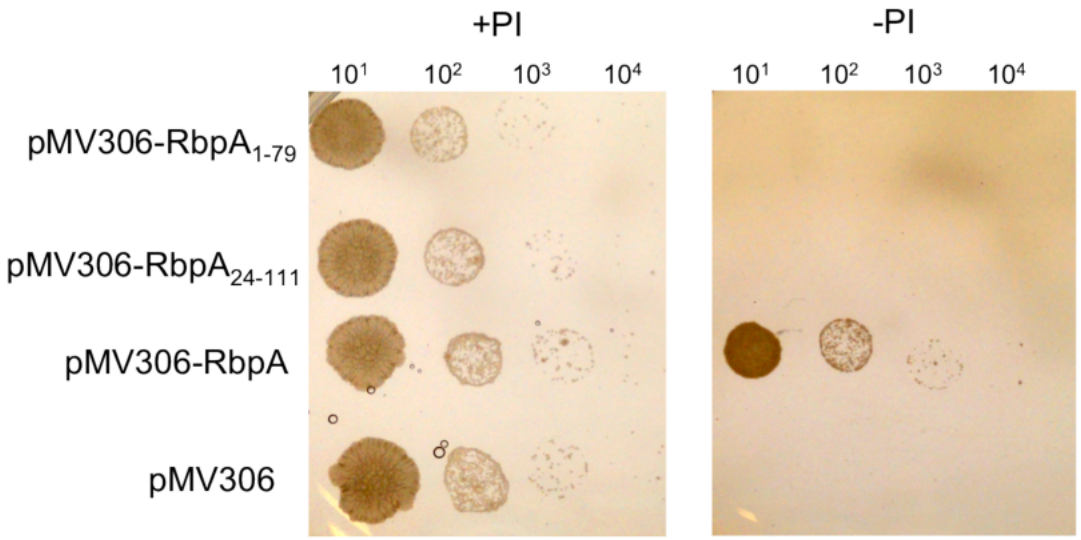


Figure 4.3. Complementation of the *M. tuberculosis* Δ*Rv2050* strain. Liquid cultures of the conditional mutant strain transformed with pMV306, pMV306-RbpA, pMV306-$RbpA_{1-79}$ or pMV306-$RbpA_{24-111}$ were diluted (from 10^1 to 10^4 fold) and spotted onto 7H10 agar with or without PI. Plates were placed on a white light box for photography.

4.4 Conclusions

The data reported in this chapter prove that the growth inhibition of the *M. tuberculosis* Δ*Rv2050* strain is due to lack of expression of RbpA. Moreover, the observation that the pMV306-RbpA$_{1-79}$ and pMV306-RbpA$_{24-111}$ constructs do not complement the *M. tuberculosis* Δ*Rv2050* strain suggests that both the N- and the C- terminus of RbpA are important for the function of the protein. However, it should be observed that the impossibility of the pMV306-RbpA$_{1-79}$ and pMV306-RbpA$_{24-111}$ constructs to complement the growth defect of the *M. tuberculosis* Δ*Rv2050* strain could also be due to the instability or to the poor expression of the truncated versions of the protein. Indeed, the expression level and the stability of RbpA$_{1-79}$ and RbpA$_{24-111}$ in *E. coli* were comparable to RbpA full-length but in *M. tuberculosis* have not been checked. It was somewhat surprisingly, given the interaction observed *in vitro* between RbpA$_{24-111}$ and the σ-subunit (section 3.3.14.3), that RbpA$_{24-111}$ did not complement the growth defect of the *M. tuberculosis* Δ*Rv2050* strain. However, the N-terminal region of RbpA is relatively well conserved (figure 2.1, panel C) and NMR minimal shifts experiments showed that this region of the protein is affected by the binding with the σ-subunit (section 3.3.13). Taken together these observations suggest that residues 1-23, despite their dispensability in forming a stable complex with the σ-subunit, could be important for the function of RbpA.

The data presented in this chapter suggest that both the N- and C- terminal regions of RbpA are important for its function. The termini of RbpA are the most conserved regions of the protein (figure 2.1, panel C) and are both involved in the binding with the σ-subunit (section 3.3.13). It is tempting at this point to speculate that the essentiality of RbpA in *M. tuberculosis* could be linked to the interaction with the σ-subunit. Despite

the recent progresses in the characterization of the protein, it is not clear yet, which is the essential function performed by RbpA in *M. tuberculosis.* Hu *et al*. (97) proposed that RbpA is essential because it could increase the binding affinity of the primary σ-subunit (σ^A) with the RNAP-core over alternative σ-subunits. Indeed, inside the cell different σ-subunits compete for the binding with a limited pool of RNAP-core and it was noticed that σ^A and σ^F of *M. tuberculosis* are required in similar concentrations for *in vitro* transcription (97). Thus, RbpA could play an essential role ensuring the formation of $E\sigma^A$ and the transcription of housekeeping genes essential for the growth of the bacteria, especially under stress conditions when the concentration of alternative σ-factor inside the cell is higher. The findings that RbpA can bind σ^A and σ^B (sections 3.3.2 and 3.3.4), that the termini of RbpA are involved in the interaction with the σ-subunit (sections 3.3.13), and that both N- and C- terminal regions are important for RbpA function (section 4.3.2) seem to support the hypothesis proposed by Hu *et al.* (97).

Chapter 5

Conclusions and future work.

Tight gene expression control is thought to be key for a successful infection cycle of *M. tuberculosis*. In fact, during different stages of infection the bacterium undergoes drastic changes in its metabolism that allow it to persist in the host despite immunosurveillance. RbpA is a member of the *M. tuberculosis* gene expression control network. Indeed, with the work of Hu *et al*., (97) RbpA became recognised as part of a repertoire of small RNAP binding proteins that modulate transcription. High density mutagenesis studies (98) and the construction of an *Rv2050* conditional mutant strain of *M. tuberculosis* (99) revealed that the expression of RbpA is essential for *M. tuberculosis* viability. This information combined with the knowledge that RbpA is confined to the actinobacteria led to RbpA being proposed as a potential anti-TB drug target (97,100). Furthermore, RbpA was reported to protect the RNAP from the inhibitory effects of the first line anti-TB drug rifampicin (97,101-103). Taken together, these data, make RbpA a very interesting subject for studies.

The work described in this thesis has contributed to the ongoing effort to characterize RbpA by providing its structure and identifying the RNAP σ-subunit as its binding partner.

A yeast two-hybrid screen followed by a protein fragment complementation assay, performed before the research project described in this thesis began, provided evidence that RbpA can bind σ^B (142). This early observation was confirmed by *in vitro* interaction assays and extended to the primary σ-factor (σ^A) (chapter 3). Transcription activation induced by RbpA was shown to be σ-factor specific and an allosteric

mechanism of action was proposed to justify the specificity of RbpA toward the primary σ-factor (97). In detail, it was suggested that the binding of RbpA to the β-subunit could induce a structural rearrangement of the RNAP-core (E) that ultimately would increase the affinity for the binding with the primary σ-factor leading to the formation of a more stable and active RNAP holoenzyme (97). However, the finding that RbpA can bind to the σ-subunit led us to propose a simpler mechanism. RbpA could bind both E, on the β-subunit, and the σ-subunit stabilizing the formation of a transcriptionally competent RNAP holoenzyme. Unfortunately, from the data available it is not possible to propose which are the molecular rearrangements induced by RbpA that lead to a more stable and active RNAP holoenzyme. The transcription regulator Crl of *E. coli* might share important similarities with the function of RbpA. For instance, it was shown, by SEC interaction assays, that Crl also could bind the σ-subunit in the free form (156). This early observation was confirmed by surface plasmon resonance experiments, which revealed that the Crl•σ-subunit complex has a 1:1 stoichiometry and a K_D of 2.46 ± 0.13 μM (93). Moreover, as for RbpA, Crl was found to also bind E and to stabilize the formation of the RNAP holoenzyme increasing its transcription activity (93-95). However, there are significant differences as well between RbpA and Crl, perhaps the most important is in the binding-specificity of the two proteins toward different σ-subunits. In fact, RbpA binds to the principle σ-factor and principle-like σ-factor, whereas Crl binds to the alternative stress responsive σ^S (equivalent to σ^F of *M. tuberculosis*) and no interaction was detected between Crl and σ^{70} (the primary σ-factor of *E. coli*) (93). The role of Crl seems confined to enhancing the competitiveness of the alternative σ^S for the binding with E to sustain the expression of Eσ^S dependent genes during the stationary phase (94). The data presented in this thesis seem to support the hypothesis of Hu *et al.* (97) where RbpA plays an opposite role to Crl: it increases the

affinity of the primary σ-factor for the binding with E, over alternative σ-factors, to guarantee the expression of housekeeping genes. Furthermore, Crl is non-essential, whereas loss of RbpA leads rapidly to loss of growth and viability of *M. tuberculosis* (99).

The combination of NMR minimal shift assays with a set of SEC interaction experiments shed light on some features of the interaction mode of RbpA with the σ-subunit. For example, it was demonstrated that the first 228 residues of σ^B are sufficient to form a stable and tight complex with RbpA (figure 3.16) suggesting that the σ-subunit conserved region 4 (σ^B residues 253-312) does not play a major role in the formation of the RbpA•σ-subunit complex. Furthermore, it was shown that the central region of RbpA is not involved in the interaction with the σ-subunit, but both the N- and C-terminus are affected by the binding whereas only the C-terminal region is essential for the formation of the complex (sections 3.3.13 and 3.3.14).

The structural studies undertaken on RbpA (chapter 2) showed that the protein consists of three regions: residues 1-25 are flexible and random coil, residues 26-66 form a stably folded central domain, which is connected by an unstructured linker to a helical region, residues 80-111. Interestingly, the determination of the $RbpA_{1-79}$ structure (figure 2.17 panel A) revealed that both termini are on the same side of the structured core suggesting that the N- and C- terminal regions of RbpA could be found in close proximity. Therefore, the σ-subunit could interact with both termini of RbpA without involving the central structured domain. Although the structured core of RbpA does not take part in the interaction with the σ-subunit (figure 3.20 panel B), there is a significant hydrophobic patch on the surface, bordered by conserved charged residues, which could form a putative interaction surface, perhaps for RNAP (figure 2.18). Notably the

residues that form this patch are highly conserved in all homologues, including Trp_{54}, which is invariant.

The C-terminal region of RbpA (residues 80-111) is essential for the interaction with the σ-subunit (figure 3.25) and is responsible for the concentration dependent oligomerization of RbpA (figure 2.10). The finding that the elution volume found for RbpA during SEC analyses is dependent on the concentration of the sample used (figure 2.5) could account for some of the discrepancies encountered in the literature about the oligomeric status of RbpA (97,101,103). The stoichiometry of the RbpA•σ-subunit complex was assessed in SEC interaction assays by testing different ratios of RbpA:σ-subunit (figures 3.17 and 3.18). It was concluded that the complex occurred with a 1:1 stoichiometry that, together with the established 1:1 stoichiometry of σ-subunit binding to E, suggests that a single RbpA binds per RNAP. However, *in vivo* oligomerization of RbpA cannot be ruled out, indeed it could be relevant for other functions of the protein. For example, protection from non-specific interactions, as reported for the anti-sigma factor AsiA (152).

RbpA was found to rescue the RNAP from the inhibitory effects of the front line anti-TB drug rifampicin (97,101-103). Direct interaction between RbpA and rifampicin was assessed by NMR experiments (result not shown) and no binding was detected confirming the result reported by Dey *et al.* (102). Further experiments aimed to shed light on the mechanism of action used by RbpA to rescue RNAP from rifampicin were not performed during the research project described in this thesis. However, from the data available on RbpA (97,101-103) it is clear that *in vitro* studies on rifampicin or other *M. tuberculosis* RNAP inhibitors will need to consider RbpA as an essential component of the transcription machinery.

Finally, the availability of an *Rv2050* conditional mutant strain of *M. tuberculosis* allowed to test which regions of RbpA are essential for its functions *in vivo* (chapter 4). The results obtained (figure 4.3) show that ectopic expression of full length RbpA can complement the growth defect of the conditional mutant but not expression of the truncated versions $RbpA_{1-79}$ and $RbpA_{24-111}$. The N- and C- termini are both affected by the binding to the σ-subunit and are the most conserved regions of RbpA. These observations suggest that the binding with the σ-subunit might be essential for the function of RbpA and the viability of *M. tuberculosis.*

During the preparation of the present thesis a study about RbpA and its homologue in *S. coelicolor* was published (157). The main focus of this work was on the identification of the σ-subunit as a binding partner of RbpA and on the characterization of the RbpA•σ-subunit complex. Furthermore, the structures of both RbpA (PDB: 2M6P) and of its *S. coelicolor* homologue (PDB: 2M6O) were reported. In the following session the work of Tabib-Salazar *et al.* (157) will be compared with the results obtained during the research project described in this thesis.

One of the most important findings described in this thesis is that RbpA can bind both σ^A and σ^B (sections 3.3.2 and 3.3.4). The same interactions were detected also by Tabib-Salazar *et al.* (157) using a bacterial two-hybrid assay. Moreover, bacterial two-hybrid assays were performed to test the interaction of *S. coelicolor* RbpA with a number of σ-subunits from *S. coelicolor*: σ^{HrdB} (Group1, primary σ-factor), σ^{HrdA}, σ^{HrdC} and σ^{HrdD} (Group 2, primary-like σ-factor), σ^{WhiG} and σ^B (Group 3), and σ^E and σ^R (Group 4). Binding was detected only for σ^{HrdB} and σ^{HrdA} and the interaction between σ^{HrdB} and *S. coelicolor* RbpA was confirmed also by an SEC interaction assay (157). Overall, the set of interactions assays performed by Tabib-Salazar *et al.* (157) support

the idea that RbpA action is confined to principal or principal-like σ-factors. Indeed, no interaction was detected between Group 3 or 4 *S. coelicolor* σ-factors and *S. coelicolor* RbpA.

Tabib-Salazar *et al.* (157) tested a set of truncated versions of *S. coelicolor* RbpA for the interaction with σ^{HrdB} by bacterial two-hybrid assays showing that the C-terminal region of *S. coelicolor* RbpA (residues 73-124) is essential and sufficient for the binding. A similar approach was used to investigate which of the σ^{HrdB} conserved regions was involved in the binding with *S. coelicolor* RbpA. The residues 211-347 of σ^{HrdB}, which encompass the conserved region 2, were found to be essential and sufficient for the binding with *S.coelicolor* RbpA and this observation was confirmed by pull-down assays. Furthermore, the C-terminal region of *S. coelicolor* RbpA was subjected to an alanine-scanning mutagenesis in order to identify key residues for the interaction with the σ-subunit. The mutations R89A and R90A were found to significantly reduce the binding between *S. coelicolor* RbpA and σ^{HrdB}, which was completely abolished by the R89A/R90A double mutation. *S. coelicolor* RbpA carrying the R89A/R90A double mutation cannot complement a *S. coelicolor* Δ*rbpA* strain suggesting that the interaction with the σ-subunit is essential for the function of the protein. Some of the features, described by Tabib-Salazar *et al.* (157), about the formation of the complex between *S. coelicolor* RbpA and σ^{HrdB} share important features with the results presented in this thesis. For instance, the set of SEC interaction assays reported in section 3.3.14 shows that also the whole C-terminal region of *M. tuberculosis* RbpA is essential for the interaction with the σ-subunit. Furthermore, the finding that the *M. tuberculosis* σ^{B} conserved region 4 is not essential for the interaction with RbpA (sections 3.3.11) is in agreement with the results of Tabib-Salazar *et al.* (157) which demonstrated that the conserved region 2 of *S. coelicolor* σ^{HrdB} is sufficient for the interaction with *S.*

coelicolor RbpA. Finally, the finding that *S. coelicolor* RbpA harbouring the R89A\R90A double mutation neither binds to σ^{HrdB} nor complements the *S. coelicolor* Δ*rbpA* strain is in line with the results obtained using the *Rv2050* conditional mutant strain of *M. tuberculosis* (section 4.3.2). In fact, the complementation of the *M. tuberculosis* Δ*Rv2050* strain failed when it was carried out using RbpA constructs lacking the N- or the C- terminus and these regions were shown to be involved in the interaction with the σ-subunit (section 3.3.13). Taken together these observations suggest that the interaction mode with the σ-subunit is conserved between RbpA and its homologue in *S. coelicolor* and that the binding with the σ-subunit is essential for the RbpA function.

Protein sequence analysis led Tabib-Salazar *et al.* (157) to hypothesize that *S. coelicolor* RbpA but not *M. tuberculosis* RbpA might bind zinc. Indeed, three cysteine residues, Cys_{35}, Cys_{56} and Cys_{59}, together with His_{38} are arranged as a putative C(H/C)CC zinc-ribbon motif but in *M. tuberculosis* RbpA none of these residues, except Cys_{56}, is conserved. Mass spectrometry experiments confirmed that only *S. coelicolor* RbpA can bind zinc (157). Tabib-Salazar *et al.* (157) reported also the solution structure of the core region (residues 28-72) of both *M. tuberculosis* RbpA and *S. coelicolor* RbpA. The structures of the two core regions are virtually identical except for the presence of a zinc ion in *S. coelicolor* RbpA. The zinc ion does not have a functional role but is thought to be important in stabilizing the structure of the protein. The zinc structural role in *M. tuberculosis* RbpA is replaced by a stabilizing hydrogen bond network (157). Despite the lack of chemical shift assignments for residues outside the core region Tabib-Salazar *et al.* (157) suggested, on the basis of the chemical shift dispersion and resonance line widths, that these residues are highly disordered.

The chemical shift assignments obtained here for the amide nitrogens (N) and the amide protons (NH) of $RbpA_{1-79}$ (section 2.3.7) have been compared with the assignments reported by Tabib-Salazar *et al.* (157) for *M. tuberculosis* RbpA (figure 5.1). It should be noticed that the conditions in which Tabib-Salazar *et al.* (157) recorded their NMR experiments are different from the ones used during the research project described in this thesis. Notably, there is a difference of 1 point in the pH of the buffers employed. The histogram reported in figure 5.1 highlights that most of the assignment differences found are minor suggesting that the two sets of assignments are essentially consistent. However, three residues (Arg_{27}, Trp_{54} and Cys_{56}) have chemical shift differences greater than 1 standard deviation from the average change for the entire protein, which should be considered as significant differences (158). The NMR spectra collected for $RbpA_{1-79}$ were checked to search for potential errors in the assignments of Arg_{27}, Trp_{54} and Cys_{56} but no mistakes were found. The differences observed arise probably from the different conditions used during the NMR experiments rather than from assignment errors.

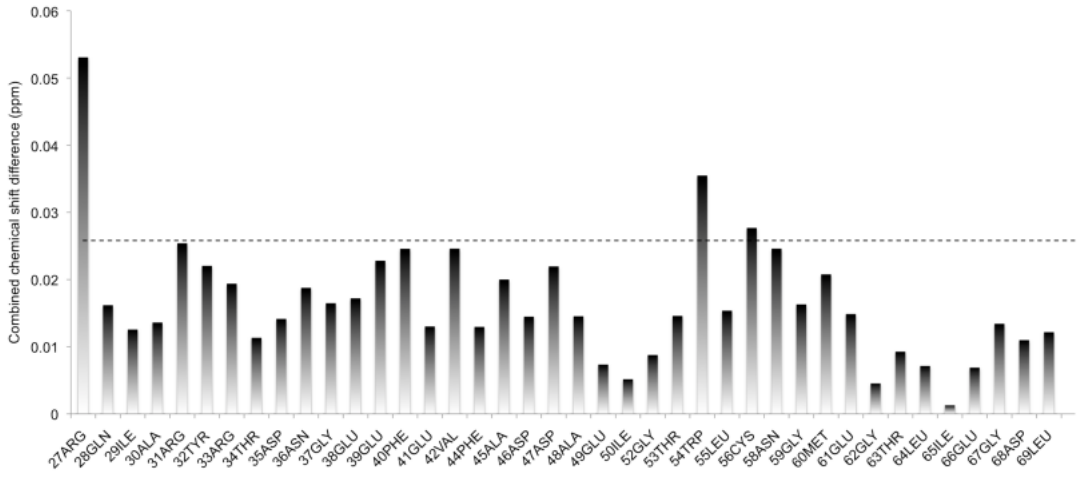


Figure 5.1. Comparison of the amide protons and amide nitrogens chemical shifts found for $RbpA_{1-79}$ with the ones reported by Tabib-Salazar *et al.* (157) for *M. tuberculosis* RbpA. The differences of the amide protons (ΔHN) and amide nitrogens (ΔN) chemical shifts found for $RbpA_{1-79}$ with the ones reported by Tabib-Salazar *et al.* (157) for *M. tuberculosis* RbpA were combined using the following equation: $\Delta\delta=\sqrt{((\Delta\delta_{HN})^2 + (\Delta\delta_N*0.2)^2}$, where 0.2 is a scaling factor required to account for differences in the range of amide proton and amide nitrogen. The horizontal dashed line indicates the threshold chosen for significance, which is the average difference in chemical shift between the two structures plus the standard deviation of the differences.

The structure of *M. tuberculosis* RbpA deposited by Tabib-Salazar *et al.* (157) (PDB: 2M6P) was compared with the structure found for $RbpA_{1-79}$ (PDB: 2M4V) during the research project described in this thesis (section 2.3.9). The superposition reported in figure 5.2 shows that the well-defined central regions of the two structures have the same fold. However, the two structures display some differences as well, which are

reflected in the relatively high backbone RMSD value of 1.16 Å (residues 27-65) found comparing the structures closest to the mean of the two bundles. The sum of the backbone RMSD values to the mean structure found for residues 27-65 in the final families of *M. tuberculosis* RbpA structures reported in section 2.3.9 (0.19 Å) and deposited by Tabib-Salazar *et al.* (157) (0.62 Å) is of 0.81 Å. This value is lower than the backbone RMSD value (1.16 Å) found comparing the two structures suggesting that the structures analysed should not be considered identical (159). However, the relatively low number of conformers deposited by Tabib-Salazar *et al.* (157) might lead to an overestimation of the precision of the structure found. The analysis of a higher number of conformers would probably reveal a higher RMSD, which would give a more reliable estimation of the precision of the structure. Furthermore, the unusually high tolerance ranges used by Tabib-Salazar *et al.* (157) for the NOE assignments (±0.05 ppm for hydrogen dimensions, ±0.5 ppm for nitrogen dimensions and ±1.3 ppm for carbon dimensions) might account for the minor structural differences observed.

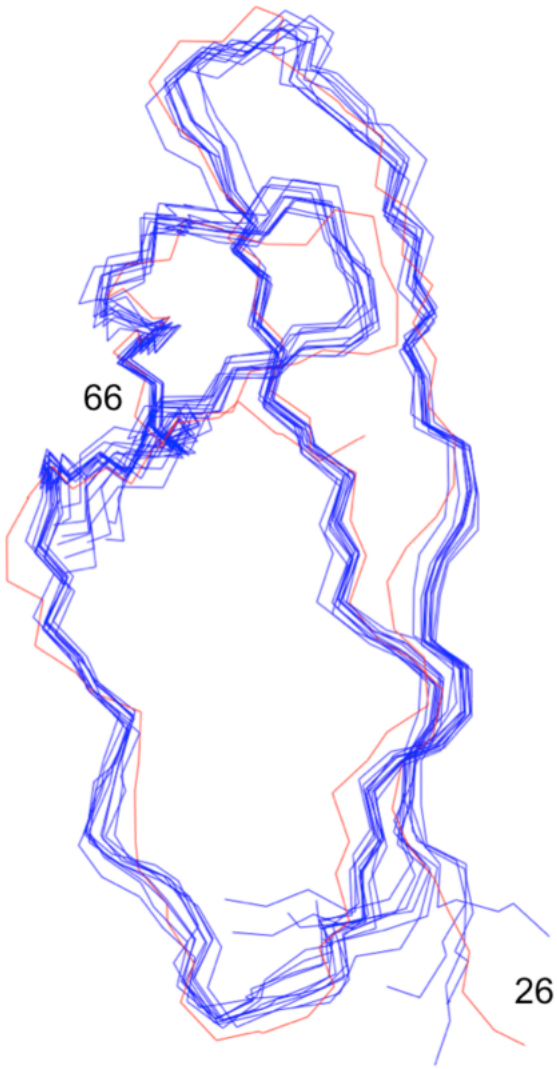


Figure 5.2. Comparison of the structure found for *M. tuberculosis* RbpA with the one deposited by Tabib-Salazar *et al.* (157). A best-fit superposition (residues 27-65) of the central structured region of the *M. tuberculosis* RbpA family of 10 structures deposited by Tabib-Salazar *et al.* (157) (PDB: 2M6P, shown in blue) and of the $RbpA_{1-79}$ structure reported in section 2.3.9 (PDB: 2M4V, shown in red). Within the family of 35 converged structures found for $RbpA_{1-79}$ (section 2.3.9) the conformer with the lowest backbone RMSD value to the mean structure was used for the superposition.

Overall, the data described in this thesis and the recently published work of Tabib-Salazar *et al.* (157) support each other in establishing RbpA as a binding partner of the σ-subunit, providing important insights in the characterization of this protein and laying the foundation for future work.

Further studies will be needed to give exhaustive answers to the many questions that still remain about RbpA function(s). Perhaps, one of the most important unanswered questions is: why RbpA is essential for *M. tuberculosis* viability? Hu *et al*. (97) postulated that RbpA is essential to increase the binding affinity of σ^A for the binding with E in order to guarantee the expression of the essential housekeeping genes. The availability of a *M. tuberculosis ΔRv2050* conditional mutant strain could be exploited to assess this hypothesis. In details, complementation of the growth defect phenotype displayed by the *M. tuberculosis ΔRv2050* conditional mutant strain ((99) and section 4.3.1) could be attempted by overexpressing σ^A or engineering σ^A to increase its binding affinity for E. In these conditions σ^A would have an additional advantage, over alternative σ-factors, for the binding with E which should allows *M. tuberculosis* to expresses essential housekeeping genes without relying on RbpA.

It was shown that RbpA in *S. coelicolor* is not essential, indeed Δ*rbpA S. coelicolor* mutant strains are viable but display a striking slow growth phenotype (101,157). This phenotype can be complemented providing ectopic expression of *S. coelicolor* RbpA (101,157). Interestingly, attempts of complementation performed using *S. coelicolor* RbpA carrying a R89A/R90A double mutation, which was shown to abolish the binding with σ^{HrdB}, failed suggesting that the interaction with the σ-subunit is important for the function of the protein (157). A double mutation of the equivalent residues of *M. tuberculosis* RbpA (Arg_{88} and Arg_{89}) is likely to disrupt the complex formed between *M.*

tuberculosis RbpA and σ^A or σ^B. Such a mutant could be tested for its ability to complement the *M. tuberculosis ΔRv2050* conditional mutant strain. Failure of the complementation would strongly support the idea that to perform its essential function *M. tuberculosis* RbpA needs to interact with the σ-subunit. In chapter 3 it was shown that RbpA could form a stable complex with both σ^A and σ^B and the work of Tabib-Salazar *et al.* (157) confirmed these observations. Because σ^B was shown to be dispensable for growth *in vitro* and in infection models (65,69) the essential function of RbpA is likely to be related to the activation of transcription driven by essential σ^A (26,65).

A further important question, which remains unanswered about RbpA function, is: which is the mechanism of action used by RbpA to activate transcription? Precise knowledge of the molecular rearrangements induced by RbpA on the RNAP holoenzyme that lead to transcription activation would clearly help to fully understand the RbpA mechanism of action. In this prospect, the identification of the σ-subunit residues key for the interaction with RbpA could be highly informative. Indeed, the σ-factors are well conserved and characterized proteins. Thus, the identification of the σ-subunit binding site for RbpA could be combined with the large amount of data available from the σ-factors literature to propose the effects of RbpA binding on the σ-subunit function. Tabib-Salazar *et al.* (157) showed that the conserved domains $\sigma_{1.1}$, σ_3 and σ_4 of σ^{HrdB} are dispensable for the binding with *S. coelicolor* RbpA. On the other end, residues 211-347 of σ^{HrdB}, which encompass the σ_2 domain, are essential and sufficient for the binding with *S. coelicolor* RbpA. The σ_2 domain confers promoter specificity for the -10 element and is conserved also in alternative σ-factors which do not bind to RbpA. A multi-sequence alignment could highlight regions within residue 211-347 that are conserved only between primary and primary-like σ-factors but not in

alternative σ-factors. Such regions might be considered as potential binding sites for RbpA and their role in the formation of the RbpA•σ-subunit complex could be easily assessed by performing point mutations on the σ-subunit.

RbpA was reported to bind also to the β-subunit of the RNAP and is not clear yet which is the region of RbpA responsible for this interaction. The analysis of the structure of $RbpA_{1-79}$ revealed the presence of a significant hydrophobic patch on the surface of the protein, which harbours highly conserved residues (section 2.3.9) and that could form a putative interaction surface, perhaps for the β-subunit. This hypothesis could be tested by designing mutations to target the residues that are part of this patch (Arg_{27}, Val_{42}, Phe_{44}, Ala_{45}, Asp_{47}, Ala_{48}, Glu_{49} and Trp_{54}) and assessing the resulting mutant versions of RbpA for the interaction with the β-subunit.

Some of the results obtained during the research project described in this thesis suggest that the N-terminal region of RbpA (residues 1-25) might play an important role for the function of the protein. Firstly, the impossibility for the $RbpA_{24-111}$ construct to complement the growth defect phenotype of the *M. tuberculosis ΔRv2050* conditional mutant strain (section 4.3.2). Secondly, the observation that the N-terminus of RbpA is a relatively well conserved region of the protein (figure 2.1 panel C). Lastly, the cluster of chemical shift perturbations observed between the RbpA residues 11-20 following the binding with His_6-σ_B^{1-228} (figure 3.20 panel B). In light of these data, it would be interesting to perform experiments aimed to investigate the function of the N-terminal region of RbpA. In fact, despite the SEC interactions assays which showed that residues 1-23 of RbpA are not essential for a tight interaction with the σ-subunit (section 3.3.14.3), the N-terminal region could play important roles in the function of the protein. For example, the N-terminus of RbpA could increase the affinity of the binding between RbpA and σ-subunit or could be essential for the mechanism used by RbpA to activate

transcription. To test these hypotheses it would be necessary to estimate and compare the K_D of the binding of RbpA and $RbpA_{24\text{-}111}$ with the σ-subunit and with the RNAP holoenzyme. Moreover, it would be important to test $RbpA_{24\text{-}111}$ in transcription assays to investigate if the first 23 residues of the protein have a significant role in activating the transcription activity of the RNAP.

A further potential approach to reach a comprehensive understanding of the function and mechanism of action of RbpA could be the structural characterization of the whole RNAP holoenzyme in complex with RbpA. Such an approach, despite obvious technical challenges, is potentially highly informative because it would provide also the structure of a mycobacterial RNAP whose structure has not been reported yet.

Appendix

Culture media

LB Broth

Per litre, 10 g tryptone, 5 g yeast extract, 10 g NaCl. Made up to volume with dH_2O. Autoclaved and when cooled the desired antibiotic was added.

Minimal media

Per litre,

$(NH_4)_2SO_4$[1]	1.0 g	$MnCl_2$	16.0 mg
Na_2HPO_4	6.8 g	$FeCl_3$	5.0 mg
KH_2PO_4	3.0 g	$ZnCl_2$	0.5 mg
NaCl	0.5 g	$CuCl_2$	0.1 mg
Na_2SO_4	0.3 mM	$CoCl_2$	0.1 mg
EDTA	50.0 mg	H_3BO_3	0.1 mg

The above were dissolved in dH_2O then autoclaved. When cool, the following were added aseptically,

MgSO4	1.0 mM	Thiamine	1.0 mg
CaCl2	0.3 mM	Glucose[2]	4.0 g
d-Biotin	1.0 mg	Ampicillin	100 mg

[1] For the expression of ^{15}N or $^{15}N/^{13}C$ labelled samples the ammonium sulphate was replaced with 1 g/l $(^{15}NH_4)_2SO_4$ (^{15}N 98%, Sigma Aldrich).

[2] For the expression of ^{13}C or $^{15}N/^{13}C$ labelled samples the glucose was replaced with 4 g/l D-Glucose (^{13}C 99%, Sigma Aldrich).

High-cell-density-IPTG-induction minimal media

Per 100 ml,

$(^{15}NH_4)_2SO_4$	0.1 g	$CoCl_2$	0.5 μM
Na_2HPO_4	0.71 g	$CuCl_2$	0.5 μM
KH_2PO_4	0.34 g	Na_2SeO_3	0.5 μM
NaCl	10 mM	H_3BO_3	0.5 μM
$MgSO_4$	5 mM	$NaMoO_4$	0.5 μM
$CaCl_2$	0.2 mM	$NiCl_2$	0.5 μM
$FeCl_3$	12.5 μM	$^{13}C_6$ glucose	1 g
$MnCl_2$	2.5 μM	Ampicillin	10 mg
$ZnSO_4$	2.5 μM	BME vitamins (Sigma)	250 μl

The above were dissolved in D_2O (99.8% atom D, Sigma Aldrich), made up to volume and filter-sterilized.

2YT

Per litre, 16 g tryptone, 10 g yeast extract, 5 g NaCl. Made up to volume with D_2O (99.8% atom D, Sigma Aldrich) and filter-sterilized.

ADC

5% albumin, 2% dextrose, 145 mM NaCl. Made up to volume with dH_2O and filter-sterilized.

References

1. Crubezy, E., Ludes, B., Poveda, J.D., Clayton, J., Crouau-Roy, B. and Montagnon, D. (1998) Identification of *Mycobacterium* DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III*, **321**, 941-951.
2. Harries, A.D. and Dye, C. (2006) Tuberculosis. *Ann Trop Med Parasitol*, **100**, 415-431.
3. Roberts, C.A. and Buikstra, J.E. (2008) In Davies, P. D. O., Barnes, P. F. and Gordon, S. B. (eds.), *Clinical tuberculosis.* Hodder Arnold, London, pp. 3-19.
4. Raviglione, M.C. and Pio, A. (2002) Evolution of WHO policies for tuberculosis control, 1948-2001. *Lancet*, **359**, 775-780.
5. World Health Organization. (2012) *Global tuberculosis report 2012* World Health Organization, Geneva.
6. Pedrazzoli, D., Fulton, N., Anderson, L., Lalor, M., Abubakar, I. and Zenner, D. (2012) Tuberculosis in the UK: 2012 report.
7. Palomino, J.C., Leão, S.C. and Ritacco, V. (2007) *Tuberculosis: from basic science to patent care*. BourcillerKamps.com., [S.l.].
8. Taylor, G.M., Stewart, G.R., Cooke, M., Chaplin, S., Ladva, S., Kirkup, J., Palmer, S. and Young, D.B. (2003) Koch's bacillus - a look at the first isolate of *Mycobacterium tuberculosis* from a modern perspective. *Microbiology*, **149**, 3213-3220.
9. Alderwick, L.J., Birch, H.L., Mishra, A.K., Eggeling, L. and Besra, G.S. (2007) Structure, function and biosynthesis of the *Mycobacterium tuberculosis* cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets. *Biochem Soc Trans*, **35**, 1325-1328.
10. Russell, D.G. (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol*, **5**, 39-47.
11. Dye, C., Scheele, S., Dolin, P., Pathania, V. and Raviglione, M.C. (1999) Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA*, **282**, 677-686.
12. Sutherland, I. (1976) Recent studies in the epidemiology of tuberculosis, based on the risk of being infected with tubercle bacilli. *Adv Tuberc Res*, 1-63.
13. Ehlers, S. (1999) Immunity to tuberculosis: a delicate balance between protection and pathology. *FEMS Immunol Med Microbiol*, **23**, 149-158.
14. Tiemersma, E.W., van der Werf, M.J., Borgdorff, M.W., Williams, B.G. and Nagelkerke, N.J. (2011) Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS One*, **6**, e17601.
15. Hurt, R. (2004) Tuberculosis sanatorium regimen in the 1940s: a patient's personal diary. *J R Soc Med*, **97**, 350-353.
16. Iseman, M.D. (2002) Tuberculosis therapy: past, present and future. *Eur Respir J Suppl*, **36**, 87s-94s.
17. Shi, W., Zhang, X., Jiang, X., Yuan, H., Lee, J.S., Barry, C.E., 3rd, Wang, H., Zhang, W. and Zhang, Y. (2011) Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science*, **333**, 1630-1632.
18. Sacchettini, J.C., Rubin, E.J. and Freundlich, J.S. (2008) Drugs versus bugs: in pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nat Rev Microbiol*, **6**, 41-52.

19. World Health Organization. Stop TB Dept. (2010) *Treatment of tuberculosis: guidelines*. 4th edition. ed. World Health Organization, Geneva.
20. World Health Organization. (2011) *Guidelines for the programmatic management of drug-resistant tuberculosis - 2011 update*. World Health Organization, Geneva.
21. Fine, P.E. (1995) Variation in protection by BCG: implications of and for heterologous immunity. *Lancet*, **346**, 1339-1345.
22. Barry, C.E., 3rd, Boshoff, H.I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson, R.J. and Young, D. (2009) The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol*, **7**, 845-855.
23. Wayne, L.G. (1994) Dormancy of *Mycobacterium tuberculosis* and latency of disease. *Eur J Clin Microbiol Infect Dis*, **13**, 908-914.
24. Wayne, L.G. and Sohaskey, C.D. (2001) Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annu Rev Microbiol*, **55**, 139-163.
25. Murphy, D.J. and Brown, J.R. (2007) Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. *BMC Infect Dis*, **7**, 84.
26. Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L. and Smith, I. (2004) Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. *J Bacteriol*, **186**, 895-902.
27. Browning, D.F. and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol*, **2**, 57-65.
28. Ebright, R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*, **304**, 687-698.
29. Archambault, J. and Friesen, J.D. (1993) Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol Rev*, **57**, 703-724.
30. Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863-1876.
31. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H. and Severinov, K. (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A*, **98**, 892-897.
32. Zhang, G. and Darst, S.A. (1998) Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science*, **281**, 262-266.
33. Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell*, **98**, 811-824.
34. Blatter, E.E., Ross, W., Tang, H., Gourse, R.L. and Ebright, R.H. (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell*, **78**, 889-896.
35. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. and Darst, S.A. (2000) A structural model of transcription elongation. *Science*, **289**, 619-625.
36. Hampsey, M. (2001) Omega meets its match. *Trends Genet.*, **17**, 190-191.
37. Saecker, R.M., Record, M.T., Jr. and Dehaseth, P.L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding,

isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol*, **412**, 754-771.
38. Sensi, P., Margalith, P. and Timbal, M.T. (1959) Rifomycin, a new antibiotic; preliminary report. *Farmaco Sci*, **14**, 146-147.
39. Shinnick, T.M. (1996) *Current Topics in Microbiology and Immunology*. Springer-Verlag.
40. David, H.L. (1970) Probability distribution of drug-resistant mutants in unselected populations of *Mycobacterium tuberculosis*. *Appl Microbiol*, **20**, 810-814.
41. Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell*, **104**, 901-912.
42. Ramaswamy, S. and Musser, J.M. (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis*, **79**, 3-29.
43. Floss, H.G. and Yu, T.W. (2005) Rifamycin-mode of action, resistance, and biosynthesis. *Chem Rev*, **105**, 621-632.
44. Jin, D.J. and Gross, C.A. (1988) Mapping and sequencing of mutations in the *Escherichia coli* rpoB gene that lead to rifampicin resistance. *J Mol Biol*, **202**, 45-58.
45. Wegrzyn, A., Szalewska-Palasz, A., Blaszczak, A., Liberek, K. and Wegrzyn, G. (1998) Differential inhibition of transcription from sigma70- and sigma32-dependent promoters by rifampicin. *FEBS Lett*, **440**, 172-174.
46. Artsimovitch, I., Vassylyeva, M.N., Svetlov, D., Svetlov, V., Perederina, A., Igarashi, N., Matsugaki, N., Wakatsuki, S., Tahirov, T.H. and Vassylyev, D.G. (2005) Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell*, **122**, 351-363.
47. Feklistov, A., Mekler, V., Jiang, Q., Westblade, L.F., Irschik, H., Jansen, R., Mustaev, A., Darst, S.A. and Ebright, R.H. (2008) Rifamycins do not function by allosteric modulation of binding of Mg2+ to the RNA polymerase active center. *Proc Natl Acad Sci U S A*, **105**, 14820-14825.
48. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C. and Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res*, **24**, 4420-4449.
49. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141-147.
50. Smith, I. (2003) Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*, **16**, 463-496.
51. Wosten, M.M. (1998) Eubacterial sigma-factors. *FEMS Microbiol Rev*, **22**, 127-150.
52. Rodrigue, S., Provvedi, R., Jacques, P.E., Gaudreau, L. and Manganelli, R. (2006) The sigma factors of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev*, **30**, 926-941.
53. Lonetto, M., Gribskov, M. and Gross, C.A. (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol*, **174**, 3843-3849.
54. Paget, M.S. and Helmann, J.D. (2003) The sigma70 family of sigma factors. *Genome Biol*, **4**, 203.

55. Barne, K.A., Bown, J.A., Busby, S.J. and Minchin, S.D. (1997) Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J*, **16**, 4034-4040.
56. Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L. and Darst, S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell*, **9**, 527-539.
57. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285-1290.
58. Dombroski, A.J., Walter, W.A. and Gross, C.A. (1993) Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev*, **7**, 2446-2455.
59. Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, **417**, 712-719.
60. Murakami, K.S., Masuda, S. and Darst, S.A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science*, **296**, 1280-1284.
61. Ishihama, A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol*, **54**, 499-518.
62. Maeda, H., Fujita, N. and Ishihama, A. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res*, **28**, 3497-3503.
63. Hughes, K.T. and Mathee, K. (1998) The anti-sigma factors. *Annu Rev Microbiol*, **52**, 231-286.
64. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544.
65. Gomez, M., Doukhan, L., Nair, G. and Smith, I. (1998) *sigA* is an essential gene in *Mycobacterium smegmatis*. *Mol Microbiol*, **29**, 617-628.
66. Wu, S., Howard, S.T., Lakey, D.L., Kipnis, A., Samten, B., Safi, H., Gruppo, V., Wizel, B., Shams, H., Basaraba, R.J. *et al.* (2004) The principal sigma factor *sigA* mediates enhanced growth of *Mycobacterium tuberculosis in vivo*. *Mol Microbiol*, **51**, 1551-1562.
67. Collins, D.M., Kawakami, R.P., de Lisle, G.W., Pascopella, L., Bloom, B.R. and Jacobs, W.R., Jr. (1995) Mutation of the principal sigma factor causes loss of virulence in a strain of the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*, **92**, 8036-8040.
68. Collins, D.M., Kawakami, R.P., Buddle, B.M., Wards, B.J. and de Lisle, G.W. (2003) Different susceptibility of two animal species infected with isogenic mutants of *Mycobacterium bovis* identifies phoT as having roles in tuberculosis virulence and phosphate transport. *Microbiology*, **149**, 3203-3212.
69. Fontan, P.A., Voskuil, M.I., Gomez, M., Tan, D., Pardini, M., Manganelli, R., Fattorini, L., Schoolnik, G.K. and Smith, I. (2009) The *Mycobacterium tuberculosis* Sigma factor B is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for in vivo growth. *J Bacteriol*, **191**, 5628-5633.
70. Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F.R. and Smith, I. (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol*, **31**, 715-724.

71. Hu, Y. and Coates, A.R. (2001) Increased levels of sigJ mRNA in late stationary phase cultures of *Mycobacterium tuberculosis* detected by DNA array hybridisation. *FEMS Microbiol Lett*, **202**, 59-65.
72. Hu, Y. and Coates, A.R. (1999) Transcription of two sigma 70 homologue genes, *sigA* and *sigB*, in stationary-phase *Mycobacterium tuberculosis*. *J Bacteriol*, **181**, 469-476.
73. Dubnau, E., Fontan, P., Manganelli, R., Soares-Appel, S. and Smith, I. (2002) *Mycobacterium tuberculosis* genes induced during infection of human macrophages. *Infect Immun*, **70**, 2787-2795.
74. Lee, J.H., Karakousis, P.C. and Bishai, W.R. (2008) Roles of SigB and SigF in the *Mycobacterium tuberculosis* sigma factor network. *J Bacteriol*, **190**, 699-707.
75. Vallecillo, A.J. and Espitia, C. (2009) Expression of *Mycobacterium tuberculosis* pe_pgrs33 is repressed during stationary phase and stress conditions, and its transcription is mediated by Sigma factor A. *Microb Pathog*, **46**, 119-127.
76. Sachdeva, P., Misra, R., Tyagi, A.K. and Singh, Y. (2010) The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J*, **277**, 605-626.
77. Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D. *et al.* (2001) Massive gene decay in the leprosy bacillus. *Nature*, **409**, 1007-1011.
78. Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell, B.G. and Cole, S.T. (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis*, **79**, 329-342.
79. Haugen, S.P., Ross, W. and Gourse, R.L. (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol*, **6**, 507-519.
80. Borukhov, S., Lee, J. and Laptenko, O. (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol*, **55**, 1315-1324.
81. Nudler, E., Kashlev, M., Nikiforov, V. and Goldfarb, A. (1995) Coupling between transcription termination and RNA polymerase inchworming. *Cell*, **81**, 351-357.
82. Komissarova, N. and Kashlev, M. (1997) Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A*, **94**, 1755-1760.
83. Toulme, F., Mosrin-Huaman, C., Sparkowski, J., Das, A., Leng, M. and Rahmouni, A.R. (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J*, **19**, 6853-6859.
84. Laptenko, O., Lee, J., Lomakin, I. and Borukhov, S. (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J*, **22**, 6322-6334.
85. Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W. and Darst, S.A. (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell*, **114**, 335-345.
86. Sosunova, E., Sosunov, V., Kozlov, M., Nikiforov, V., Goldfarb, A. and Mustaev, A. (2003) Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc Natl Acad Sci U S A*, **100**, 15469-15474.
87. Batada, N.N., Westover, K.D., Bushnell, D.A., Levitt, M. and Kornberg, R.D. (2004) Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc Natl Acad Sci U S A*, **101**, 17361-17364.

88. Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W. and Gourse, R.L. (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, **118**, 311-322.
89. Magnusson, L.U., Farewell, A. and Nystrom, T. (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol*, **13**, 236-242.
90. Potrykus, K. and Cashel, M. (2008) (p)ppGpp: still magical? *Annu Rev Microbiol*, **62**, 35-51.
91. Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I. and Vassylyev, D.G. (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell*, **118**, 297-309.
92. Stallings, C.L., Stephanou, N.C., Chu, L., Hochschild, A., Nickels, B.E. and Glickman, M.S. (2009) CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell*, **138**, 146-159.
93. England, P., Westblade, L.F., Karimova, G., Robbe-Saule, V., Norel, F. and Kolb, A. (2008) Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. *J Biol Chem*, **283**, 33455-33464.
94. Typas, A., Barembruch, C., Possling, A. and Hengge, R. (2007) Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J*, **26**, 1569-1578.
95. Gaal, T., Mandel, M.J., Silhavy, T.J. and Gourse, R.L. (2006) Crl facilitates RNA polymerase holoenzyme formation. *J Bacteriol*, **188**, 7966-7970.
96. Paget, M.S., Molle, V., Cohen, G., Aharonowitz, Y. and Buttner, M.J. (2001) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol*, **42**, 1007-1020.
97. Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.P. and Brodolin, K. (2012) *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the sigma A-containing RNA polymerase holoenzyme. *Nucleic Acids Res*, **40**, 6547-6557.
98. Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, **48**, 77-84.
99. Forti, F., Mauri, V., Deho, G. and Ghisotti, D. (2011) Isolation of conditional expression mutants in *Mycobacterium tuberculosis* by transposon mutagenesis. *Tuberculosis (Edinb)*, **91**, 569-578.
100. Provvedi, R., Boldrin, F., Falciani, F., Palu, G. and Manganelli, R. (2009) Global transcriptional response to vancomycin in *Mycobacterium tuberculosis*. *Microbiology*, **155**, 1093-1102.
101. Newell, K.V., Thomas, D.P., Brekasis, D. and Paget, M.S. (2006) The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. *Mol Microbiol*, **60**, 687-696.
102. Dey, A., Verma, A.K. and Chatterji, D. (2010) Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. *Microbiology*, **156**, 873-883.
103. Dey, A., Verma, A.K. and Chatterji, D. (2011) Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA. *Microbiology*, **157**, 2056-2071.
104. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

105. Papadopoulos, J.S. and Agarwala, R. (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, **23**, 1073-1079.
106. Livingstone, C.D. and Barton, G.J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci*, **9**, 745-756.
107. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.
108. Cole, C., Barber, J.D. and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*, **36**, W197-201.
109. Spizizen, J. (1958) Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *PNAS*, **44**, 1072-1078.
110. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. and Li, Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci*, **18**, 936-948.
111. Braunschweiler, L. and Ernst, R. (1983) Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**, 521–528.
112. Macura, S. and Ernst, R.R. (1980) Elucidation of cross relaxation in liquids by two-dimensional n.m.r. spectroscopy. *Mol. Phys.*, **41**, 95-117.
113. Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, **69**, 185-189.
114. Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.*, **111**, 1515-1517.
115. Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Formankay, J.D. (1993) A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in H2O Samples of Proteins
. *J. Magn. Reson.*, **101**, 333-337.
116. Grzesiek, S. and Bax, A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR*, **3**, 185-204.
117. Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J Magn Reson*, **89**, 496-514.
118. Bax, A. and Ikura, M. (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J Biomol NMR*, **1**, 99-104.
119. Grzesiek, S. and Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114**, 6291-6293.
120. Grzesiek, S. and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.*, **99**, 201-207.
121. Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28**, 6150-6156.

122. Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H. and Kay, L.E. (1993) A Gradient 13C NOESY-HSQC Experiment for Recording NOESY Spectra of 13C-Labeled Proteins Dissolved in H2O. *J. Magn. Reson. Ser.*, **102**, 317-321.
123. Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993) Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson. Ser*, **102**, 241-245.
124. Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, **6**, 135-140.
125. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*, **44**, 213-223.
126. Herrmann, T., Guntert, P. and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*, **319**, 209-227.
127. Guntert, P. and Wuthrich, K. (1991) Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J Biomol NMR*, **1**, 447-456.
128. Tsui, V. and Case, D.A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275-291.
129. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J Comput Chem*, **26**, 1668-1688.
130. Muskett, F.W., Thouta, S., Thomson, S.J., Bowen, A., Stansfeld, P.J. and Mitcheson, J.S. (2011) Mechanistic insight into human ether-a-go-go-related gene (hERG) K+ channel deactivation gating from the solution structure of the EAG domain. *J Biol Chem*, **286**, 6184-6191.
131. Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*, **14**, 51-55, 29-32.
132. Doreleijers, J.F., Sousa da Silva, A.W., Krieger, E., Nabuurs, S.B., Spronk, C.A., Stevens, T.J., Vranken, W.F., Vriend, G. and Vuister, G.W. (2012) CING: an integrated residue-based structure validation program suite. *J Biomol NMR*, **54**, 267-283.
133. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res*, **38**, W545-549.
134. Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.
135. Holzwarth, G. and Doty, P. (1965) The Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc*, **87**, 218-228.
136. Saito, H. (1986) Conformation-dependent 13C chemical shifts: A new means of conformational characterization as obtained by high-resolution solid-state 13C NMR. *Magn Reson Chem*, **24**, 835-845.
137. Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol*, **222**, 311-333.

138. Spera, S. and Bax, A. (1991) Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *J. Am. Chem Soc.*, **113**, 5490-5492.
139. Pardi, A., Wagner, G. and Wuthrich, K. (1983) Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *Eur J Biochem*, **137**, 445-454.
140. Williamson, M.P. (1990) Secondary-structure dependent chemical shifts in proteins. *Biopolymers*, **29**, 1423-1431.
141. Osapay, K. and Case, D.A. (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. *J Biomol NMR*, **4**, 215-230.
142. Bortoluzzi, A., Muskett, F.W., Waters, L.C., Addis, P.W., Rieck, B., Munder, T., Schleier, S., Forti, F., Ghisotti, D., Carr, M.D. *et al.* (2013) *Mycobacterium tuberculosis* RNA polymerase binding protein A (RbpA) and its interactions with sigma factors. *J Biol Chem*.
143. Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*, **4**, 363-371.
144. Farmer, B.T., 2nd, Constantine, K.L., Goldfarb, V., Friedrichs, M.S., Wittekind, M., Yanchunas, J., Jr., Robertson, J.G. and Mueller, L. (1996) Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat Struct Biol*, **3**, 995-997.
145. Williamson, R.A., Carr, M.D., Frenkiel, T.A., Feeney, J. and Freedman, R.B. (1997) Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry*, **36**, 13882-13889.
146. Muskett, F.W., Frenkiel, T.A., Feeney, J., Freedman, R.B., Carr, M.D. and Williamson, R.A. (1998) High resolution structure of the N-terminal domain of tissue inhibitor of metalloproteinases-2 and characterization of its interaction site with matrix metalloproteinase-3. *J Biol Chem*, **273**, 21736-21743.
147. Hartkoorn, R.C., Sala, C., Magnet, S.J., Chen, J.M., Pojer, F. and Cole, S.T. (2010) Sigma factor F does not prevent rifampin inhibition of RNA polymerase or cause rifampin tolerance in *Mycobacterium tuberculosis*. *J Bacteriol*, **192**, 5472-5479.
148. Waters, L.C., Veverka, V., Bohm, M., Schmedt, T., Choong, P.T., Muskett, F.W., Klempnauer, K.H. and Carr, M.D. (2007) Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterization of its interaction with eIF4A. *Oncogene*, **26**, 4941-4950.
149. Veverka, V., Henry, A.J., Slocombe, P.M., Ventom, A., Mulloy, B., Muskett, F.W., Muzylak, M., Greenslade, K., Moore, A., Zhang, L. *et al.* (2009) Characterization of the structural features and interactions of sclerostin: molecular insight into a key regulator of Wnt-mediated bone formation. *The Journal of biological chemistry*, **284**, 10890-10900.
150. Williamson, R.A., Muskett, F.W., Howard, M.J., Freedman, R.B. and Carr, M.D. (1999) The effect of matrix metalloproteinase complex formation on the conformational mobility of tissue inhibitor of metalloproteinases-2 (TIMP-2). *The Journal of biological chemistry*, **274**, 37226-37232.
151. Gao, G., Semenchenko, V., Arumugam, S. and Van Doren, S.R. (2000) Tissue inhibitor of metalloproteinases-1 undergoes microsecond to millisecond motions at sites of matrix metalloproteinase-induced fit. *Journal of molecular biology*, **301**, 537-552.
152. Gilmore, J.M., Bieber Urbauer, R.J., Minakhin, L., Akoyev, V., Zolkiewski, M., Severinov, K. and Urbauer, J.L. (2010) Determinants of affinity and activity of the anti-sigma factor AsiA. *Biochemistry*, **49**, 6143-6154.

153. Pauling, L., Corey, R.B. and Branson, H.R. (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A*, **37**, 205-211.
154. Pauling, L. and Corey, R.B. (1951) Configuration of polypeptide chains. *Nature*, **168**, 550-551.
155. Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F. *et al.* (1991) New use of BCG for recombinant vaccines. *Nature*, **351**, 456-460.
156. Bougdour, A., Lelong, C. and Geiselmann, J. (2004) Crl, a low temperature-induced protein in *Escherichia coli* that binds directly to the stationary phase sigma subunit of RNA polymerase. *J Biol Chem*, **279**, 19540-19550.
157. Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R.A., Ghosh, S., Parsy, M.L., Simpson, P.J., O'Dwyer, K., Matthews, S.J. and Paget, M.S. (2013) The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucleic Acids Res*, **41**, 5679-5691.
158. Marintchev, A., Frueh, D. and Wagner, G. (2007) NMR methods for studying protein-protein interactions involved in translation initiation. *Methods Enzymol*, **430**, 283-331.
159. Bertini, I., McGreevy, K.S. and Parigi, G. (2012) *NMR of biomolecules : towards mechanistic systems biology*. Wiley-VCH, Weinheim, Germany.

Published work

Bortoluzzi, A., Muskett, F.W., Waters, L.C., Addis, P.W., Rieck, B., Munder, T., Schleier, S., Forti, F., Ghisotti, D., Carr, M.D. *et al.* (2013) *Mycobacterium tuberculosis* RNA polymerase binding protein A (RbpA) and its interactions with sigma factors. *J Biol Chem*.