

MOLECULAR FEATURES AND PROPERTIES OF  
MYCOBACTERIAL PROTEINS LINKED TO  
TUBERCULOSIS PATHOGENESIS

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Dariush Ilghari MSci (Isfahan)

Department of Biochemistry

University of Leicester

March 2009

## MOLECULAR FEATURES AND PROPERTIES OF MYCOBACTERIAL PROTEINS LINKED TO TUBERCULOSIS PATHOGENESIS

Dariusz Ilghari, March 2009

**Abstract:** The *Mycobacterium tuberculosis* genome codes for 11 pairs of CFP-10/ESAT-6 proteins (Esx family) as well as the apparatus required for secretion of these proteins. The core machinery for the secretion of ESAT-6 and CFP-10 is encoded by their surrounding genes. Recent studies also identified a distant region, the *Rv3612c-Rv3616c* operon, which is essential for the CFP-10/ESAT-6 secretion. Constructs carrying *Rv3613c*, *Rv3614c*, *Rv3615c* and *Rv3616c* coding regions were produced and used to express the corresponding proteins. However, only *Rv3614c* and *Rv3615c* were expressed using an *E. coli*-based expression system. Analysis using a range of spectroscopic techniques on the purified proteins revealed that both *Rv3615c* and *Rv3614c* contain stable secondary structure, but little if any stable tertiary structure and exist in a molten globule-like state. This suggests the proteins probably undergo folding upon binding with possible functional partners. Yeast-two hybrid studies showed no intermolecular interaction between the proteins encoded by the *Rv3616c-Rv3612c* operon, perhaps suggesting the formation of a higher order multi-protein complex.

Together with CFP-10 and ESAT-6, *Rv0287* and *Rv0288* are the members of the Esx family which are clearly implicated in *M. tuberculosis* pathogenesis. The expression vectors carrying *Rv0287* and *Rv0288* coding regions were constructed and used to express the proteins. Analysis using a range of spectroscopic techniques on the purified proteins showed that *Rv0288* contains up to 30 % helical secondary structure, but little if any stable tertiary structure and exists in a molten globule-like state. In contrast, *Rv0287* has been found to form an unstructured, random coil polypeptide. The work reported here also shows that *Rv0287* and *Rv0288* form a tight 1:1 complex which is predominantly helical. Furthermore, the *Rv0287-Rv0288* complex was found to be significantly more stable to thermal denaturation than CFP-10-ESAT-6. The high resolution solution structure reported here reveals that both proteins, *Rv0287* and *Rv0288*, adopt an elongated helix-turn-helix hairpin structure in which the proteins lie antiparallel to each other, forming a stable four helix bundle. Comparison of the CFP-10-ESAT-6 and *Rv0287-Rv0288* complexes also revealed that the overall backbone fold for the complexes is very similar although they display significantly different surface features.

## **Acknowledgements**

Firstly I would like to thank the Iran ministry of health and medical education for the sponsorship and funding which made this project possible. I would like to acknowledge my supervisor, Professor Mark D. Carr, for the opportunity to work on an interesting project. I greatly appreciate his constant support and direction he has provided me throughout the project.

I am also grateful to Dr. Bernard Burke, Dr. Peter R. Nielsen and Dr. Stephen Gordon who served on my committee, providing enthusiastic feedback. My thanks go to the members of our group for their support and enthusiasm, which sustained and encouraged me throughout the project; especially to Dr. Kirsty L. Lightbody, Dr. Lorna C. Waters, Dr. Philip S. Renshaw, Dr. Vaclav Veverka, Alister, Catherine, Ian, Ojay, Sami and Sarah. My gratitude extends to Dr. Wei-Cheng Huang for his assistance. His guidance has been greatly appreciated.

Finally to my Mum, Dad, brothers, Reza and Mohammad reza, and sister, Parisa, whose belief in me has supported me throughout my studies.

Dariush Ilghari, March 2009

## Table of Contents

<b>Abstract</b> .....	i
<b>Acknowledgements</b> .....	ii
<b>Table of contents</b> .....	iii
<b>Abbreviations</b> .....	ix
<b>Chapter 1- Introduction</b> .....	1
<b>1.1. Tuberculosis: The current situation</b> .....	1
<b>1.2. The basics of bacteriology</b> .....	3
1.2.1. Classification of TB agents.....	3
1.2.2. Microscopic morphology.....	4
1.2.3. Cell envelope structure.....	5
1.2.4. Generation time.....	7
<b>1.3. Microbe-macrophage interaction</b> .....	7
<b>1.4. Process of the disease</b> .....	10
<b>1.5. Tuberculosis prevention</b> .....	15
<b>1.6. Treatment of tuberculosis</b> .....	17
<b>1.7. Immunity to tuberculosis</b> .....	20
<b>1.8. Mycobacterium tuberculosis genome</b> .....	21
<b>1.9. M. Tuberculosis virulence factors</b> .....	22
1.9.1. ESAT-6-CFP-10 protein family.....	23
<b>1.10. Protein secretion systems in mycobacteria</b> .....	26

1.10.1. General secretion pathway (GSP or Sec secretion system).....	27
1.10.2. Twin arginine transporter (Tat).....	27
1.10.3. Snm (ESX) secretion systems.....	28

**Chapter 2- Molecular features and properties of proteins encoded by the Rv3616c-Rv3612c gene cluster.....** 29

<b>2.1. ESX-1 Secretion pathway.....</b>	<b>29</b>
2.1.1. Regulation of ESX-1 secretion pathway.....	31
2.1.2. Role of ESX-1 secretion system in pathogenesis of M. tuberculosis.....	35
2.1.3. Molecular mechanism of ESX-1 secretion pathway.....	36
2.1.3.1. <i>Yeast two-hybrid system</i> .....	38
2.1.4. Aims.....	41
<b>2.2. Materials and methods.....</b>	<b>42</b>
2.2.1. DNA Analysis by agarose electrophoresis.....	42
2.2.2. SDS- polyacrylamide gel electrophoresis.....	42
2.2.3. Western blotting.....	43
2.2.4. Amplification of Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions.....	43
2.2.5. Zero blunt® PCR cloning.....	45
2.2.6. Subcloning of Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions into pET28a (+) and pET23a (+).....	45
2.2.7. Protein expression trials.....	46
2.2.8. His6-tagged Rv3614c expression and purification.....	47
2.2.9. Rv3615c expression and purification.....	48
2.2.10. Protein concentration.....	49

2.2.11. Far UV circular dichroism spectroscopy.....	49
2.2.12. Thermal denaturation studies.....	50
2.2.13. Fluorescence spectroscopy and chemical denaturation.....	50
2.2.14. Nuclear magnetic resonance spectroscopy for Rv3614c.....	51
2.2.15. Strains, plasmids and media used for yeast two-hybrid analyses.....	51
2.2.16. Preparation of yeast two-hybrid vectors.....	52
2.2.17. Transformation of <i>S. cerevisiae</i> .....	52
2.2.18. $\beta$ - galactosidase assays.....	53
2.2.18.1. Colony filter lift assays.....	53
2.2.18.2. Quantitative liquid assays (Assay of cell extract).....	54
<b>2.3. Results and discussion.....</b>	<b>56</b>
2.3.1. Cloning of Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions.....	56
2.3.2. Protein expression trails.....	57
2.3.3. His <sub>6</sub> -tagged Rv3614c expression and purification.....	61
2.3.4. Structural characterization of the Rv3614c protein.....	64
2.3.4.1. Fluorescence analysis of Rv3614c folding.....	64
2.3.4.2. Chemical denaturation of Rv3614c.....	65
2.3.4.3. Far UV circular dichroism spectroscopy for Rv3614c.....	66
2.3.4.4. The effect of thermal variation on Rv3614c.....	66
2.3.4.5. 1D <sup>1</sup> H NMR analysis of Rv3614c.....	68
2.3.4.6. Discussion.....	69
2.3.5. Rv3615c expression and purification.....	70
2.3.6. Structural characterization of the Rv3615c Protein.....	74
2.3.6.1. Fluorescence analysis of Rv3615c folding.....	74

2.3.6.2. <i>Chemical denaturation of Rv3615c</i> .....	75
2.3.6.3. <i>Far UV circular dichroism spectroscopy for Rv3615c</i> .....	76
2.3.6.4. <i>The effect of thermal variation on Rv3615c</i> .....	76
2.3.6.5. <i>Discussion</i> .....	78
2.3.7. Complex formation between proteins encoded by the Rv3616c-Rv3613c operon.....	79
2.3.7.1. <i>Discussion and conclusion</i> .....	85
<b>Chapter 3 – Solution structure of the complex formed between Rv0287 and Rv0288</b> .....	<b>87</b>
<b>3.1. Introduction-Part 2</b> .....	<b>87</b>
3.1.1. Region 3, a similar copy of the Region 1 (RD1) with functional genes.....	88
3.1.2. Region 3 as an important gene cluster in M. tuberculosis pathogenesis.....	89
3.1.3. Region 3 contains a set of genes encoding highly immunogenic antigens.....	90
3.1.3.1. <i>Epitope mapping of the immunodominant antigen TB10.4 (Rv0288)</i> .....	90
3.1.3.2. <i>Rv0287 and Rv0288 as two subunit vaccine candidates</i> .....	91
3.1.3.3. <i>Rv0287 and Rv0288 as protein tools for diagnosis of TB</i> .....	92
3.1.4. Region3 and its biological function.....	93
3.1.5. Structure and molecular function of the ESX-like proteins encoded by Rv0288 and Rv0287 genes.....	94
3.1.6. Aims.....	97
<b>3.2. Methods and materials- Part 2</b> .....	<b>98</b>
3.2.1. Ligation independent cloning (LIC-PCR).....	98
3.2.1.1. <i>Amplification of Rv0288 and Rv0287 coding regions</i> .....	98

3.2.1.2. Ligation independent cloning (LIC-PCR) of the Rv0288 and Rv0287 coding Regions.....	99
3.2.1.3. Transformation of DH5- $\alpha$ competent cells with the recombinant plasmids	100
3.2.1.4. Colony PCR screening.....	101
3.2.2. Subcloning of Rv0287 coding region into pET23a (+) expression vector....	101
3.2.3. Protein expression trials.....	102
3.2.4. Purification Rv0287 and Rv0288 proteins.....	102
3.2.5. Expression and purification of His <sub>6</sub> -tagged Rv0288 and His <sub>6</sub> -tagged Rv0287 proteins.....	103
3.2.6. Preparation of ESAT-6-CFP-10 complex.....	104
3.2.7. Preparation of the single- and double-labelled Rv0287 and Rv0288 proteins.....	104
3.2.8. Protein concentration.....	105
3.2.9. Circular dichroism and fluorescence spectroscopy.....	105
3.2.10. NMR spectroscopy.....	105
3.2.11. Sequence specific assignments.....	106
3.2.12. Secondary structure determination.....	107
3.2.13. Determination of slowly exchanging amide protons.....	107
3.2.14. Structural calculations.....	108
3.2.15. Comparative analysis between Rv0288-Rv0287 and its structurally homologue proteins.....	109
<b>3.3. Results and discussion- Part2.....</b>	<b>110</b>
3.3.1. Ligation independent cloning (ILC-PCR) of Rv0287 and Rv0288 coding regions.....	110
3.3.2. Subcloning of the coding Rv0287 region into pET23a (+).....	111

3.3.3. Protein expression trails for the recombinant Rv0287 and Rv0288 with removable His6 and GST tags.....	112
3.3.4. Expression and purification of Rv0288-Rv0287 complex.....	115
3.3.5. Expression and purification of the His6-Rv0287-His6-Rv0288 complex.....	119
3.3.6. Initial structural characterization of the Rv0287-Rv0288 complex.....	123
3.3.6.1. <i>Far UV circular dichroism spectroscopy</i> .....	123
3.3.6.2. <i>The effects of thermal variation on Rv0288, and the Rv0288-Rv0287 and CFP-10-ESAT-6 complexes</i> .....	124
3.3.6.3. <i>Fluorescence analysis of Rv0288-Rv0287 complex</i> .....	125
3.3.6.4. <i>1D <sup>1</sup>H NMR analysis of the Rv0288-Rv0287 complex</i> .....	126
3.3.6.5. <i>Discussion</i> .....	128
3.3.7. High-resolution structure determination of the Rv0287-Rv0288 complex...	130
3.3.7.1. <i>Sequence specific assignments</i> .....	130
3.3.7.2. <i>Secondary structure determination</i> .....	139
3.3.7.3. <i>Determination of slowly exchanging amide protons</i> .....	140
3.3.7.4. <i>Structural calculations for Rv0288-Rv0287 complex</i> .....	141
3.3.7.5. <i>Discussion</i> .....	144
3.3.7.6. <i>Conclusion</i> .....	169
<b>Appendices</b> .....	171
<b>Appendix-1. Reagents and culture media</b> .....	171
1.1. Selected reagents.....	171
1.2. Bacterial culture media.....	172
1.3. Yeast culture media.....	174
<b>Appendix-2. Competent cell preparation, transformation protocol</b> .....	175

2.1. Preparation of competent cells.....	175
2.2. Transformation of competent cells.....	176
<b>References.....</b>	<b>177</b>
<b>Published work.....</b>	<b>199</b>

Ilghari, D., Waters, L. C., Veverka, V., Muskett, F. W., and Carr, M. D. (2009)  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  resonance assignments and secondary structure determination of the Rv0287-Rv0288 complex. *Biomolecular NMR Assignments* (Accepted).

Lightbody, K.L., Ilghari, D., Waters, L.C., Carey, G., Bailey, M.A., Williamson, R.A., Renshaw, P.S., and Carr M.D. (2008) Molecular features governing the stability and specificity of functional complex formation by Mycobacterium tuberculosis CFP-10/ESAT-6 family proteins. *J Biol Chem.* 283(25):17681-17690.

## Abbreviations

Ag	Antigen
AIDS	Acquired immunodeficiency syndrome
Amp	Ampicillin
APC	Antigen presenting cell
ATP	Adenosine-5'-triphosphate
BAC	Bacterial artificial chromosome
B-cell	Bursa dependent lymphocyte
BCG	Bacille Calmette-Guerin
bp	Base pair
CD	Circular dichroism
CFP-10	Culture filtrate protein of 10 kD
Da	Dalton
DBD	DNA binding domain
DC	Dendritic cell
DOTS	Directly observed therapy short-course
DTT	1, 4-dithiothreitol
EDTA	Ethylene diaminetetraacetate
ESAT-6	Early secreted antigen target of 6 kDa
FPLC	Fast protein liquid chromatography
Gdn-HCl	Guanidine hydrochloride
GST	Glutathione-S-transferase
HIV	Human immunodeficiency virus

HSQC	Heteronuclear single quantum coherence
INF $\gamma$	Interferon gamma
IL	Interleukin
IPTG	Isopropyl-thio- $\beta$ -D-galactoside
Kan	Kanamycin
kDa	Kilo Dalton
LB	Luria-Bertani
LDS	Lithium dodecyl sulphate
LiAc	Lithium acetate
LIC	Ligase independent cloning
MDR	Multi-drug resistant
M	Marker
MW	Molecular weight
ms	Milli-second
NH	Backbone amide group
Ni <sup>2+</sup> -NTA	Nickel-nitrilotriacetic acid
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD	Optical density
ONPG	<i>ortho</i> -nitrophenyl- $\beta$ -D-galactopyranoside
ONP	<i>ortho</i> -nitrophenol
ORF	Open reading frame
PAGE	Polyacrilamide gel electrophoresis
PCR	Polymerase chain reaction

PE	Proline-Glutamic acid sequence
<i>Pfx</i>	<i>Pyrococcus sp. Platinum® Pfx</i>
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
PPE	Proline-Proline-Glutamic acid sequence
RD	Region of difference
RPM	Revolutions per minute
Rv0287/Rv0288	Refers to Rv0287 and Rv0288 as two proteins
Rv0287-Rv0288	Refers to the 1:1 complex
SD	Synthetic dropout
SDS	Sodium dodecyl sulphate
TAD	Transcription activation domain
TALOS	Torsion angle likelihood obtained from shift and sequence similarity
TB	Tuberculosis
T-cell	Thymus dependent lymphocyte
TNF $\alpha$	Tumor necrosis factor alpha
TOCSY	Total correlation spectroscopy
UAS	Upstream activating sequence
UV	Ultra violet
v/v	Volume per volume
WATERGATE	Water suppression by gradient-tailored excitation
w/v	Weight per volume
WHO	World health organisation
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside
XDR	Extensively drug resistant

YPD	Yeast extract, peptone, dextrose
$\Phi$	Phi, the amide nitrogen-alpha carbon dihedral angle
$\Psi$	Psi, the alpha carbon-carboxyl carbon dihedral angle

### DNA bases

Adenine	A	Guanine	G
Cytosine	C	Thymine	T

### Single and three letter codes for amino acids

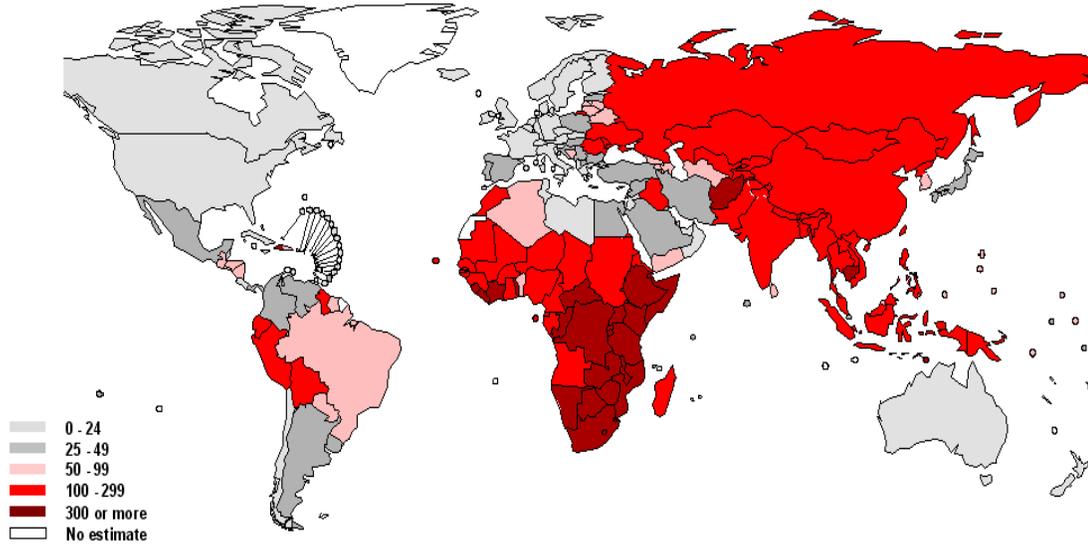
Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Asparatic acid	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamic acid	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile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

## **Chapter 1- Introduction**

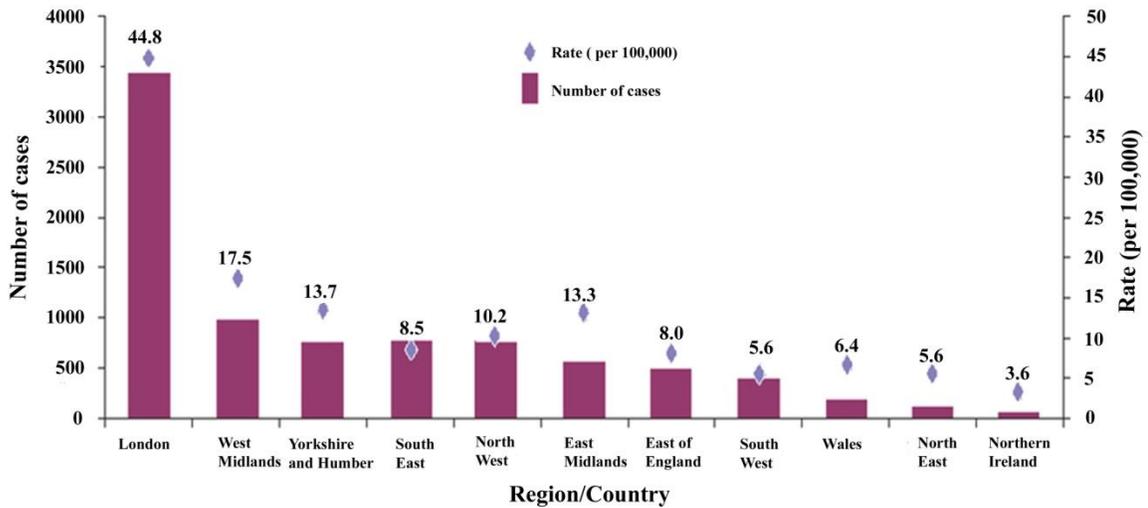
### **1.1. Tuberculosis: The current situation**

Tuberculosis (TB), once thought to be controlled, is now resurgent in many parts of the world, claiming more lives annually than any other infectious disease. Every second someone in the world is newly infected with TB Bacilli and five to ten percent of the infected individuals become infectious at some time during their life. An estimated two billion people, about one third of world's population, is currently infected with TB Bacilli, resulting in about 2 million deaths annually (1, 2).

The major burden of TB is borne by developing countries, particularly in Africa and South-East Asia (2), mainly because of the reduced socio-economic status, lack of proper public healthcare system, emergence of both multi-drug resistant (MDR) and extensively drug resistant (XDR) tuberculosis and spread of HIV/AIDS (Figure 1.1). However, in economically favored countries with outstanding clinical facilities and proper public healthcare system, other risk factors such as, stressful life, alcohol consumption and particularly increased number of immigrants from high burden counties have resulted in an increase in the number of reported TB cases (3, 4, and 5). In 2006 a total of 8497 tuberculosis cases were reported in the UK, a rate of 14.0 per 100,000 population, and the London region accounted for 40 % of the total cases and had the highest incidence (44.8 per 100,000). The majority of cases in the UK occurred in subjects aged 15-44 years, and 72% of cases were non-UK born (6) (Figure 1.2).



**Figure 1.1: Estimated global TB incidence (2004). All forms of TB estimated per 100,000 population. (From WHO Report 2005, Global tuberculosis control-surveillance, planning, financing World Health Organization, Geneva, Switzerland, 2006)**



**Figure 1.2: Tuberculosis case reports and rates by region/country, England, Wales and Northern Ireland, 2006 (6).**

## 1.2. The Basics of Bacteriology

### 1.2.1. Classification of TB agents

Bacteria of the genus *Mycobacterium* are acid-fast, aerobic, non-motile and non-spore-producing rods. They belong to the phylum *Actinobacteria* (Table 1.1) which are characterised by a high GC content DNA (61-71 %) and high lipid content in the cell wall, probably the highest among all bacteria. Even showing the above mentioned common properties, the species within the genus exhibit diversity in many aspects. Apart from few pathogen species within the genus which are host-dependent and preferentially reside in the intracellular environment of mononuclear phagocytes, most of them replicate without restraint in their natural ecosystems and rarely cause diseases in higher vertebrates. The host-dependent mycobacteria include *M. avium subsp. Paratuberculosis*, *M. lepraemurium*, *M. leprae*, and the members of *M. tuberculosis* complex (7). The members of the complex contain identical 16S rRNA sequences and share over 99 % nucleotide similarity but differ in their phenotypes, pathogenicity, zoonotic potential, and host tropisms (8). *M. tuberculosis* and the regional subtypes *M. africanum* and *M. canettii* are the major causative agents of human tuberculosis. *M. microti* and *M. bovis* are mostly pathogenic in a wide range of domesticated and wild animals, and can be transmitted to humans. In addition, there are some specific strains extracted from seals and goats that have been named *M. pinnipedi* and *M. caprae*, however sometimes they are identified as subspecies of *M. bovis* (7).

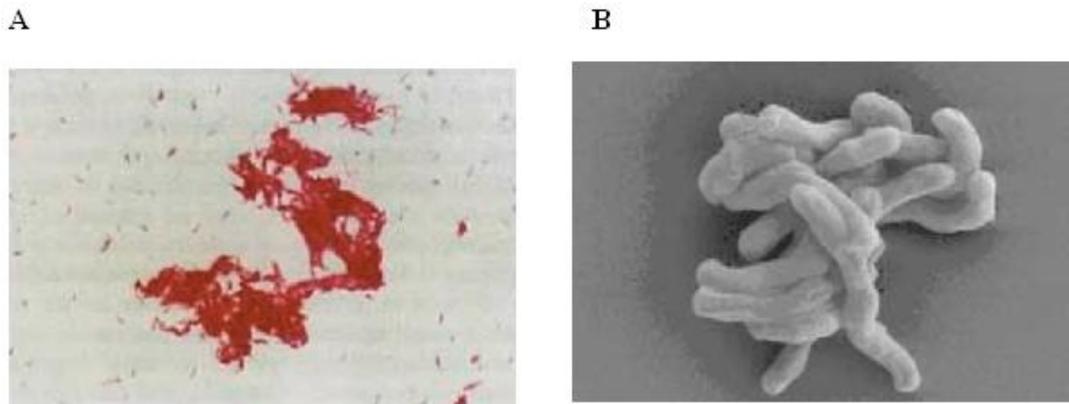
**Table 1.1: Systematic classifications of the agents of tuberculosis.**

<b>Kingdom</b>	<b>Bacteria</b>
<b>Phylum</b>	<b>Actinobacteria</b>
<b>Class</b>	<b>Actinobacteria</b>
<b>Subclass</b>	<b>Actinobacteridae</b>
<b>Order</b>	<b>Actinomycetales</b>
<b>Suborder</b>	<b>Corynebacterineae</b>
<b>Family</b>	<b>Mycobacteriaceae</b>
<b>Genus</b>	<b><i>Mycobacterium</i></b>
<b>Species</b>	<b>Unique genus</b>
	<b><i>M. tuberculosis</i></b>
	<b><i>M. bovis</i></b>
	<b><i>M. africanum</i></b>
	<b><i>M. canettii</i></b>
	<b><i>M. caprae</i></b>
	<b><i>M. pinnipedii</i></b>

### 1.2.2. Microscopic Morphology

Typically tubercle bacilli appear as straight or slightly curved rods. This microscopic morphology, however, depends on age of the culture and growth conditions and might change from short coccobacilli to long rods. *M. tuberculosis* is rarely pleomorphic compared to other actinomycetales and some fast growing mycobacteria. When examined in culture or clinical specimens, it does not appear as branched chains or elongated filaments. Intercellular bacilli found in experimental infected macrophages, however, are significantly elongated into filaments. Virulent replicating *M. tuberculosis* is strongly acid fast and tends to appear as hydrophobic bundles (Figure 1.3). In addition, the bacilli form textured rough colonies and expanded gummy veils on the surface of solid and broth media

respectively. In contrast, the contained and non-pathogenic bacteria are usually unevenly and weakly stained and observed as separated mats and more scattered on liquid and solid media respectively (7).

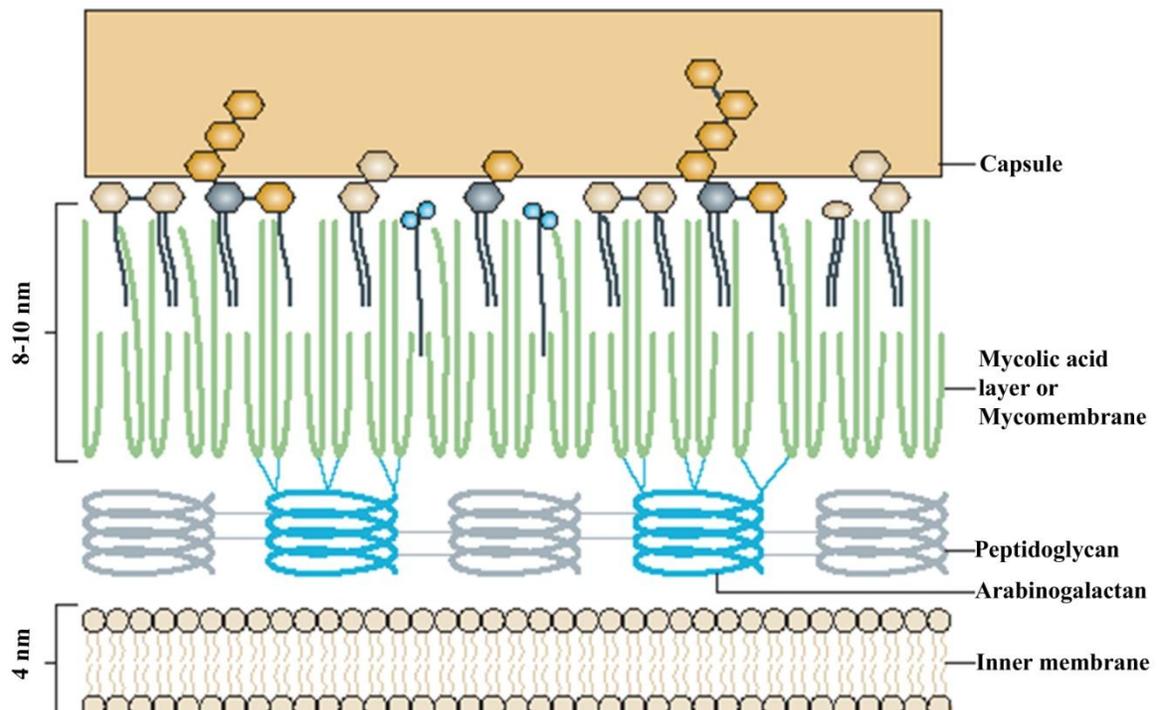


**Figure 1. 3: (A) Ziehl – Neelsen staining and (B) electronic microscopy of *Mycobacterium tuberculosis* growing in culture (7)**

### 1.2.3. Cell envelope structure

As previously mentioned, the most distinctive anatomical characteristic of mycobacteria is the cell envelope. In contrast to the gram-negative bacteria, mycobacteria do not have an extra outer membrane. Although the cell envelope of mycobacteria is structurally comparable to the gram-positive bacteria, they do not fit into this class of bacteria as the molecules contained in the cell wall are mostly lipids rather than polysaccharides and proteins(7). According to one of the current views of the mycobacterial envelop, the cell wall is primarily composed of three different types of molecules (peptidoglycan, mycolic acids and arabinogalactan) which are covalently attached. Covalent attachment of mycolic acids provides a layer known as mycomembrane which is characterized by its tremendously

hydrophobicity and low fluidity. As illustrated in figure 1.4, the outer part of the membrane consists of different lipids such as cord factor or dimycolyl trehalose and phenolic glycolipids which are intercalated with the mycolic acids. Finally, a capsule containing mainly polysaccharides (arabinomannan and glucan) forms the outer layers of the mycobacterial envelop (9).



**Figure 1.4: Schematic representation of the *M. tuberculosis* cell envelop.** The cell wall mainly contains a large cell-wall core that contains three different covalently linked structures (peptidoglycan (grey), arabinogalactan (blue) and mycolic acids (green)). The covalent linkage of mycolic acids provides an extremely hydrophobic layer known as the mycomembrane. The outer part of the mycomembrane is composed of various free lipids including cord factor or dimycolyl trehalose, phenolic glycolipids, phosphatidylinositol mannosides and sulpholipids that are intercalated with the mycolic acids. Most of these lipids are specific for mycobacteria. The capsule mainly contains polysaccharides (glucan and arabinomannan) (9).

#### 1.2.4. Generation Time

Unlike most cultivable bacteria, which replicate at regular intervals ranging from 15-60 minutes, *M. tuberculosis* duplicates in 12 to 24 hours under favorable laboratory conditions. Although the extreme impermeability of the *M. tuberculosis* cell envelope that restricts nutrient uptake is, in part, associated with the slow growth rate, it is believed that the main limiting factor for growth is the limited rate of RNA synthesis as evidenced by the studies that showed both the RNA/DNA ratio and RNA elongation rate are ten times higher in *E. coli* compared to *M. tuberculosis* (7).

### 1.3. Microbe-macrophage interaction

Initially pathogens interact with host epithelial tissues where they provide physical barriers to invasion and infection. In the next stage of host-pathogen encounter, host phagocytes and antigen-presenting cells (APCs), such as macrophages and dendritic cells (DCs), are involved, providing primary innate host immune response and also contributing to the generation of adaptive immune responses (4,10,11, 27). These myeloid cells internalize extra-cellular and intra-cellular pathogens into membrane-bound organelles called phagosomes. Phagosomes containing extra-cellular pathogens mature and fuse with lysosomes, making phagolysosomes. These acidic environments which are rich in hydrolytic enzymes, serve to degrade and kill the bacteria. Degradation of bacteria produces antigens that may elicit MHC- or CD1-restricted T cell responses (4, 10, 11, and 27).

In general, intracellular pathogens recruit two main strategies to survive and grow within the mentioned phagocytes: (I) they commonly inhibit phagolysosome biogenesis via the manipulation of host signal transduction pathways and the alteration of endocytic traffic. (II) In contrast, some groups of intracellular pathogens, such as *Listeria monocytogene* and *Shigella flexneri*, lyse the phagosomal membrane and translocate from the endocytic system into the host cytosol (27).

Analysis of infected macrophages grown in culture and those which were directly isolated from TB patients' bronchial lavage fluid has shown that the *M. tuberculosis* can survive inside the infected macrophages by inhibiting the typical maturation of phagosome, thereby limiting its acidification to pH 6.4 and restricting fusion with lysosomes (4).

Modulation of the phagosomes seems to be mediated by both mycobacterial cell-wall lipids and other bacterial effectors. *M. tuberculosis* cell-wall lipids, particularly lipoarabinomannan, appear capable of inhibiting phagolysosome fusion. Analysis of the mutants defective in expressing a serine/threonine kinase, PknG, and a secreted phosphatase, SapM, identified that these mutants are unable to block maturation of their phagosomes. SapM is able to dephosphorylate phosphoinositol species, PI3-P, which are essential for phagolysosome biogenesis (10, 12 and 13). The cytosolic  $Ca^{2+}$  concentration is also reported to be important for the maturation of the phagosome. Inhibition of sphingosine kinase 1 by *M. tuberculosis* prevents an increase in the concentration of cytosolic calcium, resulting in the inhibition of effectors, such as calmodulin and calmodulin-dependent protein kinase II, which are essential in phagolysosome maturation (14, 15)

Growth status of the bacteria in macrophages is also influenced by immune modulation of their host. Immune activated macrophages are able to reduce the effects of the bacilli on phagosome maturation and drive them to a more acidic (pH 5.2), bactericidal environment. Analysis of the effects of cytokines on the maturation of the mycobacterial phagosome has previously demonstrated that the phagolysosome fusion of *M. tuberculosis*-containing vacuoles with lysosomes is obtained by the activity of INF-inducible P47 GTPase family (4).

Despite the partial cytosolic localization with low percentages of *Mycobacterium marinum* (27), both *M. tuberculosis* and *M. leprae* were thought to be prototypical 'endosomal' pathogens. This idea was due largely to the ability of *M. tuberculosis* in blocking phagolysosome biogenesis and probably the fact that the majority of studies on localization of *M. tuberculosis* have been focused only on the first 48 hours post-infection. However, recent analyses conducted by van der Wel *et al.* resulted in a different conclusion. In these studies an extended time course was used to examine the localization of *M. tuberculosis* and *M. leprae* for up to 7 days of post-infection. These analyses showed that phagosomes containing the virulent species were fused with lysosomes in human monocyte-derived dendritic cells and macrophages, and after 2 days both bacilli progressively translocate from the phagolysosomes into the cytosol in non-apoptotic cells. This translocation occurs in a RD-1 (ESAT-6-CFP-10) and *espA* (Rv3616c) dependent manner and it does not happen simply by mycobacteria outgrowing their phagolysosomal space. In addition, the translocation is followed by replication of the bacilli in the cytosol of the infected cells resulting in considerable cell death (27).

#### **1.4. Process of the disease**

Basically TB infection is initiated by respiratory or/and gastrointestinal route of exposure.

Only a low dose of the bacilli is enough to initiate infection, however only 5-10 % of the infected patients develop to clinical disease (16).The infectious bacilli are usually inhaled as droplets or nuclei from the atmosphere. It is known the nuclei, each containing one bacterium, can remain viable in atmosphere for several hours (4). The majority of the inhaled bacilli remain in the upper respiratory tract and are expelled by the mucociliary escalator (17).

The bacteria that reach the lung are then phagocytosed by alveolar macrophages. At this stage of the disease process three possible outcomes might occur. (I) the infection is eradicated by the immune systems immediately, (II) infection progresses into primary tuberculosis, in particular in immunocompromised patients and neonates, because the host immune cells fail to mount effective innate immune response, and (III) the infection is contained by the immune system (17).

Mycobacteria produce a plethora of lipoprotein, lipid and glycolipid molecules which are recognized by macrophage receptors, stimulating different receptors on the surface of the macrophage cell membrane. These mycobacterial signals induce distinct cell-signalling pathways, regulating initial inflammatory response. As an example, when Toll-like receptors (TLR)-2 are stimulated by mycobacterial lipoproteins, a particular cell-signalling pathway is triggered, leading to pro-inflammatory response. This response can in turn result in mycobacterial killing and also promote apoptosis of the infected macrophages. In contrast, mannose-capped lipoarabinomannan result in an anti-inflammatory response

through DC-SIGN receptor on the surface of dendritic cells, thereby reducing anti-mycobacterial activity and also stimulating release of IL-10 (5). As a result, a balance in innate immune signalling determines the strength of initial inflammatory response.

It is extensively hypothesized that following phagocytosis of the bacilli by alveolar macrophages, the innate immune host cells are triggered to attack lung epithelium (18, 19). In vitro analysis has demonstrated that human and murine macrophages produce a strong pro-inflammatory response (release of IL-1, IL6, IL12 and TNF $\alpha$ ) when their toll-like receptors (TLR) are activated by antigens found abundantly on the surface of the bacteria.

Tumour necrosis factor (TNF- $\alpha$ ) and inflammatory chemokines produced by the infected macrophages result in recruitment of a successive wave of neutrophils, macrophages, dendritic cells, natural killer cells, CD4<sup>+</sup> and CD8<sup>+</sup> cells each of which in turn make its own cytokines and chemokines that increase the cellular chemotaxis and remodelling of lung tissue in the infected regions (4,18,19). The inflammatory response is replaced and controlled by a particular cellular immune response which is associated with production of INF- $\gamma$ . In fact, at this stage, the recruited cells are employed as building blocks of tubercles. In later stage, the granulomas become more mature, vascularised and an extensive fibrotic cuff of extra cellular matrix and collagens marks out a boundary between macrophages, granulocytes, giant cells and foamy cells, and lymphocytes. In the late stage, the granulomas lose their vascular appearance and become necrotic (4) (Figure 1.5).

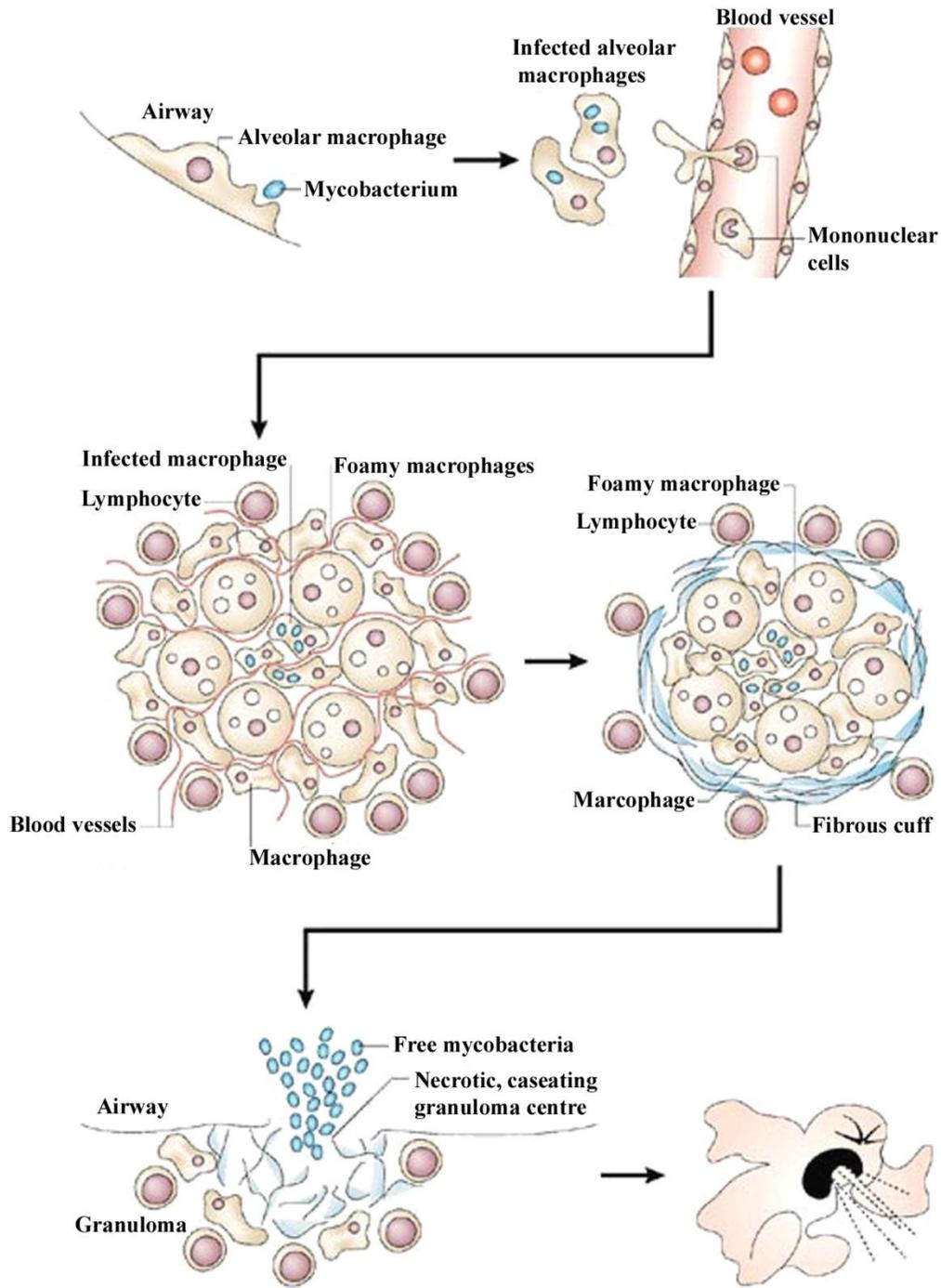


Figure 1.5: process of the development of tuberculosis infection. Following inhalation by the host, *M. tuberculosis* is taken up by alveolar macrophages and dendritic cells. Migration of the infected cells to the lymph nodes stimulates naïve T- cell populations, including CD4<sup>+</sup> and CD8<sup>+</sup> T-cells. Antigenic T-cells induce granuloma formation, in which the *M. tuberculosis* infection contained. The granuloma can contain the tubercle bacilli for many years; however infection in the immunocompromised host can lead to caseation of the granulomas. This finally results in active transmission of the bacilli to new hosts (4).

Although it is generally believed that the centre of granuloma has an important role in host - pathogen interaction, follicle-like structures outside of the tuberculomas are primary sites for the host-pathogen cross talk. Cytokine production and lymphocyte proliferation mainly occur in these structures which are composed of B-cells, antigen presenting cells (APCs), CD4<sup>+</sup> and CD8<sup>+</sup> T cells (26). In addition, recent histological studies of human tuberculomas supported the role of the follicle-like structures and have revealed that considerable number of the bacteria or bacterial products are associated with the macrophages found both bordering the central necrotic regions and outside the fibrotic margin where lymphocytes proliferate. Interestingly, the bacteria in peripheral leukocytic infiltrate were positive in isocitrate lyase expression (20, 21), which is known to be up-regulated in *M. tuberculosis* exposed to activating cytokines (22), suggesting dynamic interactions between macrophages and T cells.

Under anaerobic conditions, reduction of 2-nitroimidazole takes place which enables the compound to form covalently bound with thiol groups of intracellular proteins. In absence of hypoxia the 2-nitroimidazole adducts is not formed and the compound is rapidly cleared from the tissue. Histological analysis using the mentioned hypoxia-sensitive compound has shown at least the boundary to the central region of tubercles is hypoxic (23).

Although the development of granulomas results from a strong cellular immune response, it is believed that the bacilli can influence this response locally to guarantee persistence. Previous studies have shown that the cell wall of the bacterium trapped in *M. tuberculosis*-containing vacuole is degraded and its components, including lipoarabinomannan and

arabinomannan, are accumulated in multilamellar vesicles. These vesicles are combined in multivesicular lysosome known as MHC-class II enriched compartments. The vesicles are then exocytosed from the infected macrophages and released as exosomes; the exosomes containing the bacterial proteins and lipids are then internalised by bystander immune cells. The antigenic lipids and proteins are finally presented by MHC-II and CD1 molecules on the surface of antigen presenting cells to stimulate T cells. This kind of cross-priming could therefore result in proliferative response observed at periphery of tuberculomas. Recently, the lipid-bead granuloma model has been used to histologically analyse bioactivity of mycobacterial cell wall lipids with respect to human tuberculoma formation, immune cell recruitment, and cytokine induction. In this model, individual BCG-derived cell wall lipid as lipid coated particles were incorporated into extra-cellular matrix gel and then were inoculated to mice peritoneum. These analyses have shown that trehalose mycolates, particularly trehalose mono- and dimycolates (TDM), are the most bioactive components in mycobacterial cell wall fractions, resulting in an intense recruitment of immune cells and granuloma development. (24). Subsequent mutational analyses have shown that TDM is modified by the addition of a cyclopropane and any defect in production or transfer of the cyclopropane to TMD results in a significantly increased inflammatory response. This implies that the lipid has evolved structurally to a measured degree of pathology (4).

Clinical symptoms of TB in adults, post-primary TB, mostly result from reactivation of a chronic infection which is in turn mostly because of a decrease in function or number of CD4<sup>+</sup> T cells. This finding also explains the reduced granuloma-forming ability and, significantly diminished capacity to prevent progress of infection in AIDS patients. Different factors such as senescence, malnutrition, alcohol consumption and HIV/AIDS

make compromise the immune system, providing proper conditions for reactivation of the pre-existing infection. Following such a change in immune surveillance, caseation occurs and the bacilli are released into airways, leading to active transmission (4, 5).

### **1.5. Tuberculosis prevention**

Current control of human tuberculosis relies upon prevention through use of the effective vaccine, based on a live attenuated strain of *M. bovis*, known as *Mycobacterium bovis* Bacillus Calmette-Guerin (BCG). It is estimated that today more than 3 billion people have received BCG, making this vaccine the most widely used in the world (3). It is thought that BCG has undergone some mutations and lost its virulence upon subcultures accomplished for a period of 13 years. Comparison of the genomes of wild type *M. tuberculosis* and attenuated BCG vaccine strains have shown a number of deletions in the genomes of daughter strains. Only one of these regions, called RD1 carrying nine genes (Rv3871-Rv3879c), is deleted consistently from all types of BCG strains, but is present in all virulent species. RD1 contains genes encoding proteins such as CFP-10 and ESAT-6 which are clearly implicated in *M. tuberculosis* pathogenesis (40).

Although there is a general agreement on effectiveness of BCG against severe form of childhood tuberculosis, its efficacy is limited (from 0 to 80 %) against adult pulmonary tuberculosis. It is known that BCG vaccine gradually decreases its efficacy over period of 10 -15 years, making adolescents and the adults more susceptible to TB (3).

It has been suggested that variable efficacy of BCG results from two main reasons: (I) lack of protective antigen against virulent *M. tuberculosis* and (II) influence of prior infection with environmental mycobacteria (28, 29). In addition, the vaccine has another two

deficiencies, firstly, BCG as a live vaccine is not ideal for vaccination of subjects who suffer from anti-mycobacterial reactivity (29) and secondly, BCG vaccination of individuals makes skin test hypersensitivity unreliable (30).

As *M. bovis* BCG vaccine has shown variable efficacy both in cattle and in human, improving its effectiveness is urgent. Efforts to increase efficacy of BCG vaccine have been mainly concentrated on two areas: (I) complementation of BCG vaccine by over expression of antigens which are not expressed or expressed at very low level at the wild type BCG (28). Recent studies have shown that recombinant BCG (rBCG30) secreting large amount of *M. tuberculosis* 30kDa major secretory protein is more potent than BCG (31). In addition, other analyses have demonstrated that introduction of RD1 into BCG improves protection against tuberculosis in guinea pigs and mice and also increases the amount of Ag-specific CD8<sup>+</sup> T cells (31, 32). (II) Another way to improve efficacy of BCG vaccine is to employ heterologous prime-boost strategies. These strategies apply a combination of BCG, subunit vaccines and DNA vaccines to boost the strength of BCG-induced immunity. Recently, a recombinant adenoviral vector called rAd35-TBS has been designed to express a unique fusion recombinant protein of *M. tuberculosis* antigens Ag85A, Ag85B and TB 10.4. The vaccine, which is a type of replication-deficient adenovector, can be used either as a stand-alone or priming vaccine and has been designed for the stimulation of strong cellular immune responses in host. This protective immunity is obtained by stimulation of both INF $\gamma$ -producing CD4<sup>+</sup> and CD8<sup>+</sup> T- cells (3).

## 1.6. Treatment of Tuberculosis

DOTS (directly observed treatment, short course) has been employed for more than 10 years and millions of TB patients have been treated by this approach. DOTS includes a six month regimen comprising isoniazid, rifampicin, and pyrazinamide given for two months followed by isoniazid and rifampicin for four months. This regimen has been found the best treatment for both HIV-infected and uninfected patients with fully sensitive organisms. It is necessary to note that this regimen is effective even if the causative agent is resistant to isoniazid. Ethambutol usually is included in the initial regimen reduce possibility of drug resistance (33, 34).

However, it has been suggested that slow-acting anti-tubercular agents along with improper health-care systems have resulted in incomplete treatment, relapse and in particular the emergence of resistant strains. Indeed, these risk factors have resulted in genetic mutations, generally occurring either in the targets or the activators of the drugs, leading to a heritable loss of susceptibility to antitubercular agents.

Tables 1.2 and 1.3 summarize the most common mutations for current first- and second-line antitubercular drugs, and refer to their respective mechanisms of action and half-lives.

Drug resistant strains of *M. tuberculosis* can be generally categorized in three groups : (I) the strains which are resistant to one of the first-line anti-tuberculosis agents (Table 1.2) and have become common.(II) The second group includes multidrug-resistant (MDR) strains that are resistant to multiple anti-tuberculosis antibiotics, particularly the first-line drugs isoniazid (INH) and rifampicin, and finally (III) the extensively drug resistant (XDR)

bacilli which show greater level of drug resistance. These mycobacteria are actually MDR strains which are also resistant to fluoroquinolones and at least one injectable antibiotic such as streptomycin (35).

**Table 1.2: Mechanism of action, resistance and half-life of current first-line anti-tuberculosis agents**

<b>Antibiotic</b>	<b>Mechanism and target</b>	<b>Mutations</b>	<b>Half-life in humans</b>
Isoniazid	Inhibits mycolic acid synthesis; main target is InhA and secondary targets are Kasa, and DfrA	<i>KatG</i> ( required for drug activation; <i>inhA</i> (promoter mutation)	1-3 hours
Rifampicin	Inhibits transcription; RNA polymerase $\beta$ -subunit	<i>rpoB</i>	2-3 hours
Ethambutol	Inhibits arabinogalactan synthesis; possibly EmbB	<i>embB</i>	3-4 hours
Pyrazinamide	Unknown( possibly inhibits FAS-I or changes membrane energetics)	<i>pncA</i> (required for drug activation)	10 hours
Streptomycin	Inhibits protein synthesis; 30S subunit	<i>rpsL</i> and <i>rrs</i>	2-3 hours

**Table 1.3: Mechanism of action, resistance and half-life of current second-line anti- tuberculosis agents**

<b>Antibiotic</b>	<b>Mechanism and target</b>	<b>Mutations</b>	<b>Half-life in humans</b>
Fluoroquinolones	Inhibits DNA gyrase	<i>gyrB</i>	Maxifloxacin:12; Gatifloxacin:8 hours
Ethionamide	Inhibits mycolic acid synthesis; InhA	<i>ethA</i> (required for drug activation) and <i>inhA</i> (promoter mutations)	2 hours
Cycloserine	Inhibits peptidoglycan synthesis by blocking and use of D-Ala; Ala-racemase and D-Ala-D-Ala ligase	<i>alr</i> and <i>ddl</i> (overproduction)	10 hours
Capreomycin	Inhibit protein synthesis; methylated nucleotides in both ribosomal subunits	<i>tlyA</i> and <i>rrs</i>	4-6 hours
<i>P</i> -aminosalicylic acid	Inhibits folate metabolism; possibly dihydropteroate synthase	<i>tlyA</i>	0.75-1 hours
Kanamycin	Inhibits protein synthesis	<i>rrs</i>	2 hours
Amikacin	Inhibits protein synthesis	<i>rrs</i>	3 hours

### **1.7. Immunity to tuberculosis**

Despite extensive studies aiming to discover how host immune response shapes upon *M. tuberculosis* infection, it is still unclear exactly what constitutes a protective immune response to tuberculosis. However it has been suggested that T-cell mediated, rather than

B-cell mediated, immune response plays the main role in control of both human and animal tuberculosis. In addition, there is consensus on contribution of CD4<sup>+</sup> T cells, as the main T cells, in activation of macrophages to kill the intracellular bacilli (32, 36).

Although *M. tuberculosis* infection is followed by a robust CD8<sup>+</sup> T cell response, whether there is an absolute requirement for CD8<sup>+</sup> T cells in immunity to tuberculosis remains controversial. This question becomes more debatable by knowing that depletion of CD8<sup>+</sup> T cells did not affect the bacterial load in the lungs of infected mice. On the other hand, other analyses demonstrated that the cytotoxic activity of CD8<sup>+</sup> T cells, particularly in the late phase of infection, does play a role in the control of the *M. tuberculosis* infection. This role in control of the infection is further supported by studies showing that mice subjected to *in vivo* depletion of CD8<sup>+</sup> T cells or mice without functional CD8<sup>+</sup> T cells show a decreased control of the infection compared to immunocompetent mice (32, 36).

### **1.8. *Mycobacterium tuberculosis* genome**

The *M. tuberculosis* genome was sequenced completely about 10 years ago (37). The singular GC rich circular chromosome consists of 4.4 Mbp and contains 4006 functional protein genes (1, 37). About 52 % of these proteins have been assigned a precise or predicted function. An unusual feature of the *M. tuberculosis* genome is the occurrence of 225 genes, about 6% of the total genes, which are involved in metabolism of fatty acids (16). Another unusual characteristic of the genome is the presence of 167 genes encoding the two unrelated, acidic protein families, known as the PE and PPE proteins. These two protein families are recognized by their N-terminal proline-glutamic acid and proline-proline-glutamic acid motifs and it has been suggested that at least some of these proteins

are surface exposed and play a role in adhesion, immune modulation and antigenic variation (16, 38)

### **1.9. *M. Tuberculosis* virulence factors**

*M. tuberculosis* is one of the most successful pathogens, yet the basis of virulence and pathogenesis remains poorly understood. However, different methods of proteomics and genetic analysis including comparative genomic approaches have led to the identification of several genes and their products which are involved in *M. tuberculosis* pathogenesis. In general, *M. tuberculosis* virulence factors can be classified into four distinct groups based on the known or predicted functional properties: (I) the first class of virulence factor is associated with transcriptional regulators. Mutation analyses of particular transcriptional regulators including some sigma factors (such as Sigma E which is up-regulated upon extracellular stresses), and response regulators (such as PhoP that senses  $Mg^{2+}$  starvation) have resulted in significant attenuation for in vivo growth of *M. tuberculosis* (16). (II) The second group of virulence factors includes anaerobic respiration and oxidative stress proteins. These enzymes, generally superoxide dismutases such as SodA (the main iron-factored superoxide dismutase) and catalases such as KatG (the only recognized enzyme with catalase activity in *M. tuberculosis* which converts isoniazid (INH) to an active form) function to detoxify reactive oxygen intermediates (ROIs) including superoxides,  $H_2O_2$  and organic peroxides produced by activated host phagocyte cells during the process of infection. It is not surprising that these enzymes are important for *M. tuberculosis* virulence, as the host cells employ ROIs to kill invading pathogens (16).

(III) It has been shown that a number of cell surface components can play an important role in *M. tuberculosis* pathogenesis. As an example, phthiocerol dimycocerosate (PDIM) locus contains 13 genes encoding proteins which are involved in pathogenesis of *M. tuberculosis*. Mutation of genes in this locus, either those that are directly involved in biosynthesis of PDIMs (*fadD26* and *fadD28*) or those responsible for appropriate localization of the compound in envelope (*drrC* and *mmpL7*), have resulted in attenuated growth of the mutants in infected mice (16, 39)

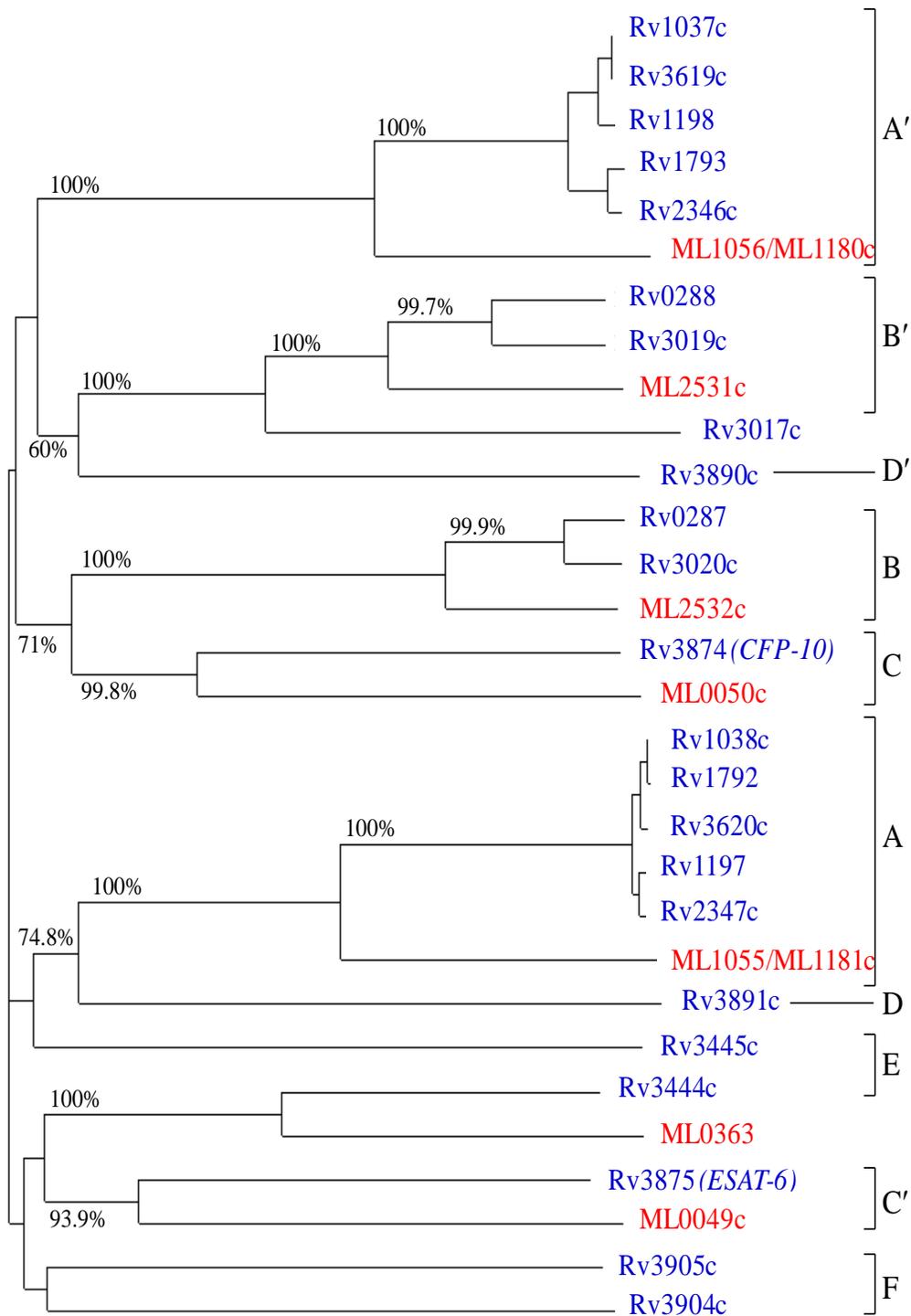
(IV) The last class of virulence factor is related to the *M. tuberculosis* culture filtrate proteins (16). This class of secreted proteins including EspA (encoded by *Rv3616c*), and ESAT-6/CFP-10 related proteins is clearly implicated in *M. tuberculosis* pathogenesis as evidenced by deletion of RD1 from virulent *M. tuberculosis*, leading to considerably reduced virulence (42, 43) and through reintroduction of the region into *M. bovis* BCG strain, resulting in increased virulence of the recombinant strains (43).

#### 1.9.1. ESAT-6/CFP-10 Protein Family

ESAT-6 (*Rv3875* or *EsxA*) and CFP-10 (*Rv3874* or *EsxB*) proteins are members of a large family known as the ESAT-6/CFP-10 (or *Esx*- like) protein family. Members of the family typically contain about 100 residues and are characterized by their arrangement as tandem pairs within the genome and also by their conserved central WXG motif (9, 40, 63, and 66).

Genomic analysis revealed that the *M. tuberculosis* genome contains 22 pairs of genes encoding for CFP-10/ESAT-6 related proteins, which are located at 11 loci (40, 63, and 66). It is known that ESAT-6 and CFP-10 form a tight 1:1 complex (40). Presence of other *esx*-like genes as tandem pairs in the *M. tuberculosis* genome suggests that their protein products could also form complexes (62, 63). This is supported by yeast two hybrid analysis, showing that other genome partners including Rv0287 and Rv0288 form heterodimer complexes (62).

The phylogenetic tree illustrated in figure 1.6 clearly shows that some members of the *Esx*-like protein family are more closely related than others. For instance, members of groups B (Rv0287 and Rv3019c) and B' (Rv0288 and Rv3020c) share a high level of amino acid sequence homology (> 90 %) (40). This raises the possibility of complex formation between non-genome partners which would potentially increase the number of functional complexes, perhaps providing greater functional flexibility. Interestingly, this cross talk has already been observed between the two closely related pairs Rv0287/Rv0288 and Rv3019c/Rv3020c (62).



**Figure 1.6: Bootstrapped phylogenetic tree for the ESAT-6/CFP-10 protein family from *M. tuberculosis* (prefixed by Rv, blue) and their *M. leprae* homologues (prefixed by ML, red). There are 23 family members of ESAT-6/CFP-10 family, 22 of which are arranged as tandem pairs in 11 loci in *M. tuberculosis* genome. Major pairing are shown by square brackets and labeled (A pairs with A' etc.) (40).**

Further genomic studies showed that ESAT-6 and CFP-10 within the RD1 region lie in a cluster of 9 genes (*Rv3871* to *Rv3879c*) (40, 65, and 66). Subsequent comparative studies showed that there are five similar copies of the RD1 region (regions 1-5) within the genome of *M. tuberculosis* (63, 65). These five loci contain a pair of *esx*-like genes, often preceded with a *pe-ppe* pair, and flanked by open reading frames (ORFs) encoding putative ABC transporters, transmembrane and membrane associated proteins (9, 28, 63, 65, and 66). It has been suggested that genes surrounding the *esat-6/cfp-10* related genes are likely to encode essential elements for a secretory apparatus responsible for export of the Esx-like proteins out of the bacterial cytoplasm (9, 63, 78, and 79).

### **1.10. Protein secretion systems in mycobacteria**

A common and conserved strategy for pathogenic bacteria is to use alternative secretion systems to secrete proteins that promote invasion of host cells and resistance to host immune responses (47). As previously explained, mycobacteria have a distinctive cell-envelope including the inner cell membrane, a unique cell wall containing peptidoglycan and arabinogalactan polymer layers surrounded by an extremely hydrophobic mycolate layer and polysaccharide capsule (9). Although this tremendously hydrophobic lipid-rich barrier protects the mycobacterium from the host immune system and external toxic molecules, it causes problem for export of bacterial products. In order to export bacterial products the bacilli take advantage of three types of secretion systems: Sec secretion system, twin arginine transporter and finally ESX alternate secretion pathways (9, 44).

### 1.10.1 General secretion pathway (GSP or Sec secretion system)

Like Gram positive and Gram negative bacteria, *M. tuberculosis* employs this kind of secretion system to secrete unfolded proteins across its plasma membrane. The GSP is composed of SecF, SecD, SecY, SecE, SecG membrane constituents and also SecA ATPase which recognises the conventional N-terminal signal sequence of the secreted substrates. Interestingly, similar to Gram positive bacteria, the mycobacterial sec system does not contain SecB chaperone which delivers the mentioned substrates in unfolded form from ribosome to SecA. How the bacilli transport the mentioned proteins with the signal sequence across the thick cell wall and mycolate layer is unclear; it is, however, believed that the bacteria probably use similar protein complexes to that found in Gram negative bacteria (Omp/YeaT complex) to direct their substrates across the cell wall(44).

### 1.10.2. Twin arginine transporter (Tat)

Mycobacteria also employ a Sec-independent secretion pathway called Tat to transfer folded protein substrates across the plasma membrane. This secretion pathway translocates the proteins with a specific N-terminal signal sequence. This signal sequence is structurally similar to the Sec signal sequence but contains a specific motif containing two arginines followed by two uncharged amino acid close to the N-terminus (44). Recent studies have shown that the mycobacterial Tat system is similar to that of Gram positive and Gram negative bacteria. These analyses have also demonstrated that the system is functional in both *M. tuberculosis* and *M. smegmatis* (45) and appears to be essential for growth of *M. tuberculosis*, prohibiting phenotypic analysis of null alleles in the pathway (46).

### 1.10.3. Snm (ESX) secretion systems

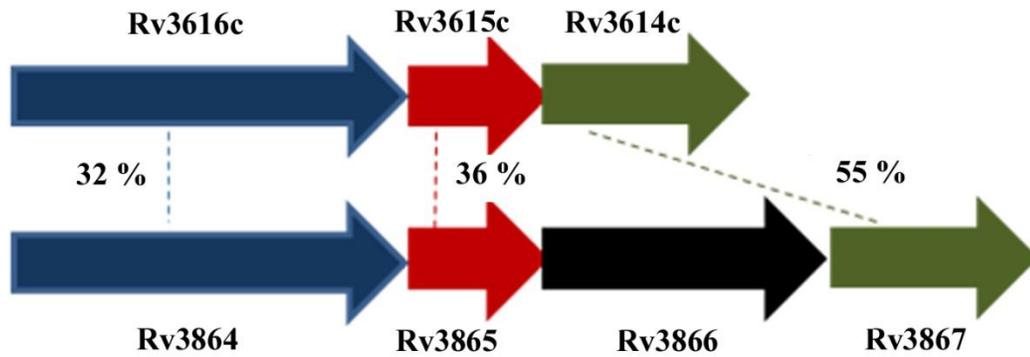
Snm secretion systems are very complicated pathways that involve multiple protein complexes encoded from different loci in the genome of *M. tuberculosis*. These specialised secretion pathways are responsible for secretion of the Esx-like proteins including the ESAT-6-CFP-10 complex, that lack conventional signal sequences (9, 44, and 63).

## **Chapter 2- Molecular features and properties of proteins encoded by the *Rv3616c-Rv3612c* gene cluster**

### **2.1. ESX-1 secretion pathway**

The most studied Snm secretion system is the ESX-1 secretion pathway. Recent genetic analyses have demonstrated that the core machinery of the ESX-1 pathway is encoded at the genomic locus known as region of difference 1 (RD1) (9, 42, 44, 48, and 79). Using *M. bovis* BCG and *M. microti* strains complemented with an extended RD1 region, *Rv3868*, *Rv3870*, *Rv3871*, *Rv3872* and *Rv3877* were found to be required for secretion of the ESAT-6-CFP-10 complex.

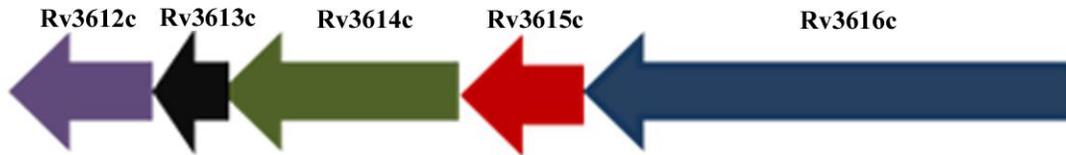
In addition to the RD1 region, another locus, *Rv3616c-Rv3612c* operon, has recently been discovered that has an essential role in the secretion pathway as mutation of three genes in the gene cluster (*Rv3616c*, *Rv3615c* and *Rv3614c*) blocks secretion of the ESAT-6-CFP-10 complex (42, 48, 79). This gene cluster has probably resulted from a gene duplication event since *Rv3616c*, *Rv3615c*, and *Rv3614c* are homologous to *Rv3864*, *Rv3865* and *Rv3867* respectively (48). These homologues are located upstream of the RD1 region (figure 2.1) (48).



**Figure 2.1:** *Rv3616c*, *Rv3615c* and *Rv3614c* have paralogues within the *M. tuberculosis* genome just upstream of the RD1 locus. Percentages indicate percent identity of amino acid sequence (48).

The *Rv3616c*, *Rv3615c* and *Rv3614c* genes are coordinately up-regulated in low pH and low iron conditions (49, 50). In addition, recent analyses conducted by Rickman *et al.* indicated that the *Rv3616c-Rv3613c* gene cluster is coordinately induced in the *M. tuberculosis* *Rv3676* mutants. In fact, deletion of the *Rv3676* gene, which encodes a member of cAMP receptor protein family of transcription regulators in the *M. tuberculosis*, results in a considerable increase in expression of the *Rv3616c-Rv3613c* gene cluster. Interestingly, this gene cluster has a *Rv3676* binding site just upstream of the *Rv3616c* gene (51). Further quantitative PCR (qPCR) analyses have shown that the expression of the *Rv3616c-Rv3612c* gene cluster is significantly induced at two hours post-infection in a *Rv3849*-dependent manner (79). This finding is further supported by microarray studies, showing that the genes contained in the *Rv3616c-Rv3612c* gene cluster were coordinately down regulated in the *Rv3849* mutant (79). The proximity and orientation of these ORFs, together with the level of induction and suppression with respect to the rest of the genome under the conditions described above and also the presence of *Rv3676* and *Rv3849* binding sites upstream of *Rv3616c*, strongly suggests that the *Rv3616c-Rv3612c* gene cluster forms

an operon transcribed from a promoter located upstream of the *Rv3616c* gene (48, 49, 50, and 51) (Figure 2.2).



**Figure 2.2:** The proximity and orientation of the *Rv3616c-Rv12c* gene cluster retrieved from Tuberculist website (<http://genolist.pasteur.fr/Tuberculist>).

The *Rv3616c*, *Rv3615c* and *Rv3614c* genes are highly conserved in *M. tuberculosis*, *M. bovis* and *M. leprae*, but absent from *M. smegmatis* and *M. avium* genomes (48). This finding, along with the fact that the genes are non-essential for the in vitro growth of *M. tuberculosis* H37Rv, strongly suggests that the gene cluster most probably play a role in the virulence of pathogenic mycobacteria (52). Recent experimental studies show convincingly that the *Rv3616c-Rv3612c* gene cluster plays an important role in the pathogenesis of *M. tuberculosis* because *M. tuberculosis* *Rv3616c*, *Rv3614c* or *Rv3614c* mutants are considerably attenuated for growth in spleen and lung of infected mice. In addition, these studies demonstrate that the mutants show significant growth defects in macrophages, and are unable to suppress macrophage proinflammatory responses (42, 48).

### 2.1.1. Regulation of ESX-1 secretion pathway

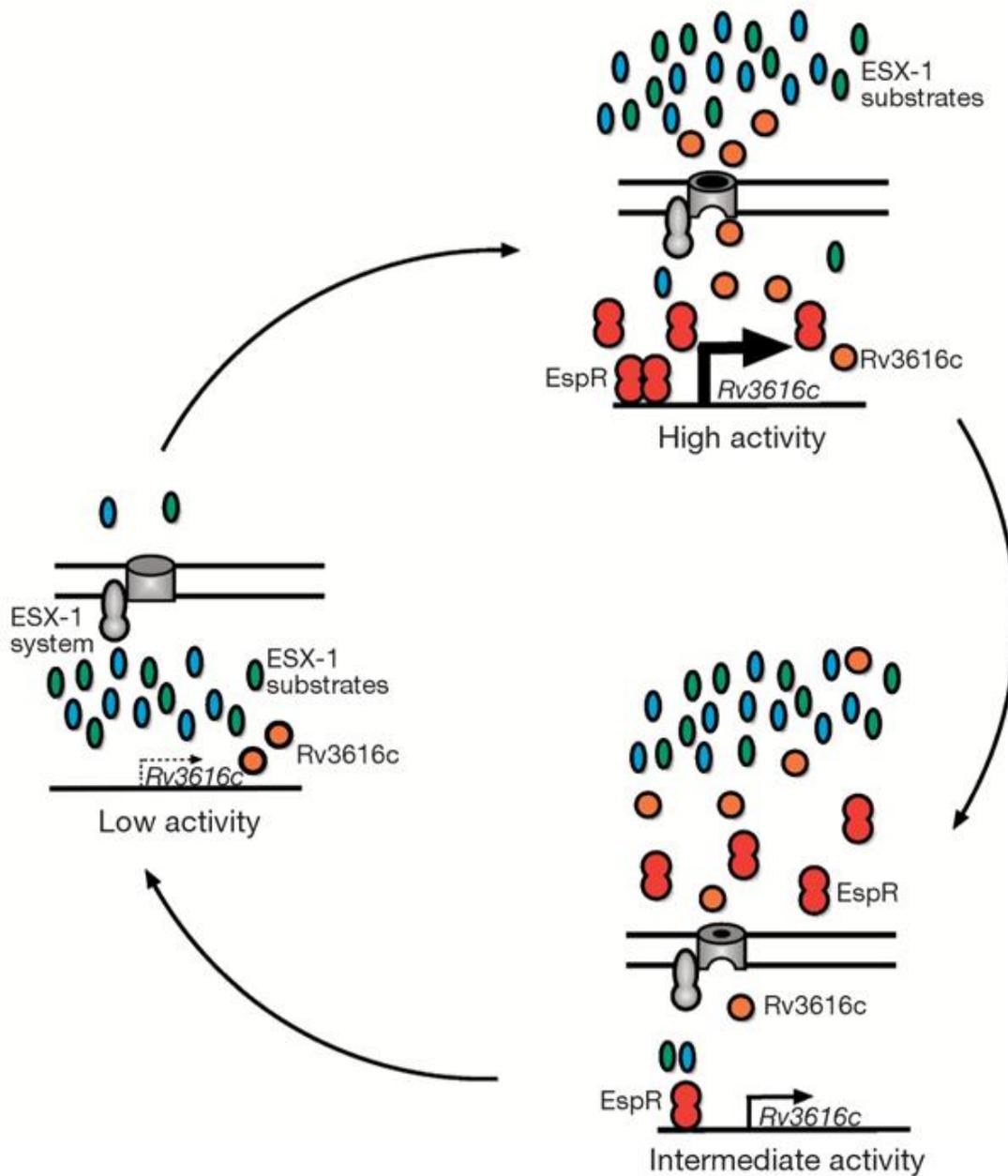
During tuberculosis infection, control of the ESX-1 secretion pathway is probably obtained via integration of multiple regulators, including *Rv3676*, and *EspR* (51, 78, and 79). Combined mutation and microarray analyses conducted by Rickman *et al.* have recently shown that deletion of *Rv3676*, a cAMP-dependent transcription regulator in *M.*

*tuberculosis*, significantly up regulates transcription of the *Rv3616-Rv3613c* gene cluster (51). Presence of Rv3676 binding site consensus at just upstream of the *Rv3616c-Rv3612c* operon (51), together with the fact that expression of proteins encoded by *Rv3616c-Rv3614c* genes is essential for secretion of ESX-1 substrates (42, 48, 78, and 79), strongly suggests a role for Rv3676 as a transcription repressor in negative regulation of ESX-1 secretion pathway. The transcription repressor probably represses expression of the gene cluster, and subsequently reduces activity of the ESX-1 secretion system.

It has been shown that Rv3676 is also a direct regulator of *rpfA*, inducing transcription of *rpfA* (51). This, along with the fact that *rpfA* codes for a resuscitation promoting factor, may imply that the reduced function of ESX-1 secretion system probably is required for reactivation of dormant *M. tuberculosis* infections.

Furthermore, recently a novel secreted substrate of the ESX-1 pathway has been discovered. It has been suggested that this secreted protein known as EspR (or Rv3849) is also directly involved in the regulation of ESX-1 secretion system (Figure 2.3). This protein, like Rv3676, is a DNA-binding transcription factor and has a DNA-binding site located upstream of the *Rv3616c-Rv3612c* operon. On the contrary, EspR induces transcription of three components of ESX-1, whose expression in turn promotes secretion of ESX-1 substrates. EspR directly binds to the promoter, activates transcription of the *Rv3616c-Rv3614c* genes and, unexpectedly, is itself secreted from the bacilli by the ESX-1 secretion pathway. Efflux of EspR leads to reduced *Rv3616c-Rv3614c* transcription, and therefore reduced ESX-1 secretion (78, 79).

It is believed that ESX-1 modulates early events during infection. This, along with the fact that the virulence factors secreted by the ESX-1 system are highly immunogenic, suggests that *M. tuberculosis* may activate the ESX-1 pathway through expression of EspR to establish itself in host cells and then probably reduce ESX-1 activity as the infection progresses via secretion of the inducer to survive future attack of host immune system. This suggestion is in agreement with results of studies showing that the number of ESAT-6 reactive T cells diminished during chronic *M. tuberculosis* infection (79).



**Figure 2.3:** Schematic presentation of EspR-dependent regulation of the ESX-1 secretion pathway. EspR transcriptional activity switches ESX-1 function between low-activity and high-activity states (left and top right, respectively), but its regulation is limited by EspR secretion. Activation of *Rv3849* (EspR) by unknown stimuli leads to increased (thick black arrow) transcription of *Rv3616c* (the corresponding protein is shown by the orange circles) and other genes in the operon (not shown). The up-regulation of *Rv3616c* results in high activity of ESX-1, probably through interactions with membrane-bound ESX-1 secretion components (grey), leading to secretion of substrates (green and blue circles) including *Rv3616c*. Secretion of EspR reduces the intracellular concentration of the activator and then decreases *Rv3616c* transcription, shown here (thin arrow, bottom right) as an intermediate-activity state. Eventually, this negative feedback loop returns the ESX-1 system back to the low-activity state transcription (dashed arrow) (79).

### 2.1.2. Role of ESX-1 secretion system in pathogenesis of *M. tuberculosis*

The roles of the secretion system can be summarized as follows: (I) It is believed that this secretion pathway modulates early events during infection of *M. tuberculosis* as evidenced by several experimental studies, showing that ESX-1 mutation results in severe attenuation of growth for the first few days of infection in cultured macrophages and infected mice which is eventually followed by a delayed growth. In fact, it has been suggested that the ESX-1 secretion system mediate early contacts with the host cells, perhaps altering the host immune response at early stage of the infection process. Initial clues that the ESX-1 machine might manipulates host immune response came from the real time studies of *M. marinum* infection in a Zebrafish model that showed that the ESX-1 mutants were unable to recruit macrophages to form granuloma, suggesting that the ESX-1 secretion system is essential for macrophage proinflammatory response which leads to macrophages aggregation and granuloma formation later during infection (44). (II) Moreover, recent studies conducted by van der Wel *et al.* showed that ESX-1 contribute to escape of the bacilli from the phagolysosome into the cytoplasm where they can replicate in infected dendritic cells. This process is dependent on EspA (Rv3616c), ESAT-6 and CFP-10(27). However, ESX-1 also inhibits phagolysosome biogenesis in macrophages via a process which is independent from the function of the known substrates secreted by the smm pathway (78, 81). These analyses, therefore, suggest that ESX-1 regulates the innate immune response of the infected host via different mechanisms, probably through its secreted substrates (78).

(III) It is known that ESX-1 is involved in secretion of some virulence factors including the ESAT-6-CFP-10 complex, Rv3615c, and Rv3616c. These potent T-cell antigens are clearly implicated in the pathogenesis of tuberculosis (42, 44, 48, and 80).

(IV) It is also thought that the machine has a role in cell to cell communication in both pathogenic and non-pathogenic mycobacteria as evidenced by experimental studies showing that the ESX-1 system is required for phagosome escape and cell to cell spread in *M. marinum*. It has also been shown that *M. smegmatis* ESX-1 mutants have an increased conjugation efficiency compared with wild-type species, further supporting the fact that ESX-1 contributes to cell to cell communication (44, 53, and 119).

### 2.1.3. Molecular mechanism of ESX-1 secretion pathway

Protein-protein interaction studies along with structural analyses have provided insight into the molecular mechanism of the S<sub>nm</sub> secretion system. It has been shown that the secreted proteins CFP-10 and ESAT-6 interact with each other to form a 1:1 tight heterodimer complex within the mycobacterium cytoplasm (1, 9, and 40). Subsequent yeast two-hybrid analyses showed that Rv3870 may interact with Rv3871 (9). It has been suggested that Rv3871 recognizes the CFP-10-ESAT-6 complex, and delivers it in an ATP-dependent fashion to Rv3870 (9). This is evidenced by yeast two-hybrid analysis, showing Rv3871 interacts with CFP-10. Analysis of solution structure of the ESAT-6-CFP-10 complex revealed that the carboxy terminus of CFP-10 is unstructured and is not involved in the complex formation (1). Interestingly, subsequent mutation studies on the C-terminal amino acids of CFP-10 revealed that the last 7 amino acids of C-terminus of CFP-10 are crucial for both the interaction with Rv3871 and the secretion of ESAT-6-CFP-10 complex (120).

As this region of CFP-10 is unstructured and is not required for the interaction with ESAT-6 (1), Rv3871 can easily bind to the unstructured region and thereby target the ESAT-6-CFP-10 complex for secretion. Interestingly, this ESX-1 signal sequence is portable and is sufficient for an unrelated soluble protein to be secreted by ESX-1 (9).

It still remains unclear how the proteins encoded by the *Rv3616c-Rv3612c* gene cluster contribute to the secretion of ESAT-6-CFP-10 complex. However, some physical interactions between the proteins encoded by the cluster and other known components of the snm secretion pathway have been shown (48). Yeast two-hybrid analysis has recently revealed interaction between snm10 (encoded by *Rv3614c*) and snm7 (encoded by *Rv3882c* and predicted to be a membrane protein with a large soluble domain located in the periplasm) (48). This suggests that the Snm secretion pathway is a process which involves protein components encoded at multiple loci in the *M. tuberculosis* genome (48).

Furthermore, the protein encoded by *Rv3616c* (EspA) is a 40 KDa secreted protein without predicted signal sequence. The protein is found in culture filtrate with the other two secreted proteins, ESAT-6 and CFP-10; interestingly, secretion of ESAT-6, CFP-10 and EspA is mutually dependent. Indeed, not only *M. tuberculosis Rv3616c* mutants are unable to secrete ESAT-6-CFP-10 complex, but also the *M. tuberculosis Rv3875* (*esat-6*) or *Rv3874* (*cfp-10*) mutants inhibit secretion of EspA (42). This implies that EspA probably acts as both a constituent and a substrate of the ESX-1 secretion pathway (42).

Yeast two hybrid analyses conducted by Macgurn *et al.* have recently shown some physical interactions between some of the known components of ESX-1 secretion pathway, these

studies, however, have not included the proteins encoded by *Rv3613c* and *Rv3612c* (48). This raises the question of the possibility of complex formation between the proteins mentioned above and other known constituents of the secretion system, providing a further clue for molecular function of proteins involved in the smn secretion pathway. Yeast-two hybrid assay could be expanded to work out these possibilities.

#### *2.1.3.1. Yeast two-hybrid system*

Multiple domain proteins are composed of distinct domains of partially or completely separable function. Importantly, in many cases full function of the intact protein is reproducible by only bringing separate domains into intimate contact and therefore it is not essential that these domains be covalently linked (54). Yeast two-hybrid system has employed this basic characteristic to screen protein-protein interaction and also to evaluate affinity of the interaction in vivo (54). In fact, the yeast two-hybrid system is based on the modular nature of eukaryotic transcription factors which mostly consist of two distinct functional domains, a DNA binding domain (DBD) and a transcription activation domain (TAD). Transcriptional activation function of the transcription factors can be only achieved when the TAD and the DBD come into close proximity (54).

Gal4, Gal80 and Gal3 transcriptionally control expression of six genes (*GAL1*, *GAL2*, *PGM2*, *GAL7*, *GAL10*, and *MEL1*) involved in metabolism of galactose in yeast. Gal4, as transcription factor, activates transcription of the genes mentioned above, whereas Gal80 binds to C-terminal 30 residues of Gal4, inhibiting its transcriptional function. In the presence of galactose, Gal3 binds to Gal80, leading to conformational changes in its structure, and therefore allowing Gal4 to carry out its transcriptional activity (54).

It has been shown that activity of the Gal4 protein can be reconstituted by independent expression of functional domains of the Gal4 protein when fused with proteins that interact in vivo (54, 55). N-terminal region of Gal4 (residues 1-147) provides Gal4 DNA binding ability which recognizes four 17 base pair sites within the Gal4 upstream activating sequence (UAS<sub>G</sub>), while its transactivation activity results from C-terminal part of the protein(residues 148-196 and 768-881) (54,55).

As shown in figure 2.4, chimeric fusions are generally unable to activate transcription of reporter gene unless they are co-expressed and interact in vivo. However, in some cases, some proteins may themselves bind to DNA and/or activate transcription (auto-activation), therefore, DBD and TAD fusion proteins must be checked for auto-activation of reporter genes.

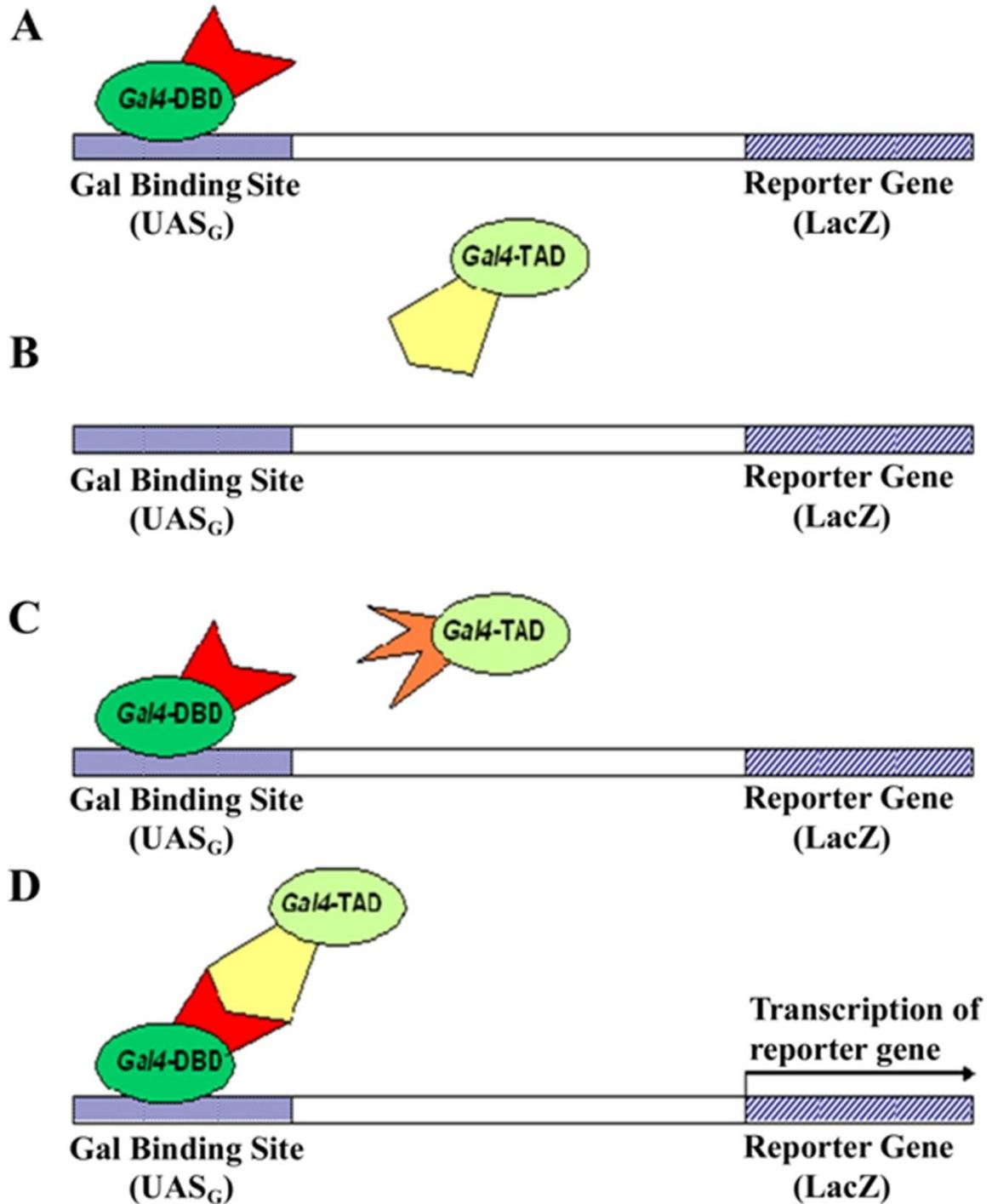


Figure 2.4: The yeast two-hybrid system. (A) A protein of interest (bait) linked to the GAL4 DBD binds to DNA, but fails to initiate transcription of *LacZ* reporter gene. (B) A possible partner (prey) of bait fused to GAL4 TAD cannot activate transcription of reporter gene, because it is unable to bind the GAL 4 binding site. (C and D) show prerequisite of interaction between bait and prey to activate transcription of the reporter gene.

#### 2.1.4. Aims

The aims of the work described in this chapter were to determine molecular features of the proteins encoded by the *Rv3616c-Rv3612c* gene cluster, and investigate the possibility of interaction between the proteins encoded by the gene cluster. Initial yeast two-hybrid studies indicated interactions between some of the known constituents of the Esx-1 secretion pathway (48). These studies, however, have not included the proteins encoded by *Rv3613c* and *Rv3612c* (48). This raises the question of the possibility of complex formation between the proteins mentioned above and other proteins encoded by the gene cluster, providing a further clue for molecular function of proteins involved in the Esx-1 secretion pathway. Yeast-two hybrid assay could be expanded to work out these possibilities.

## **2.2. Materials and methods**

### **2.2.1. DNA Analysis by Agarose Electrophoresis**

Plasmid DNA and PCR products preparations were analyzed by horizontal electrophoresis using 1 % (w/v) agarose gels run at a constant 100 V for 40 minutes. The products were prepared for analysis by mixing 2µl aliquots of DNA samples with 8µl de-ionized H<sub>2</sub>O and 2µl 6x Blue/Orange loading dye (Promega). Ethidium bromide was used to visualize the location of the plasmid DNA under ultra violet light. All gels used to analyze plasmid DNA were run with 100 bp DNA ladder and lambda DNA/*EcoRI* + *HindIII* marker (Promega) to help with identification.

### **2.2.2. SDS-polyacrylamide Gel Electrophoresis**

In order to monitor the progress of protein purification and the composition of protein samples, SDS-PAGE was performed. The protein samples were prepared for analysis by combining 20µl of aliquots of protein samples with 10µl 4x NuPAGE<sup>®</sup> LDS sample buffer (Invitrogen) and 10µl DTT (50 mM). Samples were heated at 70 °C for 10 minutes prior to loading on 4–12 % acryl amide gradient, NuPAGE<sup>®</sup> Bis-Tris Gel (Invitrogen). Electrophoresis was carried out in MES SDS Running Buffer at a constant 200 V for 35 minutes. Typically, gels were run with one lane of protein marker (Sigma MW-SDS-70L). The gels were then stained for 30 minutes with Coomassie Brilliant Blue and destained in two steps, initially by the Fast Destain buffer for 30 minutes and finally by the Destain buffer II.

### 2.2.3. Western Blotting

SDS-PAGE was run for the His<sub>6</sub>-tag recombinant protein, Rv3616c, as described in section 2.2.2. The protein was then transferred to Hybond-P PVDF membrane at a constant 35 V and 110 mA for 1.5 hours. The membrane was washed two times with 25 ml TBS and incubated in 30ml blocking solution at 4 °C for overnight. The blocked membrane was washed two times with 25 ml TBST (25 mM Tris-base, 125 mM NaCl, 0.1 % Tween 20, pH 8) and then washed for 10 minutes with 20 ml TBS (25 mM Tris-base, 125 mM NaCl, and pH8). The blot was incubated for 1hour at room temperature in 10 ml His.Tag<sup>®</sup> Monoclonal Antibody diluted 1:4000 in blocking solution (5 g non-fat dry milk in 100 ml TBST) and washed two times with 25 ml TBST. The blot was then incubated for 1 hour at room temperature with 10 ml Horseradish Peroxidase-conjugated goat anti-mouse IgG diluted 1:10000 in blocking solution. Finally the blot was washed 6 times with 25 ml TBSTT. Chemiluminescent detection using the ECL Plus western blotting detection reagents (Amersham Biosciences) was carried out according to the manufacturer's protocol. The blot was placed in an X-ray film cassette and sheets of autoradiography films placed on the top. Exposure times were 15, 30, 45 and 60 seconds.

### 2.2.4. Amplification of Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions

The Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions were amplified by PCR from the Bacterial Artificial Chromosome (BAC 149), *M. tuberculosis* H37Rv and AF2122/97 *M. bovis* DNA templates (Received from Veterinary Laboratories Agency) using Pfu DNA polymerase (Promega) and the following primers:

<b>Primer name</b>	<b>Primer(5' - 3')with underlined restricted sites</b>	<b>Restriction site</b>
<i>Rv3613c_ forward</i>	<u>GGAATTCCATATGTGCACAATGCCAAAGCTTTGGCGCGCCT</u> TCATG	<i>NdeI</i>
<i>Rv3613c_ Reverse</i>	<u>CCGGAATTCCGGCTAAAGAGCAGCACGGGCAGCGGGAGCA</u> GAACTCGG	<i>EcoRI</i>
<i>Rv3614c_ Forward</i>	<u>GGAATTCCATATGGTGGACTTGCCCCGAAATGACTTTGAC</u>	<i>NdeI</i>
<i>Rv3614c_ Reverse</i>	<u>CCGGAATTCCGGTCACCATGGATCGCTCTCGTCGTCTGGTG</u> C	<i>EcoRI</i>
<i>Rv3615c_ Forward</i>	<u>GGAATTCCATATGACGGAAAACCTTGACCGTCCAGCCCCGAG</u>	<i>NdeI</i>
<i>Rv3615c_ Reverse</i>	<u>CCGGAATTCCGGTCAGGTAAACAACCCGTCGATAGCCTTGC</u> G	<i>EcoRI</i>
<i>Rv3616c_ Forward</i>	<u>GGAATTCCATATGAGCAGAGCGTTCATCATCGATCCAACG</u>	<i>NdeI</i>
<i>Rv3616c_ Reverse</i>	<u>CCGGAATTCCGGTTAGACGACGTTTCGTACCAGCACCTTTT</u> G	<i>EcoRI</i>

PCR was performed in a 50 µl reaction volume (Promega Protocol) using the TECHNE Thermal Cycling System (TECHGENE). The protocol consisted of a 1.5 minute initial denaturation at 95 °C followed by 30 cycles of 0.5 minutes at 95 °C, 1.0 minute at 55°C for

annealing and 4.0 minutes at 72 °C for extension. The final extension cycle was 5.0 minutes at 72 °C. PCR products were identified by electrophoresis using 1% (w/v) agarose gels (Section 2.2.1).

#### 2.2.5. Zero Blunt<sup>®</sup> PCR Cloning

The individual PCR amplicons (*Rv3616c*, *Rv3615c*, *Rv3614c* and *Rv3613c*) were cloned into the pCR<sup>®</sup>-Blunt vector according to the manufacturer's protocol. 10µl of the ZeroBlunt<sup>®</sup> PCR Cloning reaction mixture was used to transform one-shot competent cells according to the manufacturer's protocol. Positive transformants were miniprep using the WIZARD PLUS SV miniprep kit (Promega) and the integrity of the purified constructs were confirmed by sequencing at PNAACL (University of Leicester). The confirmed constructs were then sequentially dual restricted with *NdeI* and *EcoRI* respectively; the digests were then gel-purified from the corresponding gels using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol and finally the concentration of the recovered restricted coding regions was determined spectrophotometrically.

#### 2.2.6. Subcloning of *Rv3616c*, *Rv3615c*, *Rv3614c* and *Rv3613c* coding regions into pET28a (+) and pET23a (+)

pET-28a (+) and pET-23a (+) expression vectors were sequentially dual restricted by *NdeI* and *EcoRI*. The linearized vectors were gel-purified and recovered using the QIAquick Gel Extraction Kit (QIAGEN). The concentration of the recovered restricted vectors was determined spectrophotometrically.

Ligation reactions were set with a >4:1 ratio of the gel-purified digests to the dual restricted expression vectors with the components first mixed in de-ionized H<sub>2</sub>O, and then heated to 50 °C for 5 minutes before being placed in the ice for 5 minutes. T4 DNA ligase (Promega) and ligation reaction buffer (Promega) were added according to the manufacturer's protocol and the ligation reactions were set at 16 °C for overnight. 8µl of the each ligation mixture was used to transform 50 µl of the competent *E. coli* DH5α cells. Positive transformants were screened for successful transformation by growth on the plates of LB agar supplemented with appropriate antibiotic (40µg/ml Kanamycin or 100 µg/ml Ampicillin). The selected transformants were then subjected to miniprep and the isolated plasmids were then analyzed for successful ligation of insert by visualizing restriction digests (*NdeI/EcoRI*) of the constructs on a 1 % (w/v) agarose gel (Section 2.2.1). The integrity of constructs was finally confirmed by sequencing at PNAACL (University of Leicester).

#### 2.2.7. Protein expression trials

Time-course expression trials were carried out using BL21 (DE3) and BL21 (DE3) - codon plus- RIL *E. coli* transformed by the respective constructs (Section 2.1.6). LB starter cultures (50 ml) containing 100µg/ml Ampicillin or 40 µg/ml Kanamycin were inoculated with one colony of each selected transformant and grown overnight at 37 °C , and 200 rpm in a shaker. The cells centrifuged for 10 minutes at 5000 rpm. The pellets were resuspended in 1ml fresh LB (without antibiotic) and subcultured into 50 ml LB broth containing 100µg/ml Ampicillin or 40µg/ml Kanamycin to final OD<sub>600nm</sub> of about 0.1. The cultures were grown at 37 °C until mid-log phase (OD<sub>600nm</sub> 0.6-0.7) and 1 ml of each culture

removed and stored at -20 °C for SDS-PAGE analysis. The cells were induced by addition of IPTG (0.5mM and 1mM for transformants transfected with pET- 23a (+) and pET-28a (+) constructs respectively) and 1 ml of the cells were harvested every hour post-induction for 4 hours. Normalized pre- and post-induction samples were analyzed by SDS-PAGE (Section 2.2.2) (and Western Blotting as required) for protein production.

In order to evaluate the solubility of the expressed proteins, the induced cultures were subjected to centrifugation at 5000 rpm for 15 minutes then re-suspended and lysed by 15 ml lysis buffer. The samples were then incubated at room temperature for 30 minutes. 20µl of the whole cell lysates were removed for SDS-PAGE analysis and the remaining samples were centrifuged at 12500 rpm for 20 minutes using Beckman Coulter Avanti J-130I centrifuge. 20µl of the clear lysates (soluble fractions) were finally removed for SDS-PAGE analysis (Section 2.2.2).

#### 2.2.8. His<sub>6</sub>-tagged Rv3614c expression and purification

BL21(DE3)-Codon plus-RIL cells carrying the pET28a(+)-*Rv3614c* construct were grown in 500ml LB broth media supplemented with 40 µg/ml Kanamycin and then harvested 4 hours post-induction as outlined in section 2.2.7. The cell pellet was resuspended in 25 ml lysis buffer. The resuspended sample was subjected to French press at 4 °C, 1280 Psi and then centrifugation at 13500 rpm, 1°C for 25 minutes to obtain the supernatant. The filtered supernatant was then diluted by lysis buffer to 50 ml final volume and finally the prepared sample was applied on the 5 ml Ni<sup>2+</sup>-NTA (Novagen) column. The column was then washed with a linear gradient of increasing imidazole concentration. The His<sub>6</sub>-tagged

Rv3614c was finally eluted in 200 mM imidazole wash. Purity of the target protein was evaluated by SDS-PAGE (Section 2.2.2). The fractions containing the target protein were then pooled together and dialyzed against 2L dialysis buffer for overnight. The dialyzed sample was finally applied on a Superdex 75, 120 ml prep grade gel filtration column (GE Healthcare). The purity of the target protein was finally assayed by SDS-PAGE (Section 2.2.2).

### 2.2.9. Rv3615c Expression and purification

BL21(DE3)-Codon plus-RIL cells carrying the pET23a(+)-*Rv3615c* construct were grown in 500ml LB broth media supplemented with 100 µg/ml Ampicillin and then harvested 4 hour post-induction as outlined in section 2.2.7. The cell pellet was resuspended in 25 ml lysis buffer. The re-suspended sample was subjected to French Press at 4°C, 1280 Psi and then sonication at the highest power, 20 seconds on and 20 seconds off for 2 minutes. Post sonicated sample were treated with 0.01 mg/ml DNase (Sigma), and then incubated on rocker at room temperature for 20 minutes. The lysate were then spun down at 13500 rpm for 20 minutes at 1°C to obtain the supernatant. The filtered supernatant was diluted to 50 ml using the same lysis buffer. The prepared sample was applied on a 5 ml Q-sepharose (Amersham) column. The column was then washed with a linear gradient of increasing NaCl concentration to elute the target protein. The protein was eluted in 200 mM NaCl wash, its purity was then evaluated by SDS-PAGE (Section 2.2.2).

The fractions containing the target protein were pooled together and dialyzed against 2L dialysis buffer containing 20 mM Piperazine at pH 5.8 for overnight. The dialyzed sample

was applied on the 5 ml Q-sepharose (Amersham) column pre-equilibrated with the buffer mentioned above. The column was then washed with a linear gradient of increasing NaCl concentration to elute the target protein. The target protein was eluted in 100 mM NaCl wash. Purity of the eluted protein was finally examined by SDS-PAGE (Section 2.2.2).

The fractions containing the target protein were pooled together and dialyzed against 2L dialysis buffer for overnight. The dialyzed sample was applied on a Superdex 75, 120 ml prep grade gel filtration column (GE Healthcare) and finally gel-filtration was carried out. The purity of the target protein was assessed by SDS-PAGE (Section 2.2.2).

#### 2.2.10. Protein concentration

The lyophilization was accomplished by flash freezing the protein samples using liquid nitrogen and then by placing the frozen samples in a lyophilizer (ThermoSavant). This allowed the frozen water in the sample to be vacuumed and to sublime directly from the solid phase to gas. The lyophilized samples were finally resolubilized using appropriate amount of distilled water to obtain the more concentrated sample. In order to maintain the final concentration of buffer components, the protein samples were dialyzed against appropriate diluted buffers prior to the lyophilization process.

#### 2.2.11. Far UV circular dichroism spectroscopy

The far UV circular dichroism (CD) spectra were used to estimate the secondary structure of the purified proteins using a Jasco J-715 spectropolarimeter. All CD spectra were

obtained from protein samples dissolved in a 100 mM NaCl and 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 with protein concentration of about 15-25 μM for a path length of 0.1 cm.

Normally, spectra were recorded from 180 to 250 nm, a scan speed of 20nm per minute, response time 1 minute second, with each spectrum representing the average of 10 accumulations. During acquisition, the sample was maintained at a regulated temperature (25 °C) in a 0.1 cm path length cell. Spectra were corrected for the buffer and converted to molar CD per residue before analysis.

#### 2.2.12. Thermal denaturation studies

The effect of increasing temperature on the structural integrity of the purified proteins was studied by following the change in the CD spectra (Section 2.2.11) over a range from 5 to 100 °C, increasing in 5 °C increments. All CD spectra were obtained from protein samples dissolved in a 100 mM NaCl and 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 with protein concentration of about 15-25 μM for a path length of 0.1 cm and allowed to equilibrate at each temperature before data collection.

#### 2.2.13. Fluorescence spectroscopy and chemical denaturation

Intrinsic fluorescence spectra of the purified protein samples were obtained using a Perkin Elmer LS50B luminescence spectrometer. The spectra were recorded at room temperature with excitation at 280 nm and fluorescence monitored from 300 to 450 nm. Final spectra were the average of 10 accumulations collected at a scan rate of 150 nm per minute. Normally, the spectra were obtained from 0.6 to 1.2 μM protein samples dissolved in a 100 mM NaCl and 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and room temperature.

In order to determine the conformational stability of the purified proteins, stock solutions of the target protein was added to a range of guanidine hydrochloride concentrations (0- 6 M) to give a final concentration of 0.6 - 1.2  $\mu$ M. Unfolding was determined by the change in emission wavelength at maximum emission ( $\lambda_{\text{max}}$ ).

Measurements were taken from an average of 10 scan passes at an excitation wavelength of 280 nm. Each sample was incubated at room temperature for 1 hour prior to analysis to allow the unfolding reaction to reach equilibrium.

#### 2.2.14. Nuclear magnetic resonance spectroscopy for Rv3614c

1 dimensional (1D) nuclear magnetic resonance (NMR) experiment was performed on 360  $\mu$ l sample of 100  $\mu$ M His<sub>6</sub>-Rv3614c dissolved in 100 mM NaCl and 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 (10 % D<sub>2</sub>O). NMR data were acquired on Bruker Avance 800 MHz NMR spectrometer at 20°C.

#### 2.2.15. Strains, plasmids and media used for yeast two-hybrid analyses

The yeast two-hybrid analyses were done using *Saccharomyces cerevisiae* strain CG1945. This strain carries both *HIS3* and *lacZ* reporter genes, encoding imidazole glycerol phosphate dehydratase (an enzyme involved in biosynthesis of histidine) and  $\beta$ -galactosidase respectively, and is auxotrophic for both leucine and tryptophan. The activation domain vector (pGAD) carries *LEU2* (and ampicillin resistance gene), and the DNA binding domain vector (pGBD) contains *TRP1* (and ampicillin resistance gene),

allowing a nutritional selection of transformants carrying both vectors. *S. cerevisiae* were grown at 30 °C on either synthetic dropout (SD) media or YPD (yeast extract/peptone/dextrose) media.

#### 2.2.16. Preparation of yeast two-hybrid vectors

Transformants (*E. coli* DH5 $\alpha$ ) containing pGAD-*Rv3613c*, pGBD-*Rv3613c*, pGAD-*Rv3614c*, pGBD-*Rv3614c*, pGAD-*Rv3615c*, pGBD-*Rv3615c*, pGAD-*Rv3616c*, and pGBD-*Rv3616c* expression vectors, were grown overnight in a shaking incubator (37 °C /200 rpm) in LB broth supplemented with 100  $\mu$ g/ml ampicillin. The overnight cultures were subjected to miniprep and finally the purified constructs were used to transform the *S. cerevisiae*.

#### 2.2.17. Transformation of *S. cerevisiae*

LiAc/ss-DNA/PEG (lithium acetate/ single- stranded carrier DNA/polyethylene glycol) method was used to transform the yeasts (56). Fresh yeast colonies were picked to inoculate 30 ml YPD broth. 15 ml of the overnight incubated culture (at 30 °C /200 rpm) was inoculated into 35 ml of fresh YPD and incubated at 30 °C /200 rpm for 5 hours. The cells were spun down for 15 minutes at 3000 rpm and 4 °C then washed in 25 ml sterile dH<sub>2</sub>O. The washing step was repeated and then the supernatant was discarded. The washed cells were re-suspended in 1 ml 100 mM LiAc before centrifugation for 1 minute at 13000 rpm and room temperature. The pellet was re-suspended in 0.4 ml 100 mM LiAc. The re-suspended cells were vortexed for a few seconds and then aliquoted into 50  $\mu$ l samples. Single stranded herring sperm DNA (Sigma) (5 mg/ml) in TE buffer was incubated at 100

°C for 5 minutes then chilled on ice prior to use. In the meantime, the cells were centrifuged for 30 seconds at 13000 rpm and room temperature. The transformation mixture containing 1 µg of each construct was added to the pellets and vortexed till thoroughly mixed. The samples were incubated at 30 °C for 30 minutes prior to heat shock at 42 °C for 30 minutes. The heat shocked cells were then centrifuged for 30 seconds at 13000 rpm and room temperature and then the transformation mixture was removed. Transformants were re-suspended in 200 µl dH<sub>2</sub>O and the re-suspended cells were plated out on the appropriate SD dropout media. The plates finally incubated at 30 °C till colonies became clearly visible (3-10 days).

## 2.2.18. β- galactosidase assays

### 2.2.18.1. Colony filter lift assays

X-gal filter lift assay was used to identify *LacZ* reporter gene activity (57). The transformants were grown on the appropriate synthetic dropout plates till growth became clearly visible. The grown yeast colonies were transferred to a circular Duralose-UV<sup>TM</sup> membranes (nitrocellulose membrane, Stratagene) prior to permeabilize by submersion in liquid nitrogen for about 20 seconds. The frozen membranes were allowed to thaw prior to repeat the action. The prepared membranes were put colony upside up on a circular Whatman filter paper pre-soaked with Z-buffer/X-gal/β-mercaptoethanol solution. The membranes were incubated at room temperature until the blue color was visible on the membranes. The reactions were stopped by transferring the developed membranes onto a Whatman filter paper pre-soaked with 1M Na<sub>2</sub>CO<sub>3</sub> for at least 10 minutes. Finally the membranes transferred to a clean petri dish and allowed to dry.

#### 2.2.18.2. *Quantitative liquid assays (Assay of cell extract)*

Specific  $\beta$ -galactosidase activity was measured by quantitative liquid assays as described by Le Douarin *et al* (58). The transformants were inoculated in 15 ml specified SD media. The overnight incubated cultures were spun down (at 3000 rpm/4 °C) and then the cell pellets were re-suspended in 1.0 ml sterile dH<sub>2</sub>O. The samples were then centrifuged at 13000 and room temperature for 15 seconds, and then the cell pellets was re-suspended in 150  $\mu$ l Z buffer. The same volume of acid washed beads (212-300  $\mu$ m and 425-600  $\mu$ m) (Sigma) were added till the beads reached a level just below the meniscus of the liquid. The samples were incubated on ice for a few minutes and then vortexed vigorously for 30 seconds before being returned to the ice. This step was repeated four times for each sample. The samples were then centrifuged at 13000 rpm and room temperature for 15 minutes. 20  $\mu$ l of the protein extract (or Z buffer for the blank) was transferred to an autoclaved micro-centrifuge tube and adjusted to 500  $\mu$ l with Z buffer. The samples were vortexed and incubated at 30 °C for 5 minutes. The reactions were started by adding 100  $\mu$ l 4 mg/ml ONPG dissolved in Z buffer and were timed until the pale yellow color appeared, and then the reactions were stopped by adding 250  $\mu$ l of 1M Na<sub>2</sub>CO<sub>3</sub>. Positive and negative control reactions were prepared exactly how the test samples prepared, but the negative reactions were stopped after 30 minutes. The OD<sub>420nm</sub> of the samples was measured and then the protein concentration of in extracts was calculated using Bradford reagent (Bio-Rad). The specific activity of the extracts was calculated as follow:

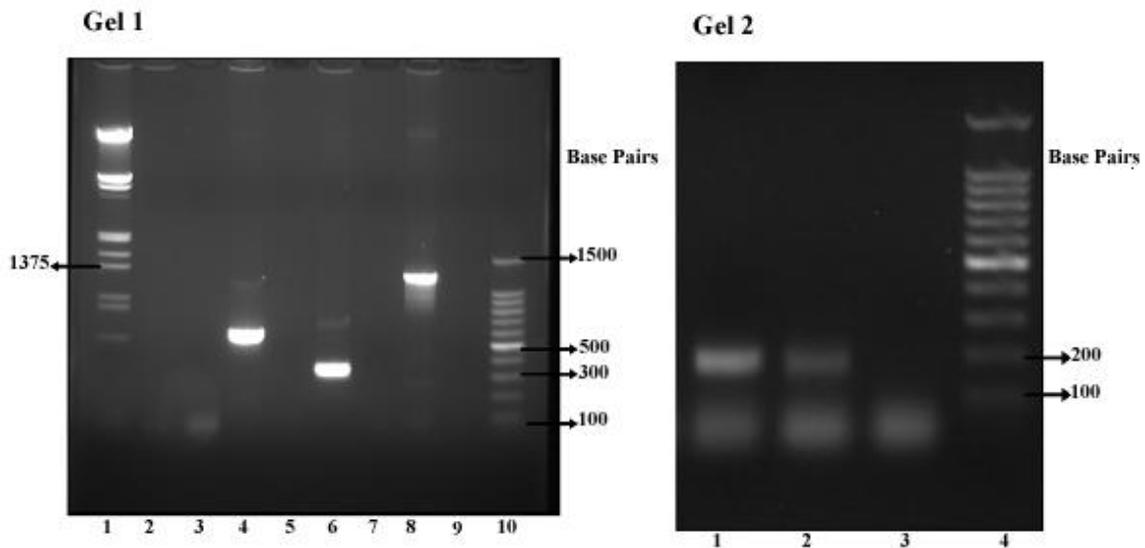
$$\text{Specific activity (nmol/mg/m in)} = \frac{\text{OD}_{420\text{nm}} \times 0.85}{0.0045 \times \text{Protein} \times \text{Volume} \times \text{Time}}$$

Where Protein is the protein concentration of the yeast extract in mg/ml, Volume is the extract volume assayed in ml, Time is the time of reaction in minutes,  $\text{OD}_{420\text{nm}}$  is the optical density of the product o- nitrophenol at 420 nm (58).

## 2.3. Results and Discussion

### 2.3.1. Cloning of Rv3616c, Rv3615c, Rv3614c and Rv3613c coding regions

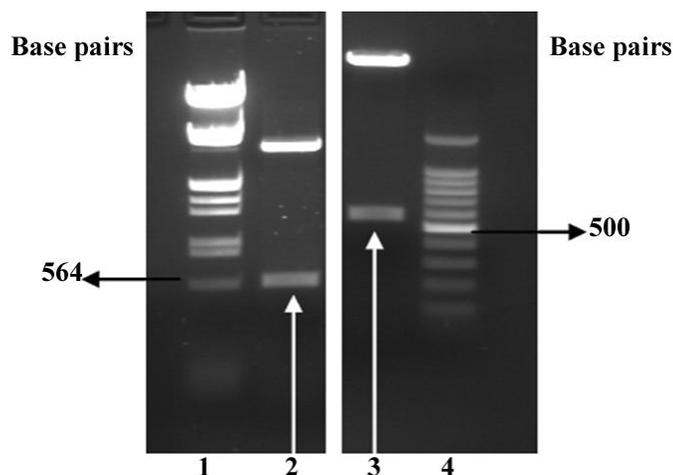
Successful PCR amplification of *Rv3616c* (1179 bp), *Rv3615c* (312 bp), *Rv3614c* (555 bp) and *Rv3613c* (162 bp) coding regions was confirmed by agarose gel analysis of PCR samples as described in the section 2.2.1. Figure 2.3.1 clearly shows amplification of DNA of expected size (Section 2.2.4).



**Figure 2.3.1:** PCR amplification of the Rv3616c-Rv3613c coding regions. Gel 1 and 2 are 1 % w/v agarose gels viewed under UV light with ethidium bromide used to visualize the products of the PCR reactions. Gel 1: Lanes 1 and 10 contain a  $\lambda$  DNA/ *EcoRI*+*HindIII* marker and 100 bp DNA ladder respectively. Lanes 3, 5, 7 and 9 are control samples of PCR reactions without template DNA. Lanes 4, 6 and 8 correspond to PCR products of Rv3614c, Rv3615c and Rv3616c coding regions respectively. Gel 2: Lanes 1 and 2 are PCR products of *Rv3613c*. Lane 4 contain a 100 bp DNA ladder and lane 3 corresponds to a control of sample of PCR reactions without template DNA.

Ligation of the PCR amplicons with the pCR<sup>®</sup>-Blunt intermediate vectors was confirmed by sequencing and also agarose gel analysis of sequentially dual (*NdeI*/ *EcoRI*) restricted plasmid samples. The digests were then gel-purified for the next step (Section 2.2.5).

The gel-purified digests were then subcloned into pET-23a (+) and pET-28a (+), and the constructs were finally confirmed by agarose gel analysis of double ((*NdeI*/ *EcoRI*) restricted plasmid samples as described in figure 2.3.2 (Section 2.2.6).



**Figure 2.3.2: Double-digestion of the ligated constructs. A 1 % agarose gel illustrates successfully ligated constructs. Lane 2 and 3 correspond to the successfully ligated pET-23a (+) and pET-28a (+) constructs respectively containing the insert, *Rv3614c* (the restricted *Rv3614c* is highlighted by arrows). Lanes 1 and 4 are 100 bp DNA ladder and  $\lambda$  DNA/ *EcoRI*+*HindIII* marker.**

### 2.3.2. Protein expression trails

As shown in the following figures, maximum expression of the recombinant proteins (*Rv3615c*, His<sub>6</sub>-tagged *Rv3615c*, *Rv3614c* and His<sub>6</sub>-tagged *Rv3614c*) occurred approximately four hours post induction (Lane 6 in figures 2.3.3, 2.3.4, 2.3.5, and 2.3.6). In all cases the protein products can be detected in both whole cell post induction fractions (Lane 7 in the gel 2, figure 2.3.3 and lane 7 in figure 2.3.6) and in the lysate supernatants (Lane 8 in the gel 2, figure 2.3.3, and lane 8 in figure 2.3.6), providing evidence that the expressed proteins are produced as the soluble products in *E. coli*-based expression system (Section 2.2.7). In addition, as illustrated in the figure 2.3.3, comparative expression analyses showed that the target protein were produced almost at the same level in both the

standard BL21 (DE3) *E. coli* strains(Lane 6, gel 1) and the BL21 (DE3) - Codon Plus- RIL (Lane 6, gel 2).

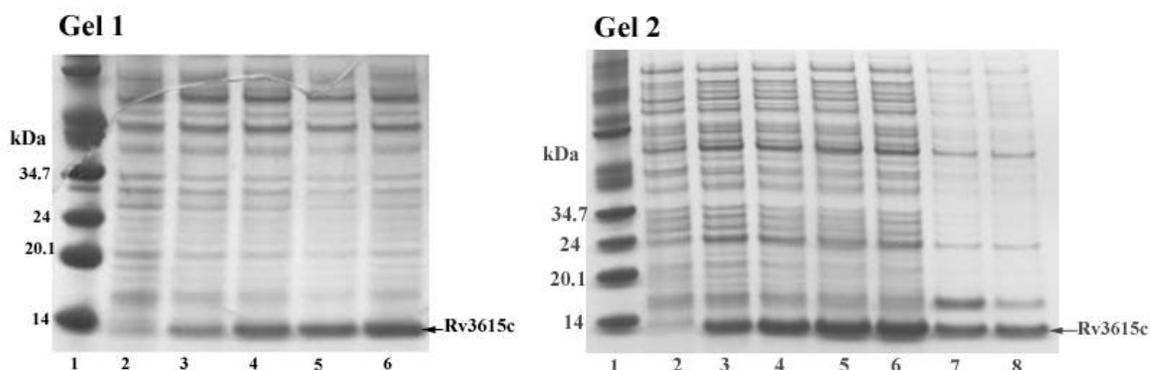


Figure 2.3.3: SDS-PAGE gels of the Rv3615c expression trails. Gel 1 shows expression of Rv3615c in BL21 (DE3) *E. coli* cells. Lanes 1 and 2 contain the high molecular weight marker and the pre-induction cell lysate fraction respectively. Lanes 3 – 6 contain post-induction cell lysate fractions taken at 1.0, 2.0, 3.0, and 4.0 hours respectively. Gel 2 shows expression of the target protein, Rv3615c, in BL21 (DE3) - Codon Plus- RIL cells. Lanes 1 and 2 contain the high molecular weight marker and the pre-induction cell lysate fraction respectively. Lanes 3 – 6 contain post-induction cell lysate fractions taken at 1.0, 2.0, 3.0, and 4.0 hours respectively. Lanes 7 and 8 contain 4 hour post-induction whole cell lysate and lysate supernatant fractions correspondingly. As shown in the figure, the apparent molecular weight of the target protein corresponds closely to that expected for full length Rv3615c (107957 Da)

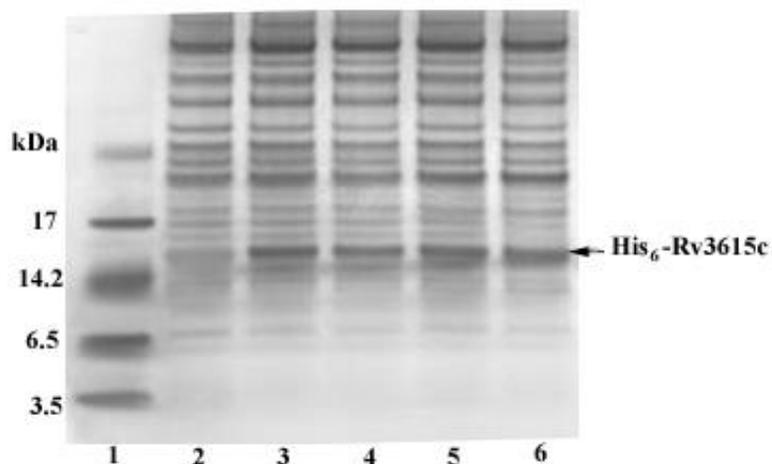
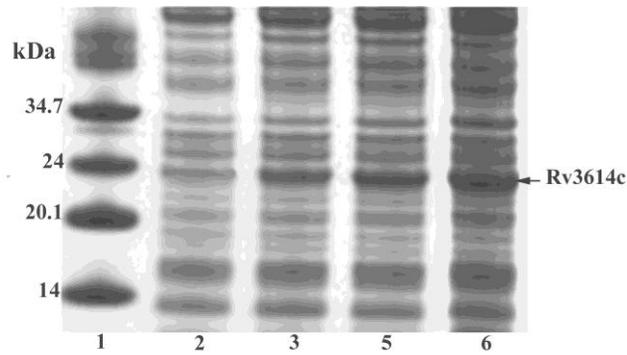
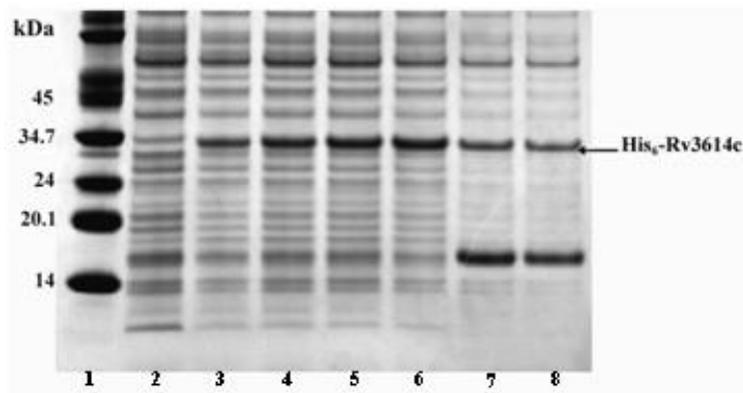


Figure 2.3.4: SDS-PAGE of the His<sub>6</sub>-tagged Rv3615c expression trails. The gel shows time course expression of the His<sub>6</sub>-tagged Rv3615c in BL21 (DE3) - Codon Plus- RIL *E. coli* cells. Lanes 1 and 2 contain a low molecular weight marker (Sigma) and the pre-induction cell lysate fraction respectively. Lanes 3 – 6 contain post-induction cell lysate fractions taken at 1.0, 2.0, 3.0, and 4.0 hours respectively.



**Figure 2.3.5: SDS-PAGE gel of Rv3614c expression.** The gel shows a time course expression of the Rv3614c in BL21 (DE3) - Codon Plus-RIL *E. coli* cells. Lanes 1 and 2 contain the high molecular weight marker and the pre-induction cell lysate fraction respectively. Lanes 3 – 5 contain post-induction cell lysate fractions taken at 1.0, 2.0, and 4.0 hours respectively. As shown, the apparent molecular mass of the full length Rv3615c protein was approximately 24 kDa larger than expected (19.8025 kDa).



**Figure 2.3.6: SDS-PAGE gels of the His<sub>6</sub>-tagged Rv3614c expression.** The gel demonstrates a time course expression of the His<sub>6</sub>-tagged Rv3614c in BL21 (DE3) -Codon Plus- RIL *E. coli* cells and also provides evidence that the protein is produced as soluble product in *E. coli*-based expression system. Lanes 1 and 2 contain the high molecular weight marker the pre-induction cell lysate fraction respectively. Lanes 3 – 6 contain post induction cell lysate fractions taken at 1.0, 2.0, 3.0, and 4.0 hours respectively. Lanes 7 and 8 contain 4 hour post-induction whole cell lysate and lysate supernatant fractions correspondingly.

The results presented here convincingly show that *E. coli*- based expression systems can be employed to produce both Rv3615c and Rv3614c. BL21 (DE 3) - Codon Plus- RIL *E. coli* cells were used as the main host to produce a significant amount of the aforementioned proteins.

In order to express Rv3616c with and without His<sub>6</sub>-tag, the BL21 (DE3)-Codon plus-RIL *E. coli* cells were transformed with the pET-28a (+)-*Rv3616c* and pET-23a (+)-*Rv3616c* constructs respectively. However, only transformants carrying the pET-28a (+)-*Rv3616c* construct could grow on the LB agar plates.

Time course expression trials were, therefore, performed using the only available transformant mentioned above. However, induction of the transformants with the different concentrations of IPTG (0.5 mM, 1 mM and 2 mM IPTG at 37 °C and 200 rpm) dramatically reduced cell density in the cultures (OD<sub>600nm</sub> was decreased considerably from 0.6 to about 0.1 when IPTG was added at the mid-log phase of bacterial growth). Moreover, protein expression trails under different conditions led to the same results. This is evidenced by a considerable decrease in the cell density of the transformants upon adding IPTG to the culture even when it was tried to induce the target protein by low concentration of IPTG (0.25mM) at low temperature (15°C). In addition, both SDS-PAGE and Western blot analyses of cell lysates obtained from the remaining cell cultures (Sections 2.2.2 and 2.2.3) showed no expression for the target protein, His<sub>6</sub>-Rv3616c. These findings, along with the fact that the host cells (BL21 (DE3) - Codon Plus- RIL and BL21 (DE3) strains) were unable to be transfected with pET-23a (+) - based constructs (discussed below), suggest that Rv3616c is likely to be toxic to the host cells.

The leaky expression of the toxic protein in the transformants carrying pET-23a (+)-Rv3616c construct probably leads to death of the cells on agar plates. In fact, pET-23a (+) and pET-28a (+) vectors provide two different stringency options at the level of the T7 promoter itself: the "plain" T7 promoter (pET-23a (+)) and the T7/*lac* promoter (pET-28a

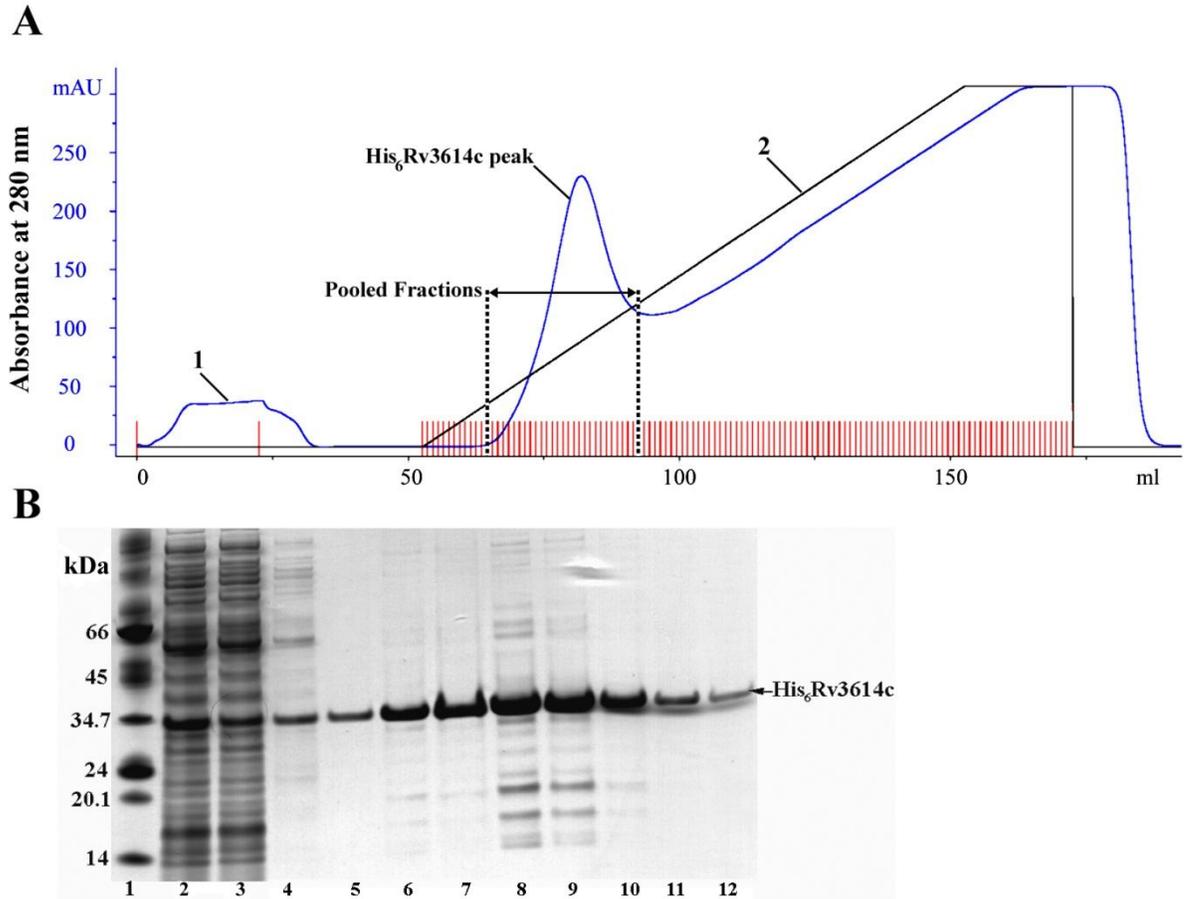
(+). The T7/*lac* promoter contains a 25 bp *lac* operator sequence immediately downstream from the 17 bp promoter region. Binding of the *lac* repressor at this site effectively reduces transcription by T7 RNA polymerase, thus providing a *lacI*-based mechanism to suppress basal expression in the BL21 (DE3) strains. The pET-28a (+) also carries its own copy of *lacI* to ensure that enough repressor is made to titrate all available operator sites.

In order to confirm that the toxicity of Rv3616c to the host cells is not strain-dependent, different strains of *E. coli* including Rosetta (DE3) - *plysS* cells (60) were transformed with the constructs. These analyses also confirmed the toxicity of the Rv3616c to all strains as evidenced by a remarkable decrease in cell density upon induction with IPTG.

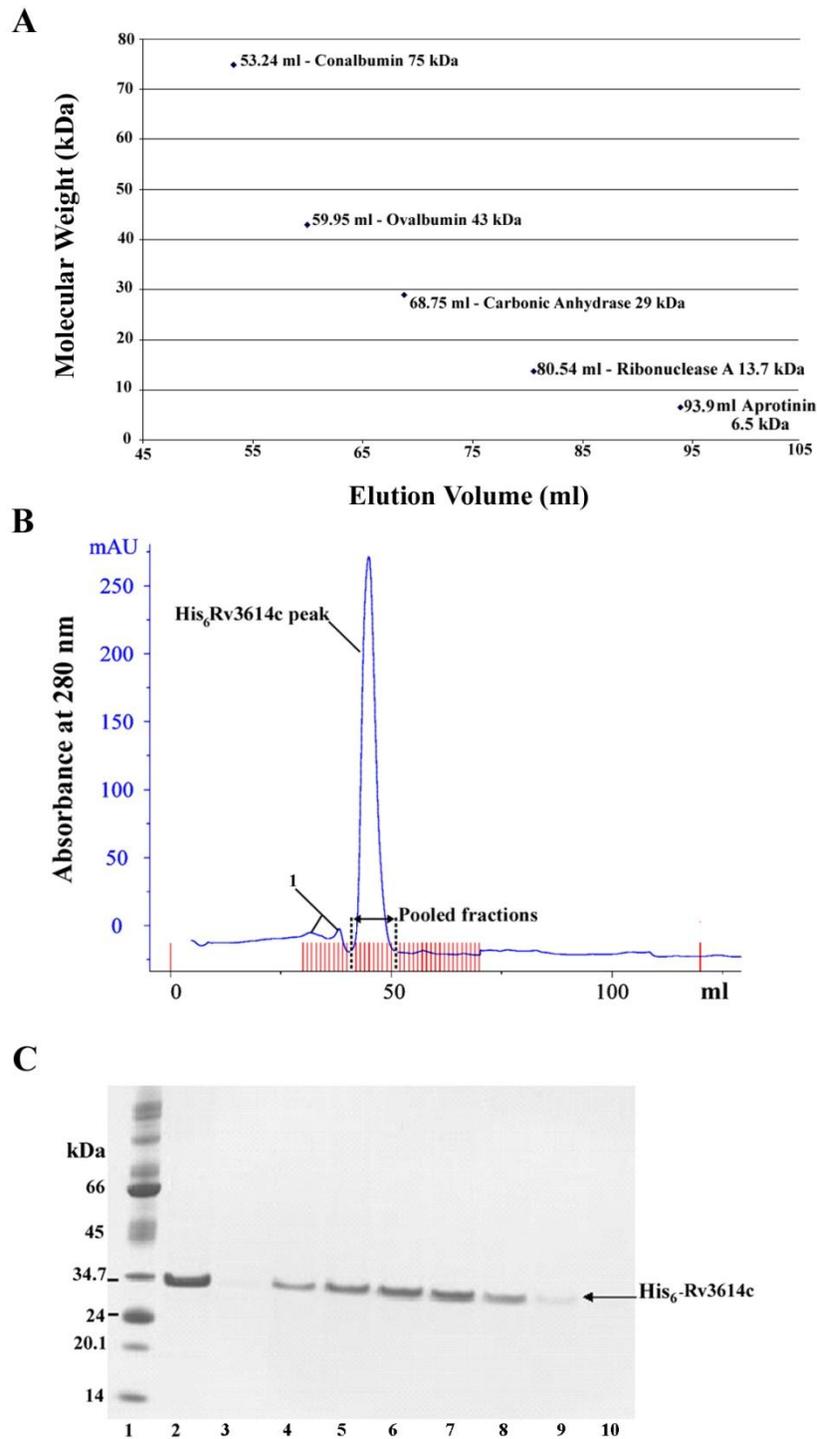
In addition, time course expression trails for Rv3613c were not successful at different conditions (different concentrations of IPTG and temperatures).

### 2.3.3. His<sub>6</sub>-Tagged Rv3614c expression and purification

A pET28a expression vector, containing the coding region of Rv3614c was used to express the protein in *E. coli*. The proteins were purified in two steps. The N-terminal hexahistidine tag allowed the proteins to be purified by affinity chromatography, as shown in figure 2.3.7. The pooled His<sub>6</sub>-tagged Rv3614c protein fractions were subjected to a final polishing purification step by gel filtration, as shown in figure 2.3.8. The final yield of Rv3615c was determined to be approximately 0.34 mg/ ml using a theoretical extinction coefficient of 30480 M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm (Section 2.2.8).



**Figure 2.3.7: Purification of His<sub>6</sub>-tagged Rv3614c.** Panel A shows a typical FPLC profile attained for the Ni-NTA column based purification of the His<sub>6</sub>-tagged Rv3614c. The Rv3614c was eluted over a linear gradient from 0 mM to 500 mM imidazole (indicated by “2”). The flow-through is indicated by “1” and the Rv3614c elution peak is labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2 and 3 show samples removed from the loaded cell lysate and column flow-through, and lanes 4-12 show samples removed from the Rv3614c elution peak. Fractions from lanes 4-12 were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).

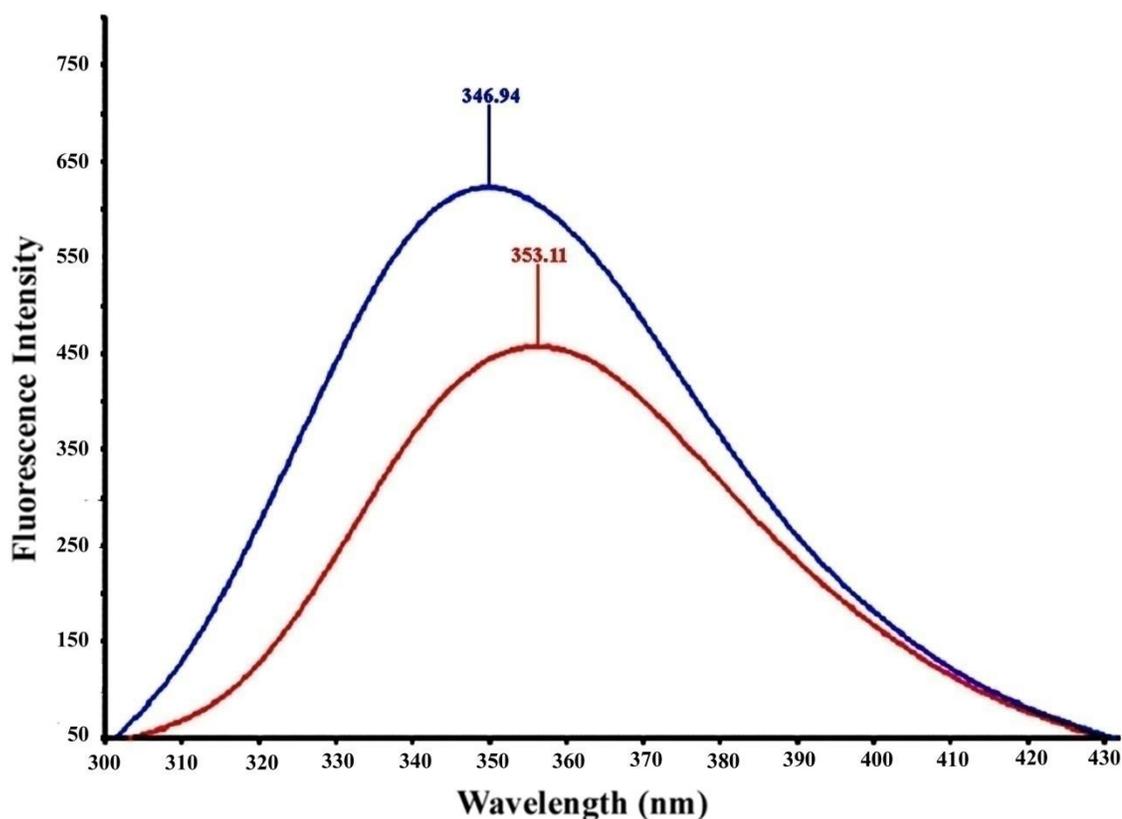


**Figure 2.3.8 Purification of Rv3614c by gel filtration.** Panel A shows a calibration curve for the standard proteins on a 120-ml Superdex 75 16/60 prepacked column. Panel B shows a typical FPLC elution profile attained by gel filtration on a Superdex 75 16/60 column. High molecular weight contaminant peaks are indicated by “1” and the Rv3614c elution peak is labelled. The pooled fractions are also highlighted. The target protein leaves the column at about 45 ml buffer, suggesting a multimeric form for the protein (the expected Mw of the protein is 19.8 kDa). Panel C shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2 shows samples taken from the load and lanes 3-10 show samples removed from the Rv3614c elution peak. Fractions from lanes 4-9 were pooled for structural studies. Lane 1 contains high molecular weight markers (Sigma).

## 2.3.4. Structural characterization of the Rv3614c protein

### 2.3.4.1. Fluorescence analysis of Rv3614c folding

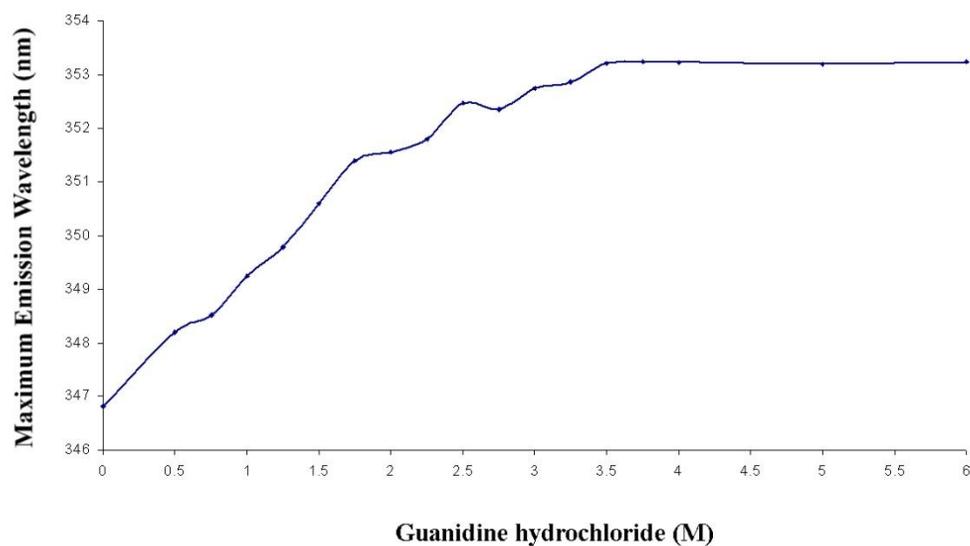
Typical intrinsic fluorescence spectra acquired for the native and denatured His<sub>6</sub>-tagged Rv3614c are shown in figure 2.3.9 (Section 2.2.13). The value corresponds to the denatured protein (353 nm) is expected for the proteins in which all the tryptophan residues are exposed to the aqueous solvent. The  $\lambda_{\text{max}}$  of the native protein is about 347 nm, which is blue shifted in comparison to the denatured protein (353 nm), indicating that at least one of the five tryptophan residues (Trp<sub>19</sub>, Trp<sub>27</sub>, Trp<sub>95</sub>, Trp<sub>150</sub> and Trp<sub>184</sub>) has become less solvent exposed.



**Figure 2.3.9:** Fluorescence analysis of His<sub>6</sub>-tagged Rv3614c. Maximum intrinsic tryptophan fluorescence spectra were obtained for the protein under native and denaturing (6 M GnHCl) conditions in 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and room temperature. The fluorescence studies were repeated at least three times.

#### 2.3.4.2. Chemical denaturation of Rv3614c

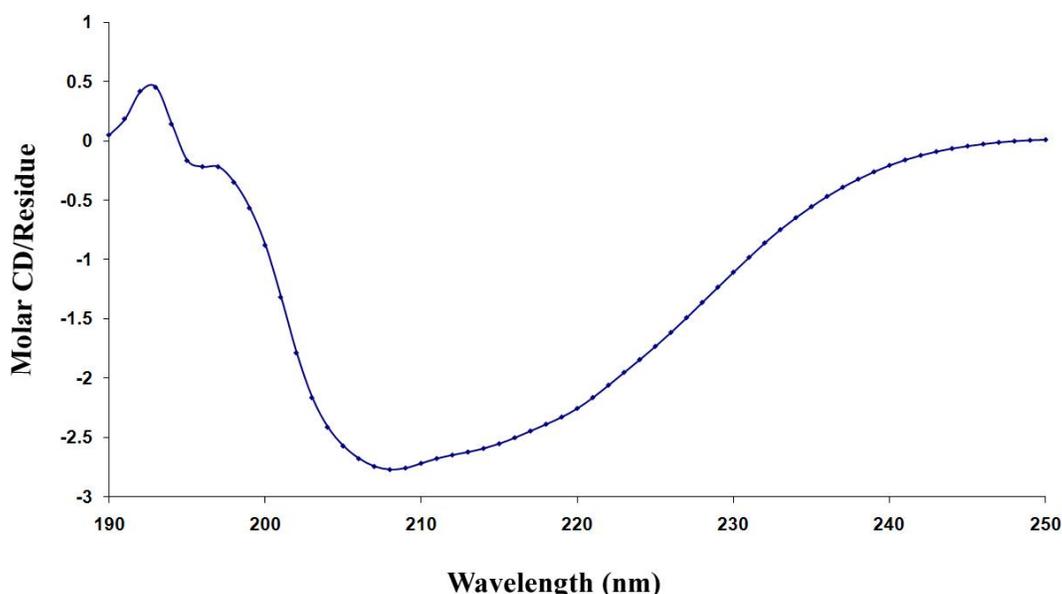
The graph shown in figure 2.3.10 demonstrates the effect of increasing guanidine hydrochloride concentrations on wavelength of maximum intrinsic tryptophan fluorescence observed for the His<sub>6</sub>-tagged Rv3614c protein. The chemical induced denaturation curve obtained for Rv3614c shows a gradual, non-cooperative unfolding of the protein, which indicates that the protein contains no stable tertiary structure (Section 2.2.13).



**Figure 2.3.10: Chemical induced denaturation of Rv3614c. Plot shows the changes in intrinsic fluorescence of Rv3614c resulted from guanidine hydrochloride induced denaturation of the protein. Data were collected in 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub>, and pH 6.5 at room temperature. Data are also based on results (the mean) from at least three independent assays.**

#### 2.3.4.3. Far UV circular dichroism spectroscopy for Rv3614c

The far UV circular dichroism spectrum shown in figure 2.3.11 is obtained from Rv3614c at room temperature (Section 2.2.11). Analysis of the spectrum using online server DICHROWEB(hosted at Brikbeck College, University of London, U.K)(59) provided the following estimates of secondary structure content of the His<sub>6</sub>-tagged Rv3614c : total helical content 14 %, total sheet 29 %, turn 25 % and 32 % unstructured.

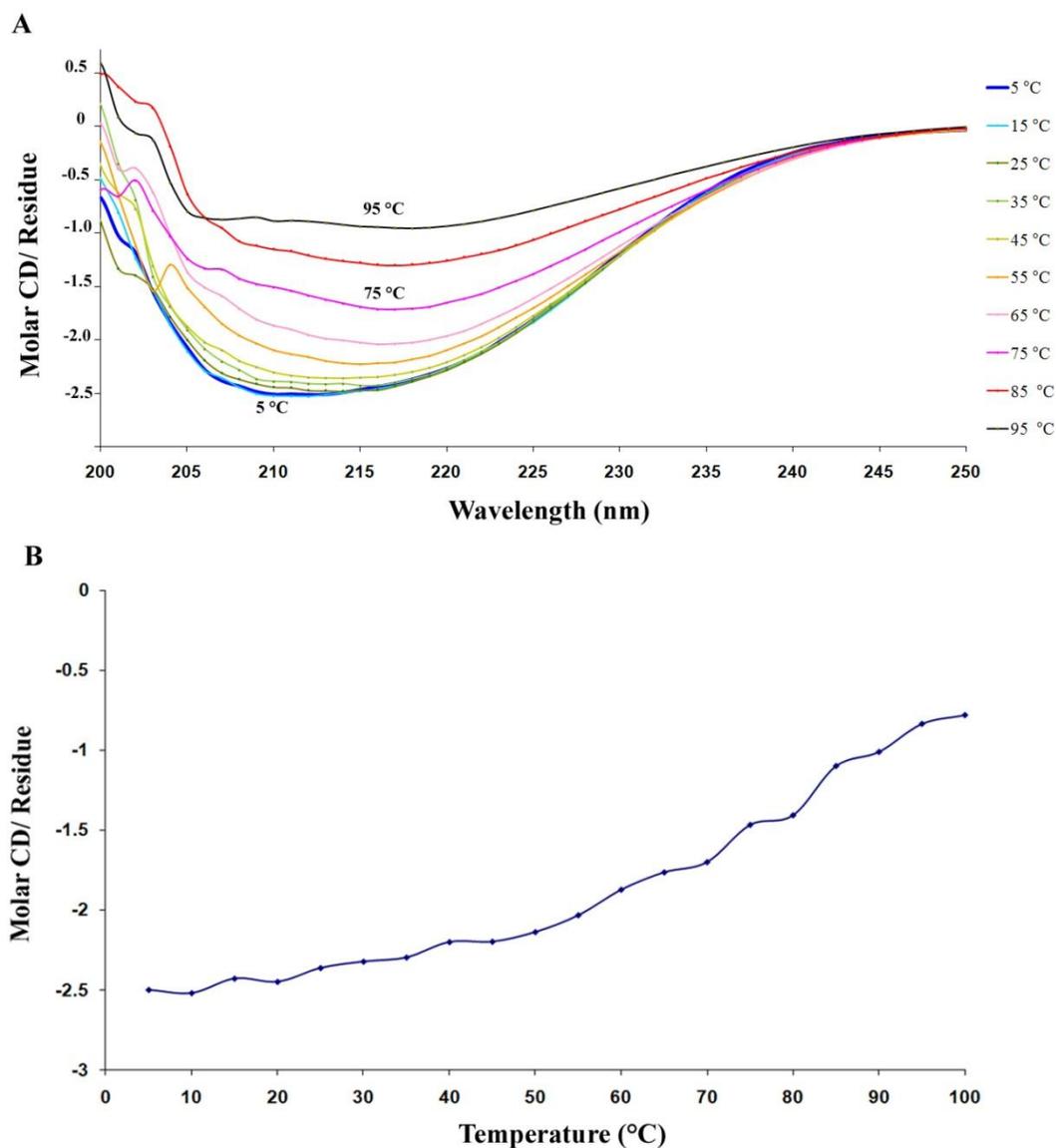


**Figure 2.3.11:** A far UV circular dichroism spectrum for Rv3614c. The CD spectrum was obtained from a protein sample dissolved in a 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and room temperature. The CD studies were repeated at least three times.

#### 2.3.4.4. The effect of thermal variation on Rv3614c

The series of far UV CD spectra recorded at increasing temperatures (Section 2.2.12) showed that Rv3614c is stable to thermal denaturation to about 20 °C ( Figure 2.3.12. panels A and B). The change in the intensity of the negative peak at 207 nm was recorded as a function of increasing temperature (panel B). The maximum negative CD peak, 207

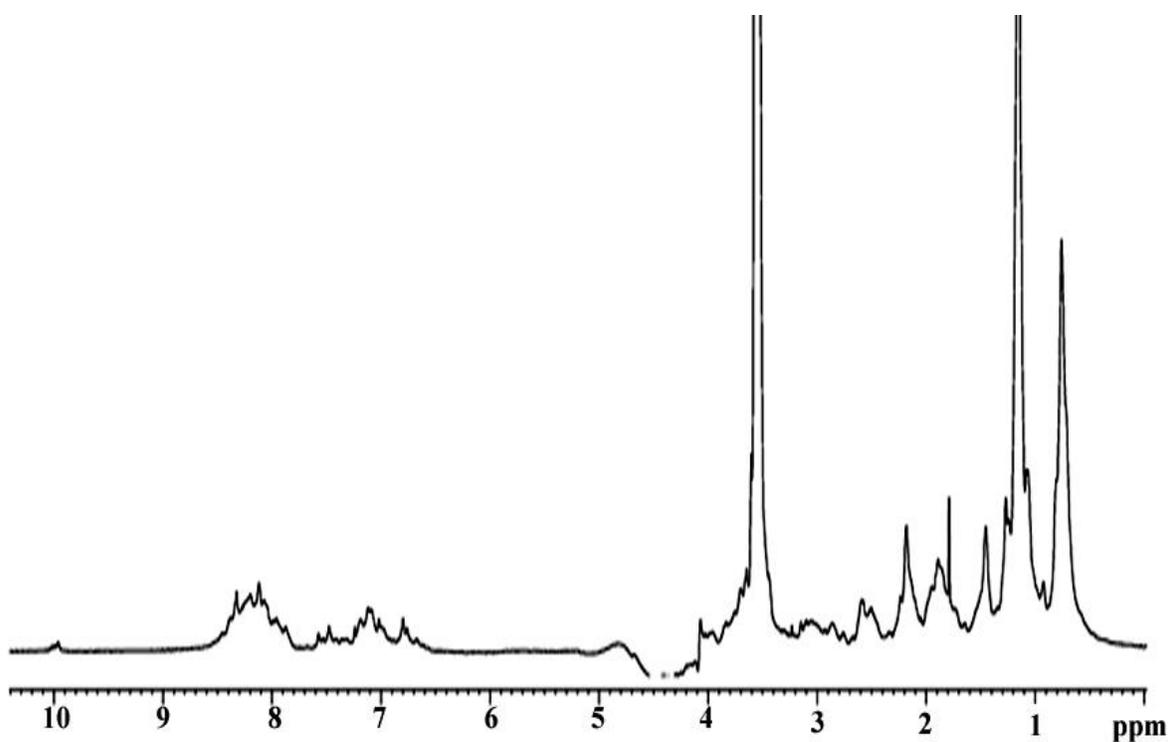
nm, was used to observe clearly the significant changes in the CD signals obtained for the protein upon increasing temperature. As shown in the panel B, the thermal denaturation curve obtained for the protein does not show a cooperative sigmoidal unfolding curve typically observed for proteins with stable tertiary structures.



**Figure 2.3.12** Far UV circular dichroism analysis of the temperature stability of Rv3614c. Panel A shows overlays of far UV spectra acquired for the protein over a range of temperatures. The spectra are the result of 10 accumulations. The spectra acquired at 5 (dark blue), 15 (cyan), 25 (dark green), 35 (green), 45 (yellow), 55 (orange), 65 (pink), 75 (purple), 85 (red) and 95°C (black) are shown. Panel B shows the effects of increasing temperature on the secondary structure of the protein, reflected as the change in CD intensity at 207 nm. The CD studies were repeated at least three times.

#### 2.3.4.5. 1D $^1\text{H}$ NMR analysis of Rv3614c

1D  $^1\text{H}$  NMR experiments were carried out at 20 °C as described in the section 2.2.14. As shown in the spectrum (figure 2.3.13), lack of signals from backbone amide groups (6.5 to 9.5 ppm) and high field shifted methyl resonances (0 to 0.6 ppm) strongly suggests that the target protein, Rv2614c, does not contain a stable tertiary structure. However, the broad  $^1\text{H}$  resonances of the protein are probably representative of a molten globule state.



**Figure 2.3.13:** One-dimensional  $^1\text{H}$  NMR spectrum of 100  $\mu\text{M}$  sample of Rv3614c in a 100 mM NaCl, 25 mM  $\text{NaH}_2\text{PO}_4$  buffer at pH 6.5 and 20 °C. The broad  $^1\text{H}$  resonances of the protein are probably representative of a molten globule state.

#### 2.3.4.6. Discussion

Circular dichroism spectra obtained for Rv3614c at 25 °C showed that the protein contains a stable mixed secondary structure, with quantitative analysis suggesting a total helical and  $\beta$ -sheet content of about 14 % and 29 % respectively (Figure 2.3.11). However, the 1D  $^1\text{H}$  NMR data acquired for Rv3614c showed that the protein does not contain a stable tertiary structure (Figure 2.3.13). This is fully consistent with the results of the chemical and heat-induced denaturation studies in which the chemical and temperature denaturation curves do not show a cooperative sigmoidal unfolding curve typically observed for proteins with stable tertiary structures (Figures 2.3.10 and 2.3.12). Taken together, presence of significant secondary structure (in total 43 %) along with an unstable tertiary structure implies that Rv3614c exists as a molten globule-structure-type structure.

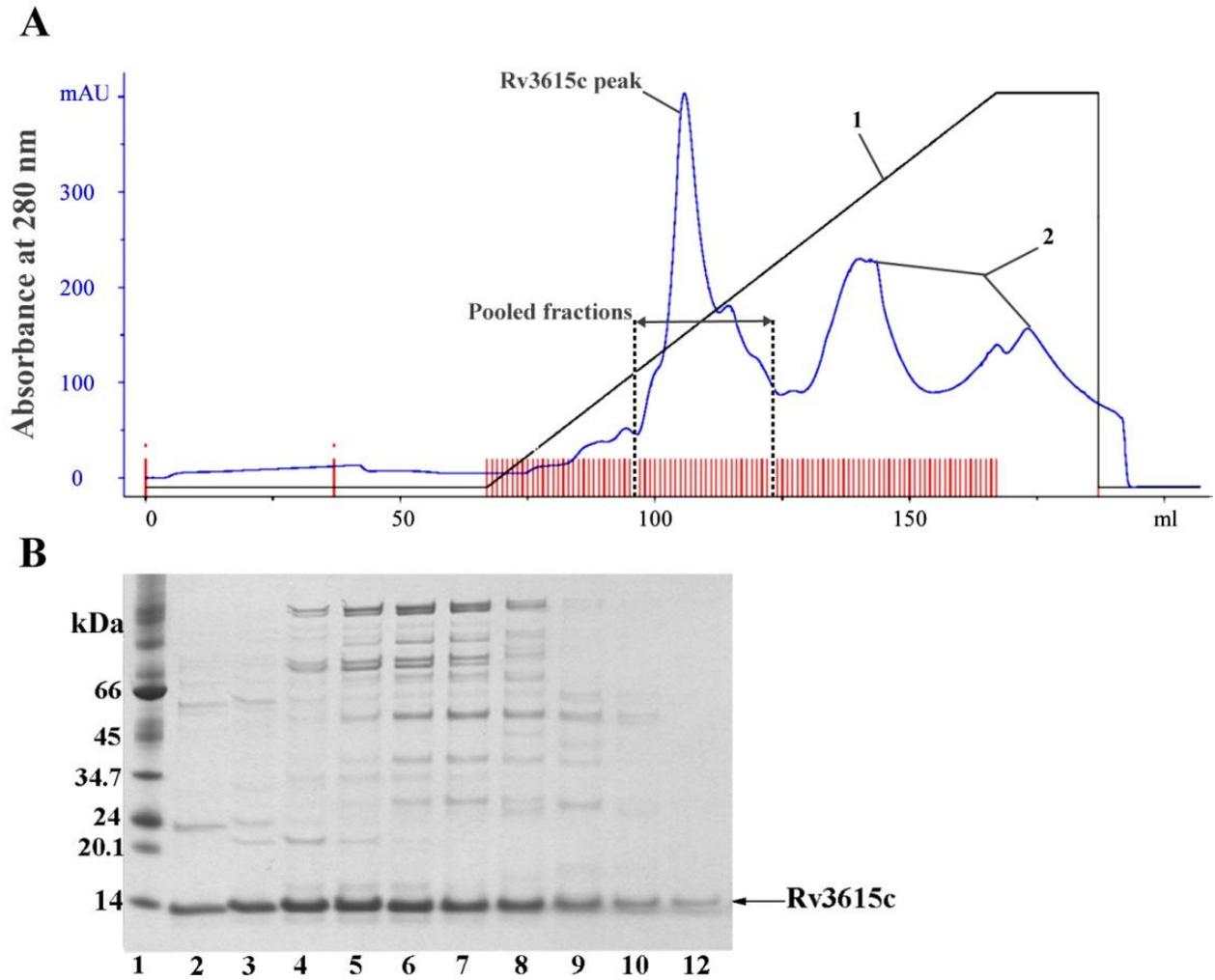
It is known that molten globules are very sticky, and show a considerable tendency to aggregate (61). Therefore, the existence of the molten globule state may explain the behaviour of the protein on gel filtration columns (the Rv3614c elution peak is at 45 ml) (Figure 2.3.8) which suggests a multimeric form for the protein. This sticky behaviour may also explain the blue shift observed in the  $\lambda_{\text{max}}$  (the wavelength of maximum intrinsic fluorescence of the native protein) (347 nm) compared to the denatured protein (353 nm) (Figure 2.3.9)

It is believed that Rv3614c is a member of the *snm* secretion pathway (48). Expression of the protein as a molten globule suggests that it may interact with other proteins involved in the secretion pathway, and probably undergoes coupled folding upon binding to its possible

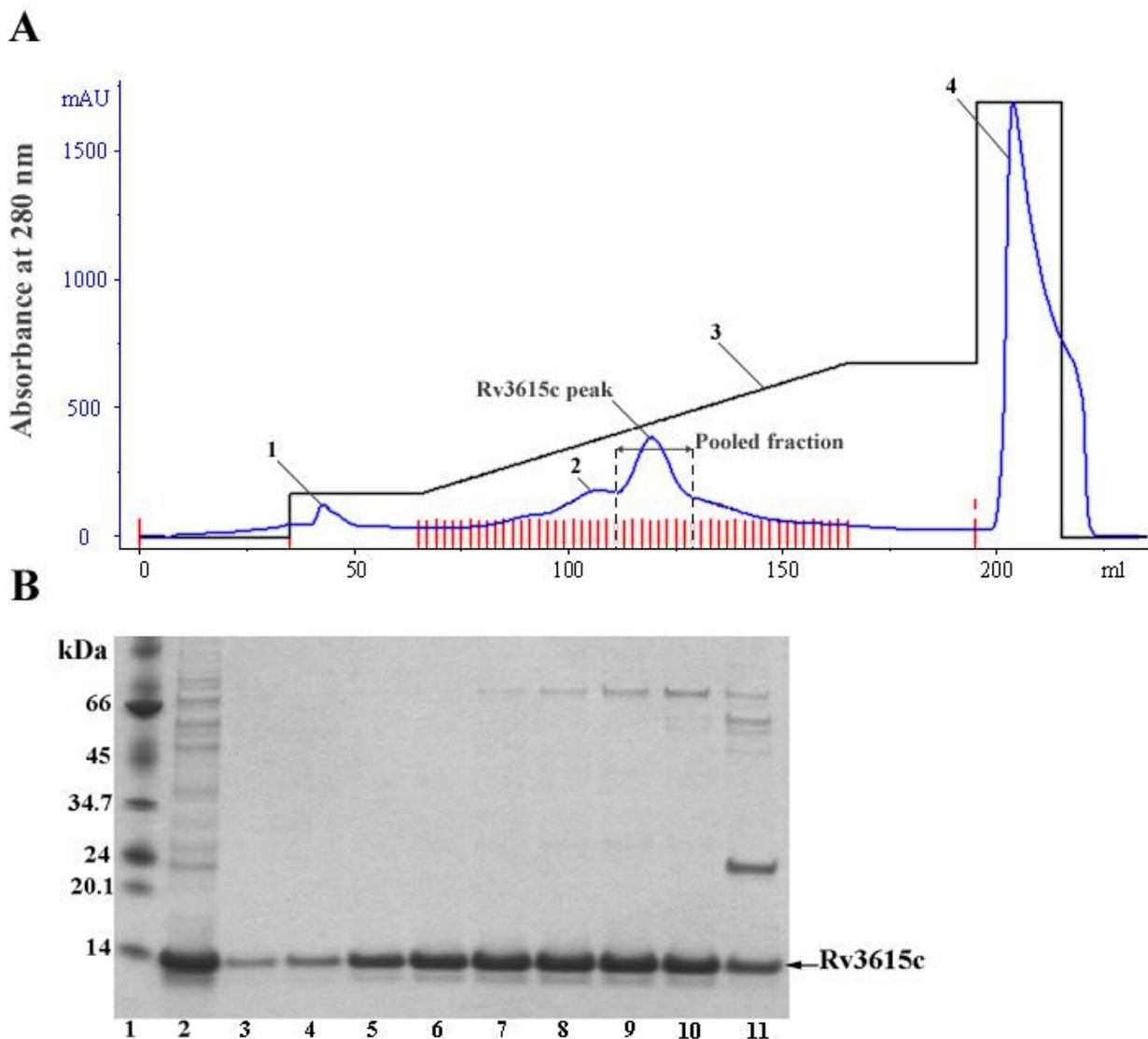
interacting partner(s). This phenomenon has already been observed for the ESAT-6-CFP-10 protein complex which is the main substrate of the ESX-1 secretion pathway (40).

### 2.3.5. Rv3615c expression and purification

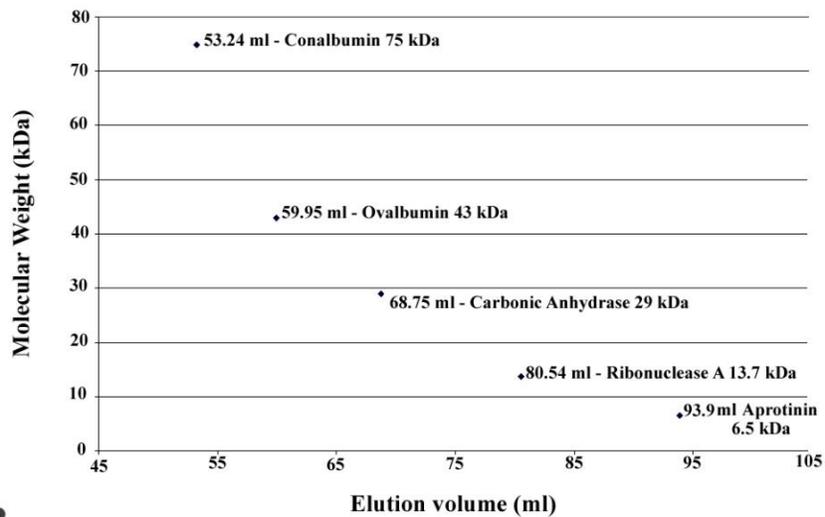
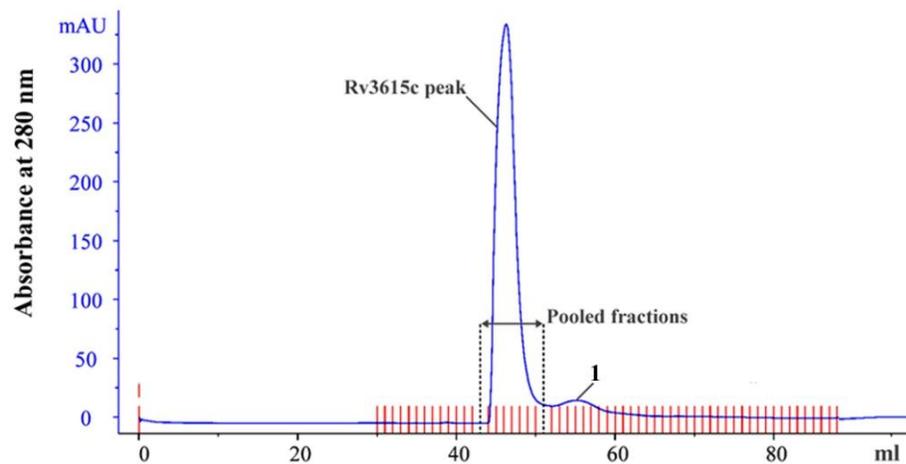
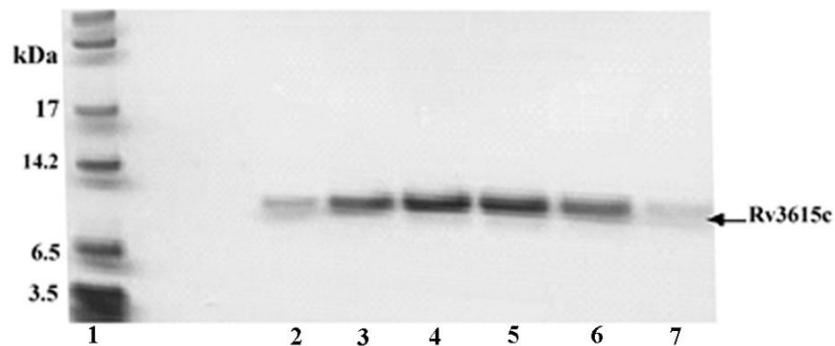
Rv3615c was produced as a soluble protein (Figure 2.3.3) and selectively purified following three chromatography steps: (I) Q-sepharose anion exchange chromatography at pH8 (Figure 2.3.14), (II) Q-sepharose anion exchange chromatography at pH 5.8 and, (III) gel filtration chromatography at pH 6.5. The final yield of Rv3615c was determined to be approximately 0.35 mg/ml using a theoretical extinction coefficient of  $9970 \text{ M}^{-1} \text{ cm}^{-1}$ , at 280 nm (section 2.2.9).



**Figure 2.3.14: Purification of Rv3615c.** Panel A shows a typical FPLC profile attained for the Q-sepharose column based purification of Rv3615c at pH 8. Rv3615c was eluted over a linear gradient from 0 mM to 500 mM NaCl (indicated by “1”). The Rv3615c elution peak is labelled. The pooled fractions are also highlighted. The contaminant peaks are indicated by “2”. Panel B shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-12 show samples removed from the Rv3615c elution peak. These fractions were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).



**Figure 2.3.15: Purification of Rv3615c.** Panel A shows a typical FPLC profile attained for the Q-sepharose column based purification of Rv3615c at pH 5.8. Rv3615c was eluted over a linear gradient from 30 mM to 500 mM NaCl (indicated by “3”). The contaminant peaks are indicated by “1, 2 and 4”. The Rv3615c elution peak and the pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying the FPLC elution fractions. Lane 2 shows the loaded sample from the previous Q-sepharose chromatography step. Lanes 3-11 show samples removed from the Rv3615c elution peak. Fractions from lanes 3-10 were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).

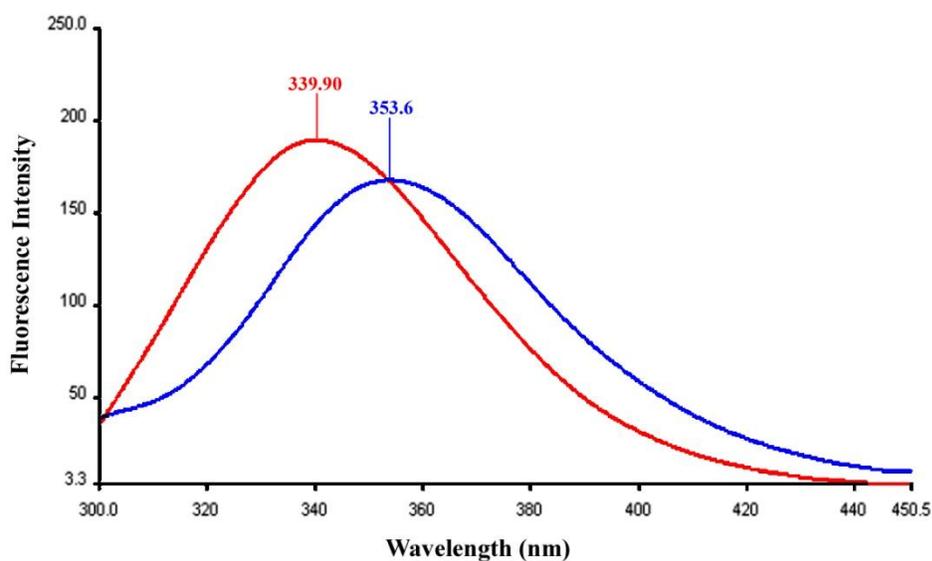
**A****B****C**

**Figure 2.3.16: Purification of Rv3615c by gel filtration.** Panel A shows a calibration curve for the standard proteins on a 120-ml Superdex 75 16/60 prepac column. Panel B shows a typical FPLC elution profile attained by gel filtration on a Superdex 75 16/60 column. A contaminant peak is indicated by “1” and the Rv3615c elution peak is labelled. The pooled fractions are also highlighted. The target protein leaves the column at about 45 ml buffer, suggesting a multimeric form for the protein (the expected Mw of the protein is 10.8 kDa). Panel C shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-7 show samples removed from the Rv3615c elution peak. These fractions were pooled for structural studies. Lane 1 contains low molecular weight markers (Sigma).

## 2.3.6. Structural characterization of the Rv3615c protein

### 2.3.6.1. Fluorescence analysis of Rv3615c folding

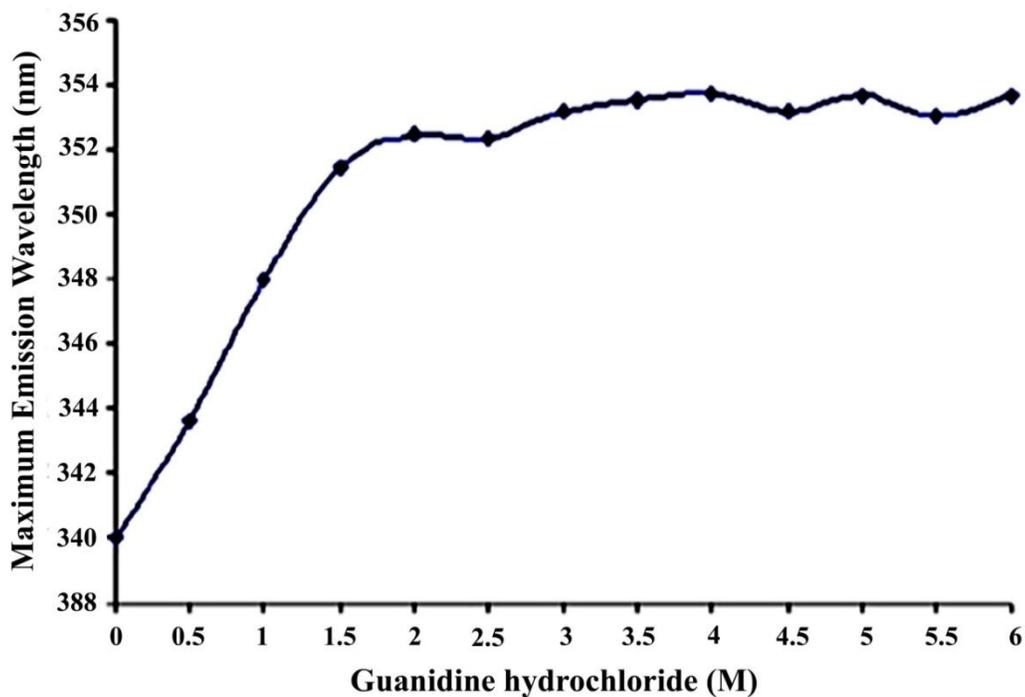
Typical intrinsic fluorescence spectra acquired for the native and denatured Rv3615c are shown in figure 2.3.17 (Section 2.2.13). The value corresponds to the denatured protein ( $353.6 \pm 0.5$  nm) is expected for the proteins in which all the tryptophan residues are exposed to the aqueous solvent. The  $\lambda_{\text{max}}$  of the native protein is about 340nm, which is blue shifted in comparison to the denatured protein (353.6 nm), indicating that the sole tryptophan residue (Trp<sub>94</sub>) has become significantly less solvent exposed.



**Figure 2.3.17:** Fluorescence analysis of Rv3615c. Intrinsic fluorescence emission spectra were obtained for the Rv3615c protein under native (red) and denaturing (blue) conditions in 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and room temperature. The fluorescence studies were repeated at least three times.

### 2.3.6.2. Chemical denaturation of Rv3615c

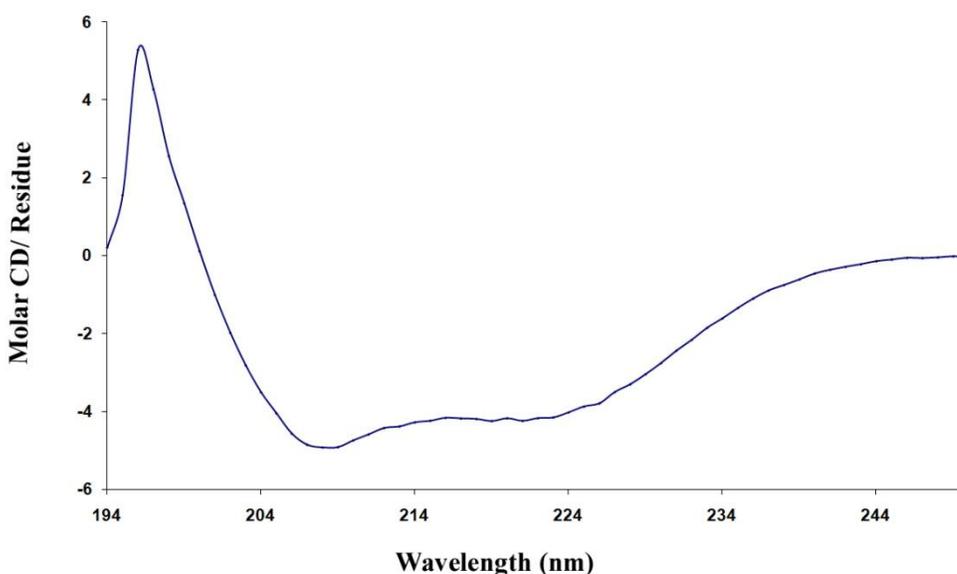
The graph shown in figure 2.3.18 demonstrates the effect of increasing guanidine hydrochloride concentrations on the wavelength of maximum intrinsic fluorescence observed for Rv3615c protein. The chemical induced denaturation curve obtained for Rv3615c shows a gradual, non-cooperative unfolding of the protein, which indicates that the protein contains no stable tertiary structure (Section 2.2.13).



**Figure 2.3.18: Chemical induced denaturation of the Rv3615c protein. The plot shows changes in intrinsic fluorescence of Rv3615c resulted from guanidine hydrochloride induced denaturation of the protein. Data are based on results (the mean) from at least three independent assays.**

### 2.3.6.3. Far UV circular dichroism spectroscopy for Rv3615c

The far UV circular dichroism spectrum shown in figure 2.3.19 is obtained from Rv3615c at room temperature (Section 2.2.11). Analysis of the spectrum using the online server DICHROWEB (59), provided the following estimates of secondary structure content of Rv3615c: total helical content 57 %, total sheet 18 %, turn 10 % and 15% unstructured.

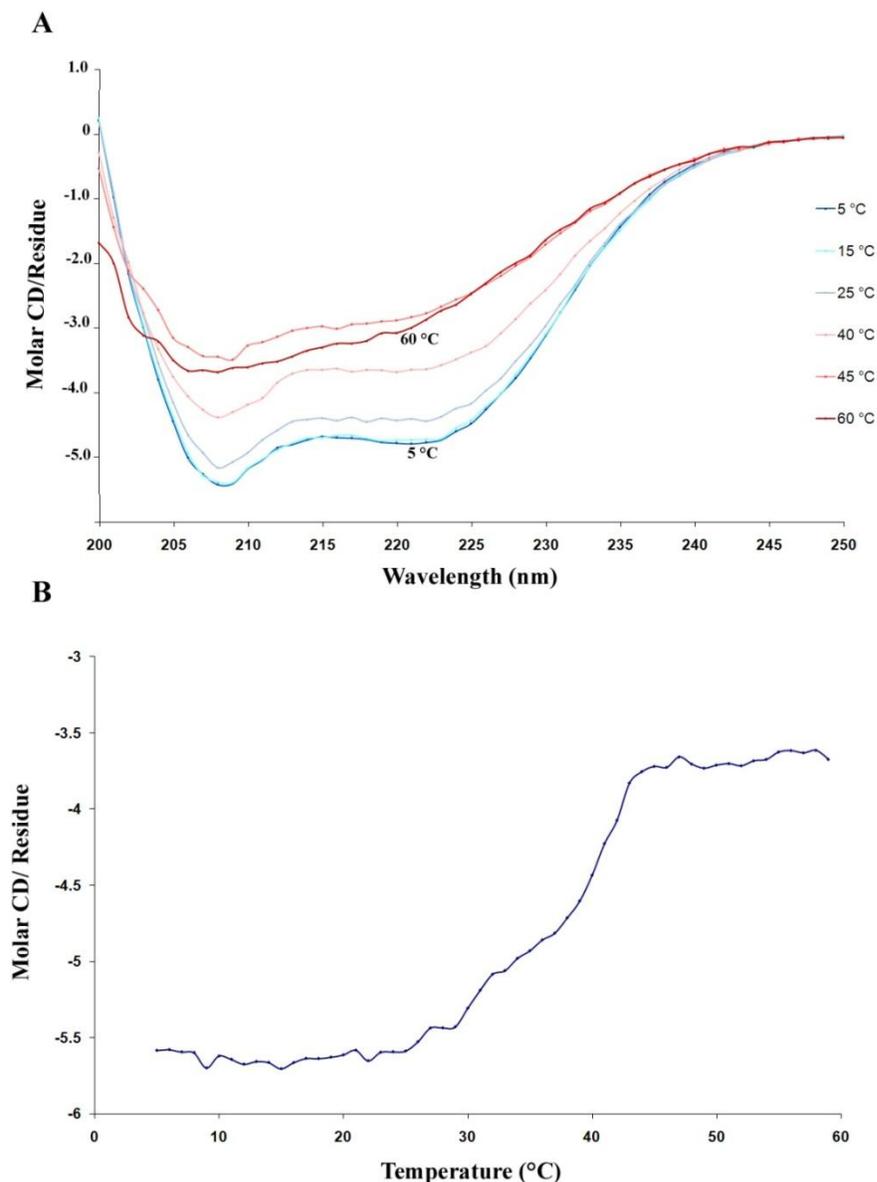


**Figure 2.3.19:** A far UV circular dichroism spectrum for Rv3615c. The CD spectrum was obtained from a 15.28  $\mu\text{M}$  protein sample dissolved in a 100 mM NaCl, 25 mM  $\text{NaH}_2\text{PO}_4$  buffer at pH 6.5 and 25°C. The CD studies were repeated at least three times.

### 2.3.6.4. The effect of thermal variation on Rv3615c

The series of far UV CD spectra recorded at increasing temperatures (Section 2.2.12) showed that Rv3615c is stable to thermal denaturation to about 20 °C ( Figure 2.3.20 panels A and B). The change in the intensity of the negative peak at 208 nm was recorded as a function of increasing temperature, as shown in panel B. The maximum negative CD peak,

208 nm, was used to observe clearly the significant changes in the CD signals obtained for the protein upon increasing temperature. The unfolding curve obtained for the protein suggests that Rv3615c is partially folded.



**Figure 2.3.20:** Far UV circular dichroism analysis of the temperature stability of Rv3615c. Panel A shows overlays of far UV spectra acquired for the protein over a range of temperatures. The spectra are the result of 10 accumulations. The spectra acquired at 5 (dark blue), 15 (cyan), 25 (light blue), 40 (pink), 45 (purple), and 60 °C (dark red) are shown. Panel B shows the effects of increasing temperature on the secondary structure of the protein, reflected as the change in CD intensity at 208 nm. The CD studies were repeated at least three times.

#### 2.3.6.5. Discussion

The circular dichroism spectra obtained for Rv3615c at 25 °C are typical of a predominantly helical protein, with characteristic negative peaks at approximately 208 and 221 nm (Figure 2.2.19)(94).

Thermal denaturation analyses suggest a partially folded structure for the protein as evidenced by the relatively low stability of the protein to thermal denaturation (up to about 20 °C) (Figure 2.3.20) and also the unfolding curve which is sigmoidal to some extent (Figure 2.3.20 panel B). Furthermore, the chemical denaturation curve obtained for the protein (Figure 2.3.18) shows a gradual, non-cooperative unfolding of the protein, which indicates that Rv3615c contains no stable tertiary structure. Taken together, the presence of significant helical structure but no apparent resistance to denaturation implies that Rv3615c exists as a molten globule.

The existence of the molten globule state may explain the behaviour of the protein on gel filtration column (Figure 2.3.16) in which Rv3615c came off in the void volume, suggesting a multimeric form for the protein. This sticky behaviour may also explain the significant blue shift observed in the  $\lambda_{\text{max}}$  (the wavelength of maximum intrinsic fluorescence) of the native protein (340 nm) compared to the denatured protein (353 nm) (Figure 2.3.17). This result is fully in agreement with previous yeast two hybrid analysis, suggesting an intermolecular interaction between Rv3615c proteins (48).

### 2.3.7. Complex formation between proteins encoded by the *Rv3616c-Rv3612c* operon

The pGAD and pGBD constructs encoding Rv3616c, Rv3615c, Rv3614c and Rv3613c (supplied by our collaborators at University of Leicester) were used in various combinations to transform *S. cerevisiae* CG1945 strains (Table 2.3.1).

In order to examine whether a yeast two hybrid system could detect the tight interactions observed in vitro for ESAT-6 and CFP-10 (1, 40), initial yeast two hybrid studies were designed. Yeast cells co-transformed with pGAD-*cfp-10* and pGBD-*esat-6* (positive control), and pGAD-*esat-6* and pGBD-*esat-6* (negative control) were assayed for *lacZ* reporter activity by filter lift assay. As expected, only yeast cells co-transformed with TAD-*cfp-10* and DBD-*esat-6* produced a functional GAL4 transcription factor, as evidenced by the development of blue colonies in the filter assay shown in figure 2.3.21 panels A, C, E, and I.

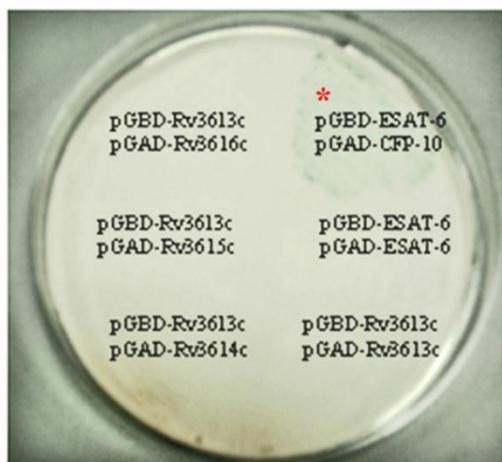
The yeast cells transformed with various combinations (Table 2.3.1) of pGBD and pGAD constructs expressing Rv3616c, Rv3615c, Rv3614c, and Rv3613c were then assayed for *lacZ* reporter activity by colony filter lift assays (Figure 2.3.21 panels A-J).

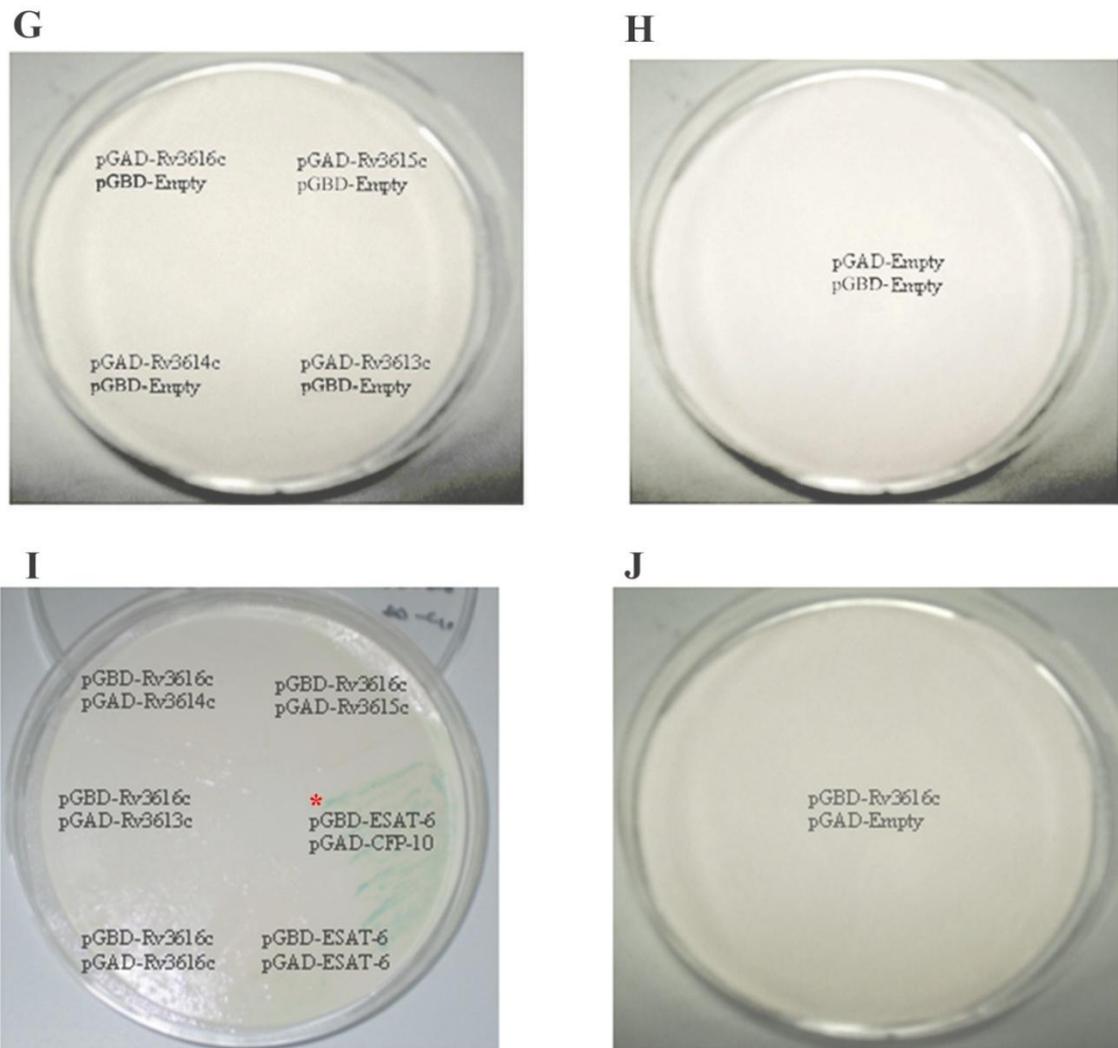
Results from the colony filter lift assays, shown in the figure 2.3.21 panels A and I, revealed that neither Rv3616c nor Rv3613c interact with itself or other protein encoded by the gene cluster in the yeast two-hybrid system. Subsequent filter lift assays, shown in figure 2.3.21 panels D and F, revealed that the yeast cells transformed with pGBD-

*Rv3614c* or pGBD-*Rv3615c* alone turn blue in the presence of X-gal, suggesting auto-activation of the *lacZ*-reporter. There was no evidence to suggest auto-activation by pGBD-*Rv3613c* (panel B), pGAD-*Rv3613c* (panel G), pGAD-*Rv3614c* (panel G), pGAD-*Rv3615c* (panel G), pGBD-*Rv3616c* (panel J), or pGAD-*Rv3616c* (panel G) (Figure 2.3.21).

**Table 2.3.1: Combinations of pGBD and pGAD constructs used to transform the *S. cerevisiae* to investigate interactions between proteins encoded by the *Rv3616c-Rv3613c* gene cluster.**

pGBD Construct	pGAD Construct
pGBD- <i>Rv3616c</i>	pGAD- <i>Rv3616c</i>
pGBD- <i>Rv3616c</i>	pGAD- <i>Rv3615c</i>
pGBD- <i>Rv3616c</i>	pGAD- <i>Rv3614c</i>
pGBD- <i>Rv3616c</i>	pGAD- <i>Rv3613c</i>
pGBD- <i>Rv3616c</i>	pGAD-Empty
pGBD-Empty	pGAD- <i>Rv3616c</i>
pGBD- <i>Rv3615c</i>	pGAD- <i>Rv3616c</i>
pGBD- <i>Rv3615c</i>	pGAD- <i>Rv3615c</i>
pGBD- <i>Rv3615c</i>	pGAD- <i>Rv3614c</i>
pGBD- <i>Rv3615c</i>	pGAD- <i>Rv3613c</i>
pGBD- <i>Rv3615c</i>	pGAD-Empty
pGBD-Empty	pGAD- <i>Rv3615c</i>
pGBD- <i>Rv3614c</i>	pGAD- <i>Rv3616c</i>
pGBD- <i>Rv3614c</i>	pGAD- <i>Rv3615c</i>
pGBD- <i>Rv3614c</i>	pGAD- <i>Rv3614c</i>
pGBD- <i>Rv3614c</i>	pGAD- <i>Rv3613c</i>
pGBD- <i>Rv3614c</i>	pGAD-Empty
pGBD-Empty	pGAD- <i>Rv3614c</i>
pGBD- <i>Rv3613c</i>	pGAD- <i>Rv3616c</i>
pGBD- <i>Rv3613c</i>	pGAD- <i>Rv3615c</i>
pGBD- <i>Rv3613c</i>	pGAD- <i>Rv3614c</i>
pGBD- <i>Rv3613c</i>	pGAD- <i>Rv3613c</i>
pGBD- <i>Rv3613c</i>	pGAD-Empty
PGBD-Empty	pGAD- <i>Rv3613c</i>
PGBD-Empty	PGAD-Empty

**A****B****C****D****E****F**



**Figure 2.3.21:** Filter lift assays of *S. cerevisiae* CG1945 cells transformed with the constructs as described in the table 2.3.1 and shown on the plates. Expression of the *lacZ* reporter gene is detected by the formation of blue colonies (indicated by \*), indicating interaction between ESAT-6 and CFP-10 (panels panels A, C, E, and I) and also auto-activation in transformants transformed with either pGBD-*Rv3614c* or pGBD-*Rv3615c* (panels C, D, E, and F). These results also show neither *Rv3613c* nor *Rv3616c* interacts with itself to form a multimer or interact with other members of the proteins encoded by the *Rv3616c-Rv3613c* gene cluster (panels A and I respectively). Panel G shows that pGAD-*Rv3613c*, pGAD-*Rv3615c*, pGAD-*Rv3615c*, and pGAD-*Rv3616c* do not auto-activate transcription of the *lacZ* reporter gene. pGBD-*Rv3613c* and pGBD-*Rv3616c* also do not auto-activate transcription of the *lacZ* reporter gene as shown in panels B and J respectively. Panel H shows that the transformants co-transformed with pGAD and pGBD could not produce functional GAL4 transcription factor.

As mentioned above, the strains transformed with either pGBD-*Rv3614c* or pGBD-*Rv3615c* alone auto-activated transcription of the *lacZ* reporter gene. Quantitative liquid assays (Figure 2.3.22, panels A and B) were then performed to compare the level of  $\beta$ -

galactosidase activity between the yeast cells transformed with pGBD-*Rv3614c* alone and those co-transformed with pGAD-*Rv3614c*/pGBD-*Rv3614c* or pGAD-*Rv3615c*/pGBD-*Rv3614c*. A comparison was also done between the cells transformed with pGBD-*Rv3615c* alone and those co-transformed with pGAD-*Rv3615c*/pGBD-*Rv3615c* or pGAD-*Rv3614c*/pGBD-*Rv3615c*. These assays were done to determine whether Rv3614c (or Rv3615c) interacts with itself to form a multimer (or homodimer) and also determine if there is any interaction between Rv3614c and Rv3615c in the yeast two-hybrid system (Section 2.2.18.2). As shown in figure 2.3.22 panel A, the quantitative liquid assays did not show a significant change in the level of  $\beta$ -galactosidase activity in the yeast cells co-transformed with pGAD-*Rv3614c*/pGBD-*Rv3614c* or pGAD-*Rv3615c*/pGBD-*Rv3614c* as compared to those transformed with pGBD-*Rv3614c* alone. This suggests that Rv3614c neither interacts with itself (multimer or homodimer) nor Rv3615c in the yeast two-hybrid system.

However, subsequent liquid assays, indicate that Rv3615c interacts with itself to form a multimer in the yeast two-hybrid system. This is evidenced by the significant change observed in the level of  $\beta$ -galactosidase activity in the yeast cells co-transformed with pGAD-*Rv3615c*/pGBD-*Rv3615c* as compared to those transformed with pGBD-*Rv3615c* alone (Figure 2.3.22 panel B). There is, however, a very small change in the level of  $\beta$ -galactosidase activity in the yeast cells co-transformed with pGAD-*Rv3615c*/pGBD-*Rv3614c* compared to those transformed with pGBD-*Rv3615c* alone, suggesting no intermolecular interaction between Rv3614c and Rv3615c in the yeast two-hybrid system (Figure 2.3.22 panel B).

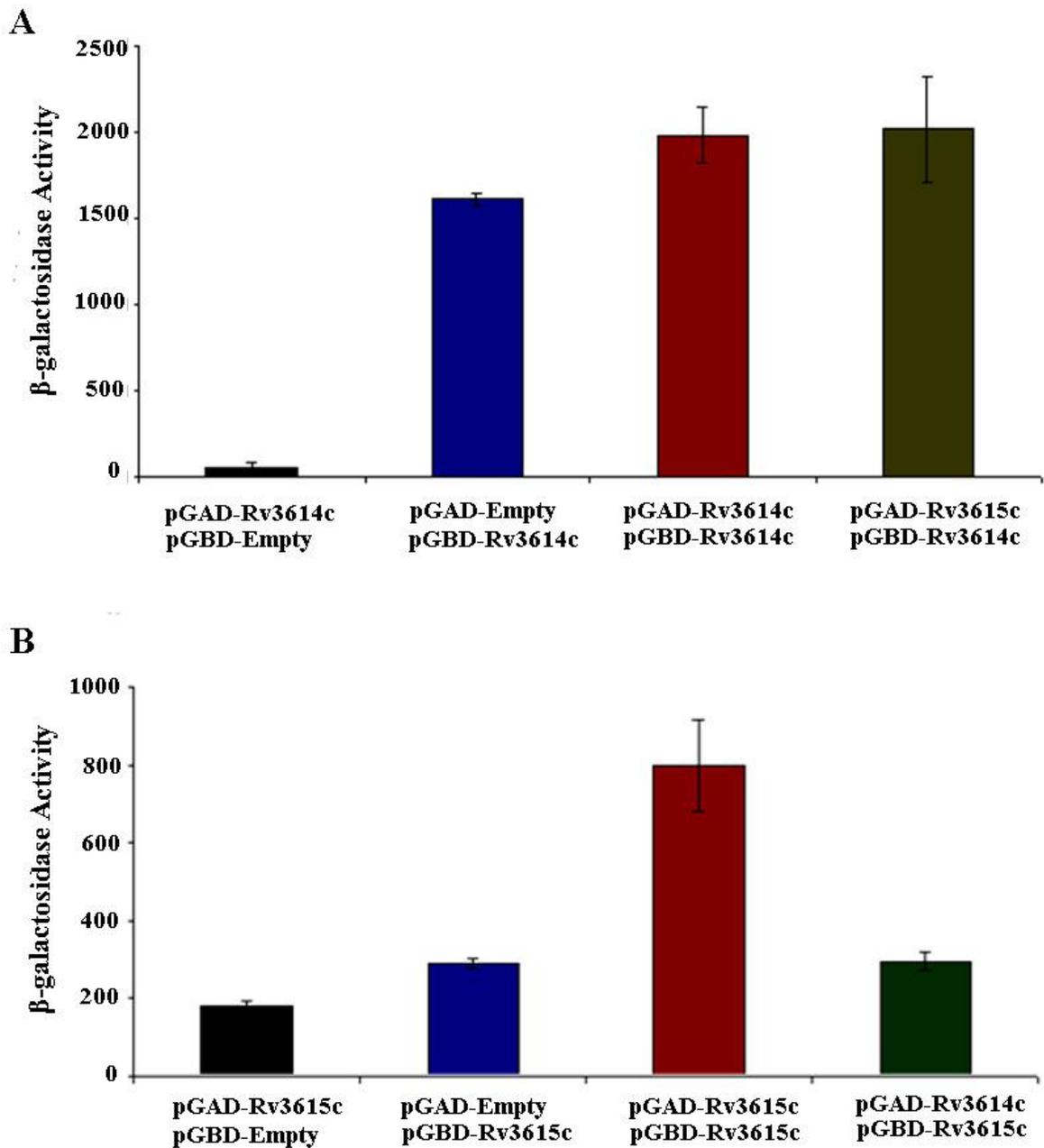


Figure 2.3.22: Quantitative liquid ( $\beta$ -galactosidase) assays to identify complex formation between the proteins encoded by Rv3614c and Rv3615c coding regions. As illustrated by the bar charts, panel A shows that there is not a significant change in the level of  $\beta$ -galactosidase activity in the yeast cells transformed with pGBD-Rv3614c alone as compared to those co-transformed with pGAD-Rv3614c/pGBD-Rv3614c or pGAD-Rv3615c/pGBD-Rv3614c. Panel B shows that the level of  $\beta$ -galactosidase activity in the yeast cells co-transformed with pGAD-Rv3615c/pGBD-Rv3615c has been significantly changed as compared to that of the cells transformed with pGBD-Rv3615c alone. Results are based on data from three independent assays.

### *2.3.7.1. Discussion and conclusion*

Initial structural analysis reported in this chapter indicated that Rv3614c and Rv3615c exist as molten globules, suggesting the proteins probably undergo coupled folding upon binding with a possible interacting partner(s). It is also known that the C-terminus of Rv3615c, like CFP-10, is required for secretion of the protein (9). The molecular features of Rv3615c described above, together with the fact that Rv3616c, like ESAT-6, does not contain any typical secretion signal sequence, raises the possibility of the interaction between these two secreted proteins (Rv3615c and Rv3616c). In fact, it is possible that Rv3615c interacts with Rv3616c and facilitates secretion of its binding partner. This phenomenon has already been observed for the ESAT-6-CFP-10 protein complex which is the main substrate of the ESX-1 secretion pathway (9, 40). In order to answer this question and also to find out the possibility of interactions between other proteins encoded by the gene cluster, yeast two hybrid assays were performed.

Comprehensive yeast two hybrid studies, however, showed no intermolecular interaction between the proteins encoded by the operon (the yeast-two hybrid analysis involving Rv3612c was done by Dr. Kirsty Lightbody at University of Leicester). The results of quantitative  $\beta$ -galactosidase assays, however, showed that Rv3615c is able to interact with itself to form a multimer. This is fully in agreement with the results from previous yeast two-hybrid studies (48), and also supported by the behavior of the protein on the gel filtration column in which the protein comes off in the void volume, suggesting a multimeric form for the protein (Figure 2.3.16).

The results obtained from the yeast-two hybrid analyses, however, cannot rule out the possibility of interactions between the proteins expressed by the gene cluster because these proteins might form a multi-protein complex. In fact, a single protein may be strongly bound to a multi-subunit complex by interacting simultaneously with several different constituents of the complex, while interaction of the protein with individual partners in the multi-protein complex might be at low affinity insufficient to be detected by typical yeast-two hybrid system (54).

## Chapter 3 - Solution structure of the complex formed between Rv0287 and Rv0288

### 3.1. Introduction-Part 2

Comparative studies identified five similar copies of the RD1 (regions 1-5) in the genome of *M. tuberculosis* (63, 65). These five loci contain a pair of *esx*-like genes, often preceded with a *pe-ppe* pair, and flanked by ORFs encoding putative ABC transporters, potential membrane proteins (9, 28, 63, 65, and 66). As illustrated in figure 3.1, there is a remarkable similarity between region 1 and 3 (from *Rv0282- Rv0292*) (63, 65, 66). Similar to the RD1 region, another couple of *esx*-like genes (*Rv0287* and *Rv0288*) are located in region 3 as a tandem pair, preceded with *pe-ppe* genes (63, 65, and 66). It has been suggested that the genes surrounding the *esat-6/cfp-10* related genes are highly likely to encode essential elements for a secretory apparatus which is responsible for export of the Esx-like proteins out of the bacterial cytoplasm (9, 63, 78, and 79).

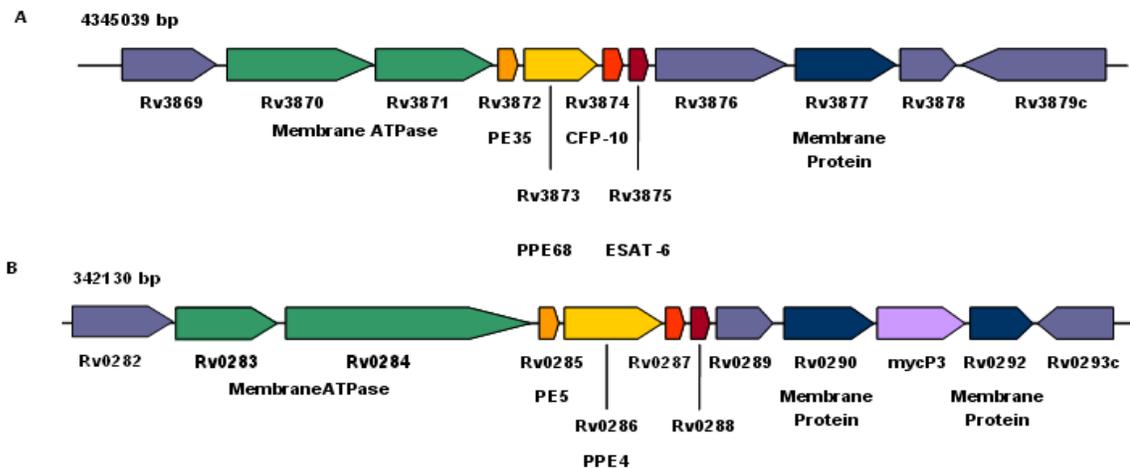


Figure 3.1: Schematic representation of *M. tuberculosis* CFP-10/ESAT-6 regions 1 (RD1) (A) and 3(B). The *esx*-like genes are shown by red and dark red colors. The *pe* and *ppe* genes are highlighted by orange and yellow colors respectively. The putative membrane ATPases are highlighted by green color. The putative membrane associated proteins are shown by dark blue color.

### 3.1.1. Region 3, a similar copy of the region 1 (RD1) with functional genes

The genes contained in the region 3 appear to be functional as evidenced by the fact that most of the sequences of the genes belonging to the region contain no frameshifts or multiple stop codons (65). This is significant when placed in the context of a mycobacterium like *M. leprae* as it is believed that the genome of *M. leprae* contains the minimal gene set for a pathogenic mycobacterium and that some functional genes once present in the *M. leprae* genome became pseudogenes through multiple stop codon mutations and frameshifts (65, 71).

Microarray-based analyses have also shown that the *esx*-like genes contained in the region, *Rv0287* and *Rv0288*, are transcribed in *M. tuberculosis* H37Rv (67). *Rv0287* and *Rv0288* are 291 and 288 bp respectively and are separated by a 35-bp DNA fragment. The short intergenic space suggests that the ORFs are co-transcribed. Reverse transcriptase-PCR (RT-PCR) analyses conducted by Okkels *et al.* subsequently showed that the genes are transcribed as one mRNA unit in *M. tuberculosis* grown in vitro (66). This finding is further supported by recent bioinformatics and microarray analyses, suggesting that the genes are co-transcribed in the laboratory strain of *M. tuberculosis* (68).

Proteomics analyses including two-dimensional gel electrophoresis, mass spectrometry and immunodetection have also detected the protein products of the *esx*-like genes associated with the regions, further supporting the fact that the region contains functional genes (66, 69, and 70).

### 3.1.2. Region 3 as an important gene cluster in *M. tuberculosis* pathogenesis

There is no direct experimental evidence for Rv0288 and Rv0287 proteins contributing to virulence in *M. tuberculosis*. However, the following findings suggest a possible role for the Esx-like in *M. tuberculosis* pathogenesis:

(I) it has been suggested that *M. leprae* has undergone extensive reductive evolution, retaining minimum functional gene set required for a pathogenic mycobacterium (65,71). The presence of homologues of Rv0288 and Rv0287 in *M. leprae* genome (*ML2531c* and *ML2532c* respectively) as functional genes, therefore, suggest that the genes are likely to be involved in *M. tuberculosis* pathogenesis (63, 65).

(II) Moreover, the sequence of TB 10.4 protein encoded by the Rv0288 gene was found to be highly conserved within different clinical isolates of *M. tuberculosis* obtained from different geographical locations (72, 73).

(III) Comparison of the gene expression pattern between the two closely related species, attenuated H37Ra and virulent H37Rv, identified that Rv0288 expression is significantly down-regulated in H37Ra compared to the virulent H37Rv strain, further supporting the contribution of Rv0288 protein in the *M. tuberculosis* pathogenesis (74, 75).

(IV) Furthermore, recently in vivo induced antigen technology (IVIAT) has been used for probing *M. tuberculosis* in vivo induced antigens. In this method a *M. tuberculosis* H37Rv cDNA expression library was screened with pooled TB patient sera pre-absorbed with in

vitro grown *M. tuberculosis* H37Rv. The pre-eliminated serum allowed identification of antigens which are specifically expressed or up-regulated during growth and infection in vivo. This study identified six ORFs including Rv0287 as in vivo induced antigens (77). Taken together, these analyses indicate that the *esx*-like genes located in the region 3 probably contribute to virulence in *M. tuberculosis*.

### 3.1.3. Region 3 contains a set of genes encoding highly immunogenic antigens

#### 3.1.3.1. Epitope mapping of the immunodominant antigen TB10.4 (Rv0288)

TB10.4 has been recently described as an Ag which is expressed by both bacillus Calmette-Guérin (BCG) and *M. tuberculosis*. Recent epitope mapping analyses identified three epitopes in the protein which are strongly recognized by CD8<sup>+</sup> and CD4<sup>+</sup> T cells.

Recently, Majlessi *et al.* has shown that TB10.4<sub>20-28</sub> epitope is naturally processed and presented by MHC-I molecules of antigen presenting cells to TB10.4<sub>20-28</sub>-specific CD8<sup>+</sup> T cells in both BCG- and *M. tuberculosis*-infected mice(82).

These TB10.4<sub>20-28</sub> specific CD8<sup>+</sup> T cells comprised up to 30 % of total CD8<sup>+</sup> T cells found in lung of chronically infected mice. The T cells produced INF- $\gamma$ , TNF $\alpha$  and also express CD107A/B molecules (cytotoxic T lymphocytes markers on the cell surface), thus suggesting that the cells are likely to be involved in activating macrophages, granuloma formation, inducing apoptosis in infected cells, and probably play a role as cytotoxic T lymphocytes (CTLs) in vivo.

Furthermore, recent epitope mapping studies conducted by Billeskov *et al.* identified another CD8<sup>+</sup> T cell epitope in N-terminal part of TB10.4, TB10.4<sub>3-11</sub>. These studies showed that the TB10.4<sub>3-11</sub>-specific CD8<sup>+</sup> T cells were induced and employed initially to the site of infection where they were present throughout the infection in high numbers. The induced antigen-specific T cells, which are characterized by production of IFN- $\gamma$  and TNF- $\alpha$ , also up-regulated the expression of CD107A/B (cytotoxic T lymphocytes markers). This strongly suggests a strong cytolytic activity for the activated T cells.

Further epitope mapping of the antigen conducted by Hervas-Stubbs *et al* also identified the TB10.4<sub>74-88</sub> peptide as an immunodominant CD4<sup>+</sup> T-cell epitope in mice. These analyses have also revealed that although the antigen delivered by the CyaA vector, a newly designed vector able to deliver CD4<sup>+</sup> or CD8<sup>+</sup> epitopes to MHC class I and II molecules, is able to induce TB10.4<sub>74-88</sub>-specific T cell mediated immune responses, it does not provide protective immunity against *M. tuberculosis* infection in mice, unless it is co-administrated with in a strong cocktail of Th1-promoting adjuvant (72).

### 3.1.3.2. Rv0287 and Rv0288 as two subunit vaccine candidates

Doherty *et al.* have recently shown that the ESAT-Ag85B induces protective immunity superior to individual antigen components (83). Given the fact that ESAT-6 is considered as a powerful tool for diagnosis of tuberculosis infection (29, 88, 91, 92), discovering another vaccine as effective as ESAT-Ag85B that lacks ESAT-6 would be a priority.

TB10.4 antigen has recently received much attention since the protein, like ESAT-6, is strongly recognised by both TB patients and BCG vaccinated individuals (> 70 %) (27, 29,

36, 72, and 82). Subsequent vaccine analyses showed that fusing of TB10.4 to Ag85B antigen provides protection against tuberculosis infection in infected mice. Interestingly, this protective immunity was comparable to that of induced by both BCG and ESAT-6-Ag85B fusion protein (29).

In contrast to Rv0288, there is no experimental evidence for Rv0287 to act as a protective antigen. However, it has been shown that the antigen is highly immunogenic, and its immunogenicity is comparable to its partner, Rv0288 (63, 86)

### *3.1.3.3. Rv0287 and Rv0288 as protein tools for diagnosis of TB*

Development of proper immunodiagnostic reagents is crucial to improve performance of tuberculin-based reagents and also distinguish BCG vaccinated individuals from those infected with *M. tuberculosis*. As previously explained a large number of studies have shown that ESAT-6/CFP-10 protein family are potent T cell antigens. ESAT-6 and CFP-10 proteins, in particular, have been shown to be reliable immunodiagnostic reagents for diagnosis of tuberculosis in humans, cattle and guinea pigs (29, 88, 91, and 92). Subsequent studies have also demonstrated that Rv0288 and Rv0287 induce T cell-mediated immune responses to a large extent (27, 29, 36, 72, 85 and 86). It is also found that TB10.4 is even more strongly recognised than ESAT-6, suggesting Rv0288 may aid to the diagnosis of TB (87).

Recently a diagnostic cocktail of immunodominant peptides has been formulated from six *M. tuberculosis* antigens, Rv0288, Rv3019c, Rv3873, Rv3879c, ESAT-6 and CFP-10 (88). Further analyses conducted on naturally *M. bovis* infected cattle revealed that the cocktail was considerably more sensitive than purified protein derivative (PPD) and ESAT-6/CFP10

particularly in detection of the infected animals with false negative skin test. In addition, in contrast to PPD, specificity of the cocktail was not compromised with BCG vaccination because none of BCG-vaccinated animals showed INF- $\gamma$  responses after stimulation with either ESAT-6/CFP-10 or the immunogenic cocktail (88).

#### 3.1.4. Region3 and its biological function

Despite the clear importance of the CFP-10/ESAT-6 family proteins in mycobacterial virulence and pathogenesis, the precise molecular functions and mechanisms of action these protein remain unknown. Nutritional immunity, as a non-specific immune response, is an intense nutritional competition between host and pathogenic agents for nutrients and trace elements including Iron and zinc (89). Under this highly competitive condition the pathogenic bacteria usually use a simple strategy to circumvent this non-specific immune response. In fact, they usually recruit regulatory proteins to sense the altered levels of intracellular trace elements and other nutrients required for either their survival or growth within host cells. The regulators, subsequently, mediate required transcriptional responses to restore the corresponding nutritional changes (89). Recently, a zinc uptake regulator has been identified in *M. tuberculosis*. The gene encoding the regulator, *Zur*, is found to be co-transcribed with an upstream gene called *Rv2358*. Studies conducted on *M. smegmatis* showed that transcription of the *Rv2358-Zur* operon is controlled by *Rv2358* in a zinc-dependent fashion (90). Binding of Zinc ions to *Rv2358* weakens DNA binding affinity of the sensor significantly, allowing RNA polymerase to initiate transcription of the operon (90). Subsequent studies on the proteins encoded by the operon identified that the *Zur* protein acts as a metalloregulatory protein and its regulatory function is also zinc-

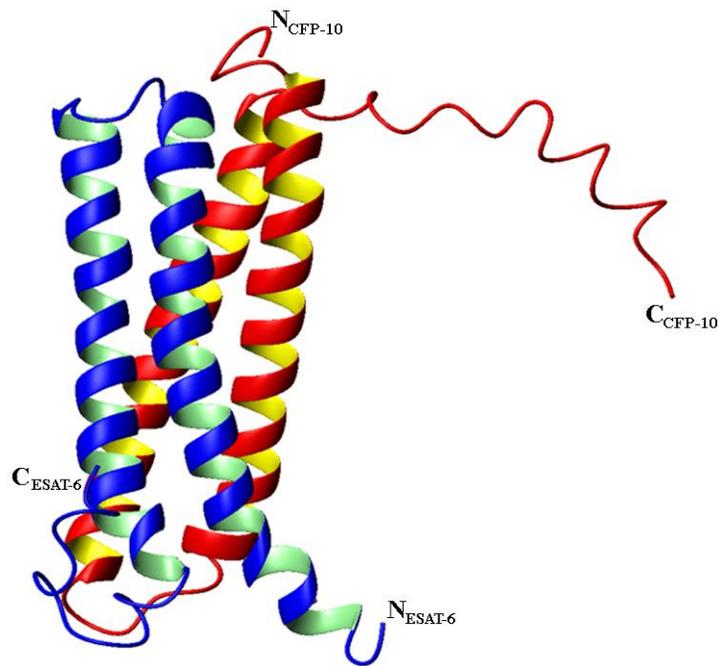
dependent. These analyses showed the regulator binds to a Zur binding site, a conserved 26 bp palindrome, and down regulates transcription of five members of ESAT-6 like proteins (Rv0288, Rv0287, Rv3019c, Rv3020c and Rv3017c) in presence of zinc; all 11 genes of cluster 3 (Rv0282-Rv0292) were induced in the *M. tuberculosis zur* mutant (89). The gene cluster is also found to be controlled by Iron and an Iron dependent regulator (IdeR) so that it is repressed by Iron and IdeR, but induced by Iron starvation (50). Consistent with these findings, two different promoters have been identified up stream of Rv0282, one overlaps IdeR binding site and another overlaps zur binding site, suggesting the cluster3- secreted proteins may be involved in zinc and/or Iron uptake or scavenging (63,89).

### 3.1.5. Structure and molecular function of the ESX-like proteins encoded by Rv0288 and Rv0287 genes

The presence of Rv0287 and Rv0288 as a tandem pair in the *M. tuberculosis* genome along with their transcription as one mRNA initially raised the question of the possibility of the interaction between the corresponding proteins (65, 66, and 68). Subsequent immunoblotting and yeast two-hybrid analyses showed that Rv0287 and Rv0288, like CFP-10 and ESAT-6, interact with each other to form a tight complex (62, 63, and 66). Comprehensive yeast two-hybrid analyses were then used to characterize complex formation between the three pairs of CFP-10/ESAT-6 family proteins (ESAT-6/CFP-10, Rv0288/Rv0287 and Rv3019c/Rv3020c) to examine the possibility of complex formation between non-genome paired members of the family. Interestingly, these analyses clearly demonstrated that the closely related Rv0287/Rv0288 and Rv3019c/Rv3020c proteins (> 80 % amino acid identity) can also form non-genome paired complexes (Rv0287-Rv3019c

and Rv0288-Rv3020c), but are not capable of binding to the more distantly related CFP-10/ESAT-6 proteins(62).

Examination of the high resolution solution structure of the ESAT-6-CFP-10 complex showed that the core of the complex contains two helix–turn–helix hairpin structures formed from the individual proteins, which have an extensive hydrophobic contact surface (about 1800 Å<sup>2</sup>) and lie anti-parallel to each other to form a four-helix bundle (figure 3.2). (1, 63).



**Figure 3.2:** A ribbon presentation of the backbone topology of the CFP-10-ESAT-6 complex which illustrates the two helix–turn–helix hairpin structures is formed by the individual proteins (ESAT-6 and CFP-10, which are depicted in blue and red respectively (1)). N and C termini of the both proteins are labelled.

Further analysis of sequence conservation across the ESAT-6 and CFP-10 related proteins reveals that over half of the residues involved in interface contacts are conserved to residue type in at least two-thirds of the sequences (figure 3.3) (1,63). This, together with the fact

that several members of the *M. tuberculosis* ESX-like proteins form complexes (62,63,66) strongly suggest that all pairs of these proteins, including Rv0288 and Rv0287, are most likely to form four-helix bundle structures similar to that of ESAT-6-CFP-10 complex (1,63).

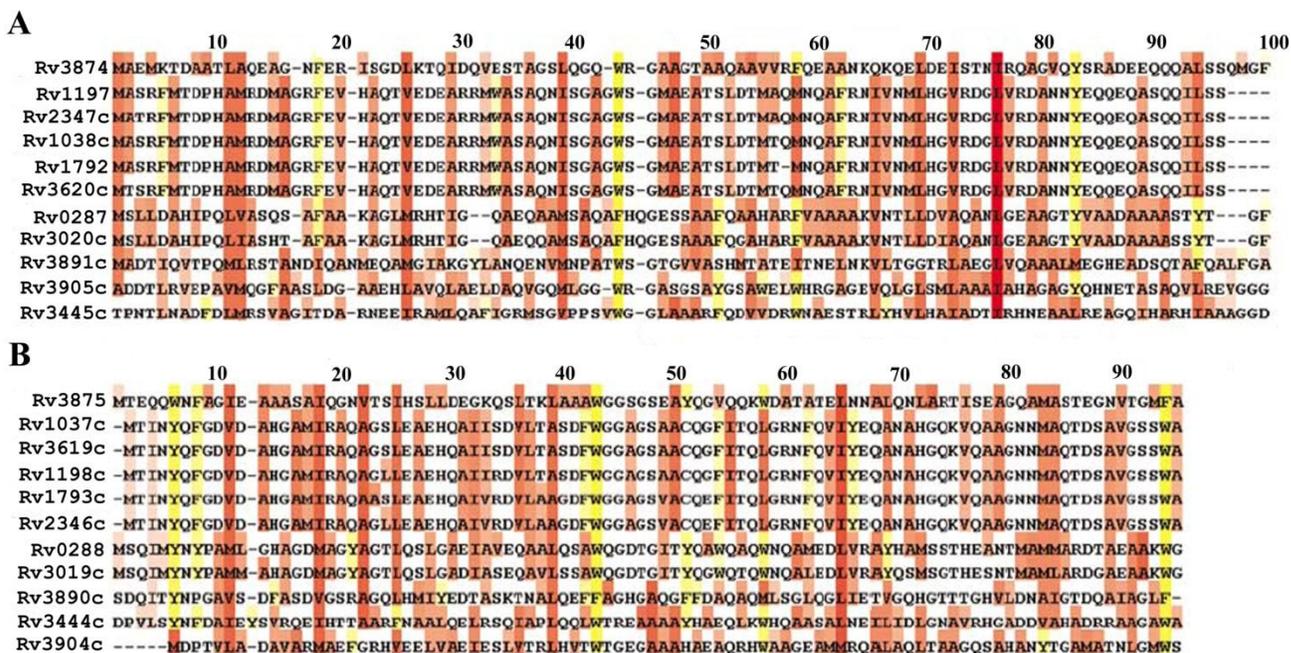


Figure 3.3: Multiple sequence alignment illustrates the conservation of the amino acid sequences of the CFP-10 (A) and ESAT-6 (B) related proteins from the *M. tuberculosis* ESAT-6 protein family. The residues are highlighted as follows: Aliphatic residues with hydrophobic side chains (Leu, Ile, Val, Met and Ala) in red, aromatic residues (Phe, Tyr and Trp) in yellow (1).

### 3.1.6. Aims

The aims of the work described in this chapter were

(I) to determine molecular features and properties of the proteins encoded by the Rv0287 and Rv0288 coding regions

(I) to characterize the complex formation between Rv0287 and Rv0288.

(II) to characterize the stability of the Rv0287-Rv0288 complex and individual proteins.

(III) to determine of high resolution solution structure of Rv0288-Rv0287 complex.

(IV) to compare the structural features of the complex to that reported for the closely related CFP-10-ESAT-6 (1)

Furthermore, analysis of the determined Rv0288-Rv0287 structure may assist in determining the possible function and mechanism of action for the complex by allowing the identification of structural similarity with proteins of known function, or through the identification of putative sites on the surfaces (117).

## **3.2. Methods and materials- Part 2**

### **3.2.1. Ligation independent cloning (LIC-PCR)**

In order to produce Rv0287 and Rv0288 proteins with removable His<sub>6</sub>-tag, the corresponding coding regions were cloned into the pLEICS-01 and pLEICS-02 expression vectors using LIC-PCR method (54). Ligation-independent cloning (also known as enzyme-free cloning) increases both the speed and efficiency of cloning of PCR products. LIC-PCR also eliminates the need to ligate PCR products to a vector, and does not rely on restriction sites (54). In this method, overlapping sequences are designed into the vector and PCR primers used to amplify the target DNA. Controlled digestion of the PCR product and the vector with a 3'-5' exonuclease such as exonuclease III is then used to create complementary protruding 3' ends. When the insert and the vector are mixed, the PCR products are annealed to the prepared vector. The chimeric molecule is then used to transform *E. coli*, and the ligase of the bacterial cells finally seals the single-stranded nicks and generates a covalently closed circular molecule (54).

#### *3.2.1.1. Amplification of Rv0288 and Rv0287 coding regions*

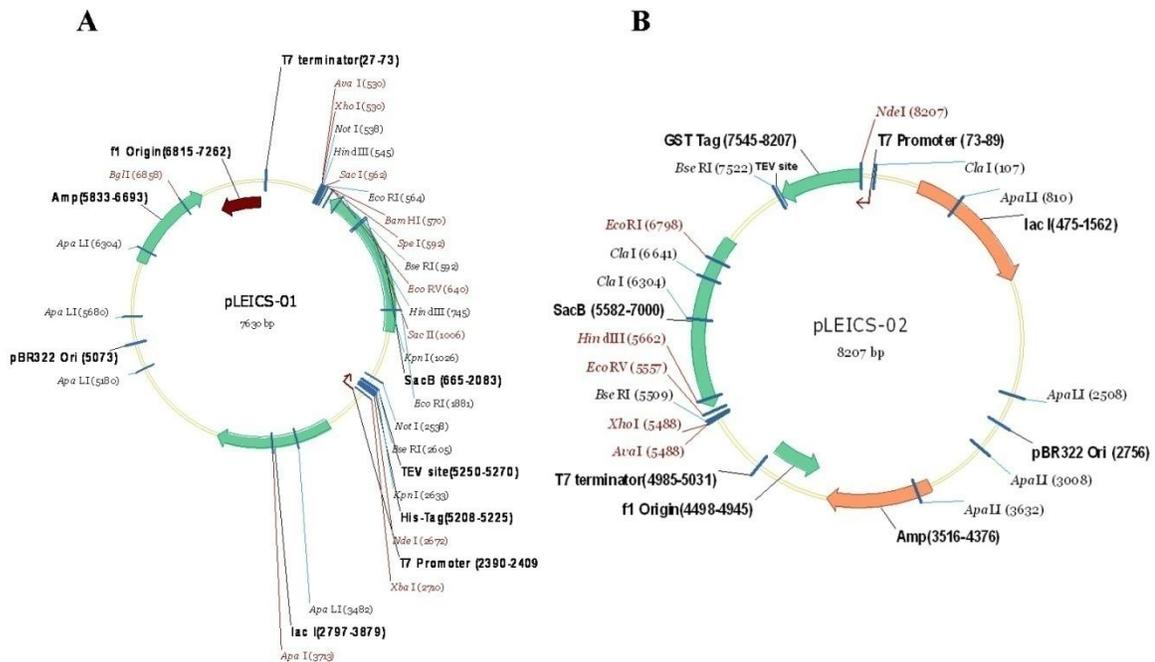
The coding regions were amplified by PCR from pET28a (+)-*Rv0288* and pET28a (+)-*Rv0287* constructs as DNA templates using Pfu (promega) and Pfx DNA polymerases (Invitrogen), and the following primers:

<b>Primer names</b>	<b>The sequences (5'- 3')</b>
<i>Rv0288</i> forward	<u>TAC TTC CAA TCC</u> ATG TCG CAA ATC ATG TCA AAC
<i>Rv0288</i> reverse	<u>TAT CCA CCT TTA CTG</u> CTA GCC GCC CCA TTT GGC GGC
<i>Rv0287</i> forward	<u>TAC TTC CAA TCC</u> ATG AGC CTT TTG GAT GCT CAT
<i>Rv0287</i> reverse	<u>TAT CCA CCT TTA CTG</u> TCA GAA CCC GGT ATA GGT CGA

Each primer contains a 5' extension (the underlined fragments) which is recognized by a 3'-5' exonuclease (discussed above). The PCR reactions were performed according to the manufacturer's protocols, using the TECHNE Thermal Cycling System (TECHGENE). The PCR products were finally identified by electrophoresis using 1 % (w/v) agarose gels (Section 2.2.1), and purified using the QIAquick PCR purification kit.

### *3.2.1.2. Ligation independent cloning (LIC-PCR) of the Rv0288 and Rv0287 coding regions*

pLEICS-01 and pLEICS-02 expression vectors (Figure 3.2.1 panels A and B) ( received from our collaborators at University of Leicester) were digested by the restriction enzyme *Bse*RI to remove the inserted lethal *SacB* gene (the *Bacillus subtilis sacB* gene encodes levansucrase that induces lethality upon exposure to 5% (wt/vol) sucrose in the growth medium) (93). The freeze-dried 3'-5' exonuclease supplied by In-fusion™ PCR cloning Kit (Clontech) was then thoroughly dissolved in 5 µl dH<sub>2</sub>O. 2 µl of dissolved enzyme, the linearized vector and the purified PCR product were mixed together and the mixture was finally subjected to a two-step incubation procedure at 30 °C and room temperature respectively.



**Figure 3.2.1: Schematic maps of pLEICS vectors. Panels A and B show the schematic maps of pLEICS-01 and pLEIC-02 expression vectors which are used to produce the proteins with N-terminal His<sub>6</sub>-tag and N-terminal GST-tag respectively. The vectors carry TEV protease cleavage site to remove the tags from the expressed proteins. The vectors also contain the T7 promoter and T7 terminator sequences as highlighted in the panels. Both vectors contain a bacterial origin of replication which is indicated by “f1 origin”. The vectors also contain the ampicillin resistance gene (indicated by “Amp”) and the lethal SacB gene (indicated by “SacB”) for the selection of transformed bacteria. As highlighted in the panels, both vectors contain two *BseRI* restriction sites which allow the removal of the lethal SacB gene.**

### 3.2.1.3. Transformation of *DH5-α* competent cells with the recombinant plasmids

50 µl of *E. coli* DH5- $\alpha$  competent cell were chemically transformed with the constructs (Section 3.2.1.2) and the transformants were incubated at 37 °C for 30 minutes prior to plating out on LB agar plates supplemented with 5 % (w/v) sucrose and 100 µg/ml ampicillin. 5 % (w/v) sucrose and 100 µg/ml ampicillin were used as markers for the selection of transformants (Section 3.2.1.2).

#### 3.2.1.4. Colony PCR screening

This method was used to rapidly screen the transformants (from the transformation step as described in the section 3.2.1.3) for the plasmid inserts (*Rv0287* and *Rv0287*) by analytical PCR done directly on the transformed bacteria. In fact, five colonies were transferred separately into five PCR tubes containing 20 µl of sterile water. In order to trace out the plasmids contained in the positive transformants, 5 µl of each inoculated sample was used to inoculate 10 ml LB broth containing 100 µg/ml ampicillin. The 10 ml inoculated cultures were then incubated at 37 °C/ 200 rpm for overnight to be subjected for miniprep (Promega). The remaining samples (15 µl) were then used for the analytical PCR. The analytical PCR were performed using T7 promoter and T7 terminator primers (the oligonucleotide primers specific for the plasmids) and pGEX5' and pGEX3' primers (the oligonucleotide primers specific for the pLEICS-02 plasmid) (Figure 3.2.1). The PCR products that yielded a single band of the right size were then identified by running the PCR products on 1 % (w/v) agarose gels (Section 2.2.1). The positive constructs identified by the colony PCR screening were finally checked by sequencing which is performed at PNAACL, University of Leicester.

#### 3.2.2. Subcloning of *Rv0287* coding region into pET23a (+) expression vector

PET-23a (+) and pET-28a (+)-*Rv0287* were sequentially dual restricted by *NdeI* and *BamHI* (promega). The linearized pET23a (+) and dual digested *Rv0287* were then gel-purified and recovered using the QIAquick Gel Extraction Kit (QIAGEN). Ligation reactions were carried out with a >4:1 ratio of the insert (the restricted *Rv0287*) over the digested pET23a (+) vector using T4 DNA ligase (Promega). The ligation mixture was

used to transform 50µl of competent *E. coli* DH5α cells. Positive transformants were then screened for successful transformation by growth on plates of LB agar supplemented with 100 µg/ml ampicillin. The selected transformants were then subjected to miniprep and the isolated plasmids were then analysed for successful ligation of insert by visualising restriction digests (*NdeI/BamHI*) of the plasmids on a 1% agarose gel (Section 2.2.1). The integrity of the construct was finally confirmed by sequencing at PNACL (University of Leicester). The right construct was finally transformed in to BL21-codon plus- RIL competent cells.

### 3.2.3. Protein Expression Trials

Time course expression trials and an examination of the solubility of the expressed proteins were carried out as described in the section 2.2.7.

### 3.2.4. Purification Rv0287 and Rv0288 proteins

The full-length coding regions for Rv0287 (*esxG*) and Rv0288 (*esxH*) were cloned into the pET23a (Novagen) and the pLEICS-01 expression vectors as described in the sections 3.2.1 and 3.2.2. The pLEICS-01 generated the construct encoding an N-terminal His<sub>6</sub>-tag and TEV protease cleavage site (ENLYFQSM) followed by the Rv0288 coding sequence. Expression was carried out in *E. coli* BL21 (DE3) cells at 37 °C, which were induced for 4 hours by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) (1 mM) at an optical density of 0.6 at 600 nm (the mid-log phase of bacterial growth). Under these expression conditions both Rv0287 and Rv0288 were found to be insoluble and were initially isolated as inclusion bodies. Rv0287 and Rv0288 were co-refolded by solubilizing the inclusion

body pellets in a 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 6 M guanidine HCl buffer (pH 7.4) at a protein concentration of 0.5 mg/ml, followed by dialysis against the same buffer without the denaturant at 4 °C. The soluble protein complex obtained was purified by nickel affinity chromatography. The Rv0288-Rv0287 complex was eluted from the column in the 20 mM Tris-HCl buffer containing 172.5 mM imidazole. FPLC fractions containing the fusion protein were pooled, and then treated with TEV protease. The His<sub>6</sub>-tag was then removed by cleavage with the His<sub>6</sub>-tagged TEV Protease and the two products (the tag and the Rv0287-Rv0288 complex) separated by a second Ni<sup>2+</sup>-NTA affinity chromatography step. The sample was finally subjected to a polishing purification step by gel filtration chromatography on a Superdex 75 16/60 pre-packed column (Amersham Biosciences). The purified protein contained in 100 mM sodium chloride, 25 mM sodium phosphate and 0.02 % NaN<sub>3</sub>, at pH 6.5 was judged to be greater than 95% pure by SDS-PAGE (Invitrogen 4-12% Bis-Tris NuPAGE gel system).

### 3.2.5. Expression and purification of His<sub>6</sub>-tagged Rv0288 and His<sub>6</sub>-tagged Rv0287 proteins

The pET28a (+) expression vectors carrying the Rv0287 and Rv0288 coding regions (supplied by Dr. Kirsty L. Lightbody at University of Leicester) were used to express full-length proteins with an N-terminal His<sub>6</sub>-tag in BL21 (DE3)-Codon plus-RIL cells. Both His<sub>6</sub>-Rv0287 and His<sub>6</sub>-Rv0288 were purified from inclusion bodies, which were initially washed four times in 50 mM Tris, 10 mM EDTA, and 0.5 % Triton X-100 buffer at pH 8.0 and then resolubilized in 25 mM sodium phosphate, 6 M guanidine hydrochloride (GdnHCl) buffer at pH 7.4. The protein samples were dialyzed against 25 mM NaH<sub>2</sub>PO<sub>4</sub>,

200 mM NaCl, 30 mM imidazole buffer at pH 7.4 and then purified by Ni<sup>2+</sup>-NTA affinity chromatography. The purified proteins were mixed together prior to a final gel filtration chromatography step on a 120-ml Superdex 75 16/60 pre-packed column (GE Healthcare).

### 3.2.6. Preparation of ESAT-6-CFP-10 complex

Lyophilized ESAT-6-CFP-10 protein sample (provided by Dr. Philip S. Renshaw at University of Leicester) was resolubilized in a buffer containing 0.02 % NaN<sub>3</sub>, 100 mM NaCl, and 25 mM NaH<sub>2</sub>PO<sub>4</sub> at pH 6.5.

### 3.2.7. Preparation of the single- and double-labelled Rv0287 and Rv0288 proteins

Uniformly <sup>15</sup>N, <sup>13</sup>C, and <sup>15</sup>N/<sup>13</sup>C labeled Rv0287 and Rv0288 were prepared from pET23a (+)- and pLEIC-01 based *E. coli* expression vectors respectively, which were grown in minimal media supplemented with 1 g/l (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and/or 2 g/l <sup>13</sup>C D-Glucose as the sole nitrogen and carbon sources. <sup>13</sup>C labelled protein samples used for <sup>13</sup>C-edited NOESY experiments (103) were prepared as described above, but with the addition of non-isotopically labeled aromatic amino acids to the minimal media (His, Phe, Trp and Tyr at 50 mg/l). The recombinant proteins were then purified as complexes of labelled/unlabelled proteins produced as described in section 3.2.4.

### 3.2.8. Protein concentration

The purified protein samples were concentrated by ultrafiltration and lyophilizing (Section 2.2.10) to give 0.35 ml NMR samples of 0.7–1mM Rv0288-Rv0287.

### 3.2.9. Circular dichroism and fluorescence spectroscopy

CD and fluorescence analyses were carried out as described in the sections 2.2.11, 2.2.12, and 2.2.13.

### 3.2.10. NMR spectroscopy

NMR spectra were acquired from 0.35 ml labelled samples of 0.7-1.0 mM Rv0288-Rv0287 in a 100 mM NaCl and 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.02 % (W/V) NaN<sub>3</sub>, and 0.1 mM AEBSF buffer at pH 6.5 containing either 10 % D<sub>2</sub>O / 90 % H<sub>2</sub>O, or 100 % D<sub>2</sub>O as appropriate. All NMR data were acquired at 35 °C on 600 MHz Bruker Avance/DRX spectrometers or 800-MHz Bruker Avance system. The two-dimensional (2D) and three-dimensional (3D) spectra recorded to obtain sequence specific assignments for Rv0287 and Rv0288 in the Rv0288-Rv0287 complex as follows: <sup>1</sup>H-<sup>1</sup>H TOCSY with a mixing time of 30 and 50 ms (95), <sup>1</sup>H-<sup>1</sup>H NOESY with an NOE mixing time of 85 ms (96). <sup>15</sup>N/<sup>1</sup>H HSQC (97); NOESY-HSQC with an NOE mixing time of 85 ms (98, 103); TOCSY-HSQC with a mixing time of 45 ms (99); <sup>13</sup>C/<sup>1</sup>H HSQC; HCCH-TOCSY with a mixing time of 12 ms(100); <sup>13</sup>C-edited NOESY with an NOE mixing time of 85 ms (103) and <sup>13</sup>C/<sup>15</sup>N/<sup>1</sup>H HNCACB (101), CBCA(CO)NH (102), TROSY HN(CO)CA (103), and HNCA(103). Typical acquisition times in F<sub>1</sub> and F<sub>2</sub> for three-dimensional experiments were 20 ms for <sup>15</sup>N, 8.5 ms for <sup>13</sup>C, and 15 ms for <sup>1</sup>H (indirect proton dimension) and an acquisition time of

55 ms in  $F_3$  ( $^1\text{H}$ ). The majority of the 3D spectra were collected over about 88 hours, 2D  $^1\text{H}$  experiments over 24 hours,  $^{15}\text{N}/^1\text{H}$  and  $^{13}\text{C}/^1\text{H}$  HSQC over 30 minutes. Typical acquisition times in 2D experiments were either 60 ms ( $^{15}\text{N}$ ), 10 ms ( $^{13}\text{C}$ ) or 48-75 ms ( $^1\text{H}$ ) in  $F_1$  and 77-85 ms in  $F_2$  ( $^1\text{H}$ ). The WATERGATE method (104) was employed to suppress the water signal as required. The 3D NMR data were processed using Topspin with linear prediction used to extend the effective acquisition times by up to 1.5-2 fold in  $F_1$  and  $F_2$ . The spectra were analysed using the Sparky package (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco).

### 3.2.11. Sequence specific assignments

Sequence-specific backbone resonance assignments (N, NH,  $\text{C}\alpha$  and  $\text{C}\beta$ ) were obtained for the Rv0288-Rv0287 complex from the identification of intra and inter-residue connectivities in HNCACB (101), CBCA(CO)NH (102), TROSY HN(CO)CA (103), HNCA (103) and  $^{15}\text{N}/^1\text{H}$  NOESY-HSQC (98) spectra. Initially, intra and inter-residue amide nitrogen/proton to  $\text{C}\alpha$  and  $\text{C}\beta$  peaks were identified in the triple resonance spectra and used to search for signals from neighboring residues in the sequence. The identification of signals from adjacent residues was confirmed wherever possible by the observation of NH to NH NOEs in the NOESY-HSQC spectra (98).

Assignments were then extended to the side chain signals using correlations observed primarily in  $^{15}\text{N}/^1\text{H}$  TOCSY-HSQC (99) and  $^{13}\text{C}/^1\text{H}$  HCCH-TOCSY (100), with additional supporting evidence provided in some cases by  $^{15}\text{N}/^1\text{H}$  NOESY-HSQC (98) and  $^{13}\text{C}$ -edited

NOESY(103) spectra.  $^1\text{H}/^1\text{H}$  TOCSY (95) and  $^1\text{H}/^1\text{H}$  NOESY (96) spectra were used to assign aromatic side chain protons.

### 3.2.12. Secondary structure determination

The Chemical Shift Index (CSI) method was used to determine the secondary structure of Rv0287 and Rv0287 in the complex on the basis of their  $\text{C}\alpha$  chemical shifts (111). In addition, the pattern of sequential and medium range NOEs involving backbone amide signals, observed in the  $^{15}\text{N}/^1\text{H}$  NOESY-HSQC spectrum, was used to confirm the locations of helical regions (109).

### 3.2.13. Determination of slowly exchanging amide protons

In order to identify amide protons which exchange with solvent at very low rates (shielded from the solvent),  $^{15}\text{N}/^1\text{H}$  HSQC was carried out on a 0.7 – 1 mM complex containing either  $^{15}\text{N}$ -labelled Rv0288 or  $^{15}\text{N}$ -labelled Rv0287 as described in section 3.2.10. The protein samples were prepared in 100 %  $\text{D}_2\text{O}$  (100 mM NaCl and 25 mM  $\text{NaH}_2\text{PO}_4$ , 0.02 % (W/V)  $\text{NaN}_3$ , and 0.1 mM AEBSF, pH 6.5) and incubated overnight at room temperature prior to the experiment (one  $^{15}\text{N}/^1\text{H}$  HSQC spectrum was collected for each protein sample after 24 hour incubation).

### 3.2.14. Structural calculations

The family of converged Rv0288-Rv0287 structures was determined in a two stage process using the program CYANA (105). Initially, the combined automated NOE assignment and structure determination protocol (CANDID) (106) was used to automatically assign the NOE cross peaks identified in three dimensional  $^{15}\text{N}$ - and  $^{13}\text{C}$ -edited NOESY spectra to produce preliminary structures of the protein. After which several cycles of simulated annealing combined with redundant dihedral angle constraints (REDAC) were used to produce the final converged Rv0288-Rv0287 structures (107).

The input for the CANDID stage comprised of essentially complete  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  resonance assignments for the non-exchangeable groups in the Rv0288-Rv0287 complex, four manually picked three-dimensional NOE peak lists corresponding to all NOEs involving amide protons (1538) and all NOEs between aliphatic protons (2222), and two manually picked two-dimensional NOE peak list corresponding to all NOEs involving aromatic side chain protons (127). The CANDID stage also included 264 torsion angles constraints for the complex determined by the protein backbone dihedral angle prediction program TALOS (108). The observed slowly exchanging amide protons (Section 3.2.13) were used to identify hydrogen bonds contained in the complex (discussed in section 3.3.7.3). 24 hydrogen bond constraints obtained from the six identified H-bonds were also included in the CANDID stage. CANDID calculations were carried out using the default parameter settings in CYANA with chemical shift tolerances set to 0.04 ppm (direct and indirect  $^1\text{H}$ ) and 0.4 ppm ( $^{15}\text{N}$  and  $^{13}\text{C}$ ).

The final converged Rv0288-Rv0287 structures were produced from 100 random starting coordinates using a standard torsion angle-based simulated annealing protocol combined with 5 cycles of redundant dihedral angle constraints (REDAC) (107). The calculations were based upon 2323 non-redundant, NOE-derived upper distance limits assigned to unique pairs of protons using CANDID, 24 hydrogen bond constraints, and 264  $\Phi$  and  $\Psi$  torsion angle constraints derived from TALOS. Analysis of the family of structures obtained was carried out using the programs CYANA and MOLMOL (105, 110).

### 3.2.15. Comparative analysis between Rv0288-Rv0287 and its structurally homologue proteins

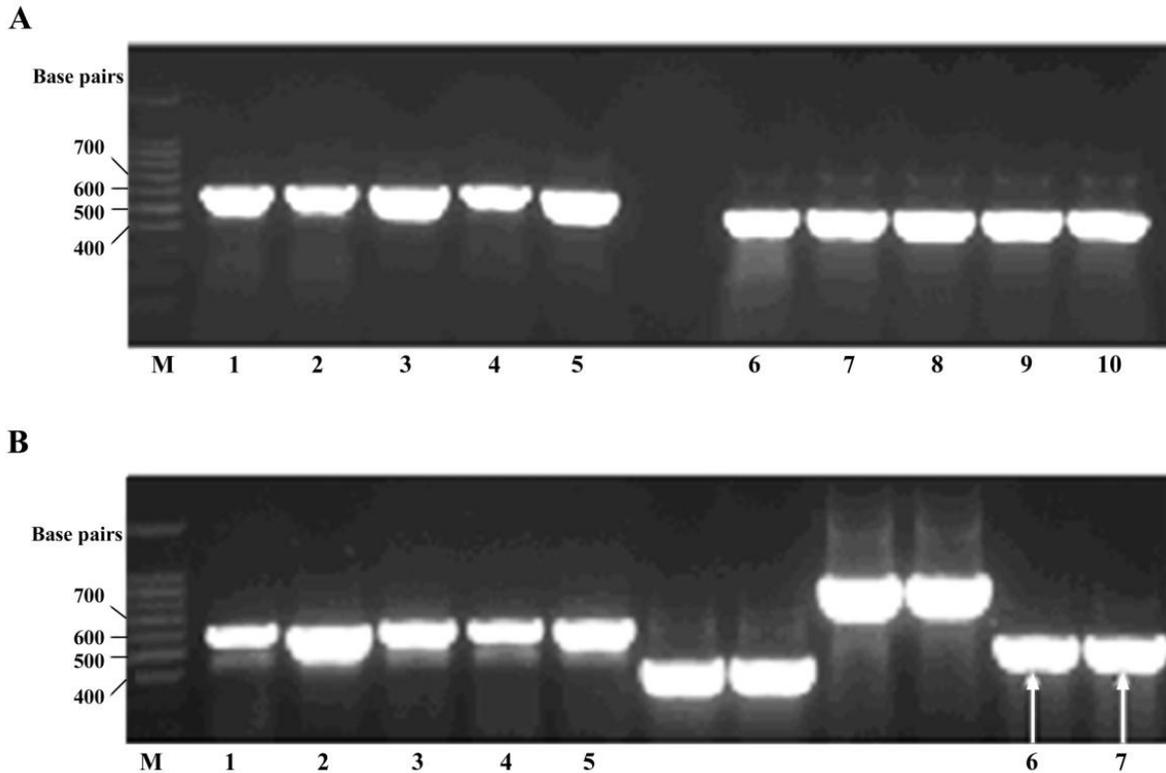
A network tool (DALI) was used for an optimal protein structure comparison. DALI (Distance matrix alIgnment) is an algorithm for optimal structure alignment using distance matrices (117). This algorithm has three basic steps. In the first step, 3D structures of proteins are described with distance matrices which stores all intramolecular distances between the C $\alpha$  atoms. Similar sub-matrices of size six (hexapeptides) in the proteins are then found by the comparison of their distance matrices. The comparison results in alignments of the hexapeptides between the proteins. The compatible alignments are finally merged to obtain larger alignments (117).

### 3.3. Results and discussion- Part2

#### 3.3.1. Ligation independent cloning (ILC-PCR) of Rv0287 and Rv0288 coding regions

The full-length coding regions for Rv0287 (291 bp) and Rv0288 (288 bp) were PCR amplified from the pET28a (+)-*Rv0288* and pET28a (+)-*Rv0287* constructs as DNA templates (Section 3.2.1.1). The amplicons were purified and then cloned into pLEICS-01 and pLEICS-02 expression vectors using the ligation independent cloning method (LIC-PCR) as described in section 3.2.1.2. The constructs were used to transform DH5- $\alpha$  competent cells (Section 3.2.1.3).

Colony PCR screening was carried out to spot the transformants carrying the right constructs as detailed in Section 3.2.1.4. The PCR products that yielded a single band of the right size were then identified by running the PCR products on 1 % (w/v) agarose gels as shown in figure 3.3.1. The expected bands for *Rv0287* and *Rv0288* are 571 and 568 bp when T7 promoter and T7 terminator primers used for the colony PCR screening of the transformants carrying the pLEICS-01 constructs. The expected bands for *Rv0287* and *Rv0288* are also 465 and 462 bp when the pGEX5' and pGEX3' primers used for the colony PCR screening of the transformants carrying the pLEICS-02 constructs. Integrity of the constructs was finally confirmed by sequencing at PNAAC, University of Leicester.

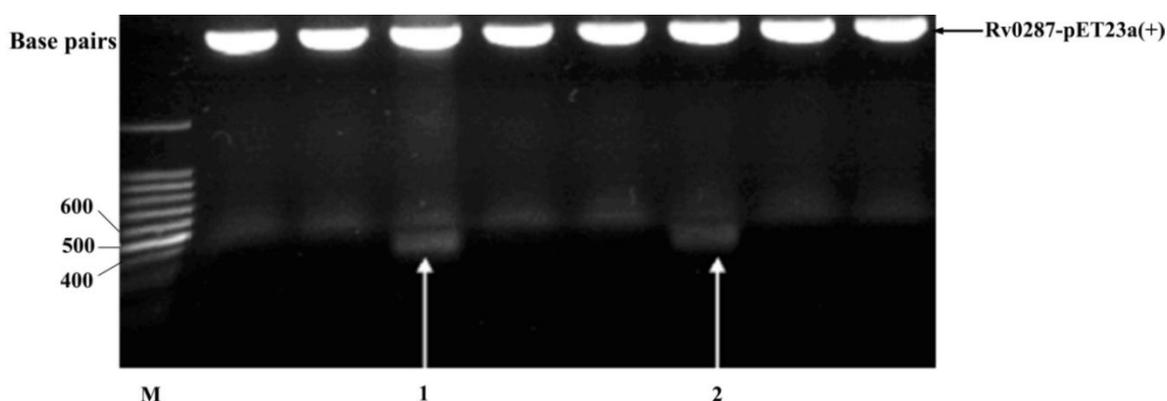


**Figure 3.3.1: Colony PCR screening of transformants.** Gel A shows the results of colony PCR screening for the transformants carrying of the right *Rv0288*-pLEICS-01 (lanes 1-5) and *Rv0288*-pLIECS-02 constructs (lane 6-10). Gel B illustrates the results of colony PCR screening for the transformed DH5- $\alpha$  cells containing of the right *Rv0287*-pLEICS-01 (lanes 1-5) and *Rv0287*-pLIECS-02 constructs (lanes 6 and 7). Lane M in (A) and (B) contain 100 bp DNA markers.

### 3.3.2. Subcloning of the coding *Rv0287* region into pET23a (+)

In order to produce *Rv0287* protein without any tag, the *Rv0287* coding region was cut (*NdeI/BamHI*) from the *Rv0287*-pET28a (+) construct and then inserted into the pET23a (+) expression vector (Section 3.2.2). As shown in figure 3.3.2, the constructs carrying the *Rv0287* coding region were confirmed by dual (*NdeI/BamHI*) restriction of the constructs followed by visualising the presence of the insert in a 1 % agarose gel. The sequence and

orientation of the insert was also checked by sequencing at PNACL, University of Leicester.



**Figure 3.3.2: Double-digestion of the *Rv0287*-pET23a (+) constructs. The 1 % agarose gel illustrates successfully ligated constructs. Lane 1 and 2 correspond to the successful ligated pET23a-(+) containing the *Rv0287* coding region. Lane M contains 100 bp DNA markers.**

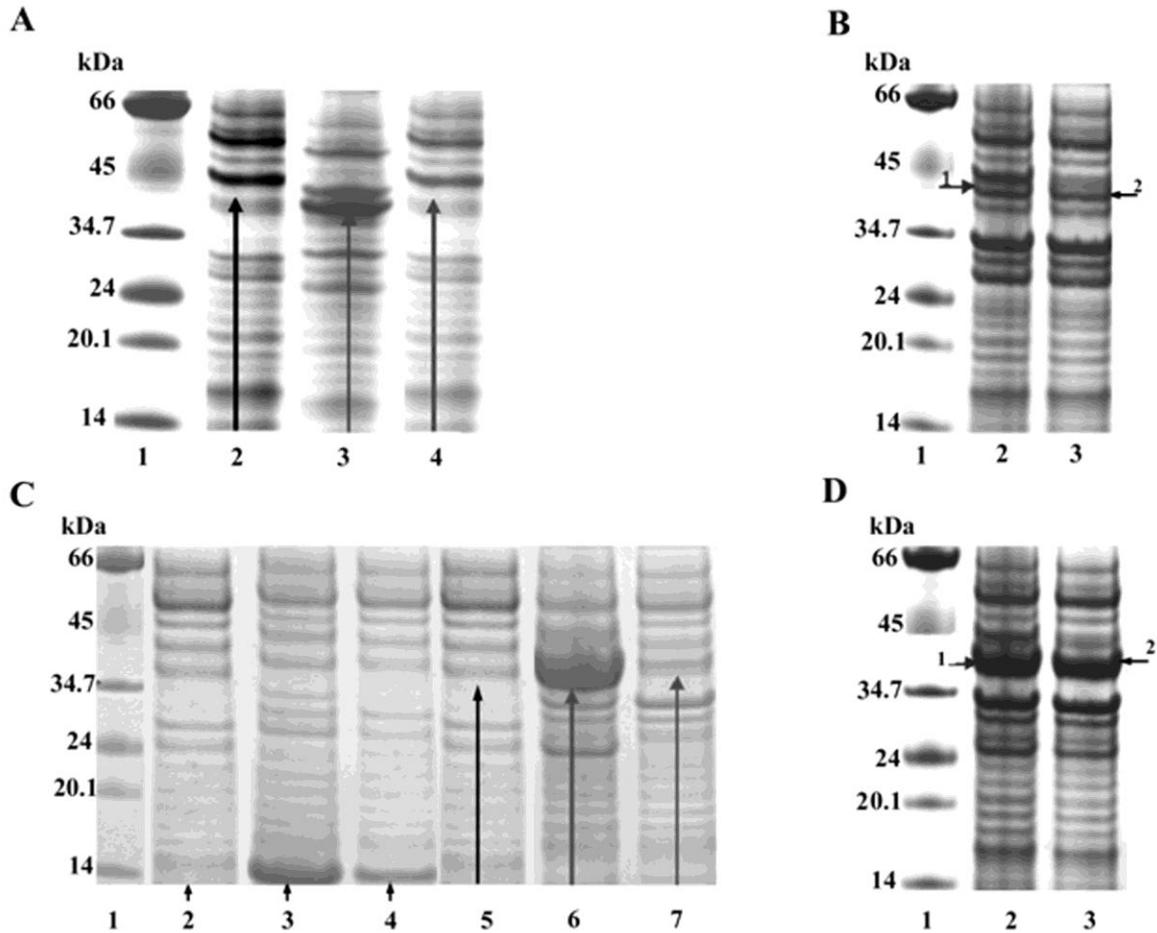
### 3.3.3. Protein expression trials for the recombinant *Rv0287* and *Rv0288* with removable His<sub>6</sub> and GST tags

The protein expression trials were carried out to evaluate the level of expression and also the solubility of the recombinant proteins with removable His<sub>6</sub>-tag and GST-tag (Section 2.2.7).

These analyses showed that with exception of His<sub>6</sub>- *Rv0287*, the other tagged proteins (His<sub>6</sub>-*Rv0288*, GST-*Rv0287*, and GST-*Rv0288*) are expressed in the *E. coli*-based expression system. As shown in figure 3.3.3, in extracts of cells expressing GST-*Rv0287* (35.778 kDa) (Lane 3 in panel A), GST-*Rv0288* (36.390 kDa) (Lane 6 in panel C) and His<sub>6</sub>-*Rv0288* (12.390 kDa) (Lane 3 in panel C), a protein of the predicted molecular weight is visible 4 hours after induction by 1 mM IPTG at 37 °C.

Analyses of the solubility of the expressed protein also indicated that the products can be only detected in the whole cell post-induction fractions (Lane 3 in panel A, and lanes 3 and 6 in panel C), but not in the supernatant fractions (Lane 4 in panel A, and lanes 4 and 7 in panel C), suggesting that the target proteins are expressed as inclusion bodies at 37 °C.

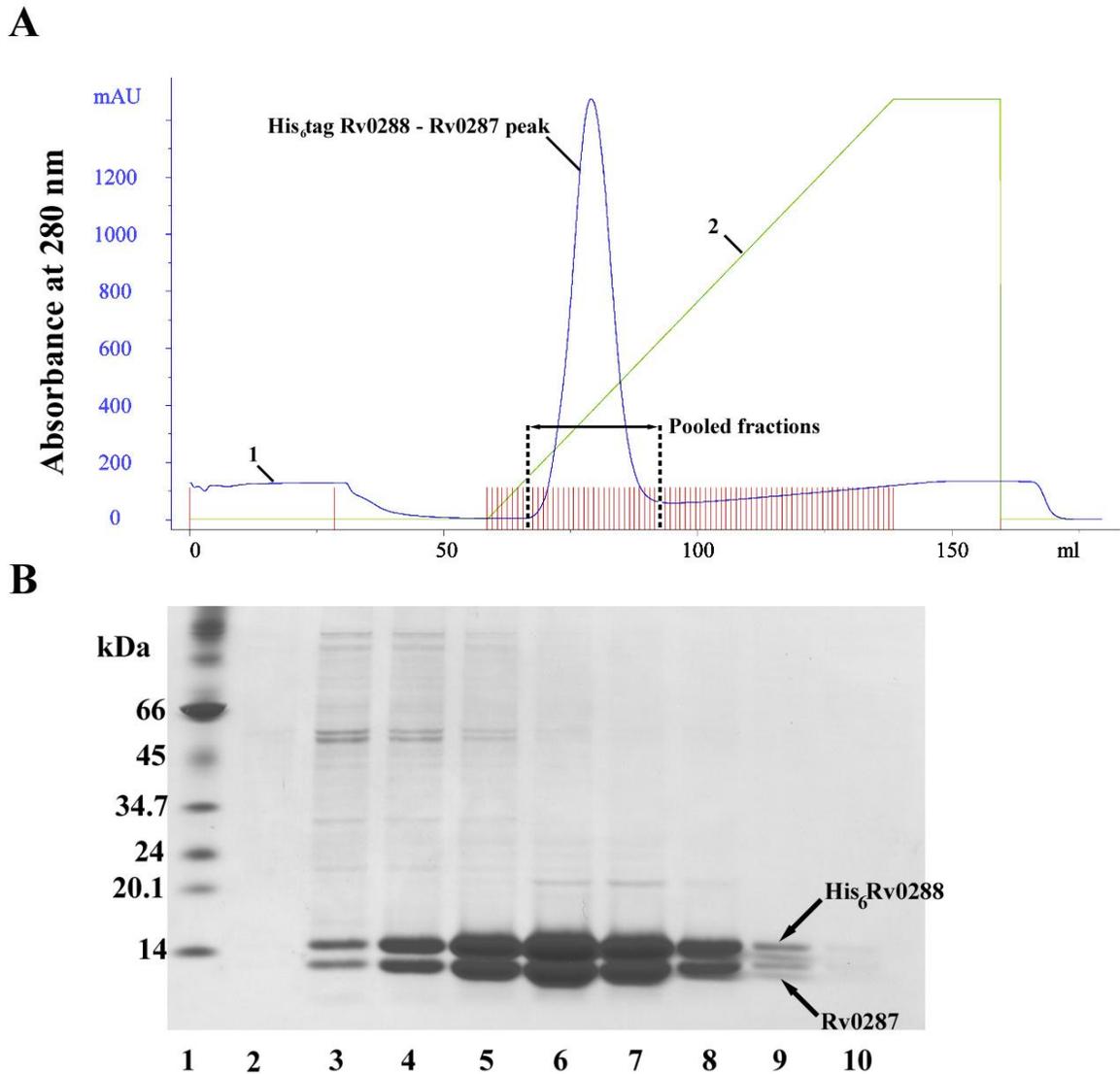
Further expression analyses under different conditions showed that soluble GST-Rv0288 and GST-Rv0287 can be produced provided that transformants expressing GST-Rv0288 and GST-Rv0287 are induced by 0.45 mM and 1 mM IPTG respectively at mid log phase of growth ( $OD_{600\text{ nm}} : 0.6$ ) and incubated at 15 °C for overnight (Figure 3.3.3 panels B and D).



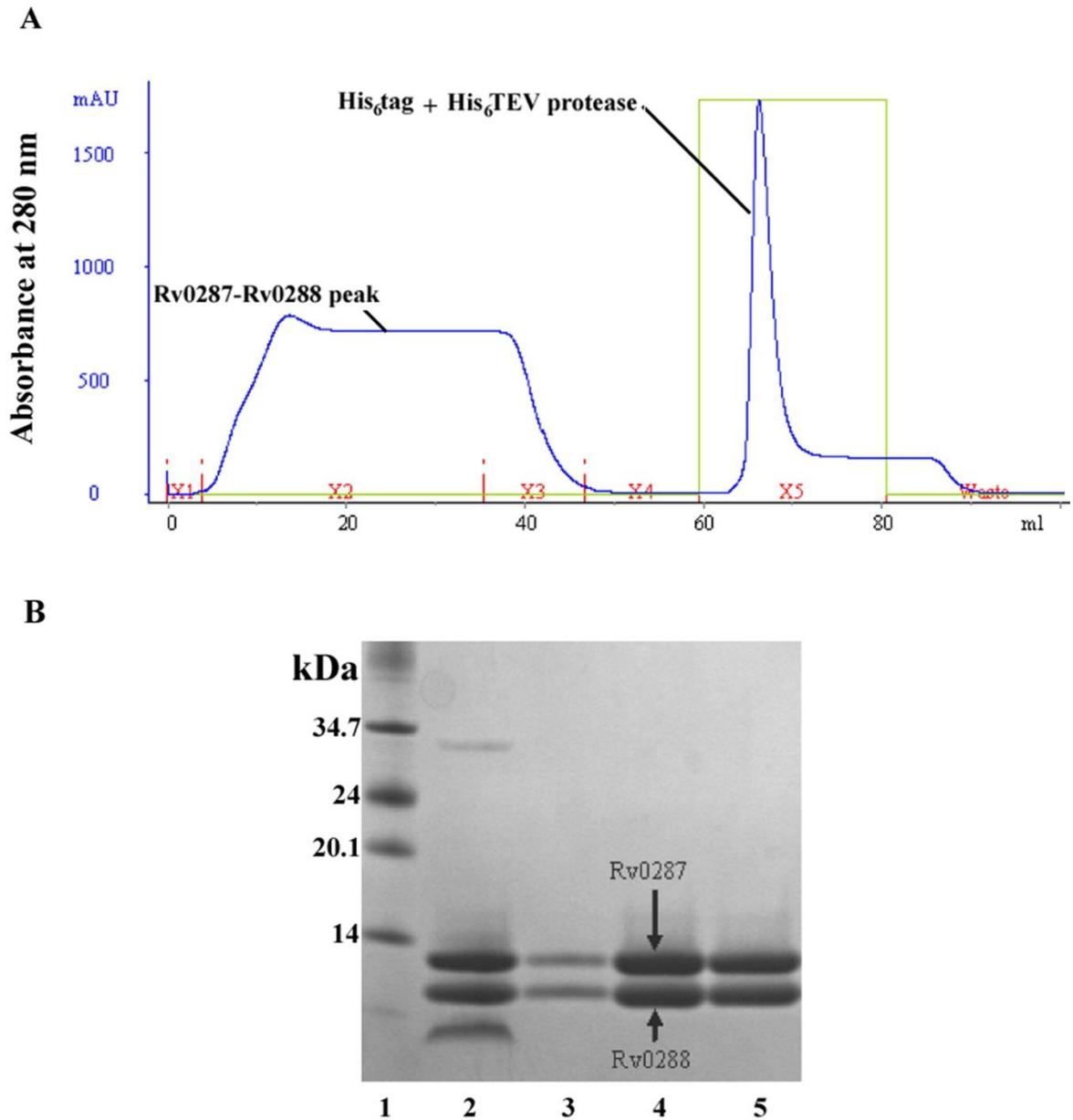
**Figure 3.3.3: SDS-PAGE gels of expression trails of the recombinant Rv0287 and Rv0288.** Gel (A) shows the expression trials of the GST-Rv0287 in BL21 (DE3) *E. coli* cells. Lanes 1, 2, 3, 4 contain a high molecular weight protein marker, the pre-induction cell lysate fraction, four hour post-induction fraction (whole cell lysate) and GST-Rv0287 supernatant fraction. The position of the GST-Rv0287 band on the gel is shown by an arrow on lane 3 (panel A). Gel (C) shows the expression trials of His<sub>6</sub>-Rv0288 and GST-Rv0288 in BL21 (DE3) *E. coli* cells. Lanes 1, 2, 3, 4, 5, 6, and 7 contain a high molecular weight protein marker, the pre-induction cell lysate fraction, four hour post-induction fraction (whole cell lysate containing His<sub>6</sub>-Rv0288), His<sub>6</sub>-Rv0288 supernatant fraction, pre-induction cell lysate fraction, four hour post-induction fraction (GST-Rv0288 whole cell lysate) and GST-Rv0288 supernatant fraction. The Position of the His<sub>6</sub>-Rv0288 and GST-Rv0288 bands on the gel are shown by the arrows on lanes 3 and 6 respectively (panel C). Panels (B) and (D) show analyses of solubility of the GST-tagged proteins when transformants expressing GST-Rv0287 (gel B) and GST-Rv0288 (gel D) are induced by 1 mM and 0.45 mM IPTG respectively at mid log phase of growth and incubated at 15 °C for overnight. Lane 1 in the gels (B) and (D) contain a high molecular weight protein marker. Lane 2 and 3 in gel (B) contain the whole cell lysate, and supernatant fractions of the transformants expressing GST-Rv0287. Lane 2 and 3 in gel (D) contain the whole cell lysate, and supernatant fractions of the transformants expressing GST-Rv0288. The arrows indicated by '1' and '2' in panel B show the protein bands corresponding to GST-Rv0287. The arrows indicated by '1' and '2' in panel D show the protein bands corresponding to GST-Rv0288.

### 3.3.4. Expression and purification of Rv0288-Rv0287 complex

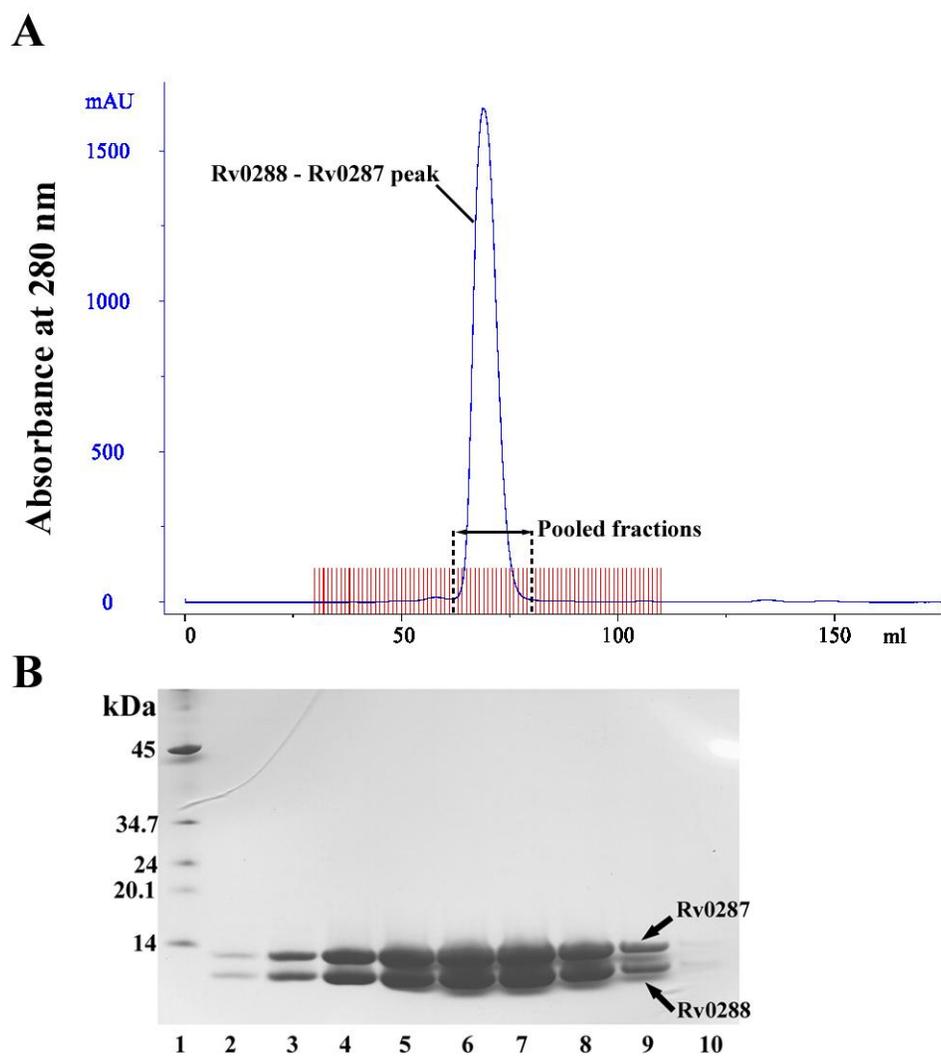
The pET23a (+)-*Rv0287* and pLEICS-01-*Rv0288* constructs were used to express Rv0287 and His<sub>6</sub>-Rv0288 in *E. coli* (Section 3.2.4). The target proteins were produced as insoluble products. The inclusion bodies were then resolubilized and co-refolded as detailed in the section 3.2.4. The N-terminal hexa-histidine tag of His<sub>6</sub>-Rv0288 allowed the His<sub>6</sub>-Rv0288-Rv0287 complex to be purified by affinity chromatography, as shown in figure 3.3.4. The N-terminal His<sub>6</sub> tag was then removed by cleavage with TEV protease, and the two products (His<sub>6</sub> tag and the Rv0287-Rv0288 complex) were separated by a second Ni<sup>2+</sup>-NTA affinity chromatography step (Figure 3.3.5). The sample was finally subjected to a polishing purification step by gel filtration chromatography as shown in figure 3.3.6. The final yield of the purified complex was determined to be approximately 1.7 mg/ml.



**Figure 3.3.4: Purification of the His<sub>6</sub>-Rv0288-Rv0287 complex.** Panel (A) shows a typical FPLC profile attained for the Ni-NTA column based purification of the complex. The complex was eluted over a linear gradient from 0 mM to 500 mM imidazole (indicated by “2”). A high molecular weight contaminant peak is indicated by “1” and the complex elution peak is labelled. The pooled fractions are also highlighted. Panel (B) shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-10 show samples removed from the complex elution peak. Fractions from lanes 3-9 were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).



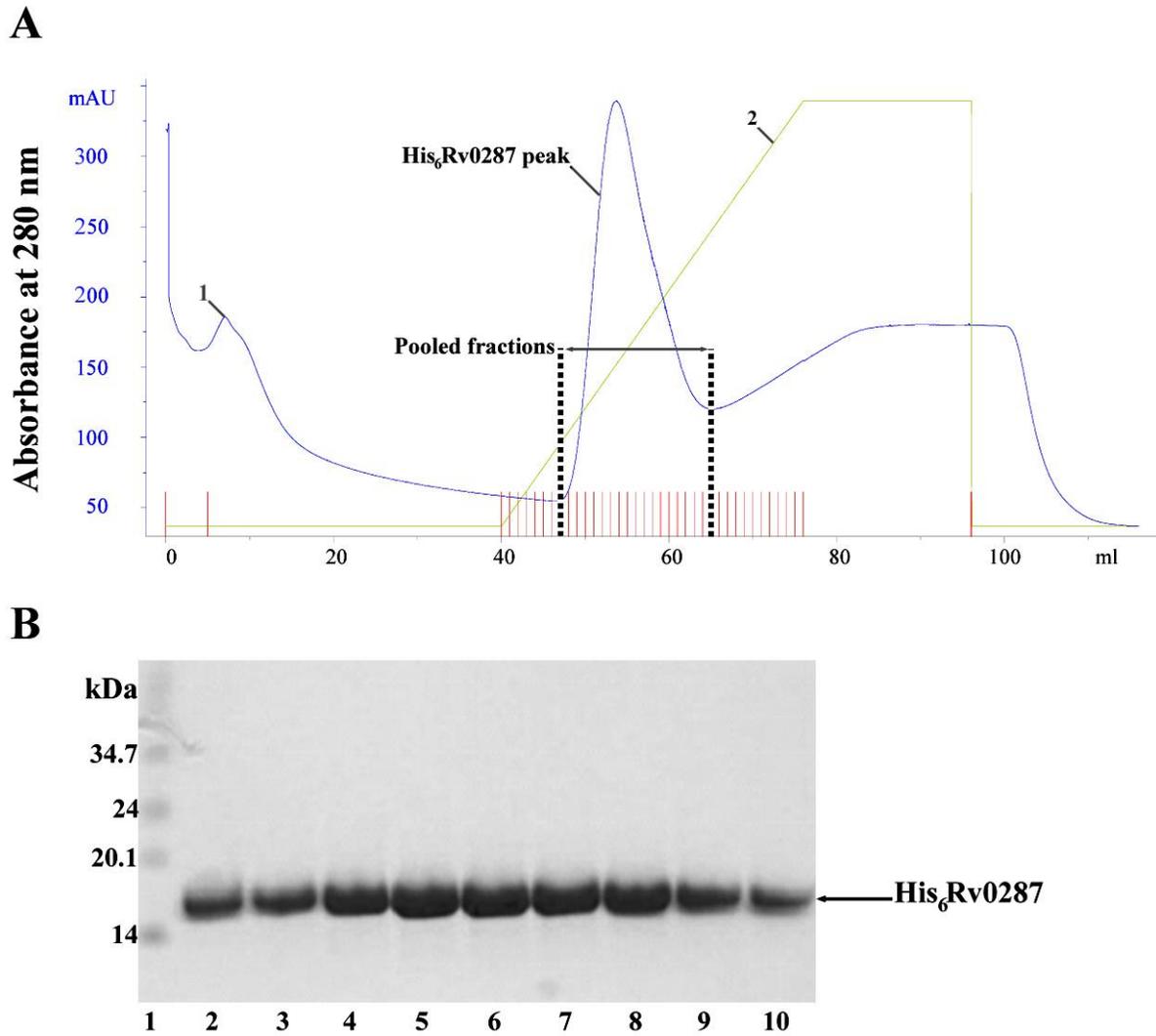
**Figure 3.3.5: Purification of the TEV-treated protein complex.** Panel (A) shows  $\text{Ni}^{2+}$ -NTA elution profile of the TEV treated protein sample. The untagged protein complex was collected in fractions X1, X2 and X3. The X5 fraction contains  $\text{His}_6$ -tagged TEV protease and  $\text{His}_6$ -tag linker which are eluted from the column in the 20 mM Tris-HCl buffer containing 500 mM imidazole (the stepwise gradient of imidazole from 0 to 500 mM is shown in green). Panel (B) shows SDS PAGE gel of the fractions collected from the second step  $\text{Ni}^{2+}$ -NTA chromatography. Lanes 1, 2 contain high molecular weight marker, the loaded sample. The first band from the bottom in lane 2 contains the cut  $\text{His}_6$ -tag linker. Lanes 3, 4 and 5 contain the untagged purified Rv0288-Rv0287 complex.



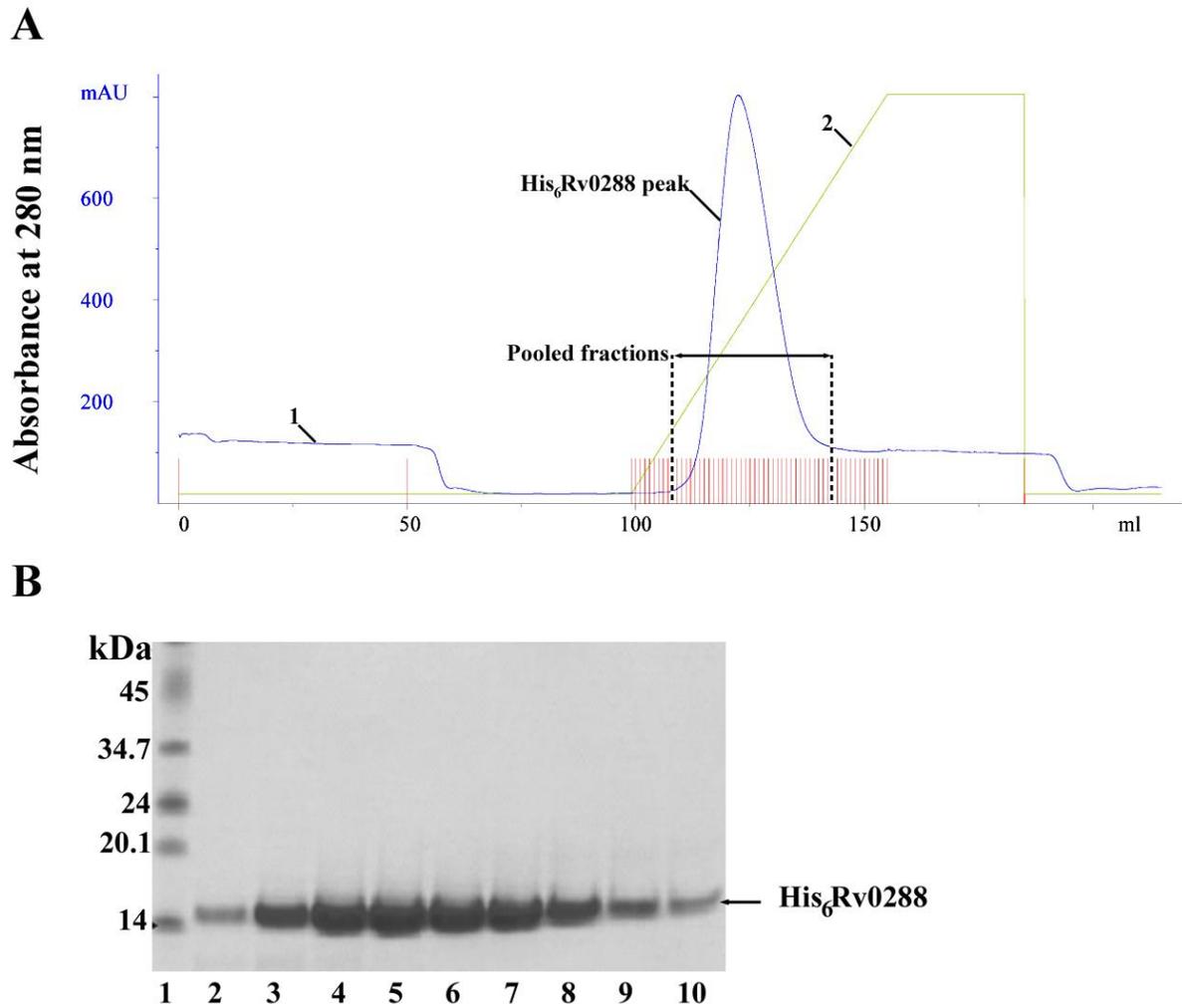
**Figure 3.3.6: Purification of the Rv0287-Rv0288 complex by gel filtration.** Panel (A) shows a typical FPLC elution profile attained by gel filtration on a Superdex 75 16/60 column. The Rv0287-Rv0288 complex elution peak is labelled. The pooled fractions are also highlighted. Panel (B) shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-10 show samples removed from the complex elution peak. Fractions from lanes 2-9 were pooled for structural studies. Lane 1 contains high molecular weight markers (Sigma). The apparent molecular weight of Rv0287 is about 10 kDa which very close to the expected molecular weight of the protein (9.778 kDa). The apparent molecular weight of Rv0288 is less than the expected molecular weight of the protein (10.390 kDa) which is typical of the ESAT-6 related proteins (40).

### 3.3.5. Expression and Purification of the His<sub>6</sub>-Rv0287-His<sub>6</sub>-Rv0288 complex proteins

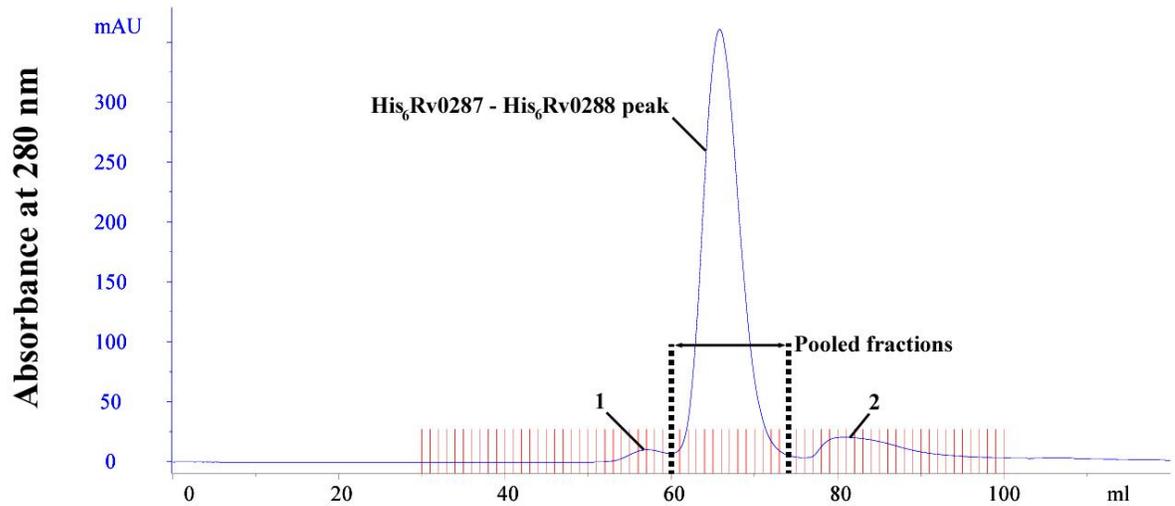
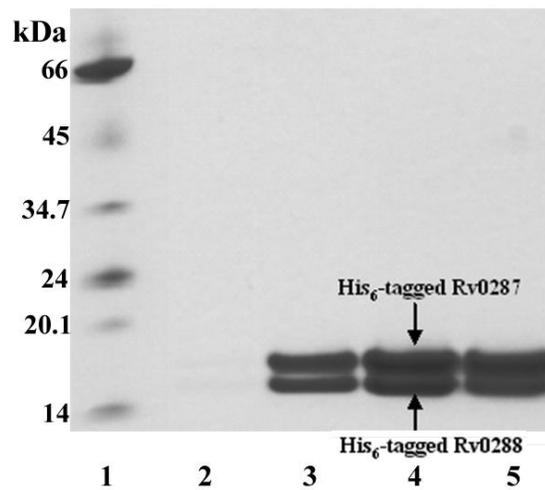
The pET28a (+) expression vectors carrying the Rv0287 and Rv0288 coding regions were used to express His<sub>6</sub>-Rv0287 and His<sub>6</sub>-Rv0288 in *E. coli* (Section 3.2.5). The target proteins were produced as insoluble products. The inclusion bodies were resolubilized and then subjected to two step chromatography as detailed in the section 3.2.5. The N-terminal hexa-histidine tag of the proteins allowed the proteins to be purified by Ni<sup>2+</sup>-NTA affinity chromatography, as shown in figures 3.3.7 and 3.3.8. The pooled protein fractions obtained from the affinity chromatography (His<sub>6</sub>-Rv0287 and His<sub>6</sub>-Rv0288) were mixed and then the resulting protein sample was subjected to a final polishing purification step by gel filtration, as shown in figure 3.3.9. The final yield of the purified complex was determined to be approximately 0.2 mg/ml.



**Figure 3.3.7: Purification of His<sub>6</sub>-Rv0287.** Panel (A) shows a typical FPLC profile attained for the Ni<sup>2+</sup>-NTA column based purification of His<sub>6</sub>-Rv0287. The protein was eluted over a linear gradient from 0 mM to 500 mM imidazole (indicated by “2”). A high molecular weight contaminant peak is indicated by “1” and the protein elution peak is labelled. The pooled fractions are also highlighted. Panel (B) shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-10 show samples removed from the complex elution peak. These fractions were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).



**Figure 3.3.8: Purification of His<sub>6</sub>-Rv0288.** Panel (A) shows a typical FPLC profile attained for the Ni<sup>2+</sup>-NTA column based purification of His<sub>6</sub>-Rv0288. The protein was eluted over a linear gradient from 0 mM to 500 mM imidazole (indicated by “2”). A high molecular weight contaminant peak is indicated by “1” and the protein elution peak is labelled. The pooled fractions are also highlighted. Panel (B) shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-10 show samples removed from the complex elution peak. These fractions were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).

**A****B**

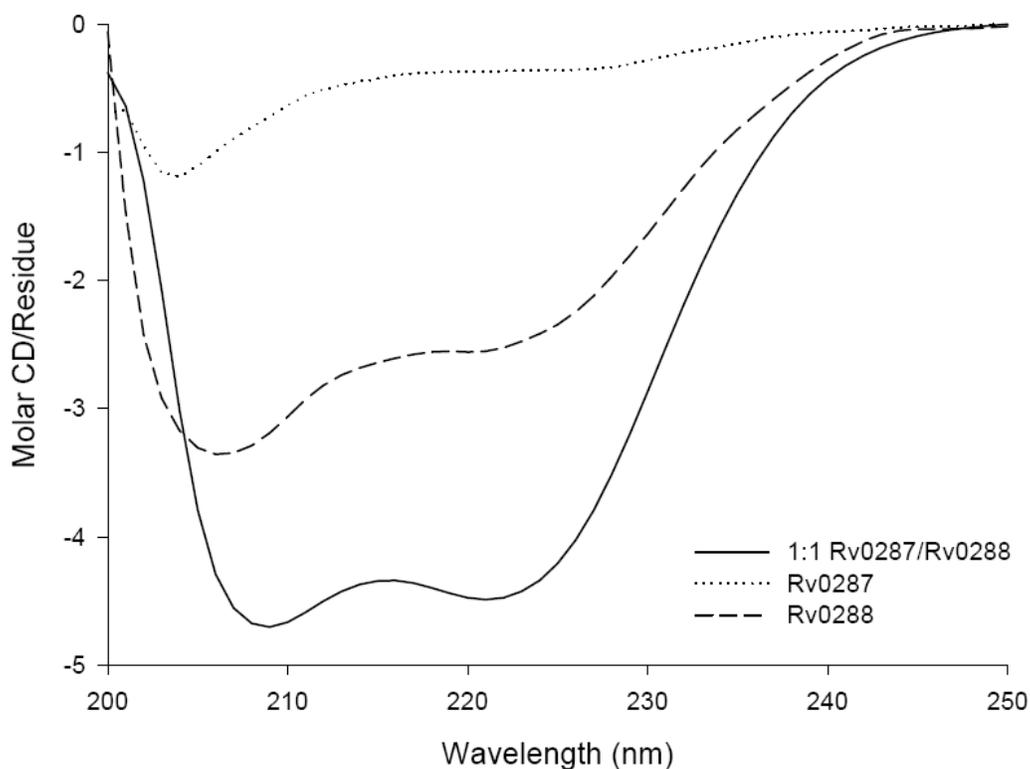
**Figure 3.3.9: Purification of the His<sub>6</sub>Rv0287- His<sub>6</sub>Rv0288 complex by gel filtration. Panel (A) shows a typical FPLC elution profile attained by gel filtration on a Superdex 75 16/60 column. Contaminant peaks are indicated by “1 and 2” and the His<sub>6</sub>Rv0287- His<sub>6</sub>Rv0288 complex elution peak is labelled. The pooled fractions are also highlighted. Panel (B) shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-5 show samples removed from the complex elution peak. Fractions from lanes 3-5 were pooled for structural studies. Lane 1 contains high molecular weight markers (Sigma). Relative staining shown on the SDS PAGE suggests that the proteins form a 1:1 heterodimer complex.**

### 3.3.6. Initial structural characterization of the Rv0287-Rv0288 complex

#### 3.3.6.1. Far UV circular dichroism spectroscopy

The secondary structure of the individual proteins and the Rv0287-Rv0288 complex were determined by far UV circular dichroism (CD) spectroscopy. The spectra obtained for the Rv0287 protein (Figure 3.3.10) are typical of an unstructured, random coil polypeptide as characterized by a single negative CD band at about 200 nm (94). In contrast, the spectra obtained for Rv0288 and the Rv0287-Rv0288 complex (Figure 3.3.10) are representative of proteins with high helical content as evidenced by two intense negative CD peaks at approximately 208 and 221 nm (94). Using the online server DICHROWEB (hosted at Birkbeck College, University of London, U.K) (59), Rv0287, Rv0288 and the complex were estimated to contain 1%  $\alpha$ -helix (11%  $\beta$ -sheet and 88 % unstructured), 24%  $\alpha$ -helix (18%  $\beta$ -sheet and 58% unstructured) and 50%  $\alpha$ -helix (23%  $\beta$ -sheet and 27% unstructured) respectively.

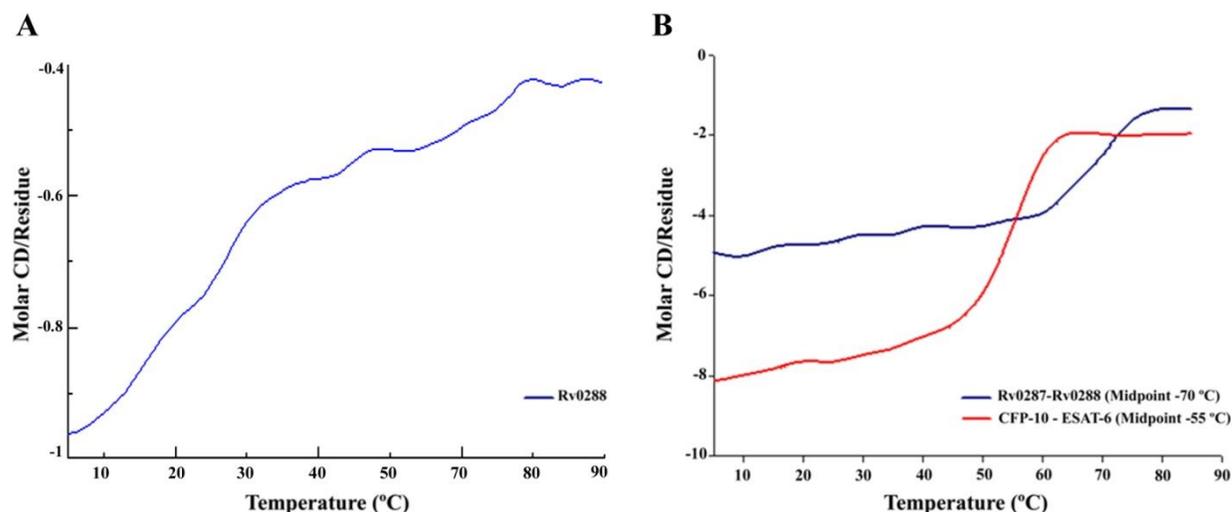
The CD spectra were, however, obtained for Rv0287 and Rv0288 which both contain an extra 20 residues at the N-terminus from the His<sub>6</sub>-tag and linker. These residues are expected to be unstructured, which clearly suggests that the helical content of the regions corresponding to Rv0288 and the Rv0287-Rv0288 complex are higher. If the secondary structures are corrected for the presence of an unstructured N-terminal His-tag linker, the helical content is about 30% for Rv0288 and 60 % for the complex.



**Figure 3.3.10: Secondary structure content of Rv0287, Rv0288 and the Rv0287-Rv0288 complex. Far UV CD spectra are shown for isolated Rv0287 and Rv0288, and for the Rv0287-Rv0288 complex at 25 °C. The spectrum obtained for Rv0287 show a single negative CD peak at around 205 nm which is typical of an unstructured polypeptide, whereas the spectra for Rv0288 and the Rv0287-Rv0288 complex show clear negative bands at about 208 and 221 nm, which is characteristic of proteins with high helical content. The CD studies were repeated at least three times.**

### *3.3.6.2. The effects of thermal variation on Rv0288, and the Rv0288-Rv0287 and CFP-10-ESAT-6 complexes*

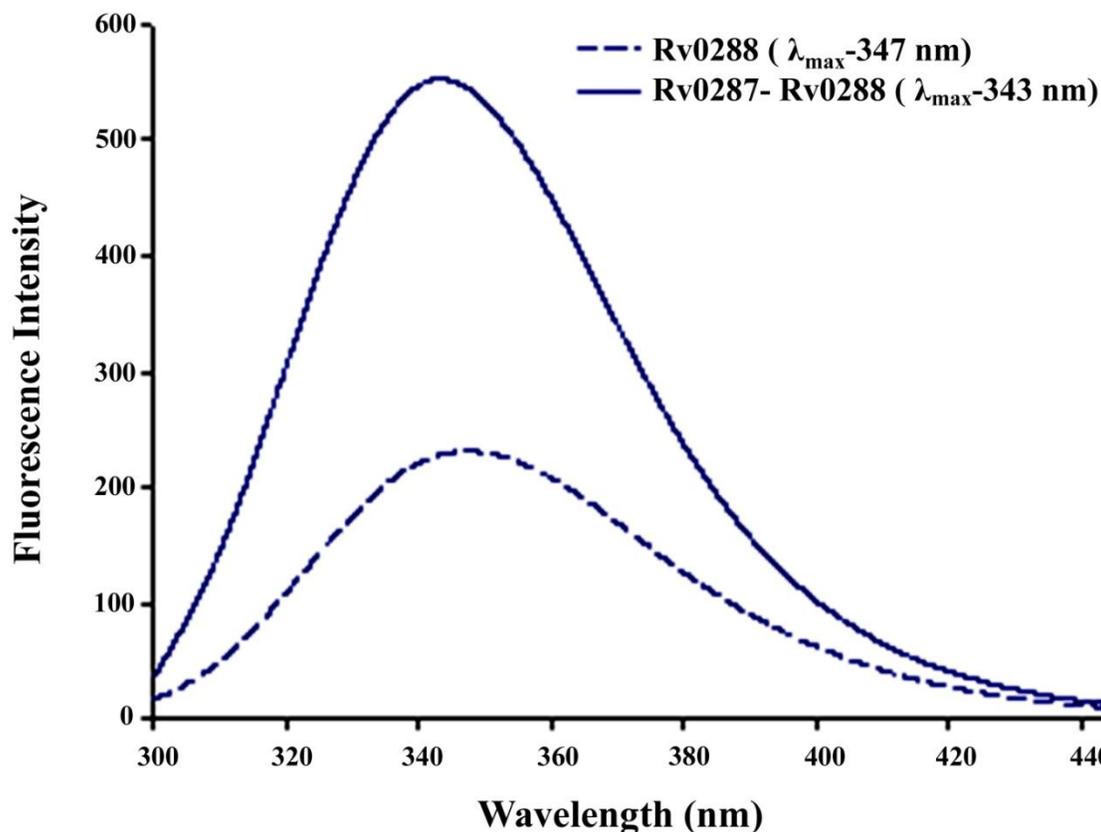
The effects of increasing temperature on the structural integrity of Rv0288, and the Rv0287-Rv0288 and CFP-10-ESAT-6 complexes were studied by following the change in the CD spectra over a range from 5 to 90 °C, increasing in 5 °C increments (Figure 3.3.11). All CD spectra were obtained from protein samples dissolved in a 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and 25 °C with protein concentrations of about 15 μM for a path length of 0.1 cm and allowed to equilibrate at each temperature before data collection.



**Figure 3.3.11: The effect of thermal variation on Rv0288, and the Rv0287-Rv0288 and CFP-10-ESAT-6 complexes. Panel (A) shows the stability of Rv0288 to heat-induced denaturation. A gradual, non-cooperative unfolding curve for the isolated protein indicates that Rv0288 contains no stable tertiary structure. Panel (B) shows the stability of the Rv0287-Rv0288 and CFP-10-ESAT-6 complexes to heat-induced denaturation. The sigmoidal cooperative unfolding curves obtained for both complexes suggest that the complexes contain stable tertiary structures. The results also show that the Rv0287-Rv0288 complex is clearly more stable than the CFP-10-ESAT-6 complex, showing resistance to at least 60 °C, with a denaturation midpoint of around 70 °C, whereas the CFP-10-ESAT-6 complex resists to only 40 °C and has a denaturation midpoint around 55 °C. The CD studies were repeated at least three times.**

### 3.3.6.3. Fluorescence analysis of Rv0288-Rv0287 complex

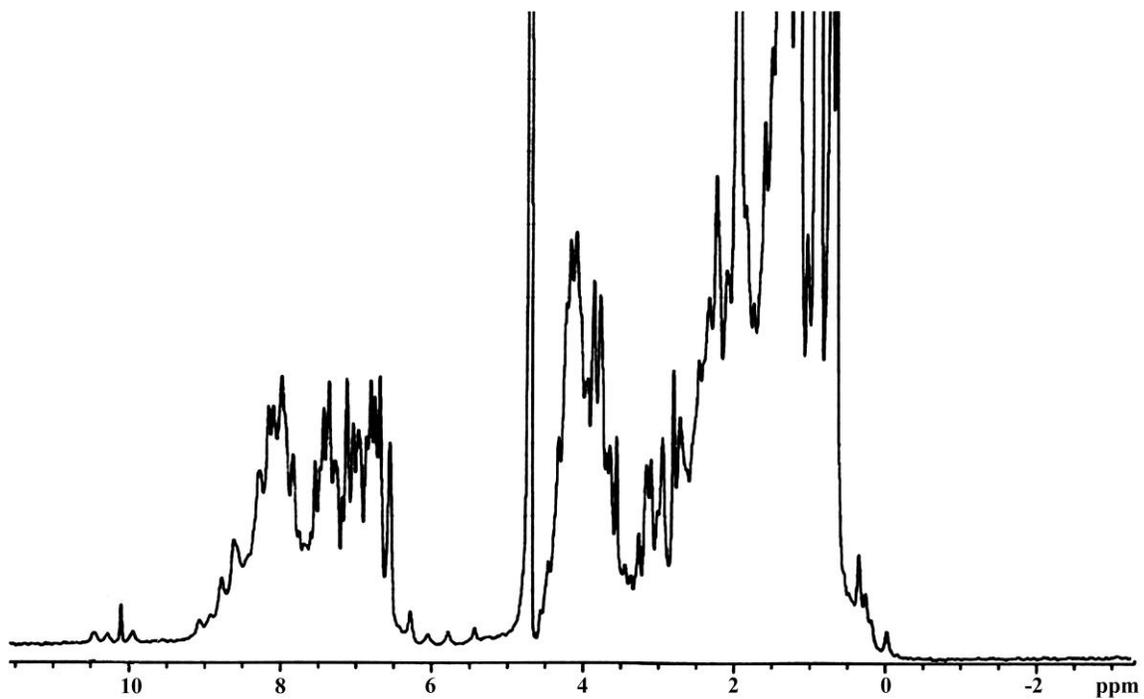
Intrinsic tryptophan fluorescence spectra recorded for Rv0288 and the Rv0287-Rv0288 complex, both of which contain four tryptophan residues, are shown in figure 3.3.12 (Section 2.2.13). No spectrum is shown for Rv0287 as this protein does not contain any tryptophan residues. The  $\lambda_{\text{max}}$  of the complex is 343 nm, which is blue shifted in comparison to the individual Rv0288 protein (347 nm), indicating that at least one of the four tryptophan residues has become less solvent exposed upon formation of the complex.



**Figure 3.3.12: Fluorescence Analysis of the individual Rv0288 and the Rv0288-Rv0287 complex.** Maximum intrinsic tryptophan fluorescence spectra were obtained for the isolated Rv0288 and the complex protein in 100 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub> buffer at pH 6.5 and room temperature. The fluorescence studies were repeated at least three times.

#### 3.3.6.4. 1D <sup>1</sup>H NMR analysis of the Rv0288-Rv0287 complex

1D <sup>1</sup>H NMR experiments were carried out at 35 °C. As shown in figure 3.3.13, the spectrum for the Rv0287-Rv0288 complex shows significant dispersion of signals from backbone amide groups (NH)(6.5 to 9.5 ppm) and a number of high field-shifted methyl-resonances (- CH<sub>3</sub>) (0 to 0.6 ppm), which are both characteristic features of a folded protein.



**Figure 3.3.13:** One-dimensional  $^1\text{H}$  NMR spectrum of a 0.7 mM sample of the Rv0288-Rv0287 complex in a 100 mM NaCl, 25 mM  $\text{NaH}_2\text{PO}_4$  buffer at pH 6.5 and 35 °C.

### 3.3.6.5. Discussion

#### ***Initial Characterization of Complex Formation between Rv0287 and Rv0288***

The previous yeast two-hybrid studies conducted by Lightbody *et al.* showed that Rv0287 and Rv0288 like CFP-10 and ESAT-6 (1, 40), interact with each other to form a tight heterodimer complex (62). Subsequent fluorescence-based binding assays indicated that the two proteins forms a tight 1:1 interaction with an estimated dissociation constant ( $K_d$ ) for the complex of 5 nM or lower (63). The stoichiometry of this interaction can also be supported by the relative coomassie blue staining of the individual proteins in the Rv0287-Rv0288 complex as shown in the figures 3.3.6 and 3.3.9.

The secondary structures of the Rv0287-Rv0288 complex and the individual proteins were determined by far UV circular dichroism (CD) spectroscopy (Figure 3.3.10). The spectra obtained for the Rv0287 protein are typical of an unstructured, random coil polypeptide (1 % helical region). In contrast, the spectra obtained for Rv0288 and the Rv0287-Rv0288 complex are typical of predominantly helical proteins, with quantitative analysis suggesting a total helical content of 30 % and 60 % respectively.

#### ***Characterization of the stability of the Rv0287-Rv0288 complex and individual proteins***

The CD analysis explained above clearly shows that the isolated Rv0287, like CFP-10 (40), exists as an unfolded random coil polypeptide, however, the stability of Rv0288 and the Rv0287-Rv0288 complex were measured by determining their resistance to heat-induced denaturation, which was monitored by changes in maximum negative CD band at 208 nm. The thermal denaturation curve obtained for Rv0288 showed a gradual, non-cooperative

unfolding of the protein (Figure 3.3.11 panel A). This suggests that Rv0288 contains no stable tertiary structure and the spectral changes observed probably reflect the unfolding of isolated helical regions. The presence of a significant helical region with no apparent resistance to denaturation suggests that Rv0288, like ESAT-6 (40), exists in the molten globule state (63). In contrast, the complex showed a significant resistance to heat-induced denaturation. The heat-induced denaturation curve obtained for the complex was a sigmoidal cooperative curve (Figure 3.3.11 panel B) which is typical of proteins with stable folded structures (63). The 1D  $^1\text{H}$  NMR spectrum of the complex, in which a clear chemical shift dispersion particularly in the regions corresponding to the protons of the backbone amide and methyl groups are observed, also suggests a stable folded structure for the complex (Figure 3.3.13). This, along with the high stability of the complex at low pH (4.5-6.5) (63), strongly suggests that the functional form of Rv0287 and Rv0288 is highly likely to be a 1:1 protein complex under physiological condition.

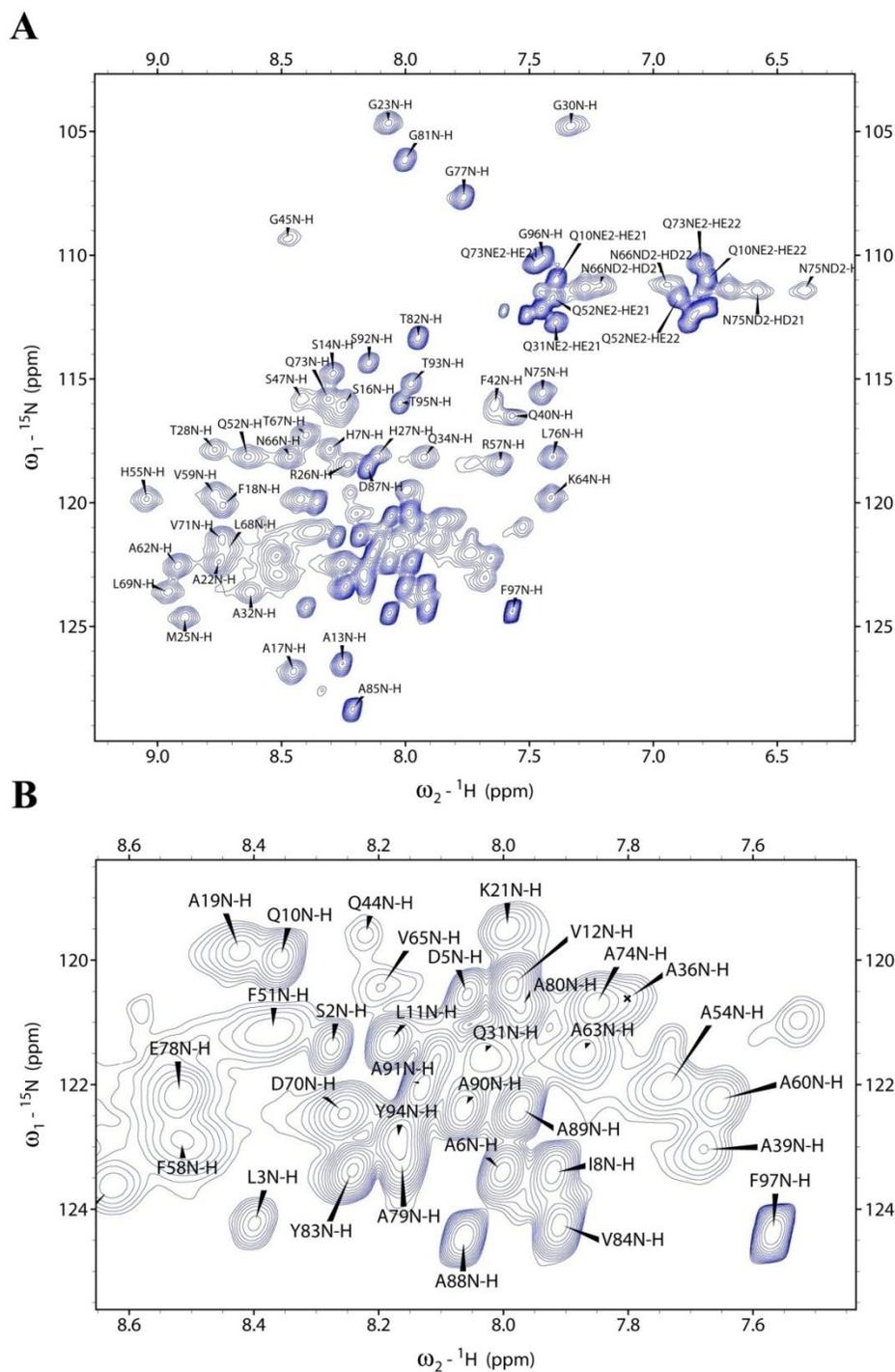
The chemical-induced denaturation analysis conducted by Lightbody *et al.* has also shown an initial stability to denaturation followed by a sigmoidal cooperative unfolding for both the CFP-10-ESAT-6 and Rv0287-Rv0288 complexes (63). These studies clearly showed that Rv0288-Rv0287 complex was more resistant to the chemical denaturation (guanidine hydrochloride (GdnHCl)) than the ESAT-6-CFP-10 complex, with midpoints of denaturation of about 1.7 M and 0.7 M GdnHCl respectively (63). This is further supported by the heat-induced denaturation curves for both complexes, indicating that the Rv0288-Rv0287 complex is significantly more resistant to thermal denaturation than the related CFP-10-ESAT-6 complex (Figure 3.3.11). The ESAT-6-CFP-10 complex is stable to around 45 °C, with a denaturation midpoint of about 55 °C, while the Rv0288-Rv0288

complex is stable to at least 60 °C, and has a denaturation midpoint of about 70 °C (Figure 3.3.11). Taken together, the spectroscopic analyses strongly suggest that the Rv0288-Rv0287 complex is significantly more stable than the related ESAT-6-CFP-10 complex.

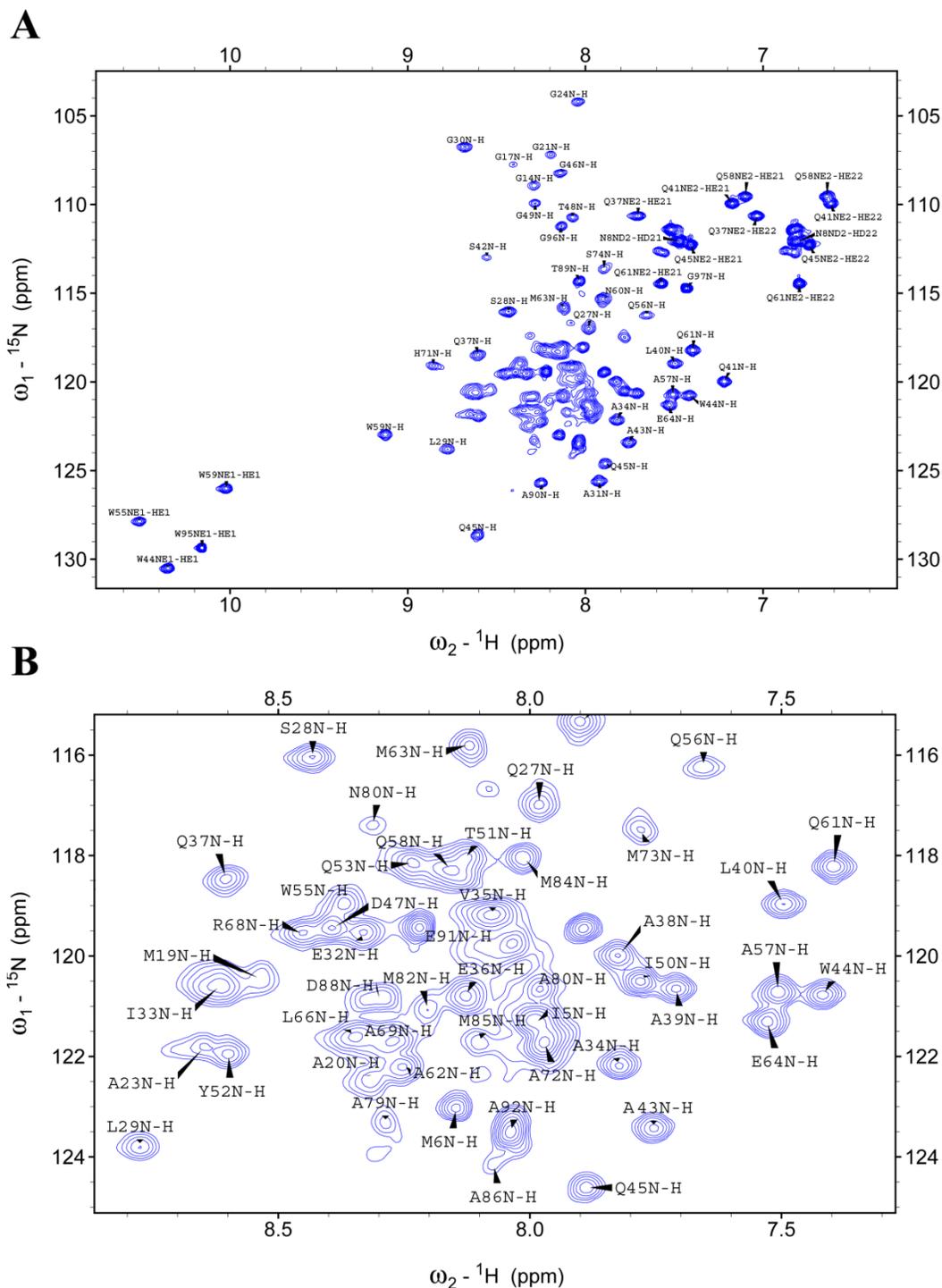
### 3.3.7. High-Resolution structure determination of the Rv0287-Rv0288 complex

#### 3.3.7.1. Sequence specific assignments

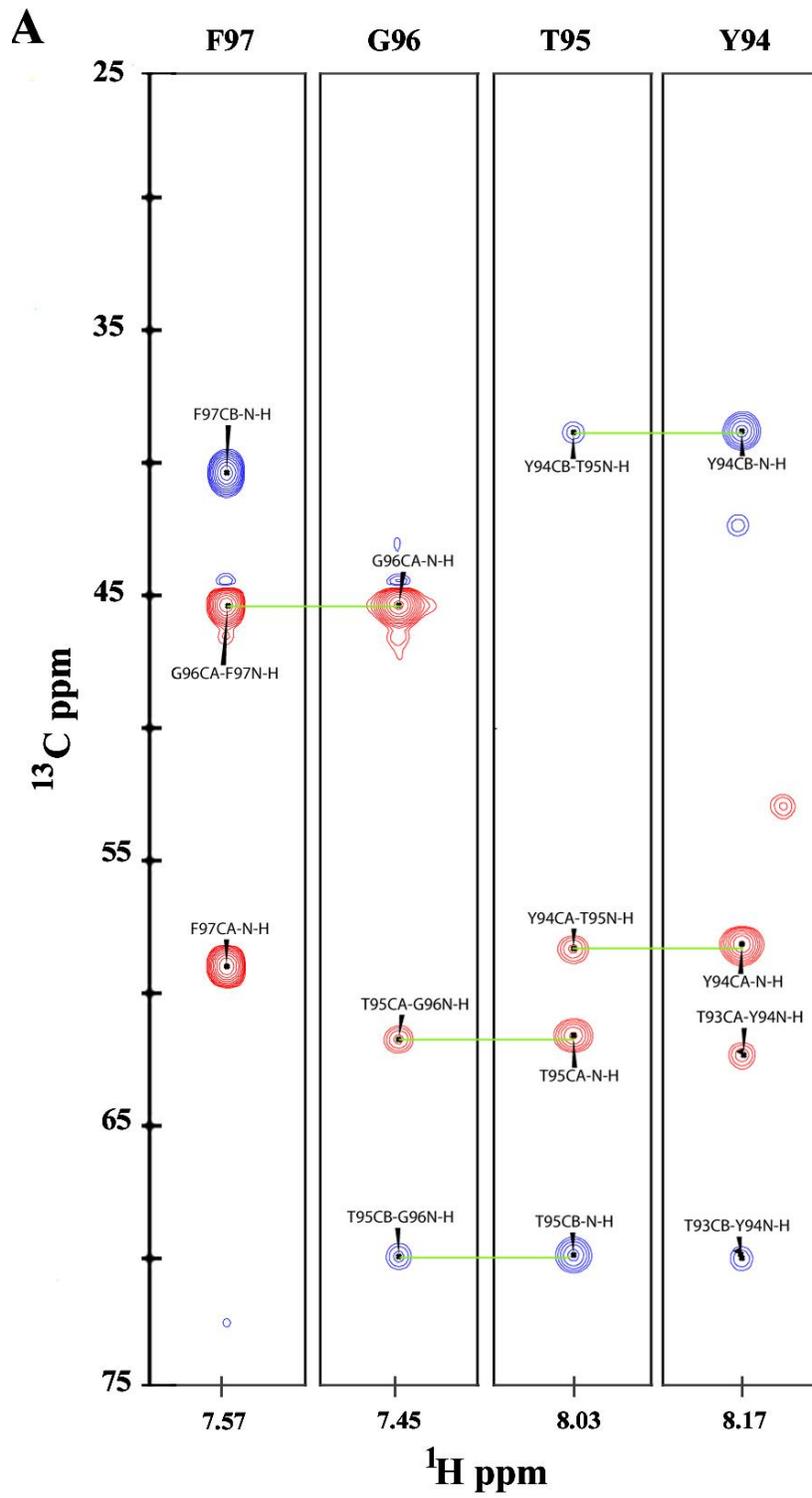
Both labelled Rv0287 bound to unlabelled Rv0288 and labelled Rv0288 bound to unlabelled Rv0287 give rise to well-resolved spectra, as illustrated by the  $^{15}\text{N}/^1\text{H}$  HSQC spectra shown in figures 3.3.14 and 3.3.15, which allowed essentially complete backbone resonance assignments to be made. Backbone amide assignments were obtained for all non-proline residues in the complex except: M1, H43, A49 and E64 in Rv0287 and M2, S75, T76, H77, E78 and A83 in Rv0288 (95 %) and for all  $\text{C}\alpha$  except: Y9 in Rv0288 (99.5 %) and  $\text{C}\beta$  except: M1, S47 in Rv0287 and S75 in Rv0288 (98.5 %). A representative set of  $^{15}\text{N}$  strips from the HNCACB and HNCA spectra used to assign the backbone atoms (NH, N,  $\text{C}\alpha$  and  $\text{C}\beta$ ) are shown in figures 3.3.16 and 3.3.17. Subsequently,  $^{15}\text{N}$  strips from the  $^{15}\text{N}/^1\text{H}$  NOESY-HSQC spectra were used to confirm the backbone assignments as described in section 3.2.11. , and shown in figure 3.3.18. Assignments were then extended to the side chain signals as described in section 3.2.11. In total 92.4 % and 96 % of the signals obtained from all the protons and carbons, contained in the Rv0288-Rv0287 complex, were successfully assigned.



**Figure 3.3.14:**  $^{15}\text{N}/^1\text{H}$  HSQC spectrum of the  $^{15}\text{N}$ -labelled Rv0287 bound to Rv0288. Panel (A) shows the assignments of the signals from backbone amide groups of Rv0287 indicated by residue type and number. The region between 6.2 and 7.5 ppm in  $^1\text{H}$  domain and 109 and 114 ppm in  $^{15}\text{N}$  domain contains, with the exception of G96, signals from the side chain amide groups. Assignments obtained for the side chain  $\text{NH}_2$  groups of Q10, Q31, Q52, N66, Q73, and N75 are indicated on the spectrum. The assignments obtained for the overlapped region between 7.45 and 8.65 ppm in  $^1\text{H}$  domain and 118.5 and 125.5 ppm in  $^{15}\text{N}$  domain is shown expanded in the panel B.



**Figure 3.3.15:**  ${}^{15}\text{N}/{}^1\text{H}$  HSQC spectrum of the  ${}^{15}\text{N}$ -labelled Rv0288 bound to Rv0287. Panel (A) shows the assignments of the signals from backbone amide groups of Rv0288 indicated by residue type and number. The region between 6.4 and 7.8 ppm in  ${}^1\text{H}$  domain and 109 and 115 ppm in  ${}^{15}\text{N}$  domain contains, with the exception of G97, signals from the side chain amide groups. Assignments obtained for the side chain  $\text{NH}_2$  groups of N8, Q37, Q41, Q45, Q58 and Q61 are indicated on the spectrum. The region between 9.8 and 10.6 ppm in  ${}^1\text{H}$  domain and 125 and 132 ppm in  ${}^{15}\text{N}$  domain contains, signals from the indole-NH group on the side chain of tryptophan residues (W44, W55, W59, and W95). The assignments obtained for the overlapped region between 7.3 and 8.8 ppm in  ${}^1\text{H}$  domain and 115.75 and 124.5 ppm in  ${}^{15}\text{N}$  domain is shown expanded in the panel B.



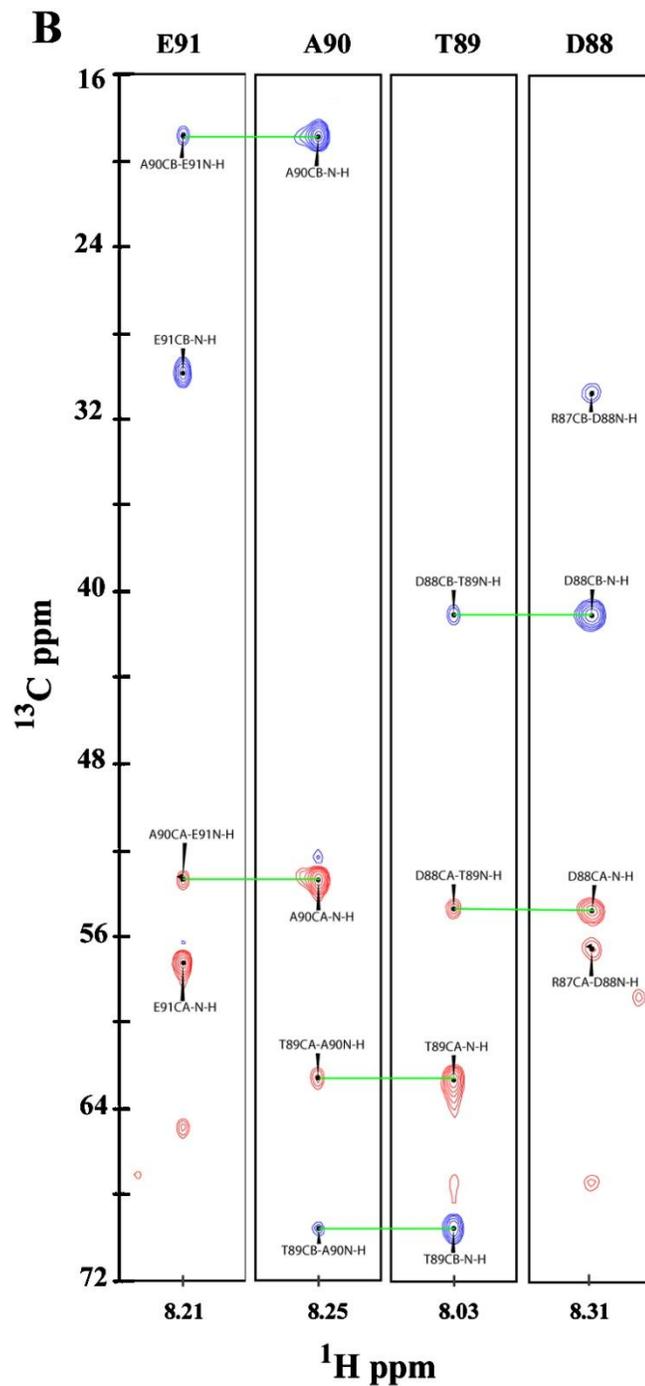
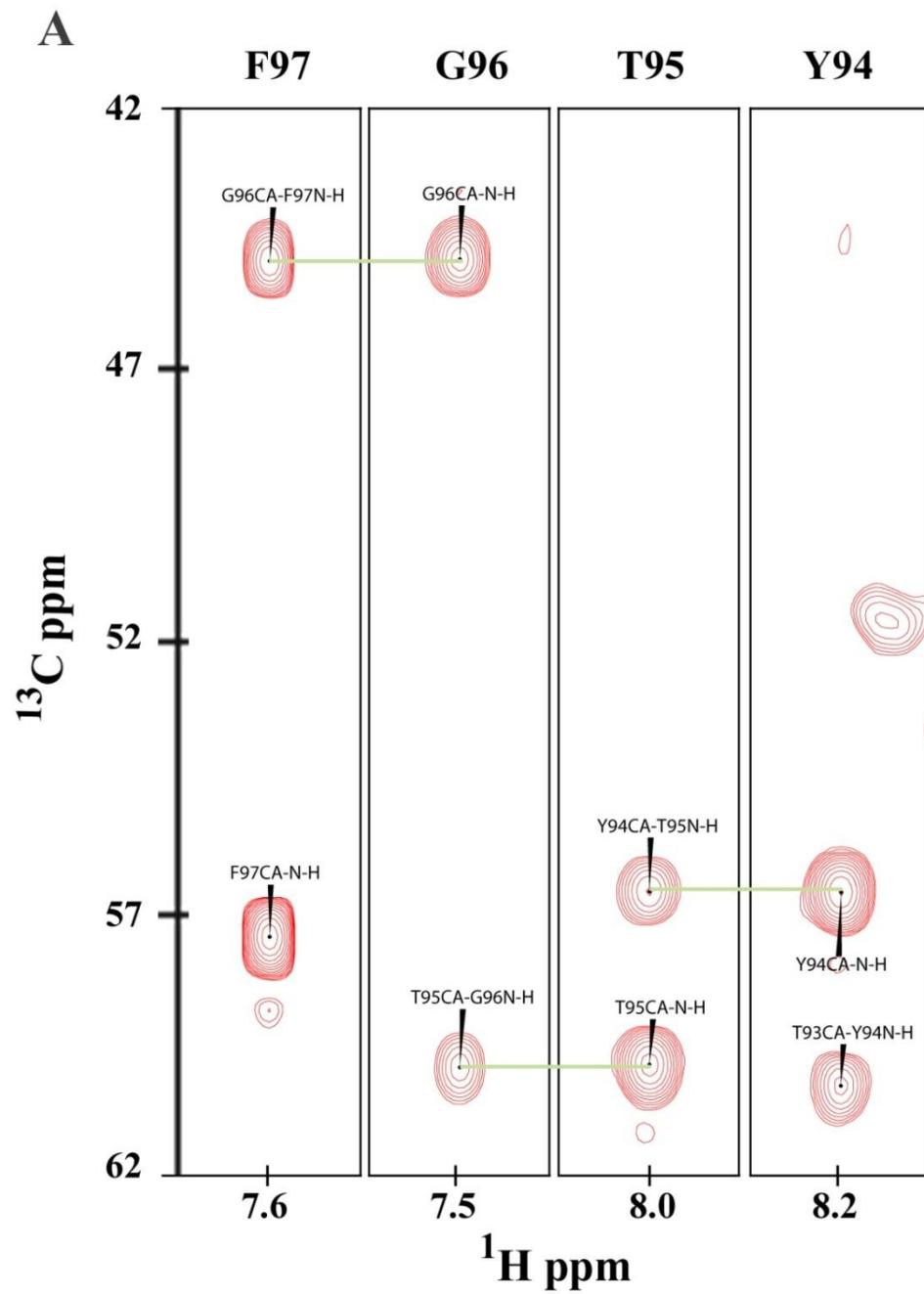


Figure 3.3.16:  $^{15}\text{N}$  strips from the HNCACB spectra of labelled Rv0287 bound to Rv0288 (A) and labelled Rv0288 bound to Rv0287 (B) illustrating the intra- and inter-residue cross peaks linking residues F97 – Y94 (in Rv0287) and E91 – D88 (in Rv0288).  $\text{C}\alpha$  signals are shown in red and  $\text{C}\beta$  in blue. Intra-residue connectivities between amide nitrogen/proton and  $\text{C}\alpha$  and  $\text{C}\beta$  show stronger signals than those observed for sequential connectivities. Ambiguities caused by the overlap of intra- and inter-residue cross peaks, or the absence of peaks were resolved using the CBCA(CO)NH and TROSY HN(CO)CA spectra. Green lines show sequential  $\text{C}\alpha$  and  $\text{C}\beta$  connections. The carbon and proton chemical shifts are highlighted. The residue name and number are also shown above the strips.



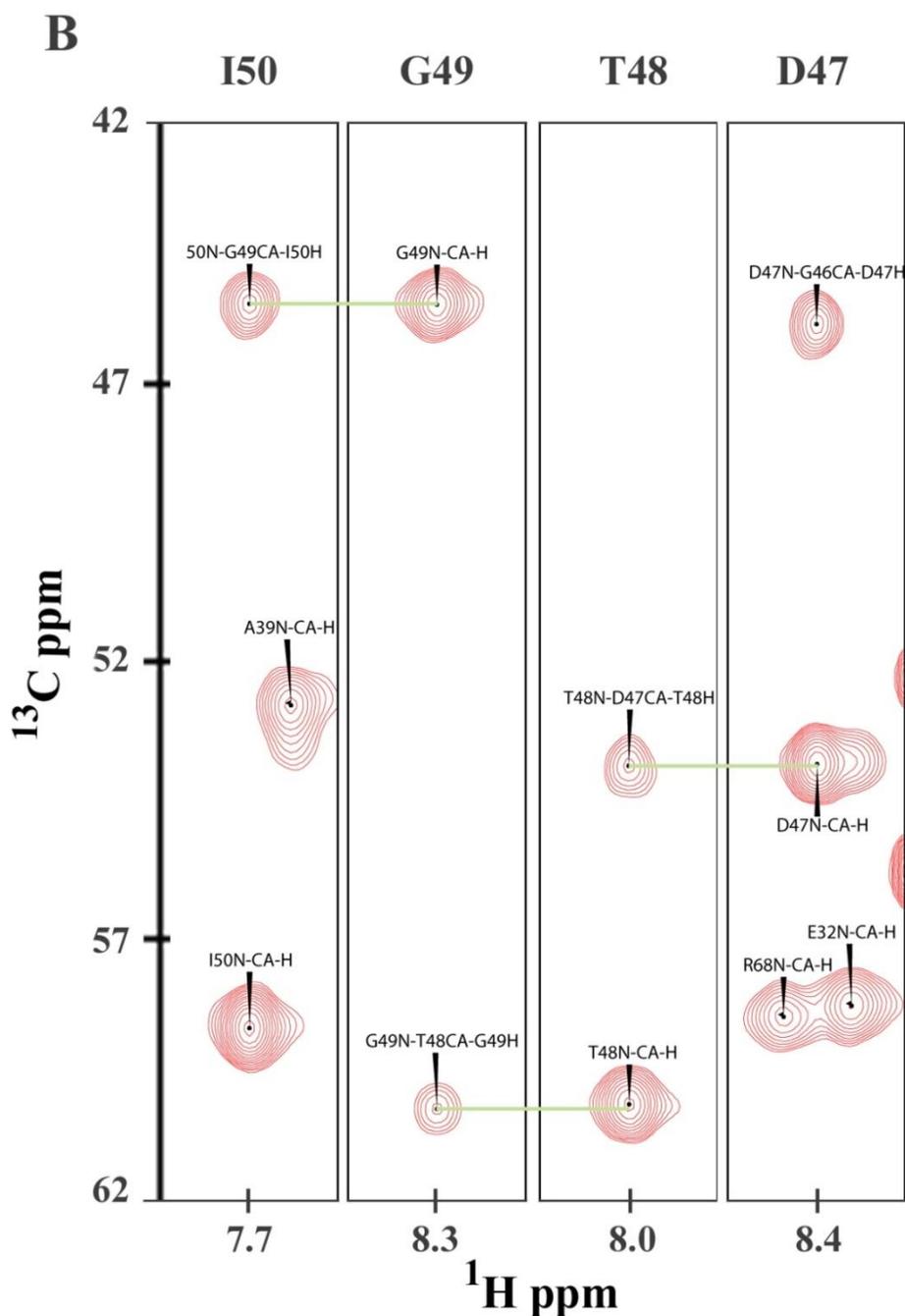
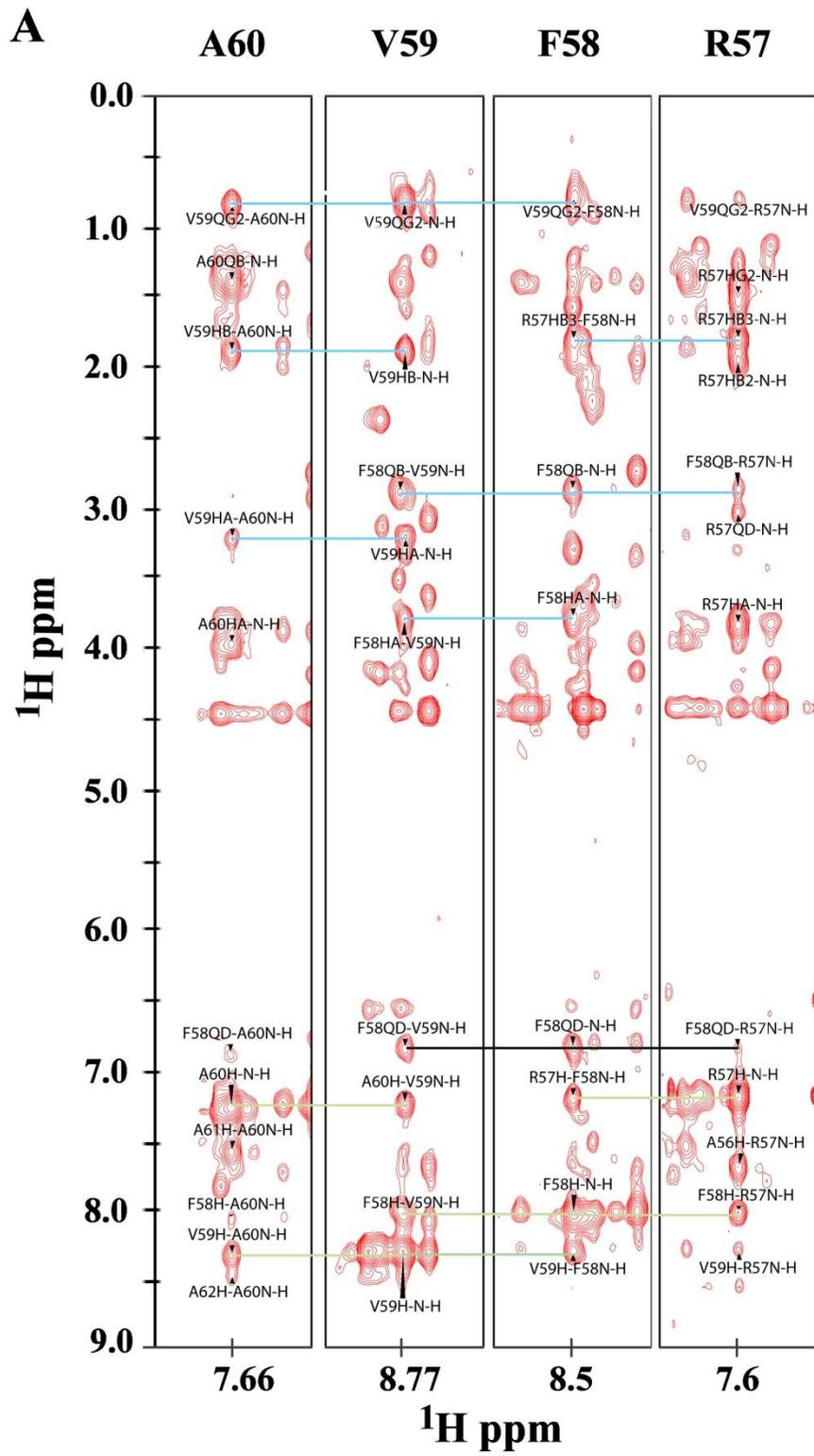


Figure 3.3.17:  $^{15}\text{N}$  strips from the HNCA spectra of labelled Rv0287 bound to Rv0288 (A) and labelled Rv0288 bound to Rv0287 (B) illustrating the intra- and inter-residue cross peaks linking residues F97 – Y94 (in Rv0287) and I50 – D47 (in Rv0288).  $\text{C}\alpha$  signals are shown in red. Intra-residue connectivities between amide nitrogen/proton and  $\text{C}\alpha$  show stronger signals than those observed for sequential connectivities. Ambiguities caused by the overlap of intra- and inter-residue cross peaks, or the absence of peaks were resolved using the CBCA(CO)NH, and TROSY HN(CO)CA spectra. Green lines indicate sequential  $\text{C}\alpha$  connections. The carbon and proton chemical shifts are highlighted. The residue name and number are also shown above the strips.



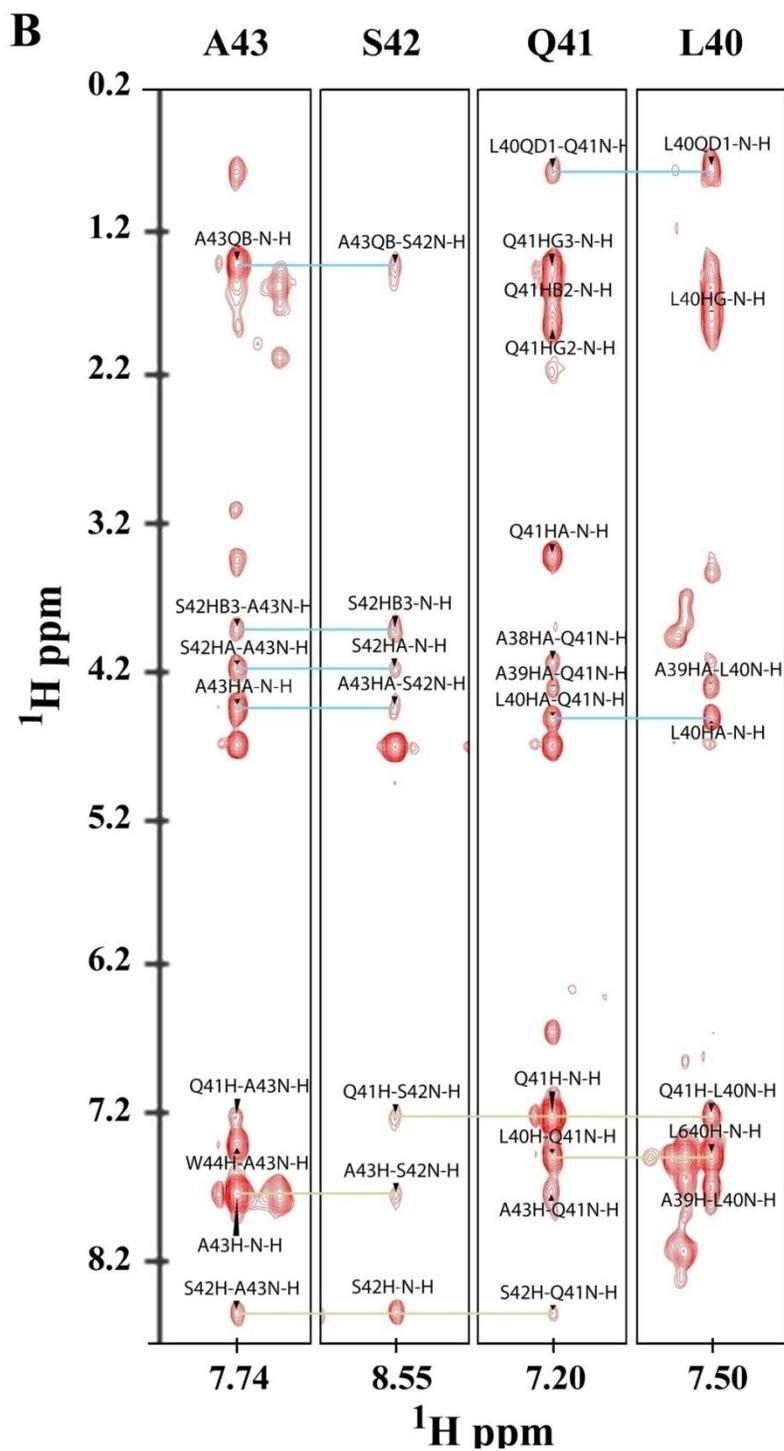
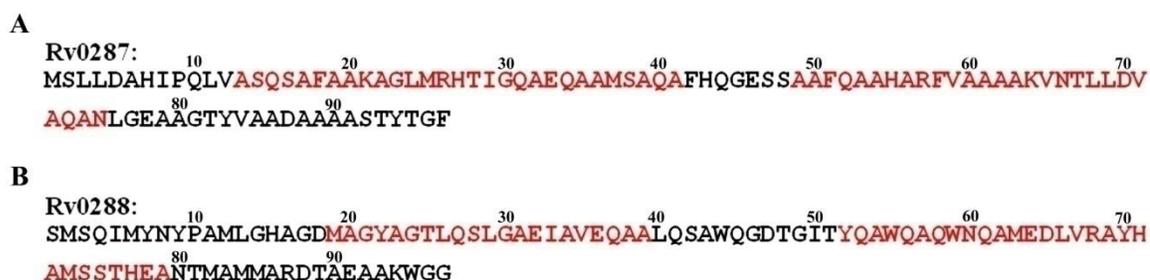


Figure 3.3.18:  $^{15}\text{N}$  strips from the NOESY-HSQC spectra of labelled Rv0287 bound to Rv0288 (A) and labelled Rv0288 bound to Rv0287 (B) illustrating typical NOE connectivities linking residues A60 to R57 (in Rv0287) and A43 to L40 (in Rv0288). Green lines indicate sequential, and  $i$  to  $i+2$  NH-NH NOEs, blue lines indicate examples of aliphatic to NH NOEs. The black line shown in panel (A) indicates an example of sequential aromatic to NH NOEs. The proton chemical shifts are highlighted. The residue name and number are also shown above the strips.

### 3.3.7.2. Secondary structure determination

The secondary structure of the Rv0288-Rv0287 complex was determined using the Chemical Shift Index (CSI) method (Section 3.2.12). This analysis was performed based on the comparison of the observed C $\alpha$  chemical shifts for the complex with the average random coil values (111). The  $^{13}\text{C}$  NMR shifts of the C $\alpha$  of all naturally occurring amino acids experience a downfield shift (indicated by a CSI value of +1) when present in a regular helical conformation and a comparable up field shift (indicated by a CSI value of -1) when located in a regular  $\beta$ -strand extended conformation (111).

The CSI results suggested that the complex contained four helices located between residues A13 - A41, A49 - N75 in Rv0287 and M19 - A39, Y52- A79 in Rv0288 as shown in figure 3.3.19.



**Figure 3.3.19: CSI analysis of Rv0288-Rv0287 complex.** The residues which represent a CSI value of +1 are highlighted in red in the primary sequences of Rv0287 (A) and Rv0288 (B). These residues are incorporated in the helical regions of the Rv0287-Rv0288 complex.

### 3.3.7.3. Determination of slowly exchanging amide protons

Slow rates of amide proton exchange result from shielding of amide protons from solvent and are most commonly caused by hydrogen bond interactions (103). Slowly exchanging amide protons contained in Rv0288 were identified (section 3.2.13) and assigned by comparing the spectrum shown in figure 3.3.20 and the previously assigned  $^{15}\text{N}/^1\text{H}$  HSQC spectrum shown in the figure 3.3.15. The assigned amide protons were then used to identify corresponding H-bonds in the complex using MOLMOL (110) and preliminary calculated structures. 24 hydrogen bond constraints obtained from the six identified H-bonds (W55-W59, A62-L66, Q58-A62, W59-M63, M63-V67, and N60-E64) were subsequently used for the final structure calculation.

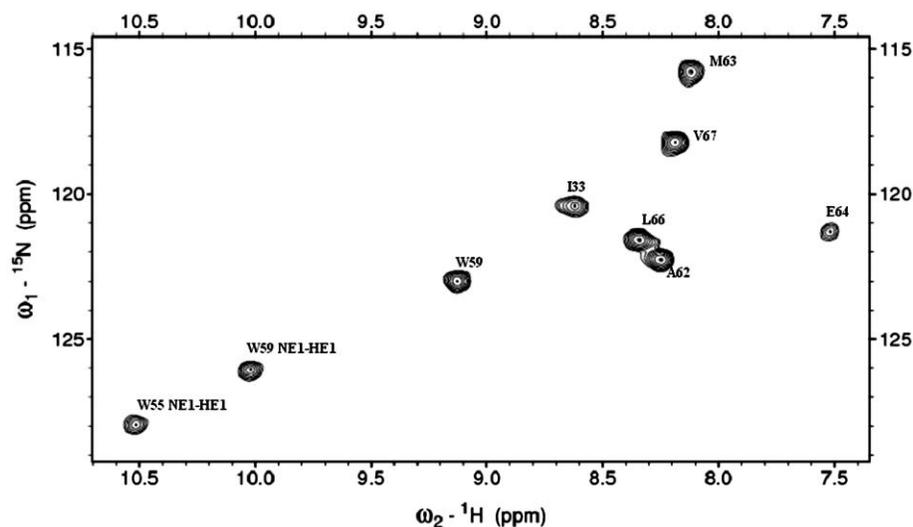


Figure 3.3.20: Amide – proton and indole group NH exchange observed in the  $^{15}\text{N}/^1\text{H}$  HSQC spectrum of the  $^{15}\text{N}$ -labelled Rv0288 bound to Rv0287. The protein solution was prepared in 100 %  $\text{D}_2\text{O}$  and incubated for overnight at room temperature, then measured at 35 °C. The counter plot illustrates the slowly exchanging protons contained in Rv0288. (No slowly exchanging amide protons were observed in the  $^{15}\text{N}/^1\text{H}$  HSQC spectrum of the  $^{15}\text{N}$ -labelled Rv0287 bound to Rv0288 when the corresponding protein sample was prepared in the way described above).

#### 3.3.7.4. Structural calculations for Rv0288-Rv0287 complex

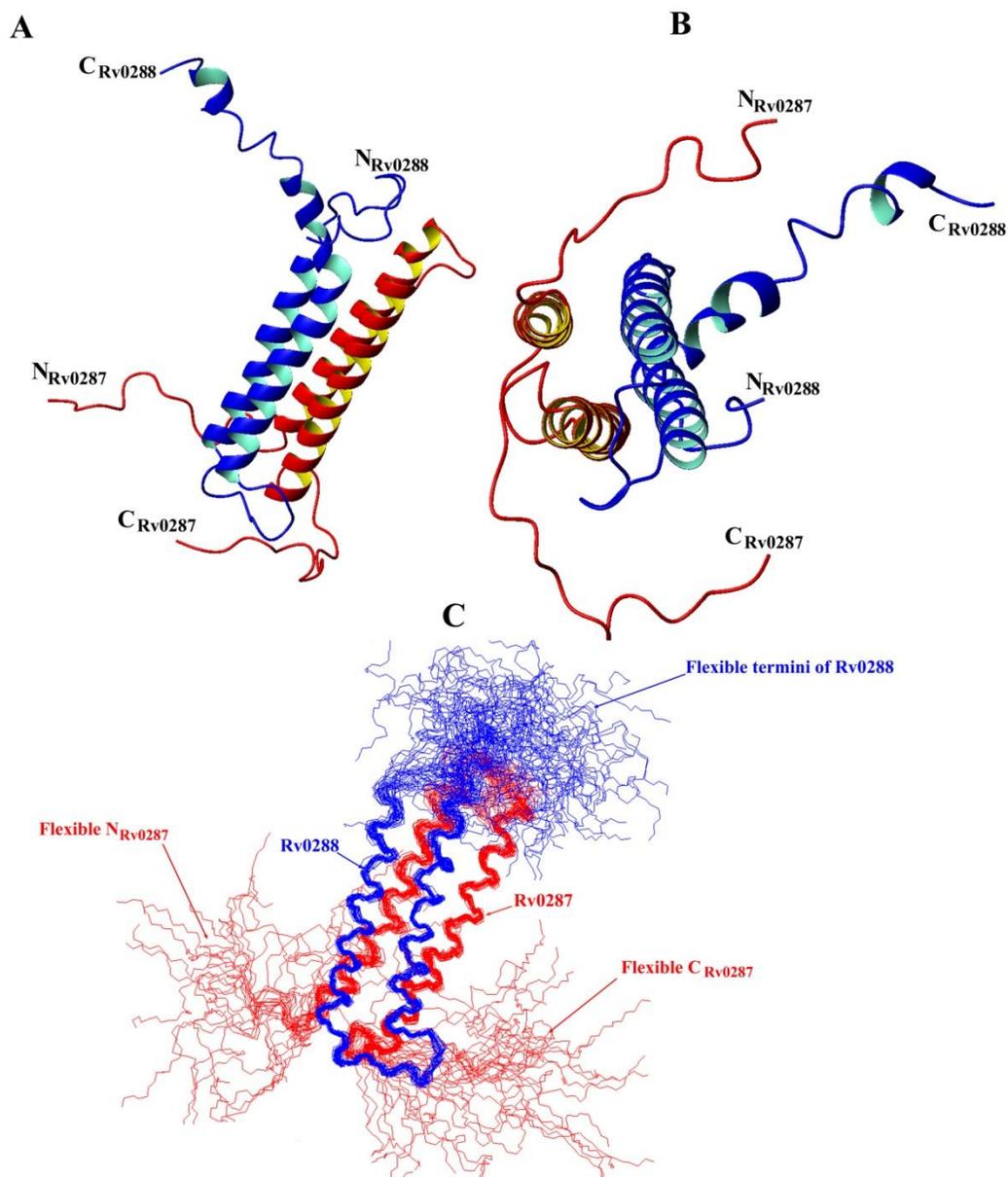
The combined automated NOE assignment and structure determination module (CANDID) (106) is known to be very effective software at determining unique assignments for the NOEs identified in the three dimensional  $^{15}\text{N}$ - and  $^{13}\text{C}$ -edited NOESY and the aromatic to aliphatic region of the two dimensional NOESY. Assignments were obtained for 87.2 % of the total NOE peaks picked in the  $^{15}\text{N}/^1\text{H}$  NOESY-HSQC spectra, the  $^{13}\text{C}/^1\text{H}$  edited-NOESY spectra and in the NOESY spectrum, which produced 2323 non-redundant  $^1\text{H}$  to  $^1\text{H}$  upper distance limits. The final family of Rv0288-Rv0287 structures was determined using a total of 2611 NMR-derived structural constraints (an average of 13.5 per residue), including 2323 NOE-based upper distance limits (544 intra residue, 598 sequential ( $i, i+1$ ), 635 medium range ( $i, i \leq 4$ ), 339 intermolecular long range ( $i, i \geq 5$ ), and 207 intramolecular long range ( $i, i \geq 5$ )), 264 backbone torsion angle constraints and 24 hydrogen bond constraints in regions of regular helical structure (table 5.2.2). Following the final round of CYANA calculations, 30 satisfactorily converged structures were obtained from 100 random starting structures. The converged structures contain no distance or van der Waals violation greater than 0.5 Å and no dihedral angle violations greater than  $5^\circ$ , with an average value for the CYANA target function of  $0.62 \pm 0.15 \text{ \AA}^2$ . The sums of the violations for the upper distance limits, lower distance limits, van der Waals contacts and torsion angle constraints were  $6.9 \pm 0.05 \text{ \AA}$ ,  $0.9 \pm 0.04 \text{ \AA}$ ,  $6.04 \pm 0.04 \text{ \AA}$  and  $52.8 \pm 0.74^\circ$  respectively. Similarly, maximum violations for the converged structures were  $0.32 \pm 0.0 \text{ \AA}$ ,  $0.1 \pm 0.0 \text{ \AA}$ ,  $0.35 \pm 0.0 \text{ \AA}$  and  $3.09 \pm 0.32^\circ$  respectively. The NMR constraints and structural statistics for the Rv0288-Rv0287 complex are summarized in table 3.3.1.

**Table 3.3.1: NMR Constraints and Structural Statistics for Rv0288-Rv0287 complex**

NMR Constraints and Structural Statistics for the Rv0288-Rv0287 complex		
a) No. of Constraints used in Final Structural Calculation		
Intermolecular NOEs	339	
Intraresidue NOEs	544	
Sequential (short range) NOEs ( $i, i+1$ )	598	
Medium-range NOEs ( $i, i \leq 4$ )	635	
Long-range NOEs ( $i, i \geq 5$ )	546	
Torsion angles	264	(132 $\Phi$ and 132 $\Psi$ )
Hydrogen bonding	24	
b) Maximum and Total Constraint Violations in 30 Converged Rv0288 /Rv0287 Structures		
Upper distance limits ( $\text{\AA}$ )	$0.32 \pm 0.00$	$6.95 \pm 0.05$
Lower distance limits ( $\text{\AA}$ )	$0.10 \pm 0.00$	$0.90 \pm 0.04$
van der Waals contacts ( $\text{\AA}$ )	$0.35 \pm 0.00$	$6.04 \pm 0.04$
Torsion angle ranges ( $^\circ$ )	$3.09 \pm 0.32$	$52.8 \pm 0.74$
Average CYANA target function ( $\text{\AA}^2$ )	$0.62 \pm 0.15$	
c) Structural Statistics for the Family of Converged Rv0288-Rv0287 Structures		
Residues within the		
favorably allowed		80.7
additionally allowed		15.9
generously allowed		3.4
disallowed		0.0
regions of the Ramachandran plot (%)		
Backbone atom r.m.s.d. for structured region (residues		
17-38 and 49-76 of Rv0287, and residues 20-39 and 52-76 of Rv0288)		$0.9 \pm 0.17 \text{ \AA}$
Heavy atom r.m.s.d. for structured region (residues		
17-38, and 49-76 of Rv0287, and residues 20-39 and 52-76 of Rv0288)		$1.3 \pm 0.18 \text{ \AA}$

The high resolution structure of Rv0288-Rv0287 complex is finally determined. This is clearly evidenced by the superposition of the protein backbone shown for the family of converged structures in figure 3.3.21 panel C (best fit for residues 17 - 38 and 49 – 76 in Rv0287, and residues 20-39 and 52 –76 in Rv0288) and is reflected in low root mean squared deviation (r.m.s.d.) values to the mean structure for both all heavy atoms and the backbone of  $1.3 \pm 0.18 \text{ \AA}$  and  $0.9 \pm 0.17 \text{ \AA}$  respectively. A best-fit superposition of the protein backbone for the family of 30 converged structures and also ribbon representations of the backbone topology of the Rv0287-Rv0288 complex are shown in figure 3.3.21. The

family of converged Rv0287-Rv0288 complex structures, together with the NMR constraints, has been deposited in the Protein Data Bank (accession code 2kg7).



**Figure 3.3.21: Solution structure of the Rv0288-Rv0287 complex. Panel (A) And (B) show ribbon representations of the backbone topology of the Rv0287(red)-Rv0288(blue) complex based on the converged structure closest to the mean, which illustrates the two helix–turn–helix hairpin structures formed by the individual proteins. The helical propensity of residues 82-97 in the flexible C-terminus of Rv0288 can be clearly seen in the top left of the figure (Panel A). Panel (C) shows a best-fit superposition of the protein backbone for the family of 30 converged structures obtained, with Rv0287 shown in red and Rv0288 in blue. The flexible N- and C-terminal arms of both proteins are highlighted.**

### 3.3.7.5. Discussion

#### **Structural Features of the Rv0288-Rv0287 complex**

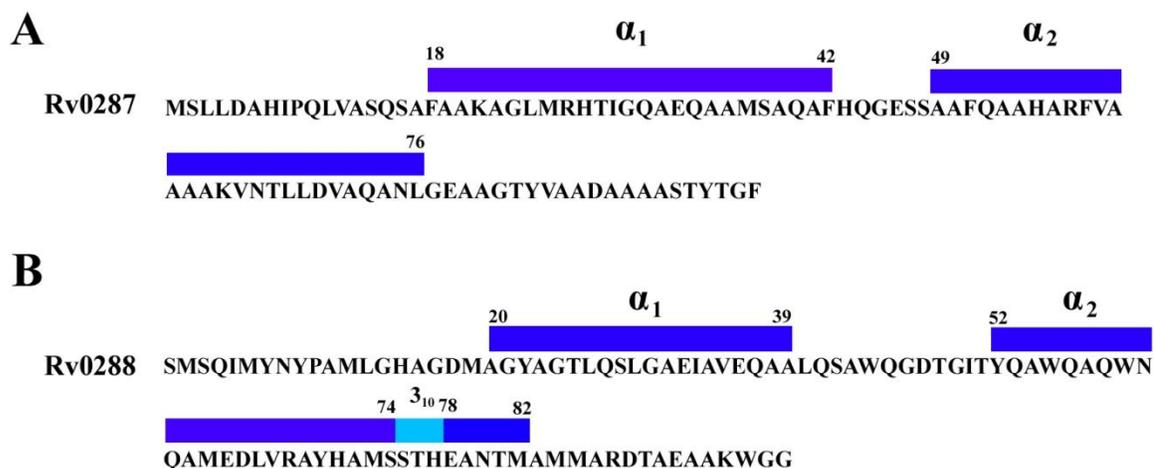
The backbone topology of the Rv0288-Rv0287 complex is illustrated by the ribbon diagram shown in figure 3.3.21 panel A and B. In agreement with the results obtained from the CSI analysis which suggests the complex contains four helices (Figure 3.3.22), the well-defined structured core of the Rv0288-Rv0287 complex contains two helix–turn–helix hairpin structures formed from the individual proteins, which have an extensive hydrophobic contact surface and lie anti-parallel to each other to form a four-helix bundle (Figure 3.3.21).



**Figure 3.3.22: CSI analysis of Rv0288-Rv0287 complex. The residues which represent a CSI value of +1 are highlighted in red in the primary sequences of Rv0287 (A) and Rv0288 (B). These residues are incorporated in the helical regions of the Rv0287-Rv0288 complex.**

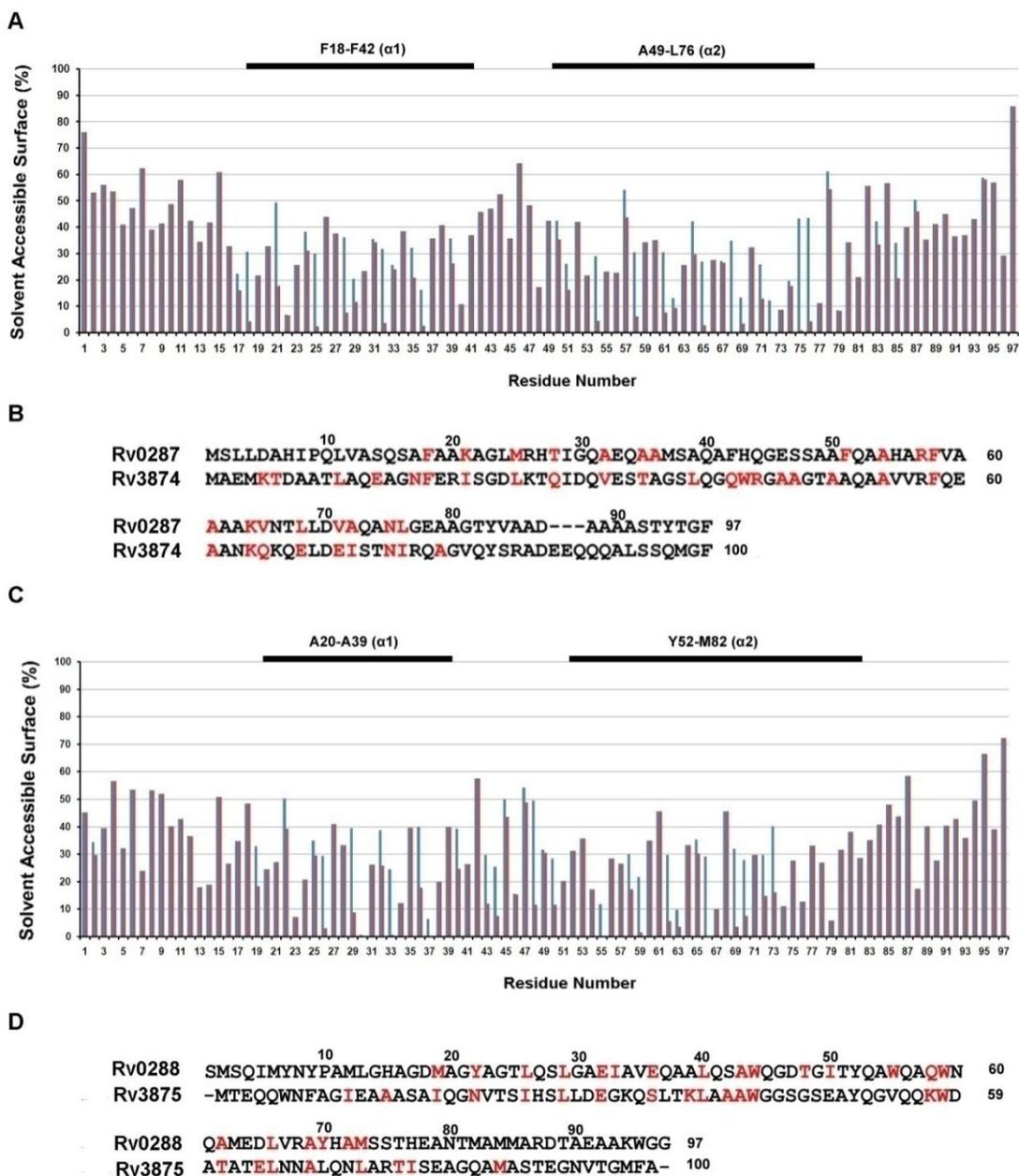
Like the ESAT-6-CFP-10 complex, a striking feature of the Rv0288-Rv0287 complex is the disordered N- and C-termini of both proteins, which form long flexible arms at both ends of the four-helix bundle part (Figur3.3.21). The two helices in the hairpin structures are formed from residues F18-F42 ( $\alpha_1$ ) and A49-L76 ( $\alpha_2$ ) in Rv0287, and A20-A39 ( $\alpha_1$ ) and Y52-M82 ( $\alpha_2$ ) in Rv0288. The helices in Rv0287 are completely  $\alpha$ -helical, whereas in

Rv0288, helix- $\alpha_2$  terminates with a single turn of  $3_{10}$  helix (S75-H77) followed by another single turn of  $\alpha$ -helix (E78-M82) as described in figure 3.3.23.



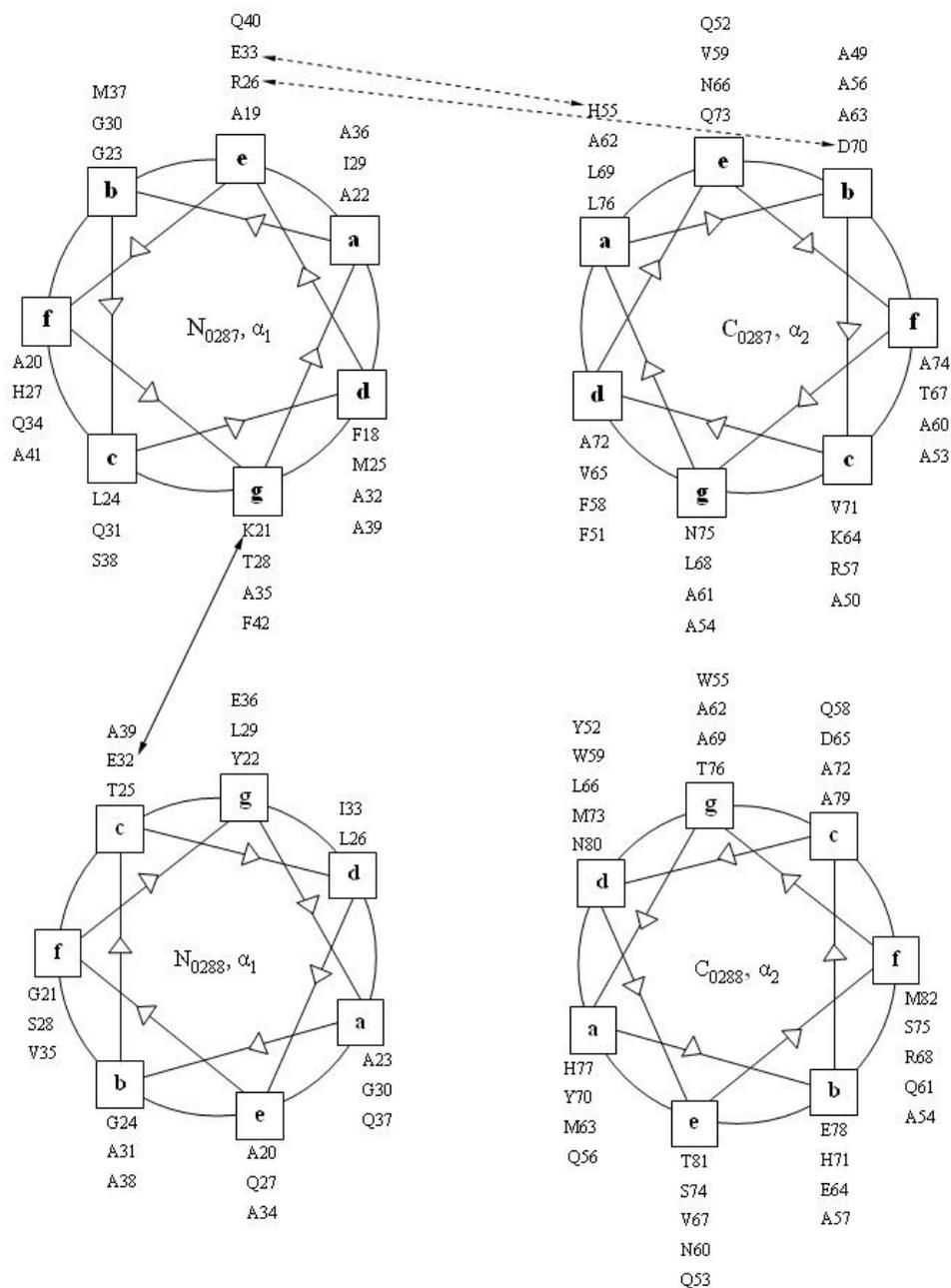
**Figure 3.3.23:** A summary of the Rv0288-Rv0287 residues present in the helical regions. Panels (A) and (B) show the positions of the helices in Rv0287 and Rv0288 respectively. Position of regular alpha helices (indicated by  $\alpha_1$  and  $\alpha_2$ ) and the  $3_{10}$  region are indicated above the sequences and highlighted in dark blue and blue respectively.

The extensive contact surface between Rv0287 and Rv0288 is essentially hydrophobic in nature and accounts for 9.1 % ( $1334.4\text{\AA}^2$ ) of the total surface area of both proteins. In the case of Rv0287, 19 residues (F18, K21, M25, T28, A32, A35, A36, F51, A54, R57, F58, A61, K64, V65, L68, V71, A72, N75, and L76) and for Rv0288, 21 residues (M19, Y22, L26, L29, E32, I33, E36, L40, A43, W44, T48, I50, W55, Q58, W59, A62, L66, A69, Y70, A72, and M73), contribute to the formation of intermolecular interface in the complex as detailed in figure 3.3.24 panels A – D. These residues were identified by significant reduction in their solvent accessibility ( $\geq 10\%$ ) upon complex formation using MOLMOL (1,110) and all of the residues present at the molecular interface are located in the helical regions of the complex. (Figure 3.3.24 panels B and D).



**Figure 3.3.24:** Analysis of the solvent accessibility of residues contained in Rv0287 and Rv0288 upon complex formation. Panels (A) and (C) show the change in solvent accessibility of individual residues contained in Rv0287 and Rv0288 upon complex formation. The solvent accessible surface (%) of individual residues in Rv0287 and Rv0288 are indicated by blue bars (for residues present in the isolated protein) and red bars (for residues when present in the complex). As shown in panels (A) and (C), a significant change in solvent accessibility can be clearly observed for the residues: F18, K21, M25, T28, A32, A35, A36, F51, A54, R57, F58, A61, K64, V65, L68, V71, A72, N75, and L76 in Rv0287, and M19, Y22, L26, L29, E32, I33, E36, L40, A43, W44, T48, I50, W55, Q58, W59, A62, L66, A69, Y70, A72, and M73 in Rv0288. This suggests the mentioned residues contribute to the formation of intermolecular interface in the Rv0287-Rv0288 complex. Panels (B) and (D) show primary sequence alignments of Rv0287 with Rv3874 (CFP-10) and Rv0288 with Rv3875 (ESAT-6) respectively and also compare the residues present in the intermolecular interface in the Rv0287-Rv0288 and CFP-10-ESAT-6 complexes. The positions of the helices present in the in Rv0287 and Rv0288 are also indicated above the sequences and highlighted in black in panels (A) and (C).

The stabilizing interactions between the helical hairpins and between their constituent helices rely almost entirely on favorable van der Waals contacts. This is evidenced by the fact that only small number of potential salt bridges and hydrogen bonds contribute to global folding of the Rv0287-Rv0288 complex (Figure 3.3.25). However, a potential intermolecular (K21-E32) salt bridge appears to stabilize the interaction between the N-terminal region of the helix- $\alpha_1$  in Rv0287 and the C-terminal region of the helix- $\alpha_1$  in Rv0288 (Figure 3.3.25). Close analysis of the structure also reveals that another two salt bridges (R26 – D70 and E33 – H55) are likely to contribute to the intramolecular interactions between the constituent helices in Rv0287 (Figure 3.3.25). The salt bridges were identified using MOLMOL by the spatial proximity of residue side chain acidic and basic groups (1, 110) which were considered to have a potential electrostatic interaction (salt bridge) when distances between them were less than 5 Å.



**Figure 3.3.25: Helical wheel representation of the Rv0287-Rv0288 complex.** The helical regions of Rv0287 and Rv0288 are depicted using helical wheel diagrams based on the structure of the Rv0288-Rv0287 complex. The potential intra-molecular salt bridges (E33 – H55 and R26 – D70) involving the residues at positions ‘e-a’, and ‘e-b’ in Rv0287 are shown by the dashed black arrows. The potential inter-molecular salt bridge (K21-E32) involving the position ‘g’ and ‘c’ is depicted by a solid black arrow. No potential hydrogen bond contributes to global folding of the complex.

There are a number of aromatic and hydrophobic residues located in the N- and C-terminal regions of Rv0287 (Y83 and A89) and Rv0288 (Y7, Y9, A90 and W95), which are highly conserved across the ESAT-6- CFP-10 protein family (Figures 3.3.26 and 3.3.27). These residues are not included in the structured core of the complex and are found on the flexible arms of the Rv0288-Rv0287 complex, implying some functional significance.

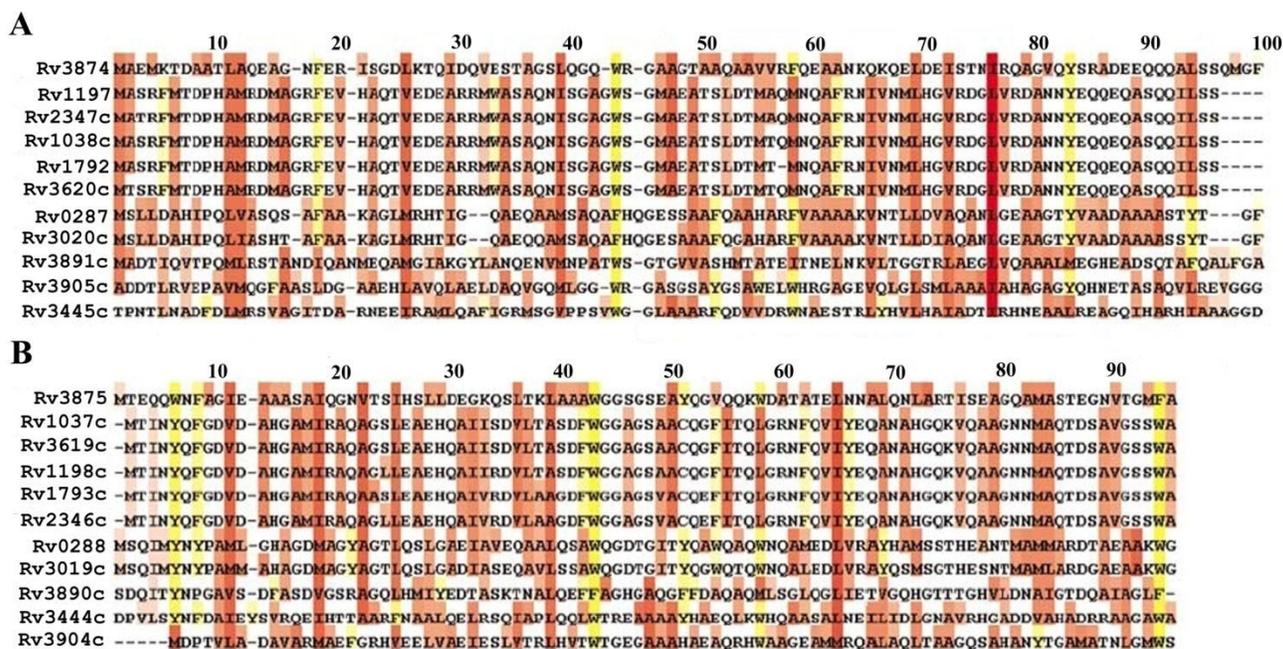


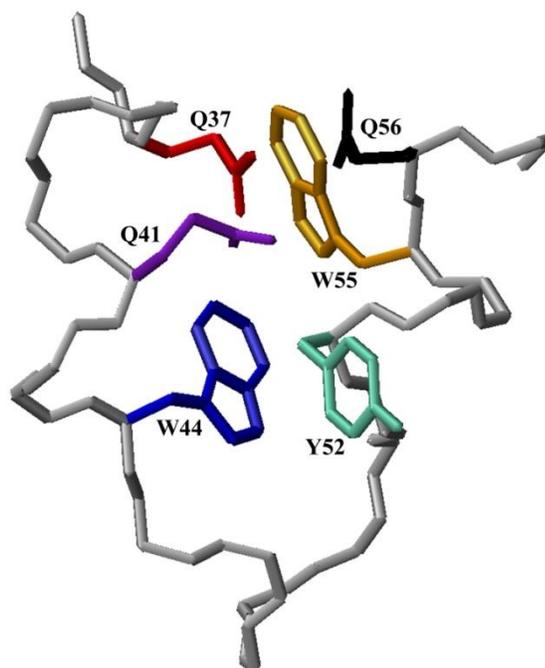
Figure 3.3.26: Multiple sequence alignment illustrates the conservation of the amino acid sequences of the CFP-10 (A) and ESAT-6 (B) related proteins from the *M. tuberculosis* ESAT-6 protein family. The residues are highlighted as follows: Aliphatic residues with hydrophobic side chains (Leu, Ile, Val, Met and Ala) in red, aromatic residues (Phe, Tyr and Trp) in yellow (1).

Both N- and C-terminal regions of Rv0287 and Rv0288 are also as well conserved between *M. tuberculosis* and *M. leprae* as the overall proteins (overall 76.84 % amino-acid sequence identity for Rv0287 / ML2532c and 70.83 % for Rv0288/ML2531c) despite the fact that they do not play any structural role in the complex, which again implies functional pressure to conserve these regions (Figure 3.3.27 panels A and B). It is, therefore, possible that, like



Rv0288 contributes to the specific binding of the Rv0287-Rv0288 complex to the surface of host cells.

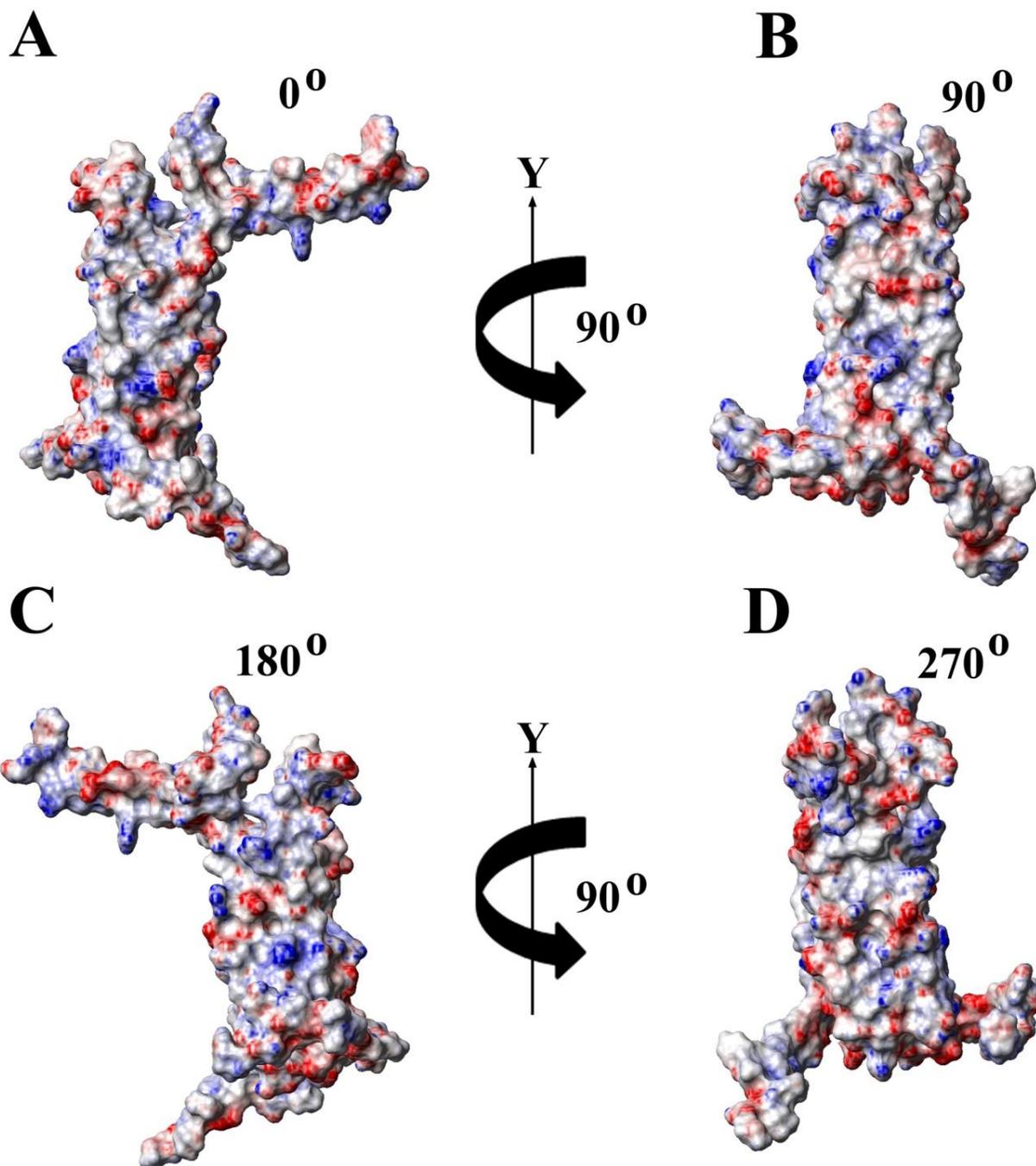
Presence of significant helical structure in Rv0288 (30 % helical content) but no apparent resistance to both heat- and chemical-induced denaturation (63) implies that Rv0288, like ESAT-6 (40, 63), exists in molten globule state (Discussed in section 3.3.6.5). Careful examination of the Rv0288-Rv0287 structure reveals that Rv0288, like ESAT-6 (1, 63), contains a mini hydrophobic core within the region forming the tight turn between the pair of the anti-parallel helices (Figure 3.3.28). The stability of this region mostly relies on the favorable van der Waals interactions between the side chains of Q37, Q41, W44, Y52, W55, and Q56 (Figure 3.3.28). The equivalent region of Rv0287 in the complex is not stabilized by a comparable network of interactions, suggesting the formation of the mini hydrophobic core is required to induce partial folding of Rv0288, and may also explain why isolated Rv0288 alone forms a molten globule, whereas Rv0287 exists as a random coil polypeptide.

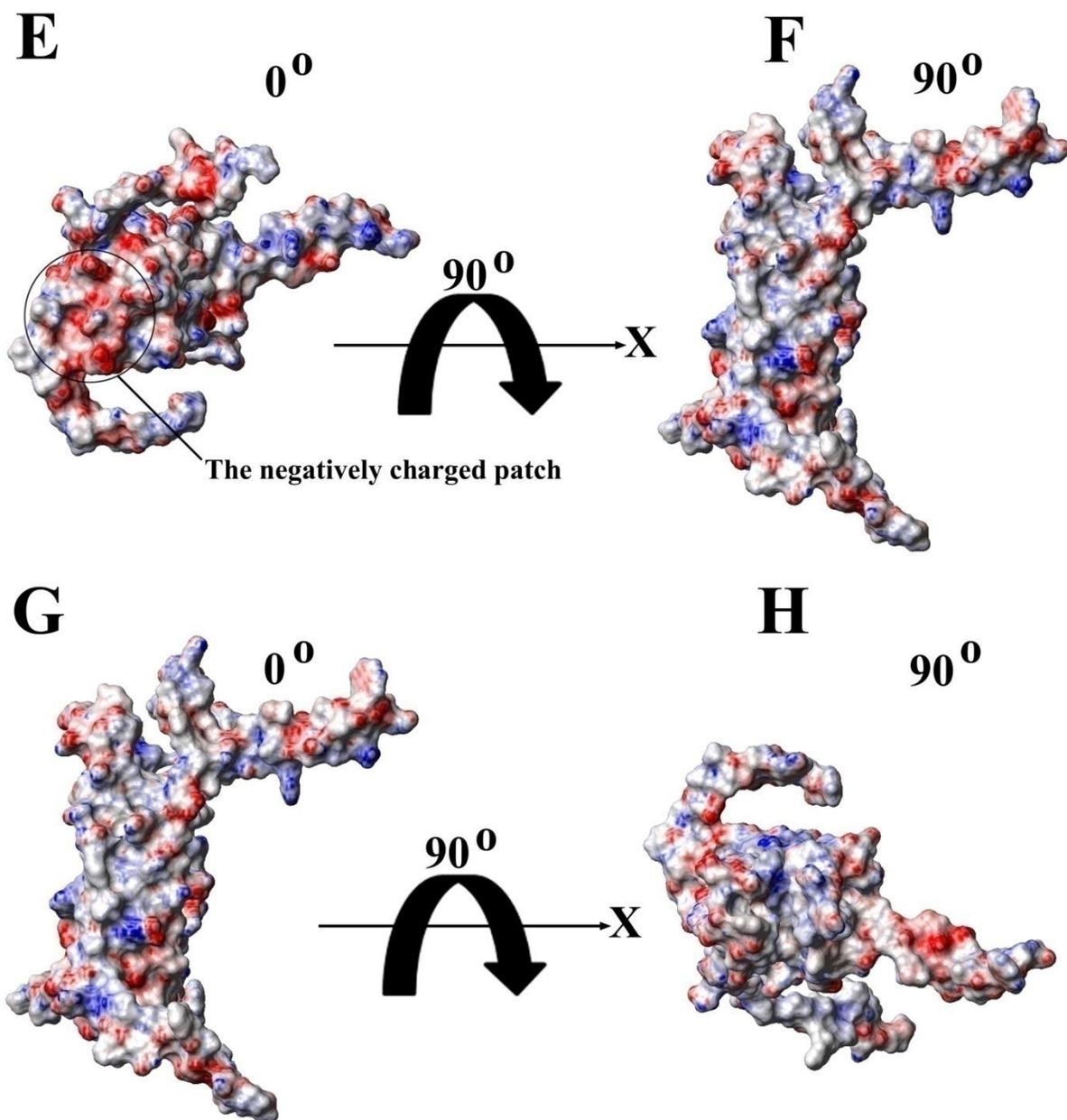


**Figure 3.3.28:** positions of key residues within the mini hydrophobic core of Rv0288. The panel shows the orientation of key amino acid side chains (Q37, Q41, W44, Y52, W55, and Q56) within the mini hydrophobic core formed at the tight turn between the two helices in Rv0288. The backbone of the protein in this region is shown in gray.

Analysis of the electrostatic surface of the complex reveals that the surface of the complex is predominantly charged with an uneven distribution of positive and negative charge (Figure 3.3.29). This, along with solubility of the complex to over 1.2 mM in aqueous solution, implies that the complex is unlikely to act as a pore forming protein.

However, the surface contains a striking acidic region, which is predominantly formed by three residues (E78 from Rv0287 and E36 and D47 from Rv0288) (Figure 3.3.29, panel E). This implies the complex may use electrostatic interaction to interact with other proteins under physiological condition.

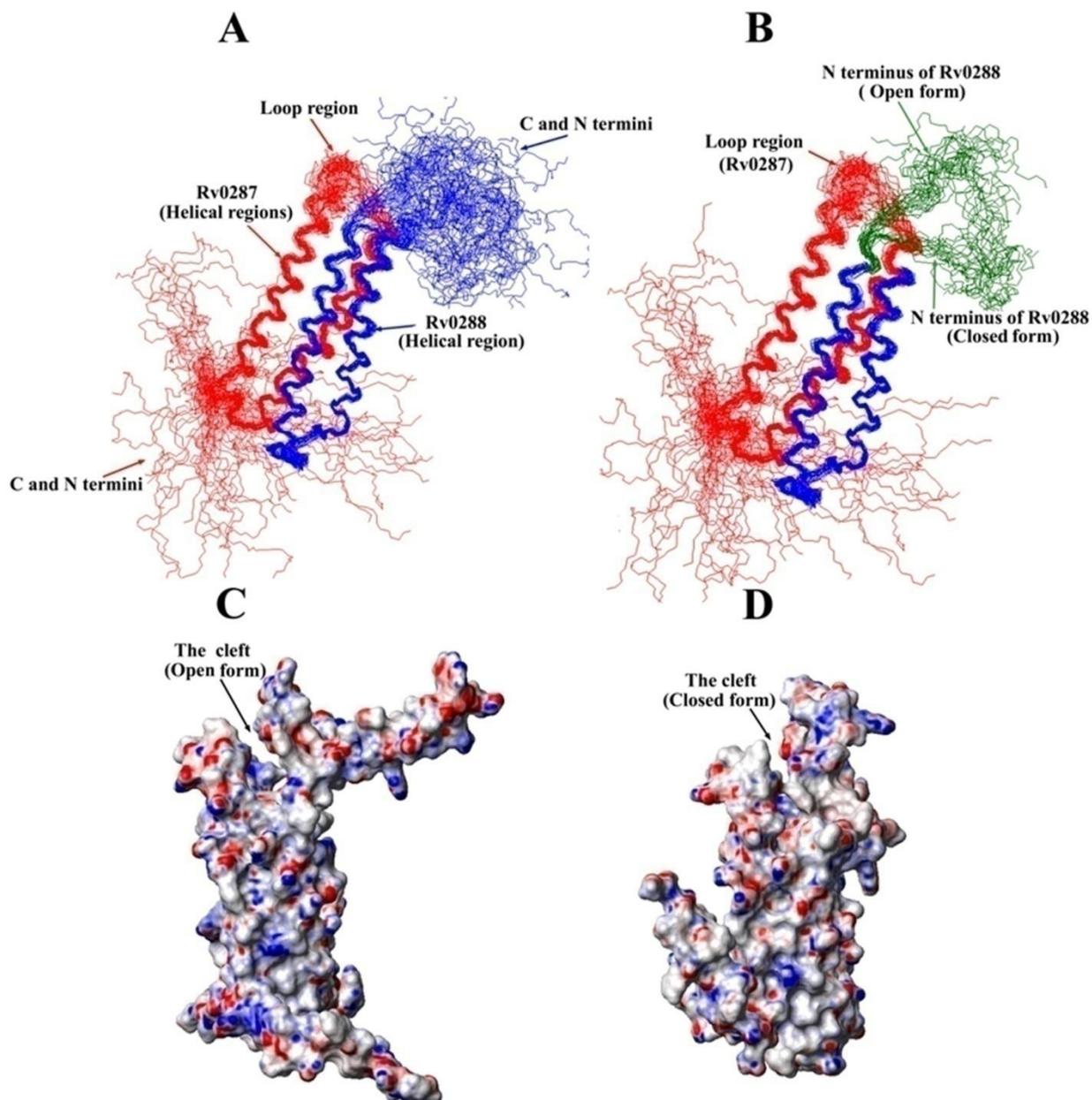




**Figure 3.3.29:** Space-filled views of the surface of the Rv0288-Rv0287 complex based on the converged structure closest fit to the mean. Panels A – D depict the different views of the complex which are rotated by  $90^\circ$  about the ‘Y’ axis, whereas panels E-H show the space-filled views of the complex rotated by  $90^\circ$  about ‘X’ axis. Panel (E) also focuses on the significant negatively charged patch of the complex where is formed predominantly by three residues (E78 in Rv0287 and E36 and D47 in Rv0288). The surface is colored according to electrostatic potential, with areas of significant negative charge in red, significant positive charge in blue and neutral as white. The electrostatic potential was calculated using MOLMOL (110), with the threshold for depicting significant areas of charge chosen to obtain a neutral representation for the fully exposed aromatic ring of W95 in Rv0288.

Another striking feature of the complex is the presence of a significant cleft/shelf in the surface of the structure, which may indicate a possible binding site. This region is predominantly formed by the area between the loop of the Rv0287 hairpin (formed by residues: A35, A36, M37, S38, A39, Q40, A41, F42, H43, Q44, G45, E46, S47, S48, A49, A50, F51, Q52, A53, and A54), the N-terminal flexible region of Rv0288 (formed by residues: S1, M2, S3, Q4, I5, M6, Y7, N8, Y9, P10, A11, M12, L13, G14, H15, A16, G17, D18, and M19) and C-terminal region of the  $\alpha_2$  helix in Rv0288 (formed by residues: A72, M73, S74, S75, T76, H77, E78, A79, N80, T81, M82) (figure 3.3.30). The above mentioned residues were identified using MOLMOL (110) by the analysis of residues forming the cleft/shelf in 10 conformers.

Interestingly, comparison of the protein backbone for the family of converged structures (30 conformers) reveals two distinct forms (the open and closed forms) for the cleft/shell which are almost equally populated (16 out 30 conformers show the open form) (Figure 3.3.30). The switch between the two forms mainly results from a change in the spatial position of N-terminal flexible region of Rv0288 as illustrated in figure 3.3.30, panel B.



**Figure 3.3.30: Stereo and space-filled views of the Rv0288-Rv0287 complex. Panels (A) and (B) show a best fit superposition of the protein backbone for the family of 30 converged structures obtained, with Rv0287 shown in red and Rv0288 in blue. Panel (B) also focuses on the regions of the structure which contribute to formation of the cleft/shelf (For clarity only the N-terminal flexible arm of Rv0288 (green), and the loop region in Rv0287 are highlighted). The change in spatial position of the N-terminal flexible region of Rv0288 (green) is also illustrated in panel B. Panels (C) and (D) illustrate the space-filled views of the open and closed forms of the cleft respectively. The cleft is indicated by the arrows. The orientation of the complex in panel (C) is identical to that shown in panel (D).**

### ***Comparative analysis between Rv0288-Rv0287 and its structurally homologue proteins***

The backbone topology determined for the Rv0287-Rv0288 complex allowed searches to be made for structural similarity with proteins of known structure and function using fold comparison packages such as DALI (Distance mAtrix aLIgnment) (117) (Section 3.2.15).

A homology search with DALI reveals the closest matches to be the ESAT-6-CFP-10 (1WA8; RMSD 1.8 Å° for 93 Cα atoms) (1), *E. coli* stress protein YciF (2GS4; RMSD 2.97 Å° for 103 Cα atoms)(112), DpsA from *Halobacterium salinarum* (1TJO; RMSD 2.97 Å° for 93 Cα atoms) (113), bacterioferritin from *Desulfovibrio desulfuricans* (1NF6; RMSD 3.07 Å° for 100 Cα atoms) (114), Antioxidant Dps protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus* (2C41, RMSD 3.091 Å° for 100 Cα atoms) (115), and MsDps protein from *Mycobacterium smegmatis* (2Z90, RMSD 2.67 Å° for 83 Cα atoms) (116).

YciF is an *E. coli* protein which is highly conserved in a range of bacterial species. It has been shown that the protein is up-regulated when the bacterium encounters stress conditions (112). Furthermore, it has been suggested that the YciF forms a dimer protein that may act as a metal-binding protein in vivo. Similar to the YciF protein, the other top hits form multimers and act as metal binding proteins, contributing to bacterial iron homeostasis (112, 113, 114, 115, and 116). Furthermore, it has been suggested that DNA-binding proteins from starved cells (the Dps proteins) protect bacteria primarily from oxidative damage (113, 115,116).

Recent studies have shown that expression of the *Rv0287/Rv0288* gene cluster (from *Rv0282* to *Rv0292*) is induced under conditions of iron and zinc starvation *in vitro* suggesting a possible role in the scavenging or uptake of iron and zinc (89). This, along with the results obtained from the homology search, may suggest that the Rv0287-Rv0288 complex acts as a metal binding protein. However, the sequence identity between the Rv0288-Rv0287 complex and the metal binding proteins described above was found to be very low (between 2 to 9 %) (Figure 3.3.31). Furthermore, the detailed sequence analysis revealed that functional residues of the metal binding proteins, which contribute to the formation of multimers and metal-binding sites, are not conserved in the Rv0288-Rv0287 complex as described in figure 3.3.31. Taken together, similarities between the Rv0288-Rv0287 complex and YciF or Dps proteins are limited only to the monomer fold, and do not extend to the overall architecture of the molecules.

**A**

```

YciF      .....XKTIeDVFIHLLSDTYSAEKQLTRALAKLARA...tSNEKLSQAFHABLE
Rv0287_Rv0288  mslldahipqLVASqSAFAAKAGLMRHTIGQAEQAAMSAQAFhggesSAAFQAAHARFVA
          * * * * *
YciF      ETHGQIERIDQVVESESNLK...ikrpkcvaxeglieeaneviestEKNEvrDAALiaAA
Rv0287_Rv0288  AAAKVNtLLDVAQANLGEAAgtyvaadaaaastytgfsmsqimynyPAML..GHAG..DM
          * * * * *
YciF      QKVEHYEIASYGTLATLAEQLG.....YRKAAKLLKETLEEEKATDIKLTDLAIN
Rv0287_Rv0288  AGYAGTLQSLGAEIAVEQAALQsawqgdgtgityQAWQAQWNQAMEDLVRAYHAMSSTHE.
          * * * * *
Rv0287_Rv0288  NVNKK.....
YciF      ANTMAMmardtaeaakwgg

```

**B**

```

DpsA      stqknarataceveqsdalrMDADraEOCVDALNADLANVYVLYHOLKKhWNVE.qAEF
Rv0287_Rv0288  .....mslldahipqlVASQ..SAFAAKAGLMRHTIGQAEQAAMSA.QAFHggESS
          * * * * *
DpsA      RDLHLFLGEEAAETAEEVADELAERVQA.LGGV.....phaspetylqaeasvdeded
Rv0287_Rv0288  AAFQAAHARFVA...AAKVNtLLDVAQANLGEAAgtyvaadaaaastytgfsmsqimynypa
          * * * * *
DpsA      vyDIRTslANDMAIYGDIIETREHTELAENLG.....dHATAHMLREGLIELED
Rv0287_Rv0288  mlGHAG...DMAGYAGTLQSLGAEIAVEQAALQsawqgdgtgityQAWQAQWNQAMEDLVR
          * * * * *
DpsA      DAHHIEHYLED..DTLV..TQAL.....
Rv0287_Rv0288  AYHAMSSTHEAntMAMMarDTAEaakwgg
          * * * * *

```

**C**

```

Bacterioferritin .....aGNREDrkAKVIEVLNKA RAMELHAIHQYMNQHYSlDDMDYGELAAANMKLI
Rv0287_Rv0288  mslldahipqLVASQ..SAFAAKAGLMRHTIGQAEQAAMSAQAF.HQGES SAAFQAAHAR
          * * * * *
Bacterioferritin AID.EMRHAENFAERIKE.LGGE.....pttqkegkvvtgqAVPviye
Rv0287_Rv0288  FVaaAAKVNtLLDVAQANLGEAAgtyvaadaaaastytgfsmsqimynypamlGHAG...
          * * * * *
Bacterioferritin SDADQEDATIEAYSQFLKVCKEQG.....dIVTARLFERIIEEQAHlTYyenIG
Rv0287_Rv0288  DMAGYAGTLQSLGAEIAVEQAALQsawqgdgtgityQAWQAQWNQAMEDLVRAYHAMSSTH
          * * * * *
Bacterioferritin ShIKNLGdtylaktpsstqtasakfv
Rv0287_Rv0288  E.ANTMA.....mmardtaeaakwgg

```

**D**

```

MsDps2     mSARRtesdiacqfhatPEFGGnLQKVLVDLIELSLQKQAHWNVV.....GSNFRDLH
Rv0287_Rv0288  .MSLL...dahipqlvASQSA.FAAKAGLMRHTIGQAEQAAMSAQafhggesSAAFQAAH
          * * * * *
MsDps2     LQLDELVDFAREGSDTIAER...MRALD...avpdgrsdtvaatttlpefpaferstaD
Rv0287_Rv0288  ARFVAAA KVNtLLDVAQANlgeAAGTYvaadaaaastytgfsmsqimynypamlghagD
          * * * * *
MsDps      VVDLiTTRINATVD TIRRVHDAVDAED....pSTADLLHGLIDGLEKQAWLIRS.ENRK
Rv0287_Rv0288  MAGY.AGTLQSLGAEIAVEQAALQSAWqgdgtgityTYQAWQAQWNQAMEDLVRAYHAMSSTH
          * * * * *
MsDps2     V.....
Rv0287_Rv0288  Eantmammardtaeaakwgg

```

Figure 3.3.31: Structural alignment of the Rv0288-Rv0287 complex with its structural homologues proteins. Panels A-D show the structural alignments of the complex with *E. coli* stress protein YciF, *Halobacterium salinarium* DpsA, Bacterioferritin, and *Mycobacterium smegmatis* MsDPs respectively. The structurally equivalent residues are in uppercase and highlighted in blue. The structurally non-equivalent residues are in lowercase. The functional residues of the metal binding proteins, which contribute to the formation of multimers and metal-binding sites, are highlighted in red. The blue asterisks show the identical residues in the alignments.

### *Comparative analysis between Rv0287-Rv0288 and CFP-10 – ESAT-6*

The solution structure of the CFP-10-ESAT-6 complex has recently been reported (1) and from comparison of the structures shown for the Rv0287-Rv0288 and CFP-10-ESAT-6 complexes it is clear that the overall backbone fold for the complexes is very similar as shown in figure 3.3.32. Comparisons of the backbone atom coordinates yield a r.m.s.d. of 1.8 Å for the superposition of residues F18-Q40 (Rv0287), A50-L76 (Rv0287), A20-A39 (Rv0288) and Y52-S74 (Rv0288) of the Rv0287-Rv0288 complex and residues F17-Q39 (CFP-10), A49-I75 (CFP-10), Q19-K38 (ESAT-6) and Y51-A73 (ESAT-6) of the CFP-10-ESAT-6 complex respectively.

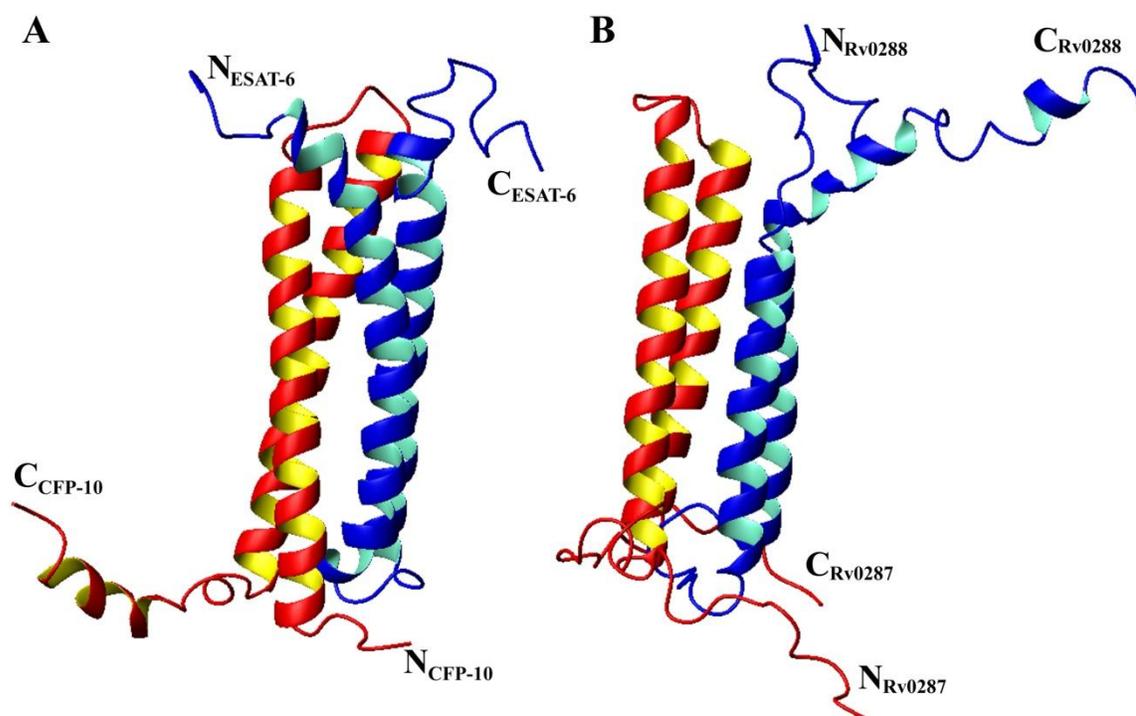


Figure 3.3.32: Comparison of the solution structures of the Rv0288-Rv0287 and ESAT-6 -CFP-10 complexes. Panels (A) and (B) show ribbon representations of the backbone topology of the CFP-10-ESAT-6 and Rv0287-Rv0288 complexes based on the converged structure closest to the mean, which illustrates the two helix–turn–helix hairpin structures formed by the individual proteins in each complex. The orientation of the CFP-10-ESAT-6 complex (A) is identical to that of the Rv0287-Rv0288 complex (B), with CFP-10 and Rv0287 in red, and ESAT-6 and Rv0288 in blue. The helical propensity of the flexible C-termini of CFP-10 (residues 85–95) and Rv0288 (residues 83–97) can be clearly seen in panels A (bottom left) and B (top right) respectively.

Furthermore, a detailed sequence analysis of the complexes reveals that 35 % of the residues which are involved in the intermolecular interface of the Rv0287-Rv0288 complex (F18, A54, F58, A61, K64 and N75 in Rv0287, and L29, E32, L40, A43, W44, W59, L66, and A69 in Rv0288) are identical to those of the CFP-10-ESAT-6 complex (Figure 3.3.33). These residues are located at the same positions on the helical wheel diagrams of the corresponding complexes (Figure 3.3.34, panels A and B), further supporting the remarkable global similarities between the complexes (similar global folds of the complexes).

### A

```

Rv0287      10      20      30      40      50
MSLLDAHIPQLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAFQAAHAREVA 60
CFP-10     MAEMKTDAATLAQEAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGTAAQAAVVFQOE 60
*           *           *           *           * * * * * * * * * * * * * *
Rv0287      70      80      90
AAAKVNTLLDVAQANLGEAAGTYVAAD---AAAASTYTG 97
CFP-10     AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
** *      ** *      * *      * *      * *      * *      * *      * *

```

### B

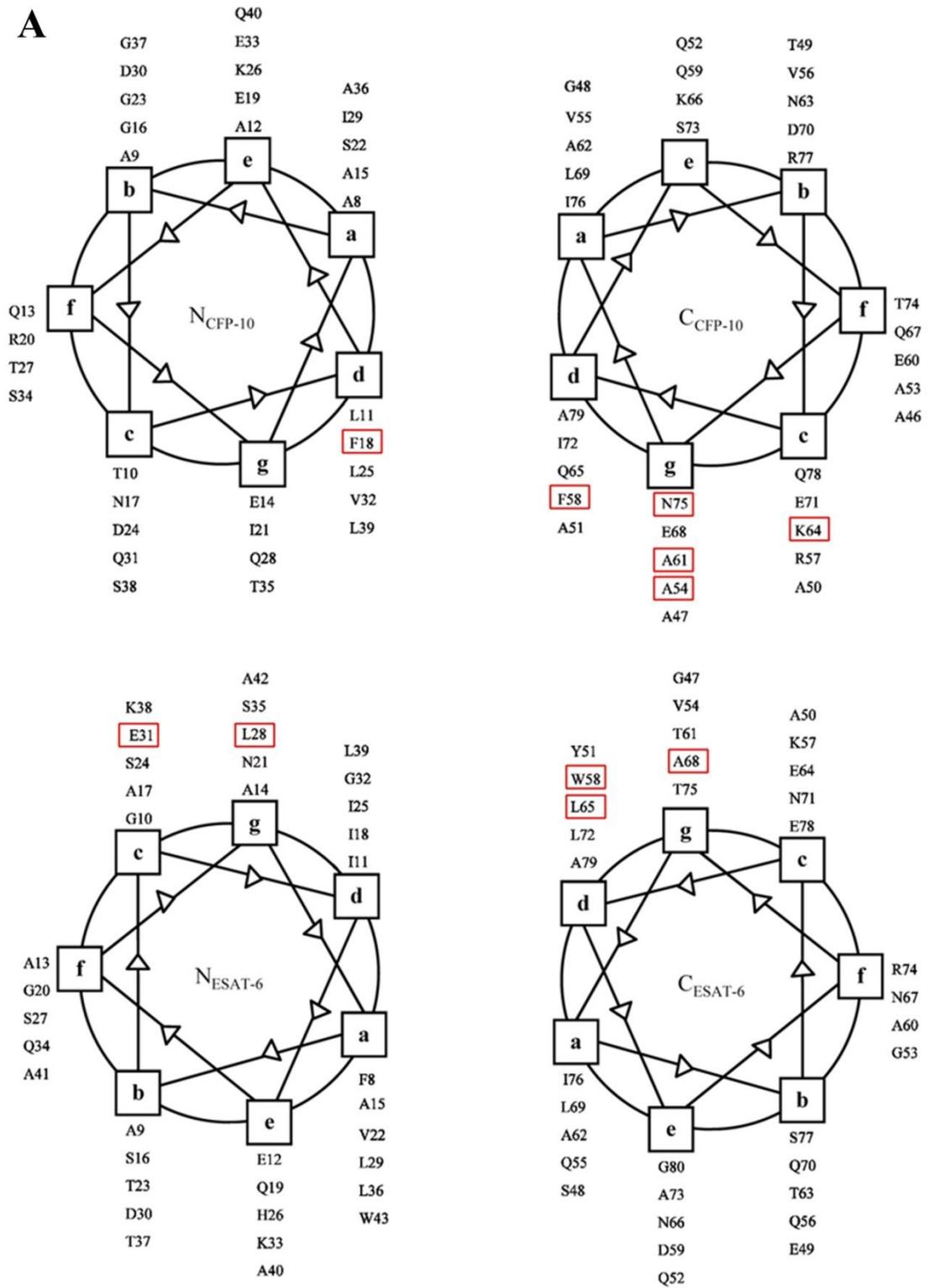
```

Rv0288      10      20      30      40      50
SMSQIMYNYPAMLGHAGDMAGYAGTLQSLGAEIAVEQAALQSAWQGDGTGITYQAWQAQWN 60
ESAT-6     -MTEQQWNFAGIEAAASAIQGNVTSIHSLLDGKQSLTKLAAAWGGSGSEAYQGVQQKWD 60
*           *           *           *           * * * * * * * * * * * * * *
Rv0288      70      80      90
QAMEDLVFRAYHAMSSTHEANTMAMMARDTAEAAKWGG 97
ESAT-6     ATATELNALQNLARTISEAGQAMASTEENVTGMFA- 95
* *      * *      * *      * *

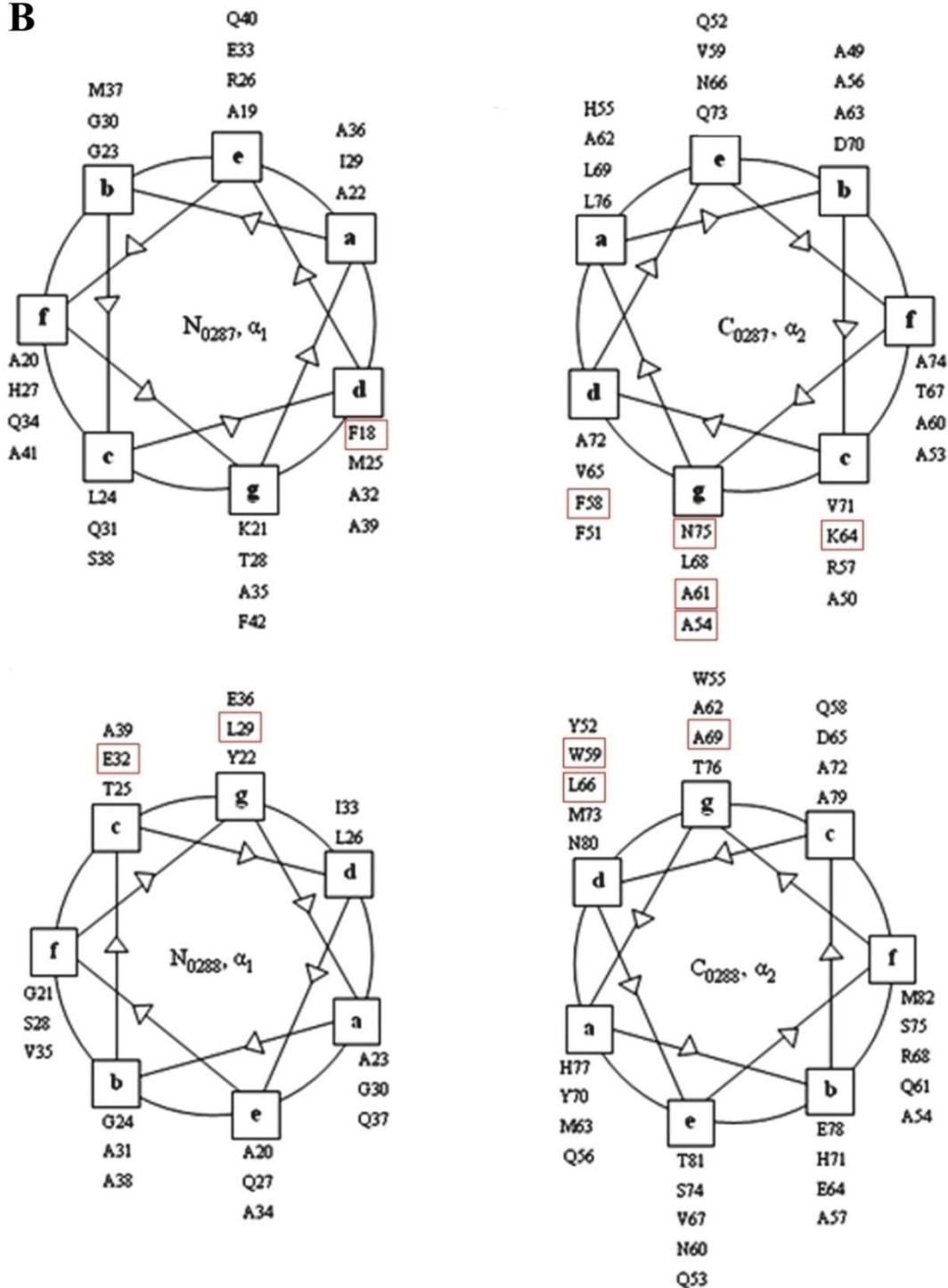
```

Figure 3.3.33: Sequence alignment of the Rv0288-Rv0287 complex with CFP-10-ESAT-6 complex. Panels (A) and (B) show primary sequence alignments of Rv0287 with CFP-10 (Rv3874) and Rv0288 with ESAT-6 (Rv3875) respectively. The panels focus on the residues which are 100 % conserved in the intermolecular interface of the Rv0287-Rv0288 and the CFP-10-ESAT-6 complexes. These residues are highlighted by red squares. The asterisks show the identical residues in the alignments. The sequences were aligned using ClustalW, with a standard Blosom30 scoring matrix, a gap opening penalty of 10 and a gap extension penalty of 1 (118).

**A**

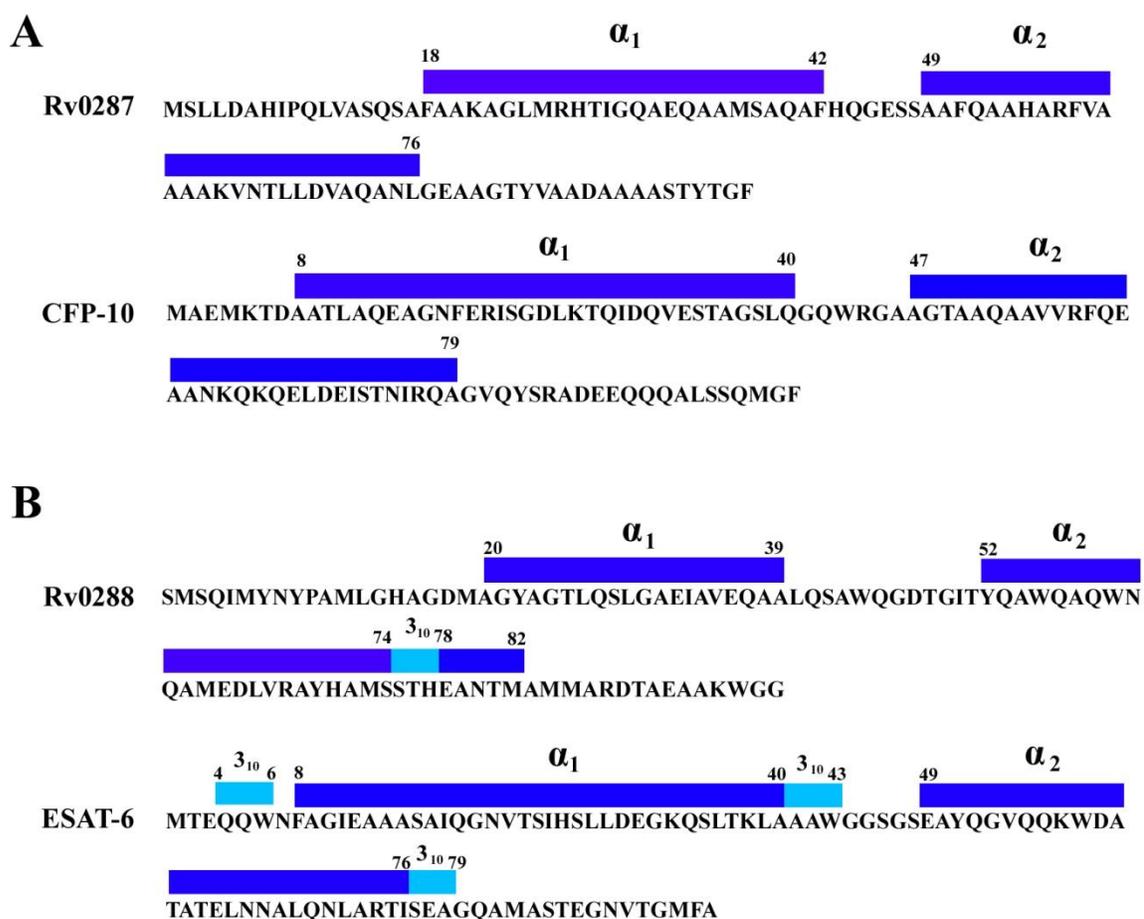


**B**



**Figure 3.3.34:** Helical wheel presentations of the CFP-10-ESAT-6 and Rv0287-Rv0288 complexes. The helical regions of CFP-10 and ESAT-6 (panel A), and Rv0287 and Rv0288 (panel B) are depicted using helical wheel diagrams based on the structures of the corresponding complexes. Panels (A) and (B) also show the positions of residues which are 100 % conserved in the intermolecular interface of the CFP-10-ESAT-6 ( F18, A54, F58, A61, K64, and N75 in CFP-10, and L28, E31, W58, L65, and A68 in ESAT-6) and Rv0287-Rv0288 ( F18, A54, F58, A61, K64, and N75 in Rv0287, and L29, E32, W59, L66, and A69 in Rv0288) complexes. These residues are highlighted by red squares.

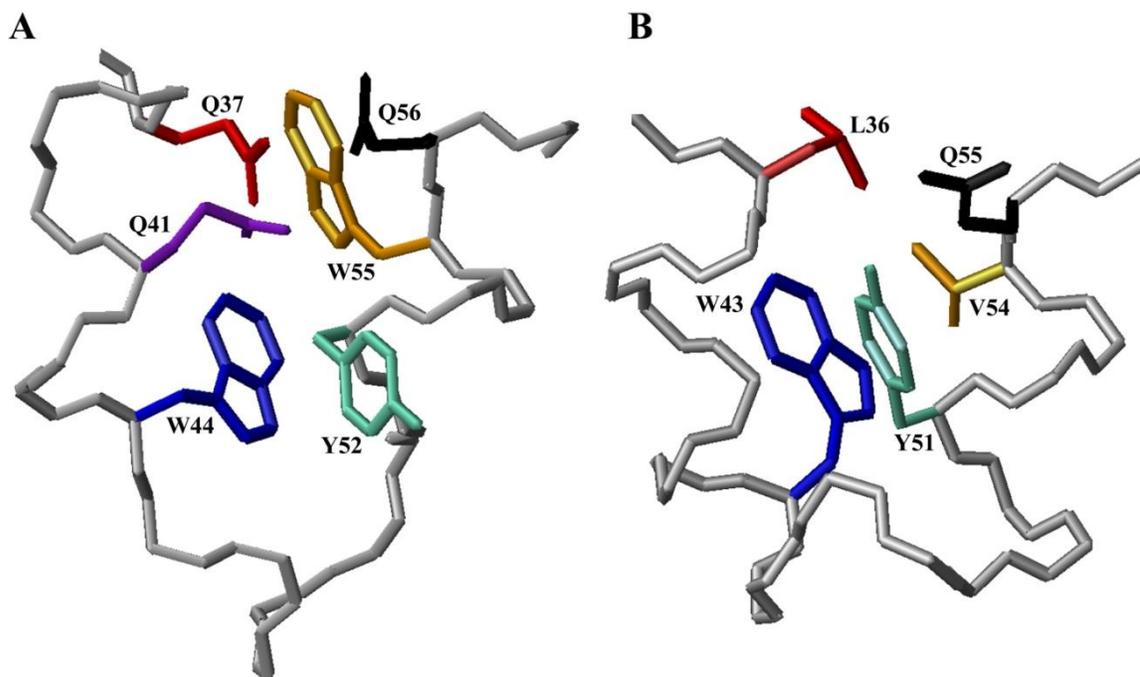
Further comparisons of the two complexes shows that the Rv0288-Rv0287 complex contains less helical content (53.6 %) than the ESAT-6-CFP-10 complex (68.2 %) (Figures 3.3.34 and 3.3.35). In addition, like CFP-10, the helices in Rv0287 are completely  $\alpha$ -helical. However, unlike ESAT-6 where both long helices terminate with a single turn of  $3_{10}$  helix, only helix- $\alpha_2$  of Rv0288 terminates with a single turn of  $3_{10}$  helix (S75-H77). The  $3_{10}$  helix in Rv0288 is in turn followed by another single turn of  $\alpha$ -helix (E78-M82) as described in figure 3.3.35.



**Figure 3.3.35: Comparison of helical regions contained in the CFP-10-ESAT-6 and Rv0287-Rv0288 complexes. Panel (A) compares the positions of the helices between in Rv0287 and CFP-10. Panel (B) compares the positions of the helices between in Rv0288 and ESAT-10. As illustrated in panels (A) and (B), the Rv0287-Rv0288 complex contains less helical structure than the related CFP-10-ESAT-6 complex. Position of regular alpha helices (indicated by  $\alpha_1$  and  $\alpha_2$ ) and the  $3_{10}$  region are indicated above the sequences and highlighted in dark blue and blue respectively.**

Similar to the ESAT-6-CFP-10 complex, a striking feature of the Rv0288-Rv0287 complex is the disordered N- and C-termini of both proteins, which form long flexible arms at both ends of the four-helix bundle core. However, unlike CFP-10 complex where the C-terminal region has a distinct tendency to adopt a helical conformation, the C-terminal region of the ESAT-6 related protein in the Rv0288-Rv0287 complex (Rv0288) has a propensity to adopt a helical conformation (Figure 3.3.32, page 160).

As previously explained, examination of the Rv0288-Rv0287 structure revealed that Rv0288, like ESAT-6, contains a mini hydrophobic core within the region forming the tight turn between the pair of the anti-parallel helices. Comparison of the mini hydrophobic regions of Rv0288 and ESAT-6 reveals the greater number of stabilizing van der Waals interactions in the mentioned region of Rv0288 compared to ESAT-6 (Figure 3.3.36). The structure of this region of Rv0288 is mostly stabilized by favorable van der Waals interactions between the side chains of six residues (Q37, Q41, W44, Y52, W55, and Q56), whereas in ESAT-6, the van der Waals contacts between the sides chains of five residues (L36, W43, Y51, V54 and Q55) stabilize the structure of related region (Figure 3.3.36).



**Figure 3.3.36:** Comparison of the mini hydrophobic cores formed at the tight turn between the two helices in Rv0288 and ESAT-6. Panels (A) and (B) shows the orientation of key amino acid side chains within the mini hydrophobic cores of Rv0288 (Q37, Q41, W44, Y52, W55, and Q56) and ESAT-6 (L36, W43, Y51, V54 and Q55) respectively. The backbone of the protein in this region is shown in grey.

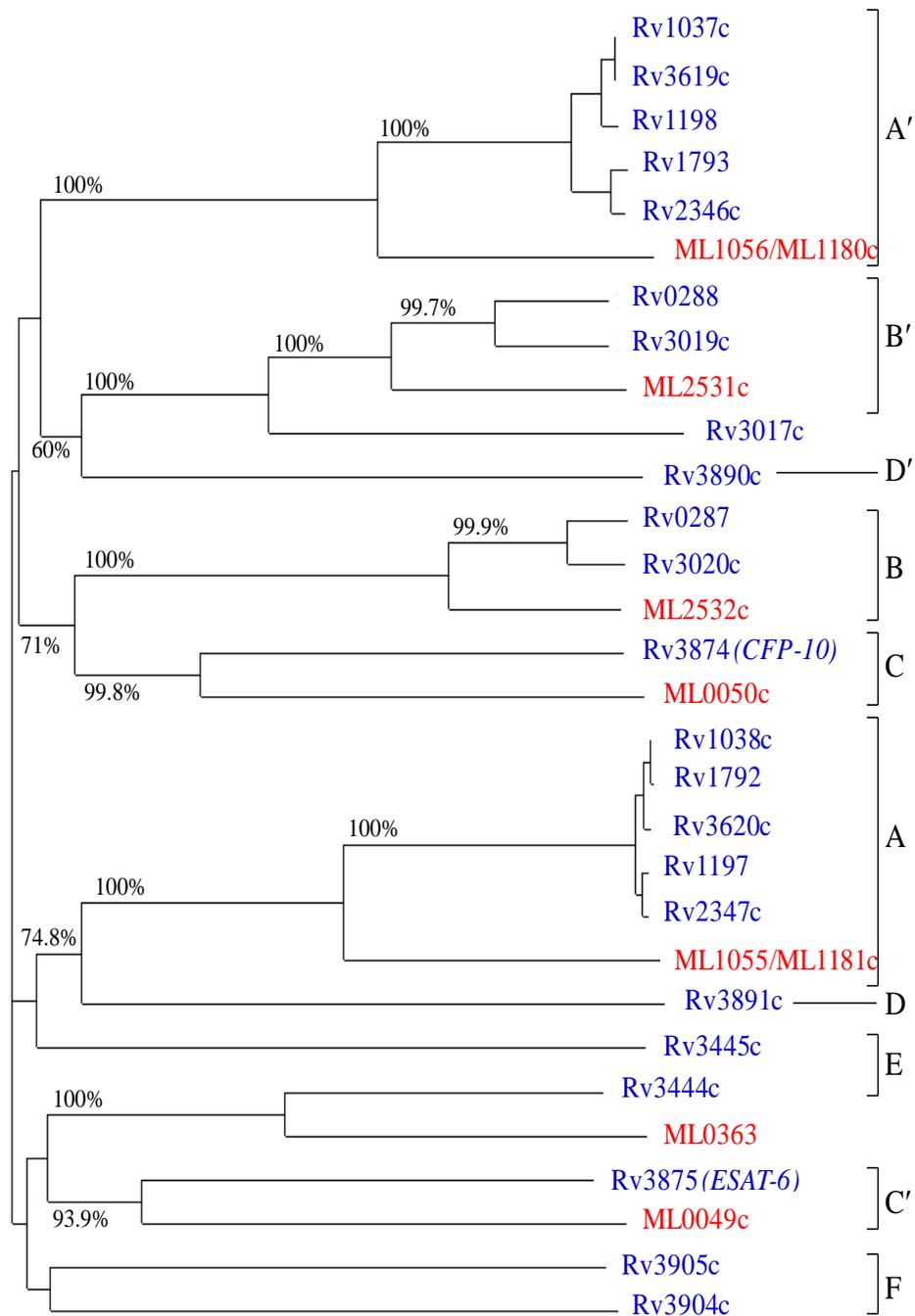
Moreover, comparative analysis of the surface of the complexes clearly shows that both Rv0287-Rv0288 and CFP-10-ESAT-6 are predominantly charged. However, unlike the CFP-10-ESAT-6 complex, the Rv0288-Rv0287 complex contains an uneven distribution of positive and negative charge (Figure 3.3.29, page 154) with a distinct negatively charge region. This may suggest the complex may use an electrostatic interaction to interact with other proteins under physiological condition.

Another striking surface feature that differentiates the two complexes is the presence of a significant cleft/shelf in the surface of the Rv0288-Rv0287 structure. This cleft/shelf is found in the two distinct forms (the closed and open forms) (Figure 3.3.30, page 156), suggesting a possible binding site for the region in the complex.

### ***Prediction of the structure of Rv3019c-Rv3020c***

*Rv3019c* and *Rv3020c* are also the members of the *esx*-like genes (*esat-6* / *cfp-10* related genes), which in common with *esat-6* and *cfp-10* genes, are found as tandem pairs in the *Mycobacterium tuberculosis* genome (1, 40, 62, 63, and 65). It is believed that the *M. leprae* contains the minimal gene sets required for a pathogenic mycobacterium (1, 63, and 65). The retention of the homologues genes for the *Rv3019c* and *Rv3020c* as functional genes in *M. leprae* (ML2531/ML2532) suggests their importance in the lifestyle of virulent mycobacteria (65) (Figure 3.3.37). However, close analysis of the of the *M. leprae* genome reveals ML2531/ML2532 seem to substitute for both the *Rv0287/Rv0288* and *Rv3019c/Rv3020c* pairs from *M. tuberculosis* (40, 65) (Figure 3.3.37). This, together with the significant sequence similarities between the two couples (*Rv0287/Rv0288* and *Rv3019c/Rv3020c*) (Figure 3.3.38), implies some redundancy within the genome. This is further supported by the fact that the *Rv3019c/Rv3020c* locus does not contain the flanking genes required for the secretion of *Rv3019c/Rv3020c* (65).

Analysis of the multiple sequence alignment (Figure 3.3.38) reveals significant similarities between *Rv0288* and *Rv3019c* (about 85 % identity), and *Rv0287* and *Rv3020c* (92 % identity). This, together with known complex formation for two pairs of ESAT-6/CFP-10 family protein (*Rv0288-Rv0287* and ESAT-6/CFP-10), strongly suggests that *Rv3019c* and *Rv3020c* are highly likely to interact with each other to form a similar four-helix bundle containing complex.



**Figure 3.3.37: Bootstrapped phylogenetic tree for the ESAT-6-CFP-10 protein family from *M. tuberculosis* (prefixed by Rv, blue) and their *M. leprae* homologues (prefixed by ML, red). There are 23 family members of ESAT-6-CFP-10 family, 22 of which are arranged as tandem pairs in 11 loci, in *M. tuberculosis* genome. Major pairing are shown by square brackets and labelled (A pairs with A' etc.) (40). As shown in the three, ML2531/ML2532 substitute for both the Rv0287/Rv0288 and Rv3019c/Rv3020c pairs from *M. tuberculosis***

**A**

```

Rv0287      MSLLDAHIPQLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAFQAAHARFVA
Rv3020c     MSLLDAHIPQLIASHTAFAAKAGLMRHTIGQAEQAAMSAQAFHQGESAAAFQGAHARFVA
***** ** *****

Rv0287      AAKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF
Rv3020c     AAKVNTLLDIAQANLGEAAGTYVAADAAAASYTGF
***** *****

```

**B**

```

Rv0288      MSQIMYNYPAMLGHAGDMAGYAGTLQSLGAEIAVEQAALQSAWQGDGTGITYQAWQAQWNQ
Rv3019c     MSQIMYNYPAMMAHAGDMAGYAGTLQSLGADIASEQAVLSSAWQGDGTGITYQGWQTQWNQ
***** ***** ** * *****

Rv0288      AMEDLVRAYHAMSSTHEANTMAMMARDTAEAAKWGG
Rv3019c     ALEDLVRAYQSMSTHE SNTMAMLARDGAEAAKWGG
* ***** ** *** ***** ** *****

```

**Figure 3.3.38: Sequence alignment of Rv0287 / Rv0288 and their closely related homologues Rv3020c / Rv3019c. Panels A and B show sequence alignments of Rv0287 / Rv3020c and Rv0288 / Rv3019c respectively. The residues present in the helical regions of the Rv0287-Rv0288 complex are highlighted in blue. The asterisks show the identical residues in the alignments. The sequence alignments are generated by ClustalW with a standard Blosum scoring matrix, a gap opening penalty of 10 and a gap extension penalty of 1 (118).**

### 3.3.7.6. Conclusion

Rv0287 and Rv0288 are the members of the *esx*-like genes (*esat-6* / *cfp-10* related genes), which in common with *esat-6* and *cfp-10*, are found as tandem pairs in the *Mycobacterium tuberculosis* genome (1, 40, and 63). The retention of the homologues of Rv0288 and Rv0287 as functional genes in *M. leprae* clearly signifies the importance of these genes in the lifestyle of the *M. tuberculosis*. Experimental studies have shown that Rv0287 and Rv0288, like ESAT-6 and CFP-10, are secreted proteins recognized by T- cells of the majority of tuberculosis patients (> 70 %) (27, 29, 36, 72, 82, and 86).

The work reported here clearly demonstrates that the Rv0287 and Rv0288 form a tight, 1:1 complex, which is considerably more stable than the related CFP-10-ESAT-6 complex. Similar to CFP-10-ESAT-6, Rv0287-Rv0288 complex formation is coupled to folding of

the constituent proteins as evidenced by the fact that the isolated proteins exist as molten globules (Rv0288) and unstructured polypeptides (Rv0287) whereas the complex forms a stable folded structure. This along with the fact the complex is stable at low pH ( pH 4.5) (63), strongly suggests that the functional forms of Rv0288 and Rv0287 are highly likely to be a 1:1 complex under physiological condition.

As previously predicted (1, 63), analysis of the solution structure of the Rv0288-Rv0287 complex shows that the complex, like the related ESAT-6-CFP-10 complex, contains a well-defined four helix bundle core. Furthermore, similar to the ESAT-6-CFP-10 complex, the disordered N- and C-termini of the constituent proteins form long flexible arms at both ends of the four-helix bundle core.

Analysis of multiple sequence alignment for the members of the EAST-6 / CFP-10 family proteins reveals that over half of the interface residues in the structure of the Rv0287-Rv0288 complex are conserved to residue type in the majority of the sequences. This, together with predicted helical structures for all members of the *M. tuberculosis* CFP-10-ESAT-6 family and known complex formation for several genome pairs, strongly suggests that all pairs of these proteins will form similar, four-helix bundle containing complexes (1, 63). However, comparative analysis of the surface structure of the Rv0288-Rv0287 and ESAT-6-CFP-10 complexes revealed distinct surface features, such as a striking positively charged patch and a distinct cleft, for the Rv0287-Rv0288 complex. This strongly suggests that the complexes formed by the distantly related pairs of the CFP-10 / ESAT-6 family proteins (Figure 3.3.37) form four-helix bundle containing complexes with different surface features.

## Appendix-1

### Reagents and Culture Media

#### 1.1. Selected Reagents

##### 1 % (W/V) Agarose

1 g agarose, made up to 100 ml with 1x TAE with the addition of ethidium bromide to a final concentration of 0.5 µg/ml.

##### 6x Blue Orange-dye

0.03 % bromophenol blue, 15 % Ficoll<sup>®</sup> 400, 0.4 % orange G, 0.03 % xylene cyanol FF, 50 mM EDTA (pH 8.0) and 10 mM Tris HCl (pH 7.5) (Promega).

##### LDS Sample buffer

0.682 g Tris base, 0.666 g Tris HCl, 4.0 g glycerol, Serva Blue G250 0.75 ml of 1 % solution, 0.8 lithium dodecyl sulphate (LDS), 0.006 g EDTA, Phenol Red 0.25 ml of 1% solution in ultrapure water to 10 ml (Invitrogen).

##### MES SDS running buffer

20x stock solution; 60.6 Tris base, 10.0 g sodium dodecyl sulphate (SDS), 97.8 g 2-(N-morpholino) ethanesulphonic acid (MES), 3.0 g EDTA in ultrapure water to 500 ml (Invitrogen).

### Coomassie brilliant blue

Dissolve 2.5 g/l coomassie brilliant blue in 40 % methanol (v/v) and 10 % acetic acid (v/v). Stir overnight at room temperature. Make up to 1 litre with de-ionized H<sub>2</sub>O and filter through 0.2 µm filter.

### β- galactosidase assay solutions

Z-buffer solution: 40 mM NaH<sub>2</sub>PO<sub>4</sub>.H<sub>2</sub>O, 60 mM Na<sub>2</sub>HPO<sub>4</sub>.7H<sub>2</sub>O, 1 mM MgSO<sub>4</sub>, 10 mM KCl in dH<sub>2</sub>O. Store at room temperature

Z-buffer/X-gal/β-mercaptoethanol solution: 5 ml Z buffer, 18 µl β-mercaptoethanol, 50 µl X-gal (50 mg/ml in Dimethylformaldehyde). Store at 4°C in the dark.

## 1.2. Bacterial Culture Media

### Luria-Bertani (LB) broth

10 g/l tryptone, 10 g/l NaCl, 5 g/l yeast extract. Make up to volume with de-ionized H<sub>2</sub>O and autoclave. Add appropriate antibiotic when cool.

### Luria-Bertani (LB) agar

LB agar prepared as described above for LB broth with addition of 15 g/l bacteriological agar (Oxoid). Flasks were autoclaved and allowed to cool, with appropriate antibiotic added immediately prior to pouring into plates.

### Minimal media

Minimal media which allowed the introduction of selective carbon and nitrogen isotopes into expressed proteins were prepared as follows;

Per litre culture

$(\text{NH}_4)_2\text{SO}_4^*$	1.0g	$\text{MnCl}_2$	16.0 mg
$\text{KH}_2\text{PO}_4$	3.0g	$\text{FeCl}_3$	5.0 mg
$\text{Na}_2\text{HPO}_4$	6.8g	$\text{ZnCl}_2$	0.5 mg
$\text{NaCl}$	0.5g	$\text{CuCl}_2$	0.1 mg
EDTA	50.0 mg	$\text{CoCl}_2$	0.1 mg
$\text{Na}_2\text{SO}_4$	0.3 mM	$\text{H}_3\text{BO}_3$	0.1 mg

were dissolved in de-ionised  $\text{H}_2\text{O}$  and autoclaved. When cool the following were added aseptically:

$\text{MgSO}_4$	1.0 mM	$\text{CaCl}_2$	0.3 mM
d-Biotin	1.0 mg	Thiamine	1.0 mg
Glucose**	4.0 g		

\* For  $^{15}\text{N}$  labelled Rv0287 and Rv0288 and also  $^{15}\text{N}^{13}\text{C}$  labelled Rv0287 and Rv0288 the ammonium sulphate was replaced with 1 g/l ( $^{15}\text{NH}_4$ ) $_2\text{SO}_4$

\*\*For  $^{15}\text{N}^{13}\text{C}$  labelled Rv0287 and Rv0288 the glucose was replaced with 2 g/l D-Glucose U- $^{13}\text{C}_6$  ( $^{13}\text{C}$  99 % +, Cambridge Isotope Laboratories, Inc).

### 1.3. Yeast Culture media

#### Yeast extract peptone dextrose (YPD) media

20 g/l glucose, 20 g/l peptone, 10 g/l yeast extract. Make up to volume with de-ionized H<sub>2</sub>O and autoclave. For YPD agar add 15 g/l bacteriological agar (Oxoid) prior to autoclaving.

#### Synthetic complete (SC) drop-out media

6.7 g/l yeast nitrogen base (minus amino acids), 20 g/l glucose, 0.6 g/l triple dropout media (-leu/-trp/-his) (BD Biosciences). Make up to 950 ml with de-ionized H<sub>2</sub>O. Add amino acid supplements as required. Adjust pH to 5.6-5.8; make up to 1 litre with de-ionized H<sub>2</sub>O and autoclave. For SC drop-out agar add 15 g/l bacteriological agar (Oxoid)

Tryptophan- Stock solution prepared at 4.8 mg /ml, use at 40 µg/ml final concentration

Leucine- Stock solution prepared at 7.2 mg /ml, use at 60 µg/ml final concentration.

Histidine - Stock solution prepared at 10 mg /ml, use at 20 µg/ml final concentration.

## Appendix-2

### Competent Cell Preparation, Transformation Protocol

#### 2.1. Preparation of competent cells

##### Reagents

LB medium- 10 ml sterile LB medium with no antibiotic

SOB medium- 2 g tryptone, 0.5 g yeast extract, 0.05 g NaCl. Make up to 95 ml with de-ionised H<sub>2</sub>O. Add 1 ml 250 mM KCl and adjust Ph to 7.0. Make up to 100 ml with de-ionised H<sub>2</sub>O, divide into 2 x 50 ml in 250 ml flasks and autoclave. Add 0.5 ml sterile 2 M MgCl<sub>2</sub> prior to use.

50 % Glycerol solution- prepare 50 % (v/v) glycerol in de-ionized H<sub>2</sub>O and sterilize by autoclaving. Store at 4 °C. Use at the final concentration of 15 %.

50 mM CaCl<sub>2</sub> solution- Prepare 50 mM CaCl<sub>2</sub> in de-ionized H<sub>2</sub>O, filter sterilize through 0.2 µm syringe filter into a sterile container.

##### Method

Streak our *E. coli* (DH5α or BL21 DE3) onto LB agar without antibiotic (Section 1.2, appendix-1). Incubate at 37 °C overnight.

Inoculate one colony from the culture plates into 10 ml sterile medium. Incubate at 37 °C with shaking at 200 rpm overnight.

Use 1 ml of this culture to inoculate 50 ml SOB. Incubate at 37 °C with shaking at 200 rpm until the absorbance at 600<sub>nm</sub> is 0.5 (1 cm path length cell). Pellet cells by centrifugation, 5000 rpm at 4 °C for 10 minutes. Resuspend cells very gently in 25 ml ice-cold 50 mM CaCl<sub>2</sub> and leave cells on ice for 30 minutes. Spin down cells as before and resuspend very gently in 2 ml ice-cold 50 mM CaCl<sub>2</sub>, keeping cells on ice. Add sterile 50 % glycerol (ice-cold) to 15 % and mix gently by inversion, do not vortex. Aliquot 100 µl cells into cryogen tubes and flash freeze in liquid nitrogen. Store at -70 °C until required.

## 2.2. Transformation of competent cells

### Reagents

Prepare LB agar plates (Section 1.2, appendix 1) with appropriate antibiotic.

Prepare LB broth with no antibiotic as described in section 1.2, appendix-1.

### Method

Aliquot 50 µl competent cells into sterile 1.5 ml eppendorf tubes. Add 4 µl of plasmid mixture and then mix gently. Incubate the mixture on ice for 30 minutes. Heat shock at 42 °C for 45 seconds then leave the sample on ice for 5 minutes. Add 300 µl LB broth (no antibiotic) and incubate for 45 minutes at 200 rpm and 37 °C, with the tubes placed horizontally in the incubator. Plate out cells in 50, 100, 200 µl aliquots onto LB agar with appropriate antibiotic and finally incubate at 37 °C overnight.

## References

1. Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S.V., Hewinson, R.G., Burke, B., Norman, J., Williamson, R.A., Carr, M.D. (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6 *EMBO J.*24(14):2491-2498.
2. Anonymous, WHO Report 2006. Global tuberculosis control-surveillance, planning, World Health Organization, Geneva, Switzerland, 2006.
3. Radosevic, K., Wieland, C.W., Rodriguez, A., Weverling, G.J., Mintardjo, R., Gillissen, G., Vogels, R., Skeiky, Y.A., Hone, D.M., Sadoff, J.C., van der Poll, T., Havenga, M., and Goudsmit, J. (2007) Protective Immune Responses to a Recombinant Adenovirus Type 35 Tuberculosis Vaccine in Two Mouse Strains: CD4 and CD8 T-Cell Epitope Mapping and Role of Gamma Interferon. *Infect Immun.* **75**(8):4105-4115.
4. Russel, D. G. (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol.* 5(1):39-47.
5. Stewart, G.R, Robertson, B.D.,Young, D.B.(2003) Tuberculosis: a problem with persistence. *Nat Rev Microbiol.* **1**(2):97-105.
6. Health Protection Agency. (2007) <http://www.hpa.org.uk/pulications/2007/TB>

7. Tuberculosis, from basic science to patient care. (2007). 93-112  
[www.TuberculosisTextbook.com](http://www.TuberculosisTextbook.com)

8. Brosch, R., Pym, A.S., Gordon, S.V. and Cole, S.T. (2001) The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol.* 452-458.

9. Abdallah, A.M., Gey, van Pittius, N.C., Champion, P.A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C.M., Appelmelk, B.J., Bitter, W. (2007) Type VII secretion--mycobacteria show the way. *Nat Rev Microbiol.* 5(11):883-891.

10. Russell, D. G., Purdy, G. E., Owens, R. M., Rohde, K. and Yates, R. M (2005) *Mycobacterium tuberculosis* Four-minute Phagosome. *ASM News* 71,459–463.

11. Russell, D. G., Dant, J., and Sturgill-Koszycki, S. (1996) *Mycobacterium avium*- and *Mycobacterium tuberculosis*-containing vacuoles are dynamic, fusion-competent vesicles that are accessible to glycosphingolipids from the host cell plasmalemma *J. Immunol.* 156, 4764–4773.

12. Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., Davis, A., de Haro, S., Naylor, J., Lee, H.H., Vergne, I. (2006) *Mycobacterium tuberculosis* inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism *Cell. Microbiol.* 8, 719–727.

13. Pethe, K., Swenson, D.L., Alonso, S., Anderson, J., Wang, C., Russell, D.G. (2004) Isolation of *Mycobacterium tuberculosis* mutants defective in the arrest of phagosome maturation. *Proc. Natl Acad. Sci. USA* **101**, 13642–13647.
14. Malik, Z.A., Thompson, C.R., Hashimi, S., Porter, B., Iyer, S.S., Kusner, D.J. (2003) Cutting edge: *Mycobacterium tuberculosis* blocks  $\text{Ca}^{2+}$  signaling and phagosome maturation in human macrophages via specific inhibition of sphingosine kinase. *J Immunol.* **170**(6):2811-5.
15. Malik, Z.A., Denning, G.M., Kusner, D.J. (2000) Inhibition of  $\text{Ca}^{2+}$  signaling by *Mycobacterium tuberculosis* is associated with reduced phagosome-lysosome fusion and increased survival within human macrophages. *J Exp Med.* **191**(2):287-302.
16. Smith, I. (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin. Microbiol. Rev.* **16**, 463-496.
17. Fenton, M.J., Vermeulen, M.W. Immunopathology of tuberculosis: roles of macrophages and monocytes. (1996) *Infect Immun.* **64**(3):683-90.
18. Ulrichs, T., Kaufmann, S.H. (2006) New insights into the function of granulomas in human tuberculosis. *J. Pathol.* **208**, 261–269.

19. Flynn, J.L., Chan, J. (2005) What's good for the host is good for the bug. *Trends Microbiol.* **13**, 98–102.
20. Fenhalls, G., Stevens, L., Bezuidenhout, J., Amphlett, G.E., Duncan, K., Bardin, P., Lukey, P.T. (2002) Distribution of IFN-gamma, IL-4 and TNF-alpha protein and CD8 T cells producing IL-12p40 mRNA in human lung tuberculous granulomas. *Immunology* **105**, 325–335.
21. Fenhalls, G., Stevens, L., Moses, L., Bezuidenhout, J., Betts, J.C., Helden, P., Lukey, P.T., Duncan, K. (2002) In situ detection of Mycobacterium tuberculosis transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. *Infect. Immun.* **70**.6330–6338.
22. McKinney, J.D., Höner zu Bentrup, K., Muñoz-Elías, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., Jacobs, W.R. Jr, Russell, D.G.(2000) Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**(6797):735-738.
23. Tsai, M.C., Chakravarty, S., Zhu, G., Xu, J., Tanaka, K., Koch, C., Tufariello, J., Flynn, J., Chan, J. (2006) Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. *Cell Microbiol.* **8**(2):218-232.

24. Rhoades E, Hsu F, Torrelles JB, Turk J, Chatterjee D., and Russell, D.G. (2003) Identification and macrophage-activating activity of glycolipids released from intracellular *Mycobacterium bovis* BCG *Mol. Microbiol.* **48**. 875–888.
25. Geisel, R. E, Sakamoto, K., Russell, D.G., and Rhoades, E.R. (2005) In vivo activity of released cell wall lipids of *Mycobacterium bovis* bacillus Calmette-Guérin is due principally to trehalose mycolates. *J. Immunol.* **174**. 5007–5015.
26. Ulrichs, T., Kosmiadi, G.A., Trusov, V., Jörg, S., Pradl, L., Mishenko, V., Gushina, N., Kaufmann, S.H. (2004) Human tuberculous granulomas induce peripheral lymphoid follicle-like structures to orchestrate local host defence in the lung. *J Pathol.* **204**(2):217-228.
27. van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., Brenner, M., and Peters, P.J. (2007) *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell*, **129** (7):1287-1298.
28. Pym, A.S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., Griffiths, K.E., Marchal, G., Leclerc, C., and Cole, S.T. (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med.* **9**(5):533-539.
29. Dietrich, J., Weldingh, K., Andersen, P. *Vet Microbiol.* (2006) Prospects for a novel vaccine against tuberculosis. **112**(2-4):163-169.

30. Fine, P.E.M., Carnerio, I.A.M., Milstien, J.B. and Clements, C.J. (1999) WHO/V&B/99.23).Geneva.
31. Horwitz, M.A. (2005) Recombinant BCG expressing Mycobacterium tuberculosis major extracellular proteins. *Microbes and Infection* **7** 947-954.
32. Billeskov, R., Vingsbo-Lundberg, C., Andersen, P., and Dietrich, J.(2007) Induction of CD8 T cells against a novel epitope in TB10.4: correlation with mycobacterial virulence and the presence of a functional region of difference-1.*J Immunol.* **179**(6):3973-3981.
33. Bass, J. B., Farer, L. S, Hopewell, P. C., Brien, R., Jacobs, R. F., Snider, D. E., Thornton, G. (1994) Treatment of tuberculosis and tuberculosis infection in adults and children. American Thoracic Society and The Centers for Disease Control and Prevention. *Am. J. Respir. Crit. Care. Med.* **149**, 1359-1374.
34. Cox, H.S., Morrow, M., and Deutschmann, P.W. *BMJ.* (2008) Long term efficacy of DOTS regimens for tuberculosis: systematic review **336**(7642):485-487.
35. Sacchetti, 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**(1):41-52.

36. Kamath, A., Woodworth, J.S., and Behar, S.M. (2006) Antigen-specific CD8+ T cells and the development of central memory during *Mycobacterium tuberculosis* infection. *J Immunol.* **177**(9):6361-6369.
37. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* **393**, 537-544.
38. Garnier, T., Eiglmeier, K., Camus, J.C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, and S.V., Hewinson, R.G. (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci.* **100**(13):7877-7882.
39. Camacho, L. R., Constant, P., Raynaud, C., Laneelle, M.A., Triccas, J. A., Gicquel, B., Daffe, M., and Guilhot, C. (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J. Biol. Chem.* **276**, 19845-19854.

40. Renshaw, P. S., Panagiotiodus, P., Whelan, A., Gordon, S. V., Hewinson, R. G., Williamson, R. A., and Carr, M. D. (2002) Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6-CFP-10 complex implications for pathogenesis and virulence. *J Bio Chem.* **277**(24):21598-21603.

41. Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., Sherman, D.R. (2003) Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guérin attenuation. *J. Inf. Dis.* **187**, 117-123.

42. Fortune, S. M., Jaeger, A., Sarracino, D. A., Chase, M.R., Sasseti, C. M., Sherman, D. R., Bloom, B. R., and Rubin, E.J. (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *PNAS.* **102**(30), 10676-10681.

43. Pym, A.S., Brodin, P., Brosch, R., Huerre, M., Cole, S.T. (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. *Mol. Microbiol.* **46**, 709-717.

44. DiGiuseppe Champion, P.A., and Cox, J.S. (2007) Protein secretion systems in Mycobacteria. *Cell Microbiol.* **9**(6):1376-84.

45. McDonough, J.A., Hacker, K.E., Flores, A.R., Pavelka, M.S.Jr, and Braunstein, M. *J Bacteriol.* (2005). The twin-arginine translocation pathway of *Mycobacterium smegmatis* is functional and required for the export of mycobacterial beta-lactamases. **187**(22):7667-7679.
46. Saint-Joanis, B., Demangel, C., Jackson, M., Brodin, P., Marsollier, L., Boshoff, H, and Cole, S.T. *J Bacteriol.* (2006) Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of *Mycobacterium tuberculosis*, increases beta-lactam susceptibility and virulence. **188**(18):6669-6679.
47. Lenz, L. L., Mohammadi, S., Geissler, A., and Portnoy, D. A. (2003) SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *PNAS*. **100**(21), 12432-12437.
48. Macgurn, J. A., Raghavan, S., Stanley, S. A., and Cox, J. S. (2005) A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **57**(6), 1653-1663.
49. Fisher, M. A., Plikaytis, B. B., and Shinnick, T. M.(2002) Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol.* **184**(14):4025-32.

50. Rodriguez, G.M., Voskuil, M. I., Gold, B., Schoolnik, G. K., and Smith, I. (2002) ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infec. Immun.* **70**, 3371-3381.
51. Rickman, L., Scott, C., Hunt, D. M., Hutchinson, T., Menezes, M. C., Whalen, R., Hinds, J., Colston, M. J., Green, J., and Buxton, R. S. (2005) A member of the cAMP receptor protein family of transcription regulators in Mycobacterium tuberculosis is required for virulence in mice and controls transcription of the rpfA gene coding for a resuscitation promoting factor. *Mol. Microbiol.* **56** (5), 1274-1286.
52. Sassetti, C. M., Boyed, D. H., and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48** (1), 77-84.
53. Volkman, H.E., Clay, H., Beery, D., Chang, J.C., Sherman, D.R., and Ramakrishnan, L. (2004) Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. *PLoS Biol* **2**: e367.
54. Sambrook, J. and Russell, D.W. *Molecular cloning a laboratory manual*, Third Edition, pp.18.6
55. Field, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340** (6230), 245-246.

56. Geitz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Meth. Enzymol.* Vol. 350, pp. 87-96.
57. Breeden, L. and Nasmyth, K. (1985) *Cold Spring Harbor Symposia on Quantitative Biology.* **50**, 643-650.
58. Le Douarin, B., Heery, D.M., Gaudon, C., Vom Baur, E. and Losson, R. (2001) Yeast Two-Hybrid Screening for Proteins that Interact with Nuclear Hormone Receptors. *Methods in Molecular Biology.* Vol. **176**, pp 227-248.
59. Kelly, S.M., Jess, T.J., Price, N.C. (2005) How to study proteins by circular dichroism. *BBA.* **1751**, 119-139.
60. Stano, N.M., and Patel, S.S. (2004) T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *J Biol Chem.* **279**(16):16136-16143.
61. Edwin, F., and Jagannadham, M.V. (1998) Sequential unfolding of papain in molten globule state. *Biochem. Biophys. Res. Commun.* **27**; 252(3):654-660.
62. Lightbody, K.L., Renshaw, P.S., Collins, M.L., Wright, R.L., Hunt, D.M., Gordon, S.V., Hewinson, R.G., Buxton, R.S., Williamson, R.A., and Carr, M.D. (2004) Characterisation of complex formation between members of the Mycobacterium

tuberculosis complex CFP-10/ESAT-6 protein family: towards an understanding of the rules governing complex formation and thereby functional flexibility. *FEMS Microbiol Lett.* **238**(1):255-262.

63. Lightbody, K.L., Ilghari, D., Waters, L.C., Carey, G., Bailey, M.A., Williamson, R.A., Renshaw, P.S., and Carr MD (2008) Molecular features governing the stability and specificity of functional complex formation by Mycobacterium tuberculosis CFP-10/ESAT-6 family proteins. . *J Biol Chem.* **283**(25):17681-17690.

64. Pallen, M.J. (2002). The ESAT-6/WXG100 superfamily - and a new Gram-positive secretion system? *Trends Microbiol.* **10**(5):209-212.

65. Gey Van Pittius, N.C., Gamielien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D. (2001) The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria .*Genome Biol.*;2(10):RESEARCH0044.

66. Okkels, L.M., and Andersen, P. (2004) Protein-protein interactions of proteins from the ESAT-6 family of Mycobacterium tuberculosis. *J Bacteriol.* **186**(8):2487-91.

67. Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., and Duncan, K. (2002) Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol Microbiol.* **43**(3):717-731.

68. Roback, P., Beard, J., Baumann, D., Gille, C., Henry, K., Krohn, S., Wiste, H., Voskuil, M.I., Rainville, C., and Rutherford, R. (2007) A predicted operon map for *Mycobacterium tuberculosis*. *Nucleic Acids Res.* 35(15):5085-5095.

69. Skjøt, R.L., Oettinger, T., Rosenkrands, I., Ravn, P., Brock, I., Jacobsen, S., and Andersen, P. (2000) Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. *Infect Immun.* 68(1):214-220.

70. Rosenkrands, I., King, A., Weldingh, K., Moniatte, M., Moertz, E., and Andersen, P. (2000) Towards the proteome of *Mycobacterium tuberculosis*. *Electrophoresis.* 21(17):3740-3756.

71. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrell, B. G. (2001) Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011.

72. Skjøt, R. L. V., I. Brock, S. M. Arend, M. E. Munk, M. Theisen, T. H. M. Ottenhoff, and P. Andersen. (2002). Epitope mapping of the immunodominant antigen TB10.4 and the two homologous proteins TB10.3 and TB12.9, which constitute a subfamily of the esat-6 gene family. *Infect. Immun.* **70**:5446-5453.

73. Hervas-Stubbs, S., Majlessi, L., Simsova, M., Morova, J., Rojas, M.J., Nouzé, C., Brodin, P., Sebo, P., Leclerc, C. (2006) High frequency of CD4+ T cells specific for the TB10.4 protein correlates with protection against Mycobacterium tuberculosis infection. *Infect Immun.* **74**(6):3396-407.

74. Rindi, L., Lari, N., and Garzelli, C. (2001) Genes of Mycobacterium tuberculosis H37Rv downregulated in the attenuated strain H37Ra are restricted to M. tuberculosis complex species. *New Microbiol.* **24**(3):289-94.

75. Rindi, L., Lari, N., and Garzelli, C. (1999) Search for genes potentially involved in Mycobacterium tuberculosis virulence by mRNA differential display. *Biochem Biophys Res Commun.* **258**(1):94-101.

76. Sasseti, C. M., D. H. Boyd, and E. J. Rubin. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**(1):77-84.

77. Deb, D.K., Dahiya, P., Srivastava, K.K., Srivastava, R., and Srivastava, B.S.(2002).Selective identification of new therapeutic targets of *Mycobacterium tuberculosis* by IVIAT approach.*Tuberculosis (Edinb)*.**82**(4-5):175-82.
78. Porcelli, S. V. (2008) Tuberculosis: Shrewd survival strategy *Nature* **454**, 702 – 703.
79. Raghavan, S., Manzanillo, P., Chan, K., Dovey, C., and S. Cox, J.C. (2008) Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature* **454**, 717 – 721.
80. Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., Eisenberg, D., Russell, R.G., Derrick, S.C., Collins, F.M., Morris, S.L., King, C.H., and Jacobs, W.R. Jr. (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A*. **100**(21):12420-12425.
81. MacGurn, J.A., and Cox, J.S. (2007) A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system.*Infect Immun*. 75(6):2668-2678.
82. Majlessi, L., Rojas, M.J., Brodin, P., and Leclerc, C. (2003) CD8<sup>+</sup>-T-cell responses of *Mycobacterium*-infected mice to a newly identified major histocompatibility complex class

I-restricted epitope shared by proteins of the ESAT-6 family. *Infect Immun.* **71**(12):7173-7177.

83. Doherty, T.M., Olsen, A.W., Weischenfeldt, J., Huygen, K., D'Souza, S., Kondratieva, T.K., Yermeev, V.V., Apt, A.S., Raupach, B., Grode, L., Kaufmann, S.H.E., and Andersen, P.(2004) Comparative analysis of different vaccine constructs expressing defined antigens from *Mycobacterium tuberculosis*. *J. Infect. Dis.* **190**(12), 2146–2153.

84. Sable, S.B., Verma, I., Behera, D., and Khuller, G.K. (2005) Human immune recognition-based multicomponent subunit vaccines against tuberculosis. *Eur Respir J.* **25**(5):902-910.

85.Sable, S.B., Kumar, R., Kalra, M., Verma, I., Khuller, G.K., Dobos, K., and Belisle, J.T. (2005) Peripheral blood and pleural fluid mononuclear cell responses to low-molecular-mass secretory polypeptides of *Mycobacterium tuberculosis* in human models of immunity to tuberculosis. *Infect Immun.* **73**(6):3547-3558.

86.Mustafa AS, Skeiky YA, Al-Attayah R, Alderson MR, Hewinson RG, Vordermeier HM. (2006) Immunogenicity of *Mycobacterium tuberculosis* antigens in *Mycobacterium bovis* BCG-vaccinated and *M. bovis*-infected cattle. *Infect Immun.* **74**(8):4566-4572.

87. Dietrich. J, Aagaard ,C., Leah, R., Olsen, A.W., Stryhn, A., Doherty, T.M., and Andersen, P. (2005) Exchanging ESAT6 with TB10.4 in an Ag85B fusion molecule-based

tuberculosis subunit vaccine: efficient protection and ESAT6-based sensitive monitoring of vaccine efficacy. *J Immunol.* **174**(10):6332-6339.

88. Cockle, P.J., Gordon, S.V., Hewinson, R.G., and Vordermeier, H.M. (2006) Field evaluation of a novel differential diagnostic reagent for detection of *Mycobacterium bovis* in cattle. *Clin Vaccine Immunol.* **13**(10):1119-1124.

89. Maciag, A., Dainese, E., Rodriguez, G.M., Milano, A., Provvedi, R., Pasca, M.R., Smith, I., Palù, G., Riccardi, G., and Manganelli, R. (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol.* **189**(3):730-740.

90. Canneva, F., M. Branzoni, G. Riccardi, R. Provvedi, and A. Milano. (2005) Rv2358 and FurB: two transcriptional regulators from *Mycobacterium tuberculosis* which respond to zinc. *J. Bacteriol.* **187**:5837-5840.

91. Sidders, B., Pirson, C., Hogarth, P.J., Hewinson, R.G., Stoker, N.G., Vordermeier H.M., and Ewer K. (2008) Screening highly expressed mycobacterial genes identifies Rv3615c as a useful differential diagnostic antigen for the *Mycobacterium tuberculosis* complex. *Infect Immun.* **76** (9) 3932-3939

92. Weldingh, K., and Andersen, P. (2008) ESAT-6/CFP10 skin test predicts disease in *M. tuberculosis*-infected guinea pigs. *PLoS ONE.* **3**(4):e1978.

93. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacterio* **164**(2):918-21.
94. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **1** (6): 2876-2890.
95. Braunschweiler, L. and Ernst, R.R. (1983). Coherence transfer by isotropic mixing application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**: 521-528.
96. Macura, S. and Ernst, R.R. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR-spectroscopy. *Molecular Physics*, **41**: 95-117.
97. Bodenhausen, G., and Ruben, D.J. (1980). Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**: 185-189.
98. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) 3-Dimensional heteronuclear NMR of N-15-labeled proteins *J. Am. Chem. Soc.* **111**, 1515-1517.
99. Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., *et al.* (1989). Overcoming the overlap problem in the assignment of  $^1\text{H}$  NMR spectra of larger proteins by use of three-dimensional hetero nuclear  $^1\text{H}$ - $^{15}\text{N}$ Hartmann-Hahn-multiple

quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28**:6150-6156.

100. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins. *J. Magn. Reson.* **88**, 425-431.

101. Wittekind, M., and Mueller, L. (1993) HNCACB, a High-Sensitivity 3D Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *J. Magn. Reson.* **101**, 201-205.

102. Grzesiek, S., and Bax, A. (1993) Amino-acid type determination in the sequential assignment procedure of uniformly C-13/N-15-enriched proteins. *J. Biomol. NMR* **3**, 185-204.

103. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M., and Skelton, N. J. (2007) *Protein NMR Spectroscopy*, 2<sup>nd</sup> Edition.

104. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-Tailored Excitation for Single- Quantum Nmr-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR* **2**, 661-665.

105. Guntert, P., Mumenthaler C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**: 283-298.

106. 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
107. 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**: 447-56.
108. Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol NMR*, **13**: 289-302.
109. Wuthrich, K. (1986). *NMR of Proteins and Nucleic Acids.: Polypeptide Secondary Structures in Proteins by NMR*. Wuthrich K (ed.). John Wiley and Sons Inc.: New York.
110. Koradi, R., Billeter, M. and Wuthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics*. **14**: 51
111. Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**: 1647-51.

112. Hindupur, A., Liu, D., Zhao, Y., Bellamy, H.D., White, M.A., and Fox, R.O. (2006) The crystal structure of the E. coli stress protein YciF. *Protein Sci.* **(11)**:2605-2611.
113. Zeth, K., Offermann, S., Essen, L.O., and Oesterhelt, D. (2004) Iron-oxo clusters biomineralizing on protein surfaces: structural analysis of Halobacterium salinarum DpsA in its low- and high-iron states. *Proc Natl Acad Sci U S A.* **101**(38):13780-13785.
114. Macedo, S., Romão, C.V., Mitchell, E., Matias, P.M., Liu, M.Y., Xavier, A.V., LeGall, J., Teixeira, M., Lindley, P., and Carrondo, M.A. (2003) The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. *Nat Struct Biol.* **(4)**:285-290.
115. Franceschini ,S., Ceci, P., Alaleona, F., Chiancone, E., and Ilari, A. (2006) Antioxidant Dps protein from the thermophilic cyanobacterium Thermosynechococcus elongatus. *FEBS J.* **273**(21):4913-4928.
116. Roy, S., Saraswathi, R., Chatterji, D., and Vijayan, M. (2008) Structural studies on the second Mycobacterium smegmatis Dps: invariant and variable features of structure, assembly and function. *J Mol Biol.* **375**(4):948-959.
117. Holm, L., and Sander C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci.* **(11)**:478-80.

118. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL\_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882
119. Flint, J.L., Kowalski, J.C., Karnati, .PK., and Derbyshire, K.M. (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *Proc Natl Acad Sci U S A*. 101(**34**):12598-603.
120. Champion, P.A., Stanley, S.A., Champion, M.M., Brown, E.J., and Cox, J.S. (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science*. 313(**5793**):1632-1636.

## Published work

Ilghari, D., Waters, L. C., Veverka, V., Muskett, F. W., and Carr, M. D. (2009)  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  resonance assignments and secondary structure determination of the *Mycobacterium tuberculosis* Rv0287-Rv0288 protein complex. *Biomolecular NMR Assignments* (Accepted).

Lightbody, K.L., Ilghari, D., Waters, L.C., Carey, G., Bailey, M.A., Williamson, R.A., Renshaw, P.S., and Carr M.D. (2008) Molecular features governing the stability and specificity of functional complex formation by *Mycobacterium tuberculosis* CFP-10/ESAT-6 family proteins. *J Biol Chem.* 283(25):17681-17690.