An NMR analysis of the structure and ligand binding characteristics of proteins from *Mycobacterium tuberculosis* implicated in dormancy.

Igor Kojic

University College London

A thesis submitted for the degree of Doctor of Philosophy

January 2008

Thesis abstract:

Two proteins putatively involved in the persistence of *Mycobacterium tuberculosis* infection have been investigated by heteronuclear nuclear magnetic resonance (NMR) spectroscopy and other methods.

The major aspect of work describes the characterization by NMR of the *Mtb* thiol peroxidase; TPx. TPx (165 residues) is a 36 kD symmetrical homodimer that catalyzes the reduction of a range of reactive oxygen species *in vitro*. Multidimensional heteronuclear NMR spectroscopy was applied to recombinant wild-type (Wt) and Cys60->Ser (C60S) TPx, both of which are clearly folded but display slightly different spectral characteristics. Triple resonance NMR experiments recorded using [^{2}H, ^{13}C, ^{15}N]-labelled C60S yielded assignments for the majority of the backbone resonances. 28 non-proline residues could not be assigned mostly due to the absence of cross peaks. TPxWt spectra showed little difference in peak number indicating that the C60S mutation itself does is not responsible for the missing signals. Consideration of related TPx crystal structures indicates that the missing cross peaks correspond to a region of the protein that must display conformational plasticity necessary for the mechanism of action suggesting that the spectrum is affected by conformational exchange processes. Efforts to modulate this exchange behaviour through the application of putative TPx ligands and chemical modification of the active site cysteine residues are described.

In a second aspect, a preliminary characterisation of the non-canonical RNA polymerase sigma factor sigJ from *Mtb* was performed. Bioinformatic analysis revealed that sigJ likely contains conserved σ2 and σ4 DNA-binding domains fused to a unique C-terminal domain. NMR spectra of recombinant sigJ recovered from inclusion bodies and in the soluble fraction differed, though both forms yielded dispersed resonances consistent with the protein containing a globular component. However further progress was hampered by signal overlap and heterogeneous line widths. sigJ has a general capability to retard the movement of DNA in an electrophoretic mobility shift assay, though specific target sequence was identified.

"If you can't explain it simply, you don't understand it well enough"

Albert Einstein

Acknowledgements:

Five years have passed since the start of this long journey. Along the way I had the privilege in meeting some extraordinary people who helped me realise what I could become and yet even before I took those first tentative steps, there are those that I owe an immeasurable debt of gratitude and thanks. I now humbly and gratefully laud you all with the credit you all so richly deserve.

I must first thank Paul Driscoll whose belief in me has made this all possible. His unwavering support during the writing of this thesis has made a life-long dream of mine turn into a tangible reality.

Deep thanks go to those who have come and gone from the Biochemistry and Molecular Biology department during my time here at UCL, many of whom have gone onto bigger and better things. From those of the Paul Driscoll group, notably Andrew Sankar – for his patience and help in the laboratory during those early days, Evangelos Dioletis and Anwar Mohamed, to the both of you for all the good spirited banter that never failed in helping me put things into perspective.

To my Mother (Dubravka Kojic) and Father (Aleksandar Kojic), to whom this thesis is dedicated, I owe you both everything. Your endless love and support, most importantly during the years of my Ph.D, kept me together and believing. I thank you both so very much, from the bottom of my heart.

I must also thank the many who were there with me every step of the way, my brother – Nenad Kojic; including his beautiful family, my dearest friends, Justyn Flanders-Brazier – and his amazing family, and Chris Owens – for his good humour and support through-out.

Angela Tomassetti, your support, love, care and constant attention has always been so important to me during the writing of this thesis. My life would not be as rich if it were not for you always being by my side.

Thesis contents:

List of Figures:

List of Tables:

Abbreviations:

ABBREVIATION	DEFINITION
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
AhpC	Alkyl hydroperoxide reductase
APS	Ammonium persulphate
ASEC	Analytical size-exclusion chromatography
BCP	Bacterioferritin-comigratory protein
BG	Background
CFP	Culture filtrate protein
C_P	Peroxidatic cysteine
C_R	Resolving cysteine
CV	Column volume
ddH_20	Distilled and deionised water
dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DOT	Directly observed therapy
DTT	Dithiothreitol
EDTA	Ethylene diamine tetracetic acid
EU	European Union
FID	Free induction decay
HSQC	Heteronuclear single quantum correlation spectroscopy
INEPT	Insensitive nucleus enhanced polarisation transfer
INH	Isoniazid
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilo Daltons
mins	Minutes
Mtb	*Mycobacterium tuberculosis*
MDR	Multi-drug resistance

ABBREVIATION	DEFINITION
NBD-chloride	7-chloro-4-nitrobenzo-2-oxa-1, 3-diazole
NEM	N-ethyl-maleimide
NMR	Nuclear magnetic resonance
NEB	New England Biolabs
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PBS	Phosphate buffered saline
PDB	Protein data bank
PFAM	Protein family [database]
Prxs	Peroxiredoxins
PrxR	Peroxiredoxin reductase
RF	Radio-frequency
R.M.S	Root mean square
SDS	Sodium dodecyl sulphate
SEC	Size-exclusion chromatography
sigJ	Sigma factor J
T7-pol	T7 RNA polymerase
TAE	Tris-acetate EDTA
TB	Tuberculosis
TBS	Tris-buffered saline
TEMED	N,N,N',N'-Tetramethylethylenediamine
TPx	Thiol peroxidase
TryP	Tryparedoxin peroxidase
TROSY	Transverse relaxation optimised spectroscopy
UV	Ultra-violet

AMINO ACID NOMENCLATURE

Cys60	Residue type using the standard three letter code (e.g. cysteine)
C60S	Mutation of Cys60 to Ser60

Chapter 1

General Introduction: *Mycobacterium tuberculosis* (*Mtb*)

1.1. Summary:

In this chapter, the causative agent of tuberculosis, the bacterium *Mycobacterium tuberculosis* (*Mtb*) is described. Details are given as to its ability to infect host organisms, proliferate, survive and elude cell mediated immune responses. Following primary infection the bacterium has the ability to enter a dormant, non-replicating state in infected individuals that is virtually impossible to detect with modern laboratory techniques. During this dormant yet viable state, a series of specific changes to the surrounding cellular environment triggers a process of reactivation within the host, sometimes many years after initial infection. The resulting impact of TB, particularly in developing countries, is currently a growing concern among many of the world health authorities and is considered to be an accelerating issue with respect to increasing cases of antibiotic resistance and phenotypic adaptation to BCG-vaccinated individuals.

1.2. Tuberculosis (TB) and *Mycobacterium tuberculosis* (*Mtb*) pathogenesis.

Much like the common cold, TB is a contagious disease that spreads through the air. Only people who are sick with pulmonary TB are infectious (Electronic Source 1, 2003). To become infected, a person needs to inhale only a small number of airborne bacilli.

Figure 1.1 shows the steps involved in tuberculosis infection, proliferation, distribution and immune response, the details of which are described below.

On inhalation of the *Mtb* bacilli into the sterile environment of the lungs, specifically the respiratory bronchioles and alveoli, the bacilli settle upon the upper epithelium (Fenton and Vermeulen, 1996). When bacilli numbers settling on the upper epithelium are high, the combined weight of these inhaled particles tends to result in the process of sedimentation where the bacilli are no longer infective. Other airborne particles, such as multi-bacilli that coagulate (known as Pflüger droplets measuring 5 to 10 μM in diameter), either sediment or are cleared by the defence mechanisms of the airways. However, upon condensation of these droplet nuclei and the loss of part of their water content, smaller particles measuring 1 to 5 μM are formed, containing approximately three tubercle bacilli each; these droplets are infective. The defence mechanisms of the upper airways (i.e., cough reflex, mucociliary system) non-specifically prevent particles measuring over 5 μM from reaching the lung parenchyma. Under infective conditions, however, some particles measuring 1 to 5 μM reach the distal airways and are deposited in the alveoli. It is believed that at least 10 to 200 of such micro-droplets must reach the alveoli in order for infection to take place. The preferential zone of arrival is the best-ventilated part of the lungs, corresponding to the sub-pleural region of the inferior lobes. Infective bacilli enter the lung capillaries that span the alveoli exogenous membrane where they may escape from the lung capillaries to the systemic arterial circulation and become deposited in various organs throughout the body (Harris, 2001). The main metastatic or target zones of such bacterial dissemination are the highly irrigated organs and tissues—the central nervous system, spongy bone, liver, kidneys, and genitals. In each of these zones, the arriving bacilli are phagocytosed by the local cells of the mononuclear phagocyte system.

In the first few days following infection of a host, a strong granulomatous response is initiated by the host to defend against the infection. This response involves an accumulation of granuloma – a tumour like mass composed of actively growing fibroblasts and capillary buds, in the area of infection. This is an essential process if dissemination and proliferation of the bacilli is to be arrested here (Fenton and Vermeulen, 1996). Within the first two weeks, mass replication of the bacilli occurs within the host (Chan *et al,* 2000). This process proceeds unabated regardless of the attempts made by the presence of the macrophages fighting to remove them. Subsequent infection of other host cells commences throughout a number of tissues (Fenton and Vermeulen, 1996). It is only after approximately two weeks that infected host antigen-specific defence develops (delayed-type hypersensitivity, DTH) and proceeds to bacilli eradication (Orme, 1993).

Over the next four to five weeks progressive infection of granulomatous focal lesions, composed of macrophage-derived epithelioid giant cells and lymphocytes, begin to form. These focal lesions envelope and contain the pathogen, immediately subduing any further growth and replication. The process for bacilli eradication is hence administered and predominantly involves the alteration of the immediate environment of engulfed bacilli, resulting in low oxygen conditions and high peroxide concentrations. The eradication of phagocytosed bacilli however is not always the end result in infected hosts. Infected macrophages in the interior of each granuloma are killed as the periphery becomes fibrotic and caseated (a form of coagulation necrosis in which the necrotic tissue resembles cheese and contains a mixture of protein and fat that is absorbed very slowly) due to the development of anaerobic and toxic fatty acid conditions. A proportion of such macrophage-infected granulomas may result in the latent, non-replicating form of the bacterium being triggered. It is at this point that a host's cell-mediated immunity (CMI) (Orme, 1993) response determines whether an infection is arrested here or progresses to the next stages (Fenton and Vermeulen, 1996). With a good CMI response, the infection may be arrested permanently at this point. If however the CMI response is insufficient or ineffective, macrophages containing ingested but viable *Mtb* bacilli may escape from the granuloma via the intrapulmonary lymphatic channels. Mycobacterial defensive mechanisms include the production of glycolipids that inhibit phagosome-lysosome fusion, while other

mechanisms alter lysosomal acid pH, thereby complicating enzyme action. Catalase is produced which destroys hydrogen peroxide, and different mycobacterial components inhibit superoxide production. This phase concludes with destruction of the alveolar macrophages by proliferating intracellular bacilli. The bacilli thereafter spread throughout the body of the host to places where CMI is inadequate resulting in pulmonary damage and further spreading of the organism until eventually effective antibiotic treatment must be administered if the patient is to survive (Fenton and Vermeulen, 1996).

There are at present a number of tuberculostatic antibiotic treatments available including a vaccine known as the Bacille Calmette-Guerin (BCG) which involves the introduction of an attenuated non-virulent form of *Mycobacterium bovis* (bovine TB) into the body of the recipient thereby initiating a cell-mediated immune response (Coates and Hu, 2001a; Chan *et al,* 2000 and Electronic Source 2, 2002).

Individuals from developing countries are at a higher risk of TB infection and reactivation because they go without the BCG vaccine or have neglected the correct antibiotic treatment regimen. This is a result of poor health care monitoring facilities, where invariably the correct antibiotic treatments may be either unavailable or poorly supported. The following sections will discuss the consequences of untreated TB-infected patients, and the strategies required to ensure that available and correct antibiotic treatment programs are provided.

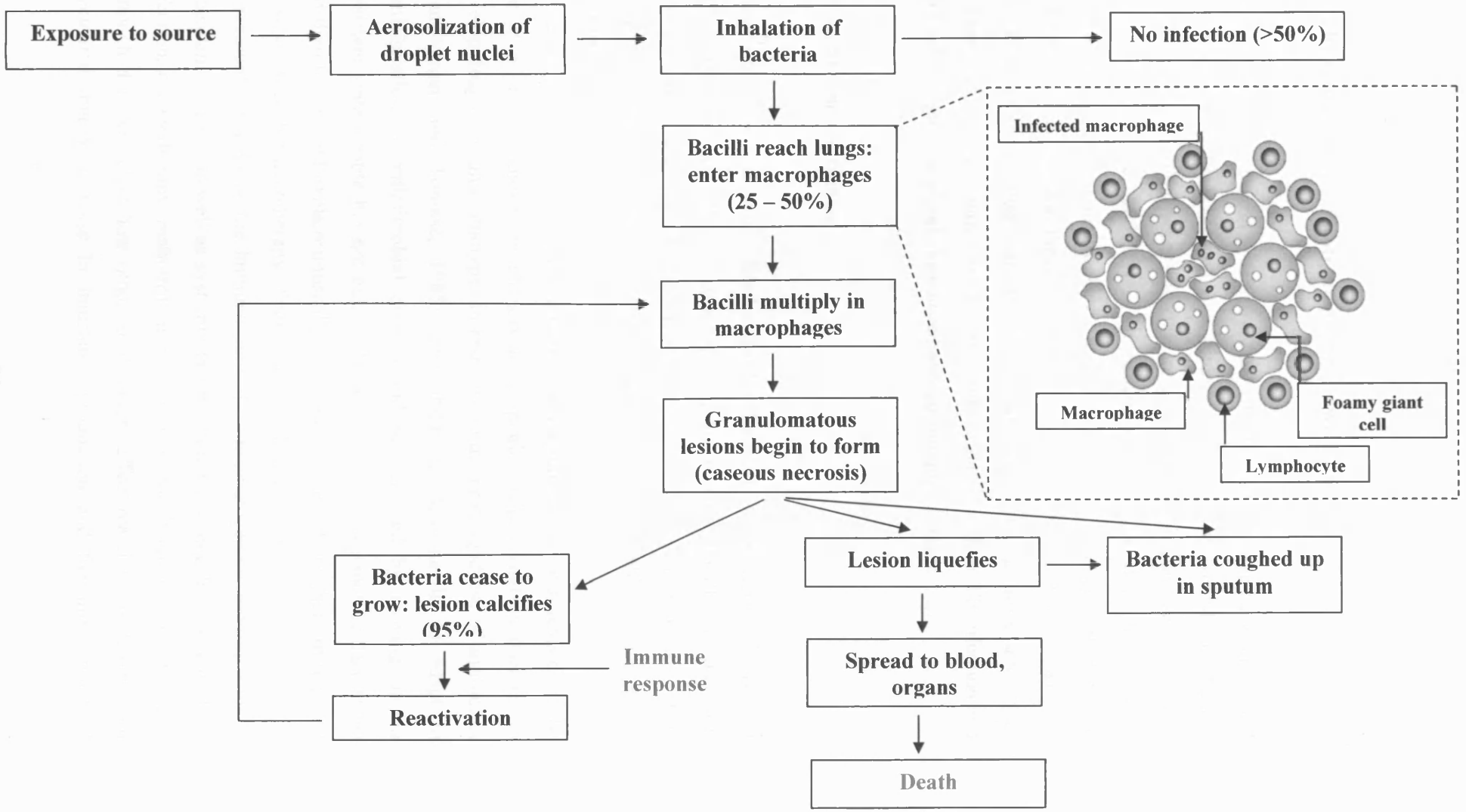


Figure 1.1: Steps in the pathogenesis of tuberculosis.

1.3. The consequences of untreated patients with TB.

Left untreated, each person with active TB will infect on average between 10 and 15 people every year. When someone's immune system is weakened, the chances of an infected individual becoming sick are greater (Electronic source 3, 2002 and Electronic Source 2, 2002). HIV infection, malnutrition, drug use, cancer, diabetes, chronic renal insufficiency and immunosuppressive drug therapy are each believed to increase the risk of reactivation TB (Bishai *et al,* 1998). The emergence of multi-drug resistant *Mtb* strains is becoming more and more prevalent, especially in much of Eastern Europe where both new TB infection and reactivation cases are closely monitored and documented (Figure 1.2).

1.4. Antibiotic resistance.

Tuberculous reactivation is defined as the development of tuberculous disease in a patient who had already been infected with the tubercle bacillus in the past. Although there are well-documented cases of exogenous re-infection, it is currently accepted that most cases of re-infection are attributable to endogenous reactivation.

The ability of *Mtb* to become tolerant to certain antimicrobial treatments could be due to a process known as phenotypic adaptation which enables bacteria to develop drug resistant phenotypes in response to the prolonged use of antibiotics (Handwerger and Tomasz, 1985). The ability of *Mtb* bacilli to adapt to administration of antimicrobial drugs could be achieved by moving into a persistent state (viable but not culturable, also known as a paucibacillary state) and therefore would explain unusually long treatment durations that are required for successful pharmacotherapy of active TB (Bishai *et al,* 1998).

In resistant individuals, the immune control of haematogenous seeding sites is dependent on local as well as systemic factors. Some systemic factors (e.g., HIV infection, corticotherapy, malnutrition) may account for tuberculous reactivation, although it is less clear how other local factors affect reactivation. It has been speculated that a decrease in interferon production and the intervention of

arabinomanans may be involved in this reduction or in the generation of a specific suppressor response. One of the known important pathogenic phenomena of reactivation is caseum liquefaction. Although not all the factors implicated in such liquefaction are known, the phenomenon has been attributed to lysosomal enzymes released by the macrophages, and to a delayed hypersensitivity reaction to mycobacterial products. The immediate consequence of caseum liquefaction is the production of an excellent growth medium for the bacteria, which begin to multiply and release products similar to tuberculin that have great toxic potential. In the case of the lungs, these released products rupture the adjacent bronchi, forming cavities and spreading the bacteria via the bronchogenic route.

Until 50 years ago, there were no medicines to cure TB. Now, strains that are resistant to a single drug have been documented in every country surveyed; what is more, strains of TB resistant to all major anti-TB drugs have emerged. Drug-resistant TB is caused by inconsistent or partial treatment, when patients do not take all their medicines regularly for the required period because they start to feel better, because doctors and health workers prescribe the wrong treatment regimens, or because the drug supply is unreliable.

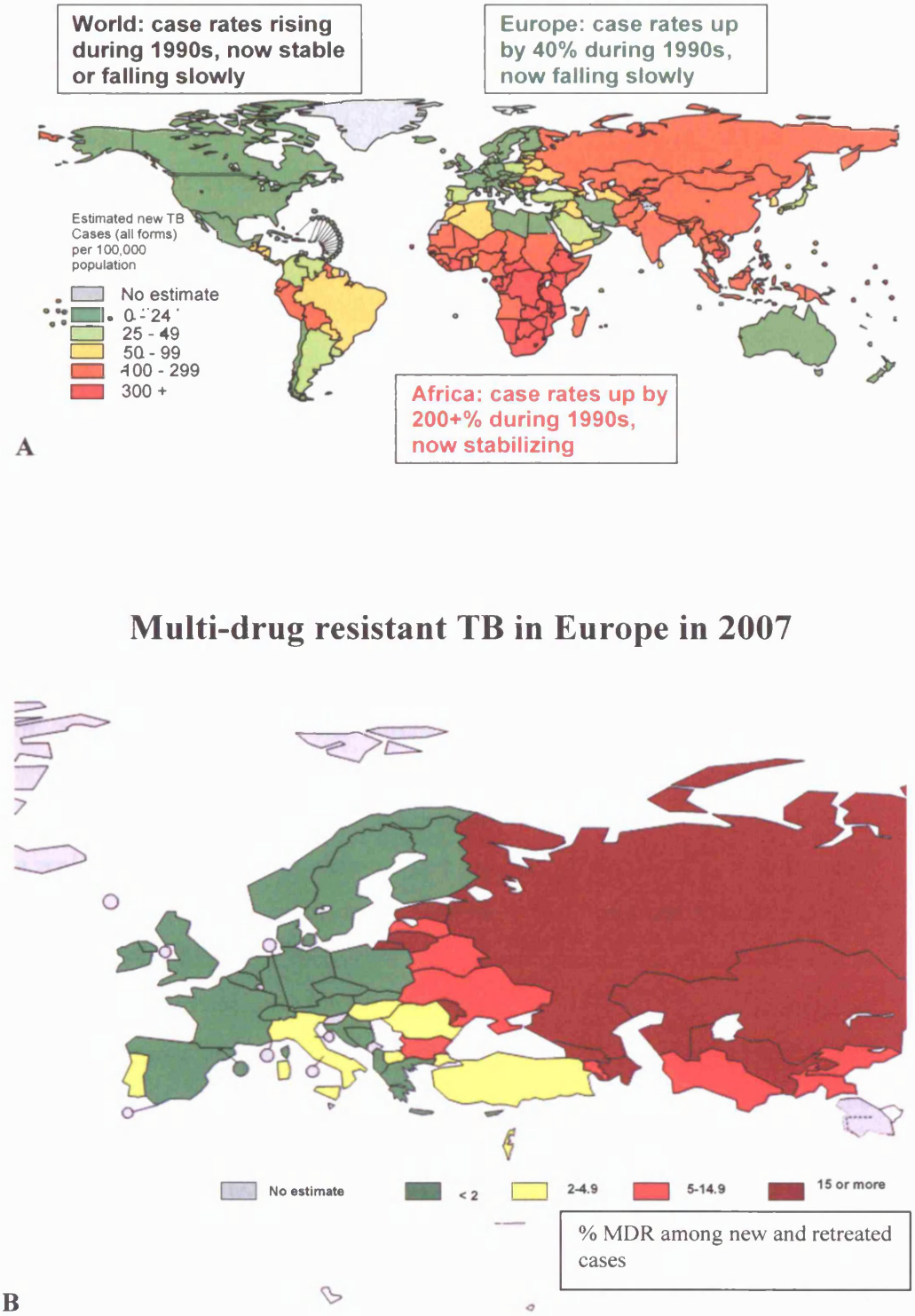


Figure 1.2: A: Global estimate of new TB cases, as shown by colour coding, and an indication upon current case trends. **B:** European scale of multi-drug resistant (MDR) cases of TB in both new and retreated patients, represented as a colour-coded % of MDR cases. Obtained from W.H.O. Global Tuberculosis Control Report 2007.

A particularly dangerous form of drug-resistant TB is multidrug-resistant (MDR) TB, which is defined as the disease caused by TB bacilli resistant to at least isoniazid and rifampicin, the two most powerful anti-TB drugs. Rates of MDR TB are high in some countries, especially in the former Soviet Union, and threaten TB control efforts.

While drug-resistant TB is generally treatable, it requires extensive chemotherapy (up to two years of treatment) with second-line anti-TB drugs which are more costly than first-line drugs, and which produce adverse drug reactions that are more severe, though manageable.

The emergence of MDR TB, particularly in settings where many TB patients are also infected with HIV, poses a serious threat to TB control, and confirms the urgent need to strengthen basic TB control and to apply the new W.H.O. guidelines for the programmatic management of drug-resistant TB.

1.5. *Mtb* dormancy.

Wayne and Sramek showed that abrupt exposure to anaerobic conditions of actively growing cultures of *Mtb* leads to cell death. However, incubation under conditions in which oxygen is depleted gradually causes *Mtb* to shift down from active replication to dormancy (Wayne and Sramek, 1994). It is in this dormant state that the bacilli are resistant to bactericidal effects of anaerobiosis and also exhibit partial or complete resistance to the bactericidal effects of the drugs isoniazid and rifampicin. *Mtb* bacilli can persist for decades in a dormant state inside a granuloma, particularly in the apical regions of the lung (Fenton and Vermeulen, 1996). Although the precise nature of the environment inside the granuloma is unknown, many factors including oxygen deprivation, nutrient depletion, low pH, toxic oxygen species, and other adverse conditions of this environment, could potentially contribute to the induction and maintenance of mycobacterial dormancy (Barry, Crane and Yuan 1996). In this state, alternate biosynthetic pathways generate cell-wall components such as mycolic acids, mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan, and the action of several of these may contribute to mycobacterial longevity, as well

as triggering inflammatory host reactions and acting in pathogenesis (Cole *et al,* 1998).

1.6. Location of latent *Mtb* in infected human hosts.

For intrapulmonary re-infection of host tissue, latent bacilli are harboured in the old pulmonary lymph nodes that comprise the Ghon complex (a scar in the lungs of TB patients who have successfully controlled the infection, where calcified lesions corresponding to the initial site of pulmonary infection are present). In the extrapulmonary cases, bacilli have been found in bones, brain, lymph nodes and other tissues. The presence of bacilli in these regions represents 15 % of latent TB reactivation cases (Bishai *et al,* 1998; Bishai, 2000 and Harris, 2001). Hence, a hypothesis by Bishai proposed that latency is maintained in widely disseminated sites and that the bacilli are seeded to these locations during the early bacteremic phase of initial exposure (Bishai *et al*, 1998). Apart from pulmonary involvement, the most common extrapulmonary locations of the disease are (in decreasing order): pleural, lymphatic, urogenital, osteoarticular, and meningeal tissues—although as has been pointed out, any organ or tissue can be affected. In immunocompetent patients, the frequency of presentation of extrapulmonary TB is no greater than 15 % to 20 %, with this figure increasing in situations of immune deficiency, as in the case of patients with AIDS, in whom extrapulmonary disease accounts for 50 % to 60 % of all TB cases.

1.7. Microscopic detection of latent TB bacilli.

Bishai *et al.* make reference to autopsy and surgical investigations into patient deaths that were a result of initial TB infection that failed to reveal, via microscopic examination, any acid-fast bacilli (Bishai *et al*, 1998). Acid-fast bacilli are a characteristic bacterial form associated with active infectious TB and is a physical property of some bacteria referring to their resistance to de-colorization by acids during staining procedures. Latent bacilli remained acid fast but were present in such low numbers that they defied microscopic detection. A

hypothetical reasoning for this phenomenon is that the bacilli enter a different developmental state such that they were no longer acid fast yet still viable.

1.8. Physiological state of the *Mtb* bacilli during latency.

The physiological state of *Mtb* bacilli during latency has, as yet, no definitive description. One view, in light of the evidence suggesting that *Mtb* is in a non-acid-fast form during latency, along with the presence of regulatory gene homologs of sporulation (DeMaio *et al,* 1996) and the ability of latent bacilli to remain essentially invisible to the immune system for such long periods of time, is that they exists in a spore-like state. There is also data suggesting that latent bacilli are metabolically active. Trials involving the use of chemoprophylaxis (the administration of an antibiotic agent to prevent an infection or to prevent an incubating infection from progressing to disease) have demonstrated that asymptomatic, non-infectious individuals have a significantly reduced risk of developing reactivation TB when treated with the antimycobacterial drugs isonazid or rifampin, as compared with their untreated counterparts (Bishai *et al,* 1998).

1.9. Reactivation of latent *Mtb*.

Work carried out by DeWit *et al.* utilising a mouse model concluded that the irregularities in bacilli numbers recovered (with respect to the *Mtb* DNA extracted from them) relative to the acid-fast forms of the bacilli seen by microscopy may be linked to the possibility that reactivation originates from non-acid-fast forms of *Mtb* (DeWit *et al*, 1995). Reactivation of these latent organisms can also lead to post-primary disease, even in people who successfully fought their initial battle against *Mtb*. One of the major causes of TB is thought to be through such reactivation (Harris, 2001). One in three individuals worldwide has latent *Mtb* infection, with a 5–10 % lifetime risk of progression to active disease (Bishai, 2000 and Electronic Source 2, 2002). The cavitary stage (the rapidly growing *Mtb* bacilli from within the Ghon complex), as a result of reactivation, is the major

milestone in the development of virulent and drug-resistant strains. The increased oxygen concentration within the Ghon complex allows large numbers of the organisms to multiply and is the major contributing factor leading to the evolutionary steps needed for the development of drug resistant mutants (Nardell, 1993).

1.10. Treatment regimens for human patients infected with TB.

A number of antibiotics are effective against tuberculosis. There are ten drugs currently approved by the U.S. Food and Drug Administration (FDA) for treating tuberculosis. Of the approved drugs, the first-line anti-TB agents that form the core of treatment regimens include isoniazid (INH), rifampin (RIF), ethambutol (EMB), and pyrazinamide (PZA).

It is common for TB infected individuals to be administered two or more types of antibiotics. This is because treatment with only one drug can sometimes be inefficient, resulting in residual bacilli becoming resistant to that drug. This kind of drug resistance attributed to adaptive bacilli to a one-drug regimen is rarely found in other bacteria. Often, a third and fourth drug are used during the initial, intensive phase of treatment to shorten the duration of treatment and to ensure success even if drug resistance exists at the outset (Beer *et al*, 2003 and Thacker, 2003).

Tuberculosis bacteria are very slow-growing; the antibiotics must be taken for a long time (usually for six months or longer), which includes two treatment phases. The 'initial phase' lasting two months is followed by the 'continuation phase' lasting either four or seven months. The continuation phase is normally extended to 28 weeks for patients who have cavitation on the initial chest film and positive sputum cultures after two months of treatment. Treatment is stopped dependent on the number of doses ingested during the treatment cycle.

Although basic TB regimens are broadly applicable, there are modifications that should be made under special circumstances (i.e. HIV infection, drug resistance, pregnancy, or treatment of children fewer than 12 years of age). Table 1.1 lists the basic regimens most commonly applied to treat immune-compromised TB infected human patients.

Preferred Regimen	Alternative Regimen	Alternative Regimen
Initial Phase: Daily INH, RIF, PZA, and EMB for 56 doses (8 weeks)	**Initial Phase:** Daily INH, RIF, PZA, and EMB for 14 doses (2 weeks), then twice weekly for 12 doses (6 weeks)	**Initial Phase:** Thrice-weekly INH, RIF, PZA, and EMB for 24 doses (8 weeks)
Continuation Phase: Daily INH and RIF for 126 doses (18 weeks) **or** Twice-weekly INH and RIF for 36 doses (18 weeks)	**Continuation Phase:** Twice-weekly INH and RIF for 36 doses (18 weeks)	**Continuation Phase:** Thrice-weekly INH and RIF for 54 doses (18 weeks)

Table 1.1: The basic antibiotic regimens, presented as preferred and alternative treatments, implemented on the Initial and Continuation phase strategy of TB treatment.

A continuation phase of once-weekly INH/RIF can be used for HIV-negative patients who do not have cavities on the chest film and who have negative acid-fast bacilli smears at the completion of the initial phase of treatment (Beer *et al*, 2003).

1.11. Directly Observed Therapy (DOT).

Many patients find it difficult to remember to take their drugs every day. Other people, for various reasons, discontinue treatment as soon as they feel better. Because of these problems, many experts recommend that people with tuberculosis receive their drugs from a health care worker under a patient-centred case management system. This is called directly observed therapy (DOT). Because DOT ensures that the person takes every dose, DOT treatments are often

less than seven days, with drugs administered two or three times during the seven days (Thacker, 2003).

In extreme cases surgery is required to remove a portion of the lung unless the patient faithfully adheres to the drug treatment plan. However, for very drug-resistant infections surgery is sometimes needed. When tuberculous pericarditis causes significant restriction of the motion of the heart, the pericardium may need to be removed surgically. A tuberculoma in the brain may also need to be surgically removed (Beers *et al*, 2003).

1.12. The global plans to control the spread of TB – targets 2006–2015.

In 2006, the World Health Organisation (W.H.O) launched the new 'Stop TB Strategy' (Thomas, 2007). The core of the TB control approach is the DOT regimen, launched by W.H.O. in 1995. Since its inception, more than 22 million patients have been treated under DOT-based services. The six components of the 'Stop TB Strategy' are summarized in Table 1.2.

The six components of the 'Stop TB Strategy'

1. The pursuit of high-quality DOT expansion and enhancement.	2. Addressing TB/HIV, MDR-TB and other challenges.	3. Contributing to health system strengthening.	4. Engaging all care providers.	5. Empowering people with TB and communities.	6. Enabling and promoting research.
The targeting of the poorest and most vulnerable individuals for access to widely available high-quality services where the requirement for DOT expansion is needed, even to the remotest areas.	Supplementing DOT implementation with individual patient care and support, essential to providing the best patient-care possible.	The implementation of TB control programmes to oversee spending, regimen and treatment planning, together with competent and informed management to disseminate information and supply.	To ensure all patients receive high-quality care, requiring the need of all participating health-care providers.	Providing communities with the ability to perform TB control tasks. Thereby mobilizing civil societies and also ensuring political support and long-term sustainability for TB control programmes.	While current tools can control TB, improved practices and elimination will depend on new diagnostics, drugs and vaccines.

Table: 1.2: The six components of the 'Stop TB Strategy' as outlined by the World Health Organisation (W.H.O) in 2006 (Thomas, 2007).

1.13. An outline of this thesis.

TB is an ongoing and continually developing area of scientific research. With many issues concerned with the bacterium's ability to elicit disease in humans with ever more increasing evidence of BCG-vaccination resistance strains, which are proving even more difficult to treat with antibiotics, understanding the mechanisms that underpin this bacterium's ever developing advantageous arsenal is proving to be a fast developing area of research. This thesis encompasses 4 years of research focused on obtaining structural and biochemical information on *Mtb* protein targets through the application of nuclear magnetic resonance (NMR) spectroscopy and use of various other biochemical techniques.

This thesis is divided into a further seven chapters as follows.

Chapter 2 describes various techniques and protocols applied throughout this project.

Chapter 3 describes a review of for the first protein investigated during this Ph.D, *Mtb* thiol peroxidase (TPx wild-type, i.e. TPxWt) (Rv1932), including its mutated C → S variant (TPxC60S). The review will include all up-to-date structural research and biochemical details explaining the TPx protein group classification and providing a comparative examination into its protein fold. This chapter looks at the protein data base (PDB) deposited crystal structures of TPxWt and TPxC60S and outlines the differences that exist between the two proteins. The chapter concludes by raising the point that both TPxWt and TPxC60S apparently mimic the oxidised and reduced states respectively, which is the result of substantial local rearrangement of secondary structure elements proximal to the active site. Such as has been suggested for a homologous peroxiredoxin.

Chapter 4 begins by detailing the early biochemical analyses carried out by independent groups to produce the two crystal structures of TPxWt and its mutated variant. A look at the two crystal forms and the differences between them is developed here in greater detail. 1D [^{1}H], [^{1}H, ^{15}N]-HSQC and [^{2}H, ^{15}N, ^{13}C]-labelled-HSQC-NMR of TPxWt and TPxC60S is described with reference to the spectral differences between the two variants and the impact this had on obtaining

the back-bone resonance assignment for the Cys60 → Ser mutant of TPx, TPxC60S.

Chapter 5 describes the efforts applied to resolve issues of incomplete back-bone assignment through the use of NMR spectroscopy. These efforts included varying experimental conditions such as temperature, sample pH and DTT concentration and the use of ligands to stabilise putative dynamic protein motion.

Chapter 6 details a summary discussion and conclusion on the results obtained from *Mtb* TPx during the course of this project and highlights the possible avenues of research that may be undertaken in resolving the issue of the incomplete back-bone assignment of TPxC60S as described in Section 5.

Chapter 7 details the background information for the final protein analysed in this thesis; an RNA polymerase sigma factor J (sigJ) (Rv3328c), a protein important in *Mtb* replication most notably in the process of bacterial transcription. The protein family (σ^{70}), and sub-division (ECF) to which sigJ belongs is described, highlighting similarities and differences between the various sigma factor groups in terms of the domain homology and organization.

Chapter 8 details the various steps taken to optimize recombinant sigJ expression and the examination into low temperature stability. The use of biochemical techniques to investigate the secondary structure of sigJ using circular dichroism (CD) and analytical size exclusion chromatography (ASEC) is also described. An NMR analysis on sigJ is also discussed with respect to the soluble and refolded forms of the protein in 1D-[^{1}H] and 2D [^{15}N]-HSQC spectra. Chapter 8 concludes with details of the various efforts applied in trying to resolve the 'poor' NMR spectra of natively folded sigJ and in the use of electrophoretic mobility shift assays (EMSA) to examine the potential for DNA binding.

Chapter 2

Materials and Methods

2.1. Chapter summary:

The proteins described in this thesis – TPxWt, TPxC60S and sigJ – have all been subject to a variety of experimental protocols and biophysical techniques. In each case the methods have been applied in essentially similar ways. Therefore, the following gives in-depth details for the experimental methodologies applied throughout the course of this project, indicating where applicable instances where the details have varied between the application to TPxWt and TPxC60S. The methods include DNA amplification and DNA sequencing preparations, bulk cell culture, protein purification and quantitation techniques, secondary structure determination and analysis tools which include circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy, and a variety of methods to assess potential ligand- or pH-induced structural changes that involve both NMR and short-wavelength spectrophotometry. The methods applied using sigJ are also presented and include the use of PCR, electrophoretic mobility shift assays (EMSA), limited proteolysis experiments and freeze/thaw studies.

This chapter concludes with a look at NMR spectroscopy and some of the many ways in which this powerful technology has been diversified and developed to aspects of protein structure in solution.

2.2. General chemicals and lab equipment.

Reagents were obtained from either Sigma-Aldrich® or Invitrogen™ unless otherwise stated. All solutions were made with the Elga® Maxima Ultra™ pure water purification system.

For all centrifugation steps, either bench-top scale or large-preparation culture scale, the following centrifuges were used: an MSE® microcentaur™ (all microfuge tubes); Eppendorf™ 5810 or Heraeus® Labofuge™ 400R (15 mL and 50 mL Falcon™ tubes); and Sorvall® RC5B centrifuges (e.g. 1L bottles).

All protein concentration steps were performed with Amicon® Centriprep®, or Vivaspin6® and Vivaspin5000® sample concentrators.

Sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using Invitrogen™ pre-cast 1.0 mm, 15 well NuPAGE™ Novex® 4-12 % Bis-Tris Gels. Electrophoresis was carried out in an Xcell SureLock™ Mini-Cell. Samples were prepared and loaded onto the gel according to the manufacturer's protocol for denaturing reduced samples. 10 μL (including 2 μL loading buffer) of sample was loaded into each gel-well following a period of heating for 5 mins at 100 °C (Invitrogen™ NuPage™ Technical Guide). Gels were stained using Coomassie blue (see Recipe 1) for 45 mins.

Recipe 1: Coomassie blue stain ingredients.

0.25 g	Coomassie blue powder
75 mL	Glacial acetic acid
500 mL	95-100 % ethanol
Double distilled water (ddH_2O) added to total volume of 1 L.	

Following electrophoresis the gels were drained and left in destaining solution (10 % ethanol, 7.5 % acetic acid, made up with ddH_2O) for 30 mins on a bench-top orbital shaker, before imaging. Imaging was carried out using an UVi DOC™ gel documentation system.

Protein concentrations were established using a Biomate-4™ spectrophotometer at 280 nm in a 1 mL quartz cuvette. Readings for each protein sample were taken twice, using two different dilutions. Protein dilutions were chosen to give an absorbance at 280 nm measurement between 0.01 and 0.5 to reflect the accurate dynamic range of such a measurement. Protein concentrations were calculated using the Beer-Lambert law, stating that the *absorbance* = $\varepsilon.c.l$ (where l is the path length, c is the protein concentration, and ε is the molar extinction coefficient). The theoretical extinction coefficients at 280 nm were obtained from the Expasy® Protparam tool (Gill and von Hippel, 1989), predicted for each protein using its primary amino acid sequence.
For preparative and analytical size exclusion chromatography an ÄKTA™ purification system was used. For the preparative size exclusion chromatography a Superdex™-75 column (Amersham® Biosciences) was employed; an analytical-Superdex™-75 was used for the analytical size exclusion chromatography.

In this thesis all references to the genetic code and amino acid nomenclature during this project are made in accordance with the details listed in Appendices A1 and A3.

2.3. General protocols.

The two main proteins described in this thesis, TPxWt and TPxC60S (including sigJ), were prepared from *E .coli* cells (BL21 – Invitrogen™) transformed with the pET22B(+) plasmid (Novagen®) containing the relevant cDNA insert in frame with the T7 promoter and including the *lac* repressor and ampicillin and kanamycin resistance genes. All proteins expressed contain an N-terminal polyhistidine (6-His) affinity tag. The details of the vector used and the relevant insert location for each of the proteins used in this project, including amino acid sequence, are provided in Appendices A2-A4 and A9. The German Research Centre for Biotechnology (GBF) supplied sigJ-, TPxWt- and TPxC60S-expressing cells on Luria broth (LB) agar. As is good practice, before expression of the protein we took steps to check the gene inserts by amplifying the relevant plasmid

DNA for nucleotide sequencing. For the sake of completeness this procedure is described below in Sections 2.3.1-2.3.1.3.

Cells from each plasmid were streaked individually onto an LB prepared agar plate (see Recipe 2), treated with the appropriate antibiotic, and incubated overnight. The resulting colonies were picked and again grown overnight in liquid culture in antibiotic-treated LB. Glycerol stocks were subsequently prepared (1 mL volume and 10 % glycerol v/v) and stored at -80 °C.

Recipe 2: Luria-Bertani (LB) media.

For 1 L of medium in double-distilled water (ddH_2O)*:	
5 g	Yeast extract
5 g	Sodium chloride (NaCl)
10 g	Tryptone
*1 L sample subsequently autoclaved and treated with the appropriate antibiotic (1 mL from a 1000x stock).	

2.3.1. DNA sequencing - sample preparation.

The following (sections 2.3.1.1-2.3.1.4) describes the methodology applied in obtaining an isolated plasmid, containing the DNA coding region of the protein of interest, required for subsequent nucleotide sequencing.

2.3.1.1. Plasmid extraction.

A 250 μL volume of BL21 (DE3) *E. coli* cells harbouring the desired expression vector were added to 25 mL antibiotic-treated LB (100 mg/mL carbenicillin - an analogue of ampicillin with increased stability) and incubated at 37 °C in a mechanical shaker overnight. The cells were spun down to a pellet in a bench-top centrifuge at 3500 rpm for 10 mins. The supernatant was discarded and the plasmid vector was isolated from the cells using a Qiagen® Qiaprep Spin Miniprep kit in accordance with the supplied protocol.

2.3.1.2. Plasmid identification and transformation.

Because recovery of plasmid DNA from BL21 (DE3) cells is known to be inefficient, amplification in a separate non-expression strain is typically required prior to DNA sequencing. Thus positive identification for the presence of a BL21 (DE3)-derived plasmid putatively containing the protein-encoding DNA insert was obtained by 0.75 % agarose gel electrophoresis (see Recipe 3), with and without restriction endonuclease digestion (see Recipe 4). The small amount of vector was transformed into *E. coli* DH5α cells (Invitrogen™) by means of heat-shock treatment and spread onto a solid LB-agarose plate. The plate was then incubated at 37 °C overnight.

Recipe 3: Agarose gel electrophoresis (0.75% agarose):

0.375 g molecular grade agarose was dissolved into 50 mL TAE buffer (see Recipe 5) and poured into 5x6 cm gel mould with a 20 μL comb in place, and left to set for 30 mins.

DNA samples for electrophoresis were prepared by adding 5 μL of sample to 2 μL bromophenol blue (1:1 ratio glycerol: water and bromophenol blue sodium crystals to give opaque brown) – migrates at same rate as 200-400 bp nucleic acid. Glycerol adds density to make sample sink to bottom of well. An additional sample of DNA marker was made; 3 μL 1 Kb DNA marker to 2 μL bromophenol blue. All references to DNA ladder refer to the 1 kDa DNA ladder obtained from Invitrogen™ composed of 12 DNA fragments ranging from 1,018 bp to 12,216 bp. The prepared gel was then submerged in the centre of the electrophoresis gel tank containing TAE buffer and the prepared samples, including DNA marker were subsequently loaded.

Electrophoresis at 50 V was then carried out and stopped once the dye front reached three quarters down the length of the gel. Staining of the gel was done with ethidium bromide for 30 mins, washed with ddH_2O, and imaged under long-wave ultra-violet (UV) light.

Recipe 4: Restriction endonuclease reaction:

HindIII reagent (1 μL HindIII enzyme, 2 μL HindIII buffer, 7 μL ddH_2O) was added to 10 μL plasmid vector and incubated in a 37°C water bath for 1 hour 30 mins, before cooling on ice. The resultant samples were then analysed by agarose gel electrophoresis (Recipe 3).

Recipe 5: TAE (Tris Acetate EDTA) buffer – 50x:

For 1 L buffer:	
242 g	Tris base
57.1 mL	Acetic acid
100 mL	0.5 M EDTA
ddH_2O added to 1 L and pH adjusted to 8.5.	

2.3.1.3. Transformation by heat shock protocol.

50 μL of *E. coli* DH5α cells was pipetted into each of 2 Eppendorf™ tubes labelled: + (nuclease reaction) and – (control). 5 μL of isolated plasmid DNA was then added to the + Eppendorf™ tube and left on ice for 30 mins. Both Eppendorf™ tubes were then placed in a pre-set 42°C water bath for 60 seconds and then transferred onto ice for 2 mins. 250 μL of LB (no antibiotic included) was then added to each tube, which were then placed into a pre-set 37 °C water bath for 1 hour. 50 μL from each Eppendorf™ tube was then spread onto a separate LB-agarose plate. The plates were then left to incubate at 37 °C overnight.

2.3.1.4. Plasmid extraction from DH5α cells.

A colony was picked and grown up as before, but this time in two separate falcon tubes containing LB medium. A Qiagen® miniprep was carried out on one batch of cells and the other was stored at -80°C in cryo-vials™ with 10 % v/v glycerol

added for future use. The plasmid DNA was incubated on the kit miniprep spin column for 5 mins before being eluted with water. A restriction endonuclease digestion was performed and a DNA gel was run to ensure correct plasmid purification. The purified plasmid DNA was submitted for forward and reverse insert sequencing by the external company, MWG-Biotech™ using standard T7 promoter and terminator oligonucleotide primers (supplied by MWG-Biotech™). The standard genetic code and nucleotide/amino acid sequences of TPxWt and TPxC60S are provided in Appendices A1, A3, and A4 respectively and Appendix A9 for sigJ. A translation of the amino acid abbreviations is also provided in Appendix A5.

2.3.2. Bulk protein production by *Escherichia coli* cell culture.

The following describes procedures adopted for the preparation of milligram quantities of TPx and sigJ proteins for NMR and other studies.
As a generic first step 250 µL of expression strain BL21 (DE3) *E. coli* cells from glycerol stocks were incubated and shaken overnight at 37 °C in 25 mL ampicillin-treated LB media.

2.3.2.1. Unlabelled protein.

The initial 25 mL culture was scaled-up to three litres of LB medium by the addition of 4 mL to each of 6 x 500 mL media in autoclaved, baffled conical flasks. (No intermediate culture size was required due to rapid *E. coli* growth in LB media.) Alternatively, for smaller scale cultures, i.e. 1 L preparations, 8 mL was added to 2 x 500 mL media in autoclaved, baffled conical flasks.

2.3.2.2. ^{15}N-labelled protein.

For the generation of single isotope-labelled protein 2 mL of the initial 25 ml LB culture of the *E. coli* expression strain was removed and added to 100 mL of feed source loaded minimal M9 media (see Recipe 6) containing 99 % ^{15}N-ammonium

sulphate (Cambridge Isotopes™). The culture was then incubated with shaking overnight at 37 °C. 4 mL aliquots of this starter culture were added to each of six baffled conical flasks containing 500 mL labelled M9 media (total culture volume: 6 L)

Recipe 6: M9 minimal media ingredients**.

For 1 L of medium in ddH_2O*:	
6.0 g	NaH_2PO_4
3.0 g	K_2HPO_4
0.5 g	NaCl
1.0 g	$(NH_4)_2SO_4$*
Feed source added:	
2 mL	1 M MgCl
10 µL	1 M $CaCl_2$
1 mL	0.01 M $FeSO_4$
1 mL	1000 x Vitamins (see Recipe 7)
1 mL	1000 x Micronutrients (see Recipe 8)
1 mL	1000 x Carbenicillin
10 mL	20 % w/v Glucose*

* Isotope-labelling for NMR spectroscopy was obtained by replacing $(NH_4)_2SO_4$ with $(^{15}NH_4)_2SO_4$, glucose with $^{13}C_6$-glucose, and ddH_2O with ^2H_2O (D_2O) as appropriate.

** The media was sterilised by autoclaving prior to the addition of the feed source components.

Recipe 7: M9 Vitamins.

1000 x Vitamins solution/L:	
0.4 g	Choline chloride
0.5 g	Folic acid
0.5 g	Pantothenic acid
0.5 g	Nicotinamide
1.0 g	Myo-inositol
0.5 g	Pyridoxal HCl
0.5 g	Thiamine HCl
0.05 g	Riboflavin
1.0 g	Biotin

Recipe 8: M9 micronutrients.

1000 x micronutrient solution/L:	
3 μM	Ammonium molybdate
400 μM	H_3BO_3
30 μM	$CoCl_2$
10 μM	$CuSO_4$
80 μM	$MnCl_2$
10 μM	$ZnSO_4$

2.3.2.3. [^{15}N, ^{13}C, ^{2}H]-labelled protein.

For the generation of triple isotope-labelled protein 2 mL of the initial 25 ml LB culture of the *E. coli* expression strain was removed and added to 100 mL of feed source loaded minimal M9 media (see Recipe 6) containing 99 % ^{15}N-ammonium sulphate and $^{13}C_6$-glucose (Cambridge Isotopes™) prepared in 99 % D_2O (Goss Chemicals™) . The culture was then incubated with shaking overnight at 37 °C. 4

mL aliquots of the overnight starter culture were added to each of four baffled conical flasks containing 500 mL labelled M9 D_2O media (total culture volume: 2 L) .

2.3.2.4. Induction of protein expression.

Large-scale cultures were carried out in a mechanical shaker operating at 37 °C. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (600 μL per 500 mL culture volume) once the optical density recorded at 600 nm (OD_{600}) had reached between 0.6 and 0.8. Expression was carried out for 12 hours for the single ^{15}N-labelled cultures and for 20 hours for the triple isotope-labelled culture.

2.3.2.5. Harvesting and homogenisation.

Following induction, cells were harvested by centrifugation for 20 mins at 6000 rpm. Protein expression was assessed by SDS-PAGE (methodology as described previously). The proteins of interest were expressed in the cytoplasm; thus the supernatant was subsequently removed and retained for sterilization and subsequent disposal, and the cell pellet retained for further processing. For purification the cell pellet were then removed from the centrifuge tube and re-suspended in 15 mL aliquots of PBS buffer (see Recipe 9).

Recipe 9: PBS Buffer.

For 1L buffer in ddH_2O:	
8 g	NaCl
0.20 g	KCl
1.44 g	NaH_2PO_4
0.24 g	K_2HPO_4
Buffer adjusted to pH 7.4 and autoclaved.	

The suspension was passed through a French Press (American Instrument Company™), five times, at a pressure of 1000 psi. The resultant liquid, containing ruptured cells, was centrifuged at 15000 rpm for 40 mins. The pellet fraction was sterilized and subsequently discarded, whilst the supernatant was retained for further fractionation as described in Section 2.3.3.

2.3.3. Ni-NTA affinity chromatography.

Uncharged nickel-nitrilotriacetic acid (Ni-NTA) agarose beads (Qiagen®) were prepared as follows: fresh resin was washed three times with 10 mL ddH_2O then incubated with 10 mL of 100 mM $NiSO_4$ for 30 mins to charge the matrix with nickel(II) ions. The charged resin was then washed three times with 10 mL ddH_2O before being equilibrated in 10 mL elution buffer (Recipe 10). Between each washing step the resin was centrifuged for 3 mins at 1250 g and the supernatant carefully removed and discarded.

The clarified cell supernatant was added to 5 mL Ni-NTA agarose beads in a 50 mL Falcon tube and rolled for 1 hour and 30 mins. The resin was pelleted by low speed centrifugation. Sequential washes of the protein-bound resin were performed using ten different elution buffers containing increasing concentrations of imidazole (see Recipe 10). Imidazole competes with the 6-His tag of the target protein for Ni-NTA binding and results in the eventual elution of the protein from the Ni-NTA beads. For each wash step the Ni-NTA resin was left to incubate at 4°C on a mechanical roller for 5 min before being centrifuged at 3000 rpm for 5 mins and the liquid above the settled beads subsequently removed and stored for further analysis.

Recipe 10: Affinity chromatography elution buffers.

For 1 L of each buffer in ddH_2O:	
6.90 g	50 mM NaH_2PO_4
5.84 g	100 mM NaCl
0.154 g	1 mM DTT
0.037 g	0.1 mM EDTA
0.292 g	Imidazole (Buffer A)
29.2 g	Imidazole (Buffer B)
Elution buffers comprised different proportions of stock phosphate buffers, buffer A (5 mM imidazole) and buffer B (500 mM imidazole). pH adjusted to 7.4.	

Elution buffer	Buffer A (v/v)	Buffer B (v/v)
E2	100 %	0 %
E2.5	90 %	10 %
E3 (applied twice)	80 %	20 %
E3.5	70 %	30 %
E4 (applied twice)	60 %	40 %
E5	40 %	60 %
E6	20 %	80 %
E7	0 %	100 %
50 mL of elution buffers made up in 50 mL Falcon tubes.		

Wash fractions were assessed by SDS-PAGE and those which contained the protein of interest were concentrated to a final volume of 5 mL using a Vivaspin5000® molecular weight cut-off tube centrifuged at 3500 rpm.

2.3.4. Size exclusion chromatography.

Preparative scale size exclusion chromatography (SEC) was carried out using an Amersham Biosciences Superdex™-75 HR16/70 column in conjunction with an ÄKTA™ Purifier 10 system of the same manufacturer. The column was

equilibrated with 2.5 column volumes (~300 mL) of SEC phosphate buffer (see Recipe 11) using a method file constructed using the UNICORN™ software provided with the ÄKTA™ system.

The concentrated protein sample was injected (5 mL) and the column was washed with 2.5 column volumes SEC buffer at a flow rate of 0.5 mL/min, collecting 5 mL fractions. The column eluate was continuously monitored by UV absorption at 260 nm and 280 nm.

Fractions giving UV absorbance at 280 nm were run on an SDS-PAGE gel. The fractions containing the protein of interest were pooled and concentrated to 500 µL using a Vivaspin5000® molecular weight cut-off tube centrifuged at 3500 rpm.

Recipe 11: SEC phosphate buffer.

For 1 L of each buffer:	
6.90 g	50 mM NaH_2PO_4
5.84 g	100 mM NaCl
0.154 g	1 mM DTT
0.037 g	0.1 mM EDTA
Made up to 1 L with ddH_2O and pH adjusted to 7.4.	

2.3.5. Analytical size exclusion chromatography.

Analytical SEC was performed using a 10/30 Superdex-75 column (Amersham Biosciences) connected to an ÄKTA™ Purifier 10 system. 100 µL of sample was injected and eluted at a flow rate of 0.5 ml/min using the SEC buffer (Recipe 11). Protein elution was monitored by UV absorption at 260 nm and 280 nm. For assessment of TPx proteins the chromatogram was separately recorded for four different dilutions of the stock protein solution in SEC buffer: undiluted (ca. 6 mM), 1-in-2, 1-in-4, and 1-in-6 (1 mM).

Note: Ion exchange chromatography (IEC) was performed using an XK 16/0 Q-Sepharose™ column (Amersham Biosciences) connected to an ÄKTA™ Purifier

10 system. The column was run in a similar manner to the SEC column with Recipe 11 being the basis for two buffers with varying salt concentration (200-650 mM) to create a linear elution gradient.

2.3.6. Circular dichroism spectropolarimetry.

Circular dichroism spectropolarimetry (CD) was performed on an Aviv™ Model 62DS CD spectrophotometer equipped using a thermoelectric cell holder and with a Hamilton™ microlab™ 500 series automatic titrator. Far-UV CD spectra were collected from 210 nm to 310 nm using a 1 mm path-length cell and a 3 second averaging time. Thermal unfolding data were collected at 222 nm from 20 °C to 95 °C. Samples were placed in a 2 mm path-length cuvette and heated in increments of 1 °C, using an equilibration time of 1 min and a 5 second measurement averaging time. Protein concentrations were kept constant at 40 μM, unless otherwise stated.

All CD measurements were corrected by subtracting the buffer spectra prior to conversion to molar ellipticity. All experiments were carried out in the same buffer at pH 7.4, unless stated otherwise: 50 mM NaH_2PO_4, 50 mM NaCl, 1 mM DTT and 0.1 mM EDTA. Data was processed using Aviv™ Software Systems V2.76. Data fitting of the CD thermal titration data was performed using Origin™ v5.0 from Microcal™ inc. USA.

2.3.7. Nuclear magnetic resonance spectroscopy.

The following (Sections 2.3.7.1-2.3.7.6) describes the methodology applied in producing a protein sample of adequate concentration fit for analysis by NMR spectroscopy. A brief description of the various NMR-based experiments is also provided below.

2.3.7.1. Sample preparation.

500 μL of protein sample was transferred to a 5 mm NMR tube and 50 μL of deuterium oxide (D_2O) was added. The addition of 10 % v/v D_2O provides a reference signal for the spectrophotometer field-frequency lock system.

2.3.7.2. NMR spectrometers.

NMR spectroscopy was carried out using Varian UnityPLUS™ spectrometers operating at nominal ^1H frequencies of 500 MHz and 600 MHz, each equipped with a triple resonance *Z*-axis pulse field gradient (PFG) probe and four radio frequency channels. Experiments were carried out at 25 °C, unless otherwise stated.

2.3.7.3. Two-dimensional [^1H, ^{15}N]-heteronuclear single quantum coherence spectroscopy.

Two-dimensional (2D) [^1H,^{15}N]-heteronuclear single quantum coherence (HSQC) spectroscopy was carried out using a pulse sequence with combined gradient coherence selection and sensitivity enhancement elements (Kay *et al,* 1992). Solvent peak suppression was carried out by using low power water flip-back pulses (Zhang *et al,* 1994). All experiments performed on the UCL NMR spectrometers were previously adapted from cited pulse sequences by staff members and were set up according to the standard guidance of Dr. R. Harris. All

spectra were recorded with 1024 complex points in the direct dimension and 100-256 complex points in the indirect dimension (unless otherwise stated). On the 500 (600) MHz spectrometer, frequency sweep widths of 4000 (4200) Hz in the ^{1}H direct and 1650 (2000) Hz in the ^{15}N indirect dimensions were employed (^{1}H transmitter set to the middle of the NH region of the spectrum). The number of scans recorded for each t_1 increment was varied depending on the protein concentration. 'Folded-in' or aliased signals such as those arising from arginine side chain NH groups were identified by manipulation of the 'f1180 flag' macro (developed in-house by Dr. R. Harris). The macro can be set to manipulate the t_1 increment timing and thereby the first order phase correction in the indirect dimension, such that the aliased peaks appear either in-phase (positive; f1180 = 'n') or 180° out of phase (i.e. negative; f1180='y'), was used with all HSQC experiments to 'fold in' Arg residues that would otherwise be located beyond the display parameters for most HSQC spectra during data processing.

2.3.7.4. Chemical shift perturbation analysis by [^{1}H, ^{15}N]-HSQC NMR titrations.

All ligands used for the titration experiments were purchased from either Sigma-Aldrich™ or Invitrogen™ and were all supplied as lyophilised powders which were dissolved into solution using the following buffer made up in ddH_2O: 50 mM NaH_2PO_4, 100 mM NaCl, 1 mM DTT (unless otherwise stated) and 0.1 mM EDTA, pH 7.4.

The interactions of the TPx proteins with putative ligands was assessed by NMR-based titrations using 2D [^{1}H,^{15}N]-HSQC spectroscopy using spectrometer parameters identical, or very similar to those described above. The temperature was kept constant at 25 °C and the sample pH maintained, before and after the addition of each aliquot of a given ligand, at pH 7.4. The experimental details for the titration series performed are listed in Table 2.1.

Table 2.1: Details of the NMR titration experiments performed with various proteins and ligands.

Ligand used	Supplier	Protein used (*Mtb*)	Total protein concentration[1] (mM)	Max. ligand concentration[2] (mM)	Titration points[3]	% Protein dilution[4]	DTT concentration[5] (mM)
(CH3COO)2Mg.4H_2O (Magnesium Acetate)	Invitrogen™	TPxC60S and TPxWt	0.50	200.00	4	10	1.00
4, 4'-dithiodipyridine (DTDP)	Sigma-Aldrich™	TPxC60S	1.00	4.60	7	12	0.00
2-[(2,6-dichlorophenyl) amino] benzeneacetic acid (Diclofenac sodium salt – "compound-40")	Sigma-Aldrich™	TPxC60S	1.00	4.00	8	10	1.00

Key: 1: Refers to the protein concentration prior to the first titration point, **2**: refers to the maximal concentration of ligand reached after the final titration point, **3**: refers to the number of titration points acquired, **4**: refers to the % of protein diluted following the final titration point, **5**: refers to the total DTT concentration in the sample during the titration experiment.

2.3.7.5. 2D [^{1}H, ^{15}N]-transverse relaxation optimised spectroscopy.

2D [^{1}H, ^{15}N]-transverse relaxation optimised spectroscopy (TROSY) (Pervushin *et al.,* 1997), including WATERGATE solvent suppression, followed an in-house implementation by Dr. R. Harris of the pulse sequences downloaded with permission from Prof. L. E. Kay (http://pound.med.utoronto.ca/pulse_reg.html) at the University of Toronto, Canada. All TROSY-based NMR experiments were carried out on the 600 MHz spectrometer. Acquisition parameters were essentially identical between the HSQC and TROSY experiments with the frequency sweep widths of 4200 Hz in the ^{1}H direct, and 2000 Hz in the ^{15}N indirect, dimension. The number of scans per increment varied depending on the protein concentration.

2.3.7.6. NMR triple resonance experiments for sequence specific resonance assignment.

A 1 mM [^{2}H, ^{13}C, ^{15}N]-TPxC60S sample (pH 7.4) was used for all heteronuclear 2D and 3D NMR experiments and recorded at 25 ^{o}C at a proton frequency of 500 MHz. The following experiments were obtained with locally implemented 'standard' pulse sequences: HNCA, HNCO, HN(CO)CA, HNCB, HN(CA)CB and CB(CACO)NH (Yamazaki *et al*, 1994A). The acquisition parameter details for each experiment are listed in Table 2.2.

Combined analysis of these 3D spectra allowed the correlation of N and H^{N} with CO, Cα and Cβ chemical shifts, to identify connected amino acid segments. The Cα and Cβ chemical shift distributions of each residue type, coupled with the probabilistic method set-up by Grzesiek and Bax (Grzesiek and Bax, 1993) facilitated the identification of connected spin systems with corresponding regions of the target amino acid sequence, thereby yielding sequence-specific resonance assignments. Section 2.10 details the general practical aspects of the assignment process while Section 4.6 describes the results of the application of this procedure to *Mtb* TPxC60S.

	^{1}H		^{13}C		^{15}N		
Experiment	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Transients**
HSQC	1024 (116 ms)	4400	N/A	N/A	256 (150 ms)	1700	8
HNCA	1024 (116 ms)	4200	64 (18 ms)	3600	32 (19 ms)	1700	8
HN(CO)CA	1024 (116 ms)	4200	64 (18 ms)	3600	32 (19 ms)	1700	8
HNCB	1024 (116 ms)	4200	64 (9 ms)	7100	32 (19 ms)	1700	16
HN(CA)CB	1024 (116 ms)	4200	64 (9 ms)	7100	32 (19 ms)	1700	16
CB(CACO)NH	1024 (116 ms)	4200	64 (8.5 ms)	7500	32 (19 ms)	1700	12
HNCO	1024 (116 ms)	4200	64 (27 ms)	1800	32 (19 ms)	1700	8

Table 2.2: Experimental details of 2D and 3D NMR experiments applied upon *Mtb* TPxC60S at 25°C using; Varian UnityPlus™ spectrometers operating at a ^{1}H frequency of 500 MHz.

2.3.8. Processing NMR data.

All raw NMR data were processed using the nmrPipe program of Delaglio and co-workers (Delaglio *et al,* 1995). Standard manipulations of data sets included application of apodization functions, zero-filling to the nearest 2^n points, Fourier transformation, first and second order phase corrections, and base-line corrections. The precise processing steps varied between data sets. Processed spectra were initially visualised in nmrDraw™ (Delaglio *et al,* 1995).

Following processing, spectra were exported into AZARA (Boucher, 2002) format. Multiple spectra were visualised and referenced in PLOT2 (Boucher, 2002). The sequential resonance assignment was performed using the ANSIG™ analysis package (Kraulis, 1991) using the contour and cross peak files generated in AZARA.

2.3.9. Measurement of protein sulfhydryls by reaction with DTDP.

The thiol group-reactive compound 4, 4'-dithiodipyridine (DTDP) was used in a series of titration experiments with TPxC60S monitored by NMR spectroscopy with concurrent UV/visible spectral analysis by absorbance changes at 250 nm and 324 nm (see Chapter 5). In the latter case the method described below was to quantify the number of reacting thiol groups within a given protein sample.

First, 5 g DTDP was mixed into 200 mL of 50 mM Na_2HPO_4, 100 mM NaCl, and 0.1 mM EDTA (pH 7.4) and allowed to stir over-night. The concentration of the dissolved mixture was then checked in a 1 mL quartz cuvette in Biomate-4™ spectrophotometer, using the same buffer as a blank. Using the Beer-Lambert law and an extinction coefficient of 21400 M^{-1} cm^{-1}, the final concentration o DTDP was assessed to be 140 mM. This stock solution was diluted as required.

To a 1 mM TPxC60S NMR sample, containing no dithiothreitol (DTT), DTDP was added sequentially to yield the following final concentrations: 200 μM, 400 μM, 1000 μM, 1600 μM, 2600 μM, 3600 μM and 4600 μM. Small aliquots (20 μL) were withdrawn from the NMR sample and analysed on the Biomate-4 spectrophotometer at two wavelengths – 250 nm and 324 nm – to determine the extent of DTDP reaction. The two wavelengths correspond to the intact DTDP chromophore, absorbing at 250 nm, and the free reacted by-product (4-thiopyridone) absorbing at 324 nm.

The following equation was used to estimate the number of reacting thiol groups for the targeted protein (Riener *et al,* 2002):

$$\text{molSH (in 0.00105 L)} = (A_{324s} - A_{324r} - A_{324p}) / (\Delta\varepsilon_{324} * 1 \text{ cm})$$

Where:

$A_{324s} - A_{324r} - A_{324p}$ yields the difference in absorbance due to reacted DTDP devoid of systematic errors;

A_{324s} = absorbance reading of 1 ml 40 μM protein plus 50 μL 0.5 mM DTDP read against H_2O;

A_{324r} = absorbance reading of 1ml protein buffer plus 50 μL 0.5 mM DTDP read against H_2O;

A_{324p} = absorbance reading of 1ml 40 μM protein plus 50 μL H_2O read against H_2O; and

$\Delta\varepsilon_{324}$ = 21,400 $M^{-1}cm^{-1}$ (effective molar absorptivity change upon reaction of cysteine with DTDP at pH 7.0; Riener *et al,* 2002).

The result of this calculation was normalised to the protein concentration determined by absorbance at 280 nm, using the predicted extinction coefficient as described above.

2.4. Polymerase chain reaction (PCR).

PCR mixtures (Recipe 12) were prepared on ice in thin-walled 200 μL PCR tubes:

Recipe 12: PCR reaction mixture.

Mixture component:	Amount added
Plasmid DNA template (50-100 ng)	30 μL
dNTP[1] mix (2 mM)	15 μL
5' primer (1 mM)	10 μL
3' Primer (1 mM)	10 μL
Pfu[1] DNA polymerase (80 μM)	3 μL
10x Pfu buffer[2]	10 μL
ddH_2O	To a final volume of 100 μL

1. An equimolar mixture of dATP, dCTP, dTTP and dGTP (Promega™)
2. Obtained from Promega™

PCR was carried out using a Primus™ thermocycler (MWG-Biotech™). The heating block of the thermocycler was pre-heated to 95 °C and the PCR samples added to the block straight from ice (a technique known as 'hot start' PCR) to improve the efficiency of the initial primer annealing event.
Samples were subjected to 30 rounds of thermal cycling (95 °C for 30 seconds, 95 °C for 30 seconds and 72 °C for 90 seconds) to amplify the required stretch of DNA.

All DNA was purchased from Invitrogen™ unless otherwise stated.

2.4.1. Primer sequences used.

HNS-*Salmonella Typhimurium* 300 bp-DNA stretch amplification,

Forward primer: 5' ATGAGCGAAGCACTTAAAATTCTGA 3'

Reverse primer: 5' TCAGAATTTTGAATGCTTCTCGCAT 3'

Upstream 300 bp-DNA sequence from sigF-*Mtb* amplification,

Forward primer: 5' GGATTTCGACGCCGTGGCCGACCTG 3'

Reverse primer: 5' CAGGTCGGCCACGGCGTCGAAATCC 3'

4 DNA constructs, OLR's (Overlapping Region), were then made from the 300 bp upstream DNA sequence from sigF-*Mtb*.

OLR1; 1 – 200 bp of 300 bp upstream DNA sequence from sigF-*Mtb*,

Forward primer: 5' GGATTTCGACGCCGTGGCCGACCTG 3'

Reverse primer: 5' GCGGTCAGGACATGCCAGCTAAAGC 3'

OLR2; 100 – 300 bp of 300 bp upstream DNA sequence from sigF-*Mtb*,

Forward primer: 5' CCGGATGCCACCCTGCGCCTGGTGG 3'

Reverse primer: 5' GCGGTCAGGACATGCCAGCTAAAGC 3'

OLR3; 1 – 120 bp of 300 bp upstream DNA sequence from sigF-*Mtb*,

Forward primer: 5' GGATTTCGACGCCGTGGCCGACCTG 3'

Reverse primer: 5' ACTTCGTCTTTTCGCGGATCGACCA 3'

OLR4; 80 – 200 bp of 300 bp upstream DNA sequence from sigF-*Mtb*,

Forward primer: 5' CCCGGTTGATTCGCTCGGCCTTGCC 3'

Reverse primer: 5' GCGGTCAGGACATGCCAGCTAAAGC 3'

2.5. Electrophoretic mobility shift assay (EMSA) experiments.

These experiments were carried out using DNA from two sources: 300 bp DNA sequence upstream of sigF (*Mtb*) (amplified and isolated from genomic *Mtb* DNA using a Midi-prep® gel extraction kit obtained from MWG-Biotech™ and carried out using the recommended protocol provided), and 300 bp fragment of *Salmonella Typhimurium* HNS (obtained from an in-house source). OLR fragments 1–4 were also used in the same manner described below. For the EMSA experiments carried out, 6 % agarose Novex® DNA precast retardation gels were used. From a 1 mM stock of sigJ, six dilutions were made; 1 mM, 0.8 mM, 0.6 mM, 0.5 mM, 0.4 mM, 0.3 mM and 0.1 mM.

The DNA concentration used for addition to each sample of the seven different protein concentrations was kept consistent (~40 ng/μL). Therefore the following initial scale volumes used for the PCR reaction mixtures are, sigJ (5 μL), DNA (5 μL), ddH_2O (2 μL), and loading dye (bromophenol blue – Invitrogen™) (3 μL) made up the final volume (15 μL) for each reaction mixture added to the 6 % agarose gel. A 1 Kb DNA ladder was also run concurrently with the samples already added to the 6 % agarose gel. A control of only DNA and no protein present was also run, with the protein replaced by 5 μL of ddH_2O in the final volume. Electrophoresis was carried out in an Xcell SureLock™ Mini-Cell at 50 V with variable amplitude for 3 hours in TBE buffer (Recipe 13). At this point the dye-front would have reached almost 1 cm before the end of the gel. The gel was then placed in ethidium bromide for staining for 10 mins and then visualised under UV detection.

Recipe 13: Tris-Borate-EDTA (TBE) buffer.

10x Tris-Borate-EDTA (TBE):	
Component:	**Amount added:**
Tris base	108 g
Boric acid	55 g
Na_4EDTA	9.3 g
Made up to 1 L using ddH_2O; no pH adjustment required	

2.6. Limited proteolysis.

Initially a 'range finding' experiment is conducted with differing protease enzymes at varying dilutions to determine the best protease type and concentration to use in the following 'time course' experiment.

All proteases were obtained from Sigma-Aldrich® and made up to 1 mg/mL in the manufacturer's recommended buffers (Recipe 14) and freeze-dried in liquid nitrogen. The subsequent samples are then kept frozen at -80 °C.

Recipe 14: Proteases (Sigma-Aldrich®) and the buffers used:

Storage and dilution buffer:	
Protease:	Final concentration and storage buffer:
Chymotrypsin	1 mM Tris, pH 8.0
Trypsin	1 mM HCL
Elastase	1 mM HCL

2.7. Limited proteolysis 'range finding' experiment.

Final sample concentrations and dilutions of proteases used:

- Chymotrypsin 1 mg/mL (1:1), 0.1 mg/mL (1:10), 0.01 mg/mL (1:100)
- Trypsin 1 mg/mL (1:1), 0.1 mg/mL (1:10), 0.01 mg/mL (1:100)
- Elastase 1 mg/mL (1:1), 0.1 mg/mL (1:10), 0.01 mg/mL (1:100)

The protein concentration was held fixed (1 mM) throughout each sample preparation. Therefore, for each reaction mixture 3 μL of protease was added to 10 μL of protein. Each reaction was left on ice for 5 mins before the protease is inactivated by adding 2 μL of loading dye (bromophenol blue – Invitrogen) and the addition of the samples to a heating block set to 100 °C.

Samples are then loaded onto an SDS-PAGE Invitrogen pre-cast NuPAGE™ Novex® 4-12 % Bis-Tris Gel 1.0 mm, 15 well gel, with protein standard markers with a control consisting of the protein minus protease. All references to 'protein marker' refer to Marker12 obtained from Invitrogen™ composed of Myosin (200 kDa), β-galactosidase (116.3 kDa), Phosphorylase B (97.4 kDa), BSA (66.3 kDa), Glutamic dehydrogenase (55.4 kDa), Lactate dehydrogenase (36.5 kDa), Carbonic anhydrase (31 kDa), Trypsin inhibitor (21.5 kDa), Lysozyme (14.4 kDa), Aprotinin (6 kDa), Insulin B-chain (3.5 kDa) and Insulin A-chain (2.5 kDa).

Electrophoresis was carried out according to the manufacturer's protocol in an Xcell SureLock™ Cell as previously referenced. Gels were stained using Coomassie blue for 45 mins. Gels were then drained and left in de-staining solution (10 % ethanol, 7.5 % acetic acid) for 30 mins on an orbital shaker, before imaging.

2.8. Limited proteolysis 'time course' experiment.

On choosing the most effective protease, a 'time course' experiment was then conducted to determine the optimum length of time to yield the most efficient digest pattern, as well as to provide details as to any fragments which may appear consistently undigested throughout the course of the experiment.

For the 'range finding' experiment the same protocol was followed as outlined above; the difference being the time at which the protease was 'deactivated' by adding loading dye (Marker12 – Invitrogen™, containing β-mercaptoethanol) to the sample and placing it in a heating block at 100 °C.

The entire 'time course' experiment was run over a two hour period, with samples containing protease being 'deactivated' after the following intervals: 5 mins, 10 mins, 20 mins, 30 mins, 1 hour and 2 hours. Electrophoresis followed by staining of the SDS-PAGE gel was then carried out, in a method as described previously for the 'range finding' experiment.

2.9. Freeze-thaw experiment.

The sensitivity of sigJ to degradation/precipitation to varying temperatures and glycerol concentrations, thereby determining the parameters by which purified sigJ could be stored under, was examined by qualitative visual assessment by SDS-PAGE (method as described above).

The glycerol percentage to store sigJ at sub-zero temperatures (-80 °C) was examined by adding the following glycerol percentages to three separate Eppendorf™ tubes, each containing 0.5 mM sigJ; 0 %, 10 % and 20 % v/v. All three Eppendorf™ tubes were then placed into a -80°C freezer for one week. Concurrently, a separate Eppendorf™ tube containing 0.5 mM sigJ containing 20 % glycerol v/v was left at room temperature (~25 °C), also for one week. All four samples were then examined by SDS-PAGE to visually determine the extent to which any degradation may have occurred.

2.10. Obtaining sigJ protein from inclusion bodies.

The process of extracting protein from *E. coli* expressed inclusion bodies, found in the insoluble cell fraction, and subsequent refolding is essentially the same as the method described in the Sections 2.3.2-2.3.4 but with the following differences.

After processing of the cell extract using the French Press and the subsequent second centrifugation step, the resulting supernatant is set aside and the cell pellet recovered. The cell pellet is re-suspended in PBS buffer containing 2 % v/v Triton X-100 and the resulting solution centrifuged at 15000 rpm for 40 mins. Following this step, the cell pellet is washed in PBS buffer only and then completely re-suspended in Buffer A (Recipe 15). The solution is then manually homogenised using a homogenising tube and pestle and then centrifuged at 15000 rpm for 40 mins.

Ni-NTA affinity chromatography is then applied in the same way as described in section 2.3.3 but using Buffer A and Buffer B (Recipe 15) to create the linear imidazole gradient. Wash fractions were assessed by SDS-PAGE and those which contained sigJ were concentrated to a final volume of 5 mL using a Vivaspin5000® molecular weight cut-off tube centrifuged at 3500 rpm. The refolding aspect of sigJ requires that the 5 mL concentrate is diluted (1:5) with Buffer A containing 1 mM DTT is then 'slow-exchanged' using a dialysis membrane against 50 mM 2-Amino-2-(hydroxymethyl)propane-1,3-diol-HCL (Tris-HCL), 300 mM NaCl and 1 mM DTT and adjusted to pH 8.0.

Recipe 15: Buffer A - sigJ refolding buffer.

For 1 L of each buffer in ddH_2O:	
6.05 g	50 mM Tris-HCL
17.52 g	300 mM NaCl
480.48 g	8 M Urea
0.292 g	Imidazole (Buffer A)
29.2 g	Imidazole (Buffer B)
Elution buffers comprised different proportions of stock phosphate buffers, buffer A (5 mM imidazole) and buffer B (500 mM imidazole). pH adjusted to 8.0.	

2.11. NMR spectroscopy structural and functional analysis of proteins.

NMR spectroscopy is the name given to the technique which exploits the magnetic properties of nuclei. A magnetic moment is intrinsic to all nuclei having either an odd or even number of nucleons. In simple terms the application of NMR exploits the alignment of such nuclei in a powerful externally-applied magnetic field and perturbing this alignment using radiofrequency electromagnetic radiation. The resulting behavior of the perturbed nuclei in the presence of the magnetic field gives rise to a response that is interpretable as a spectrum and encodes physical, chemical, and structural information about a molecule. From a practical perspective useful NMR spectra of molecules in solution requires that the observed nuclei possess magnetic quantum number spin, I = 1/2, such as ^{1}H, ^{13}C, and ^{15}N. Arguably NMR spectroscopic analysis of these nuclei provides the only method that is able to provide detailed information at the atomic level on the three-dimensional structure of biological macromolecules in solution.

One common goal in the use of NMR spectroscopy is to obtain high resolution three-dimensional structures of proteins, similar to that which can be achieved by X-ray crystallography. Such applications of NMR are mainly limited to relatively small proteins, usually smaller than 25-30 kDa, though technical advances allow ever larger structures to be solved. In addition to structural information for folded proteins, NMR can also be usefully applied to obtain high resolution information on partially or wholly unstructured proteins, as well as to characterise the global and internal dynamic properties of proteins and to monitor the interactions with other molecules.

2.11.1. 1D NMR and the need for multidimensional NMR spectroscopy.

The one-dimensional ^{1}H spectrum of an organic molecule reports on the chemical nature of the hydrogen atoms present and the resonance frequencies of the individual spectral features, routinely described as chemical shifts on a parts-per-million scale, are generally characteristic of the corresponding molecular fragment (e.g. aliphatic methyl group protons appear at ~ 1 ppm). In this context, the utility of the chemical shift is demonstrated by the fact that chemists can often use the 1D ^{1}H spectrum as confirmatory evidence of the structure of the target molecule. Complications arise in 1D

^{1}H NMR spectra when there are more than a relatively small number of protons present because of chemical shift degeneracy (resonance overlap). In addition, the spectroscopist in general wishes to derive additional information about the structure of the target by revealing connections between resonances that will yield further evidence for the chemical and spatial relationships between the observed ^{1}H atoms. The latter problem is clearly encountered in the application of NMR to biological macromolecules where the constituent 'units' of structure, e.g. amino acid residues and nucleotide bases, occur multiple times within the target molecule such that one is reliant on second order effects arising from macromolecular folding to provide some relief of the resonance overlap. Whilst the differing local environments of individual chemical groups in a protein give some dispersion to the ^{1}H NMR spectrum, deriving connectivity information between resonances requires extension of the NMR methodology into the realm of multiple dimensional NMR, including 2D, 3D and (sometimes) 4D spectroscopy. When nD NMR is combined with stable isotope labelling of the 'heavy' atoms of the polypeptide chain with spin-1/2 nuclei (^{13}C, ^{15}N) then, in principle, the spectroscopist has the scope both to resolve all resonances and reveal inter-atomic connectivity information that permits the resonance assignment and the measurement of both atom-by-atom structural and dynamic information. Over the past few decades the field of multidimensional heteronuclear NMR spectroscopy has mushroomed into a highly diversified and broadly applicable set of methodologies, for which even a relatively short summary could fill many volumes. Because much of what is practiced in the field has become essentially standard, what is described below represents a deliberately cursory description of just a few relatively straightforward elements of techniques used during the execution of the experimental work that is discussed in later sections of this thesis.

2.11.2. Heteronuclear single quantum coherence (HSQC) spectroscopy.

In the most general modes of application multidimensional NMR spectroscopy relies upon extensions of the 'pulse-Fourier transform' paradigm wherein a series of RF pulses and delays (the 'pulse sequence') is designed to yield a time varying signal response (the 'free induction decay', FID) from the sample that is digitised and 'transformed' in a computer into two or more frequency domains. Connectivities

between resonances in the spectrum are encoded in the output spectrum as 'cross peaks' that have coordinates corresponding to the connected chemical shifts, and an intensity that reports upon the magnitude of the inter-nuclear coupling. A powerful physics enables the design of pulse sequences that can provide different classes of connectivity information: these broadly fall into 'through-space' connections that rely upon inter-nuclear dipole-dipole couplings (e.g. the nuclear Overhauser effect), and 'through-bond' connections that rely upon scalar (or 'J-', or spin-spin-) coupling networks that are mediated by electrons in covalent bonds.

The pulse sequence is designed so that systematically incremented delays between pulses provide for the time domain encoding of the additional – indirectly detected – frequency domains. Thus each pulse sequence is applied multiple times, with the series time data recorded directly for each instance of the incremented time delay(s) stored separately in the computer. In general the output signal $s(t_1, t_2, t_3, \text{etc.})$ is then Fourier transformed to yield the spectrum $S(f_1, f_2, f_3, \text{etc.})$. Thus for a 3D NMR data set $s(t_1, t_2, t_3)$ the directly recorded FID information (a signal vector in variable t_3) is stored for each specific combination of t_1 and t_2. In general the directly acquired dimension corresponds to a ^{1}H frequency and has relatively high digital resolution compared to the indirectly acquired dimensions.

By way of an example of the 'through bond' category of NMR pulse sequences the following briefly describes the commonly used 2D heteronuclear single quantum coherence (HSQC) spectroscopy that connects a ^{1}H atom to its directly bonded heteronucleus (in this description, the ^{15}N atom of and N-H bond). In applications to proteins the 2D [^{15}N, ^{1}H]-HSQC pulse sequence provides for an information-rich 'fingerprint' that correlates ^{15}N and ^{1}H chemical shifts for the majority of residues along the polypeptide chain, as well as a subset of side chains.

As with many NMR pulse sequences designed to provide through-bond connectivity information the HSQC pulse sequence exploits the transfer of so-called 'transverse anti-phase magnetisation' between the J-coupled nuclei. Thus for two J-couple nuclei, it is possible – with appropriate RF pulses – to convert transverse magnetisation of one nucleus that is antiphase with respect to the other into transverse magnetisation of the second nucleus that is antiphase with respect to the first. In this way following Fourier transformation the two nuclei are seen to be coupled. In the HSQC pulse sequence a particular series of pulses and delays known as INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) is employed to transfer the superior polarisation of the ^{1}H atoms

(with a ca. 10-fold higher magnetogyric ratio γH) to the attached ^{15}N nuclei. Because phase cycling is used to suppress the intrinsic in-phase ^{15}N magnetisation following the INEPT sequence element the ^{15}N signal intensity is completely derived from the polarisation originating on the attached protons. A by-product of the INEPT module is that it leads to transverse antiphase ^{15}N magnetisation (arising out of the $^{1}J_{HN}$ scalar coupling-dependent evolution).

Following the first INEPT element, the antiphase ^{15}N magnetisation is allowed to evolve for a systematically incremented delay period of duration t_1, during which time the frequency of the ^{15}N magnetisation is effectively encoded. A second as-it-were 'reverse' INEPT sequence element is then employed to convert the transverse antiphase ^{15}N magnetisation into in-phase ^{1}H magnetisation that is recorded in the acquired FID (i.e. the output signal, $s(t_2)$). A 180° ^{1}H pulse centred in the t_1 delay and broadband decoupling of the ^{15}N nuclei during t_2 is utilised to collapse the ^{15}N and ^{1}H doublets respectively and thereby effect simplification and sensitivity enhancement of the resultant 2D spectrum.

The forgoing describes only the most basic implementation of the HSQC pulse sequence. In the applications of [^{15}N,^{1}H]-correlated spectroscopy employed in this thesis, the applied pulse sequences included additional elements that provided for further sensitivity enhancements. Examples include the use of shaped RF pulses and pulsed field gradients to achieve solvent suppression with minimal cross-saturation of the NH signals. These embellishments are well known in the field and a full description is beyond the scope of this thesis (Kay *et al*, 1992 and Zhang *et al*, 1994).

It is important to bear in mind that the application of additional elements in the pulse sequence will impact the overall sensitivity. Thus, whenever transverse magnetisation is generated within a pulse sequence (as is in effect is always the case) this corresponds to the nuclei acquiring an excited state that, although really quite stable compared to other types of spectroscopic excited states, begins to relax at a finite rate. It is important that, in order that NMR experiments are not compromised, that transverse magnetisation does not decay significantly on the timescale of the optimal delay periods for INEPT transfers, themselves determined by the magnitude of the relevant scalar coupling constant(s). In the case of the ^{15}N-^{1}H bond, the INEPT sequence employs two delay periods of $1/(4*{}^{1}J_{NH})$, where ${}^{1}J_{NH} \approx 90$ Hz, to facilitate magnetization transfer from ^{1}H to ^{15}N (and vice versa) corresponding to a total INEPT delay of 5.56 ms. For small proteins where ^{15}N T_2 values are greater than 100 ms the duration of such INEPT

transfer steps is not a limiting factor. However, for larger proteins the ^{15}N T_2 time is severely shortened and ultimately the applicability of multi-pulse experiments is limited by the relevant transverse relaxation rate constants.

In principle, the HSQC spectrum has one peak for each ^{1}H bound to the relevant heteronuclei. Thus, in the [^{15}N]-HSQC one cross peak is expected for each amino acid residue backbone amide group (except proline), along with some additional signals for certain N-containing side chains such as Trp ($N^{\epsilon}H$), Arg ($N^{\epsilon}H$, $N^{\eta}H_2$), and Asn, Gln (carboxy-NH_2). The [^{15}N]-HSQC spectrum is often referred to as the fingerprint of a protein because each protein has a unique pattern of signal positions. Analysis of the [^{15}N]-HSQC spectrum allows researchers to evaluate whether the expected number of peaks is present and thus to identify possible problems due to multiple conformations, unanticipated degradation and sample heterogeneity.

By examining the overall pattern of the chemical shift dispersion in a [^{15}N]-HSQC spectrum it is possible to ascertain qualitatively the extent of folding of a protein. Thus disordered regions of a polypeptide chain yield NH cross peaks in characteristic narrow ranges of chemical shifts, and are often disproportionally intense with narrow line widths; the relatively distribution of dispersed and 'random coil' cross peaks allows an estimation of the degree of foldedness. Whatever the conformational status of the target, where the [^{15}N]-HSQC spectrum becomes extremely useful is as a convenient means to monitor the effects of external perturbations to the protein, e.g. changes in the buffer or temperature conditions or the addition of potential binding partners. Because the chemical shifts of both the NH group ^{15}N and ^{1}H atoms is sensitive to the respective local chemical environments, changes to these parameters are interpreted to indicate that the NH is likely close to the site of interaction with the ligand, or point to a local change in the protein conformation in each of these circumstances. NMR experiments where temperature is varied are especially useful when it comes to proteins exhibiting varying extents of dynamic motion and is a useful method in probing the overall properties of a protein in solution. Performing NMR measurements at elevated temperatures carries the risk of partial or global unfolding of the protein which can be accompanied by non-reversible sample aggregation. On the other hand, a decrease in sample temperature can lead to slowing of the both tumbling rate, and exchange processes. Nevertheless, so long as the temperature is maintained below the denaturation/aggregation threshold two major effects on the sample arise upon an increase in the temperature: (a) decreasing the rotational correlation time of the protein (faster tumbling) leading to favourable

perturbations of the global average T_1 and T_2 relaxation time constants; and (b) acceleration of dynamic exchange processes: a show exchange phenomenon may become intermediate exchange with concomitant line broadening/averaging, an intermediate exchange process a fast one with line narrowing. In general therefore for large molecules such as proteins, superior resolution of the spectrum should be obtained at the highest compatible temperature. However, use of high temperatures risks the phenomenon known as 'solvent exchangeable proton signal bleaching', where the increased rate of exchange of labile protons with the bulk solvent leads to selective signal losses related to the degree of surface exposure.

The HSQC spectrum is often acquired with excellent sensitivity making it a relatively efficient means of probing the status of the protein under a variety of conditions. Moreover, the scouting conditions with the HSQC pulse sequence allows one to assess the optimal conditions for the much more time- and sensitivity-demanding triple resonance NMR experiments that are required to obtain a full resonance assignment.

2.11.3. Triple resonance experiments using double ^{13}C,^{15}N-enriched proteins.

Enriching a protein with ^{13}C as well as ^{15}N permits the recording of NMR experiments that directly and often simultaneously correlate particular subsets of atoms along the polypeptide chain. Usually the output of these experiments separately encodes the chemical shifts of the correlated ^{1}H, ^{13}C and ^{15}N nuclei in different dimensions. Many 'triple resonance' NMR pulse sequences have been developed that correlate back-bone and side-chain $^{1}H^{N}$, ^{13}C and ^{15}N spins in three or four dimensions (Gronenborn and Clore, 1995 and Cavanagh, *et al*, 1995). The acquisition of such multi-dimensional data sets permits numerous inter-atomic through-bond and through-space interactions to be individually identified in a relatively efficient and unambiguous manner.

Of the great variety of NMR pulse sequences that have been devised, the most popular can be thought of as expanding the 2D [^{1}H,^{15}N]-HSQC spectrum into a third ^{13}C dimension. These include the following: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CA)CB and CB(CACO)NH. The nomenclature of these experiments highlights the nuclei involved in the magnetization transfer pathways (Figure 2.1); spins whose chemical shifts are not explicitly encoded are denoted by parentheses. For example in

the HNCO experiment the spectrum contains a cross peak for each residue *i* at the chemical shift coordinates of N(*i*), HN(*i*) and C'(*i*-1) – in words, the [^{15}N]-HSQC cross peak can be associated with the carbonyl carbon chemical shift of the preceding residue. On the other hand, in the HN(CA)CO spectrum, for residue *i* cross peaks occur at both chemical shift coordinates representing {N(*i*), HN(*i*), C'(*i*)}and {N(*i*), HN(*i*), C'(*i*-1)} (because the $^{1}J(N^{i}C\alpha^{i})$ and $^{2}J(N^{i}C\alpha^{i-1})$ couplings have similar magnitudes). Similarly the HNCA experiment yields cross peaks at coordinates {N(*i*), HN(*i*), Cα(*i*)}and {N(*i*), HN(*i*), Cα(*i*-1)}, and the HN(CO)CA spectrum at {N(*i*), HN(*i*), Cα(*i*-1)}.

Combined analysis of these four spectra should allow, in the absence of significant chemical shift overlap, the connection of the majority of back-bone resonances, limited only by the presence of Pro residues in the polypeptide sequence (Pro has no amide proton).

In identification of stretches of connected spin systems obtained in this manner with the amino acid sequence of the target protein is aided by the recording of NMR spectra that connect side chain atoms with the Cα, H^{N}, N and C' resonances. The combination of experiments recorded using the HN(CA)CB and the HN(COCA)CB pulse sequences identifies the Cβ chemical shifts with each backbone amide group. Very often the pattern of amino acid-characteristic Cα/Cβ chemical shifts pairings can be reliably linked with the amino acid sequence of the target protein, allowing one to 'anchor' the sequential resonance assignments based on this set of triple resonance spectra.

As is illustrated in Figure 2.1 and summarized in Table 2.3, the magnetization transfers in these triple resonance NMR experiments are of the 'out-and-back' type, both originating and being detected on the back-bone amide H^{N} spin. Usually several of these experiments are required to resolve chemical shift overlaps in the ^{13}C dimension. This procedure is usually less ambiguous than the ^{1}H-NOESY-based sequential resonance assignment methods that were normally used prior to the advent of double [^{13}C,^{15}N]-isotope labelling, since it is based on unambiguous through-bond, as opposed to through-space, magnetisation transfers. In the NOESY-based methods additional cross peaks, corresponding to protons that are close in space but not belonging to the sequential residues will also appear, confusing the assignment process. When the backbone sequential assignment has been completed it is usually possible to assign the side chain ^{13}C and ^{1}H signals using, for example, the HCCH-TOCSY pulse sequence, which has the appearance of a 2D [^{1}H]-TOCSY experiment resolved in an additional ^{13}C dimension.

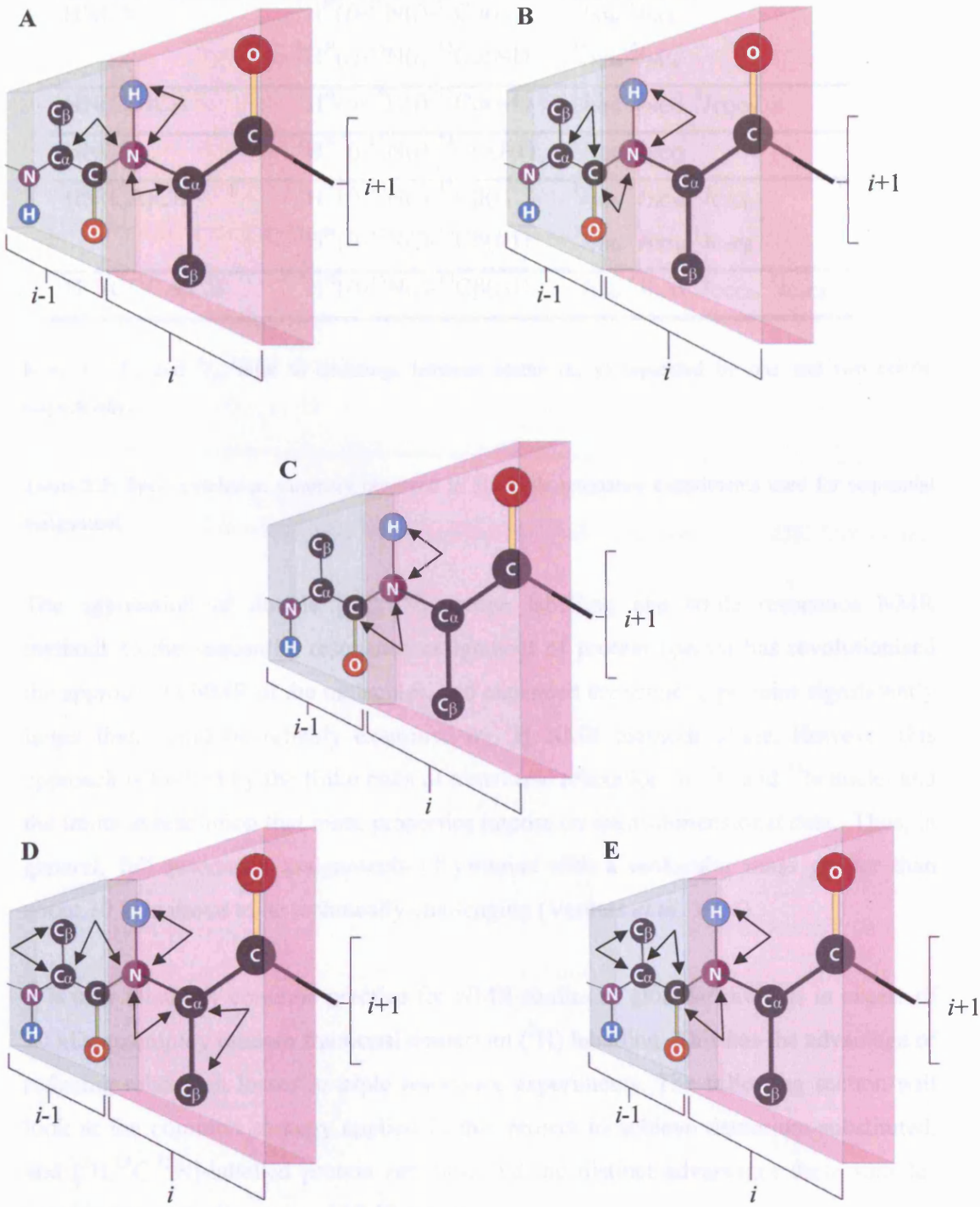


Figure 2.1: Experimental strategy applied in the resonance assignments of back-bone ^{1}H, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$ and CO nuclei in a dipeptide fragment showing residues *i* and *i*-1. The magnetization transfer pathways for each of these experiments is shown with black arrows; **A**: HNCA, **B**: HN(CO)CA, **C**: HNCO, **D**: HN(CA)CB and **E**: HN(COCA)CB. Spins whose chemical shifts are not evolved are put in parentheses in the pulse sequence name. These experiments display the 'out-and-back' description of magnetization originating and ending on the amide proton. Figure drawn using Adobe Photoshop CS2.

Experiment name	Correlation	J coupling[A]
HNCA	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\alpha(i)$	$^{1}J_{NH}$, $^{1}J_{NC\alpha}$
	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\alpha(i\text{-}1)$	$^{1}J_{NH}$, $^{2}J_{NC\alpha}$
HN(CO)CA	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\alpha(i\text{-}1)$	$^{1}J_{NH}$, $^{1}J_{NCO}$, $^{1}J_{COC\alpha}$
HNCO	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}CO(i\text{-}1)$	$^{1}J_{NH}$, $^{1}J_{NCO}$
HN(CA)CB	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\beta(i)$	$^{1}J_{NH}$, $^{1}J_{NC\alpha}$, $^{1}J_{C\alpha C\beta}$
	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\beta(i\text{-}1)$	$^{1}J_{NH}$, $^{2}J_{NC\alpha}$, $^{1}J_{C\alpha C\beta}$
HN(COCA)CB	$^{1}H^{N}(i)$-$^{15}N(i)$-$^{13}C\beta(i\text{-}1)$	$^{1}J_{NH}$, $^{1}J_{NCO}$, $^{1}J_{COC\alpha}$, $^{1}J_{C\alpha C\beta}$

Key: A: $^{1}J_{xy}$ and $^{2}J_{xy}$ refer to couplings between atoms (x, y) separated by one and two bonds, respectively.

Table 2.3: Spin correlation summary observed in 3D triple resonance experiments used for sequential assignment

The application of double [^{13}C,^{15}N]-isotope labelling and triple resonance NMR methods to the sequential resonance assignment of protein spectra has revolutionised the approach to NMR of the molecules, and expanded the scope to proteins significantly larger than could be reliably examined by ^{1}H NMR methods alone. However this approach is limited by the finite rates of transverse relaxation of ^{13}C and ^{15}N nuclei and the limits in resolution that these properties impose on multi-dimensional data. Thus, in general, full resonance assignments of proteins with a molecular mass greater than about 30 kDa prove to be technically challenging (Venters *et al,* 1996).

It is now relatively common practice for NMR studies of globular proteins in excess of 20 kDa to employ random fractional deuterium (^{2}H) labelling. This has the advantage of reducing relaxation losses in triple resonance experiments. The following section will look at the common strategy applied in this project to achieve deuterium-substituted, and [^{2}H,^{13}C,^{15}N]-labelled protein samples, and the distinct advantages these samples provide for multi-dimensional NMR spectra.

2.11.4. HSQC – the benefits of [^{2}H]-labelling.

Traditionally, the application of NMR spectroscopy has been limited to relatively small proteins or protein domains. As protein size increases, the sensitivity of heteronuclear pulse sequences is hindered because of relaxation losses in the fixed delay periods. The progressive decrease in sensitivity with increase in protein size arises primarily from a concomitant decrease in transverse ^{13}C and ^1H_N relaxation times associated with the longer rotational correlation time. The most significant contribution to ^{13}C T_2 relaxation is from the strong dipolar coupling to its directly attached proton(s) (Grzesiek *et al,* 1993). In addition, dipolar coupling to surrounding aliphatic protons has been shown to contribute significantly (~40 %) to ^1H_N T_2 relaxation (Markus *et al,* 1994). To overcome the limitations imposed by these relaxation properties two separate but ultimately mutually beneficial approaches have been developed: a new class of NMR pulse sequence denoted 'transverse relaxation-optimized spectroscopy' (TROSY); and uniform randomised substitution of non-exchangeable protons with deuterons, or per-deuteration. The latter of these two topics is discussed here.

Several strategies exist for the preparation of [^{2}H]-enriched proteins which vary in the level and specificity of substitution achieved (Sattler and Fesik, 1996 and Goto and Kay, 2000). With fractional deuterium labelling, the expression of a recombinant protein is carried out in a medium containing D_2O. During cell growth, the expression host incorporates deuterons from the medium into its amino acid reservoir and consequently into any translated protein products. The final level of deuteron enrichment achieved in a protein varies from site to site. In most deuteration strategies, the protein is subsequently purified in H_2O-based solvents, which promotes the protonation of any exchangeable deuterated groups (e.g. hydroxyl, thiol, and back bone and side chain amide, guanidino and amine groups).

In general, ^{1}H relaxation in proteins is dominated by dipolar interactions with other protons (within ca. 5 Å) and by interactions with directly bonded heteronuclei. A completely protonated 'large' protein suffers from large line-widths of the NMR signal and therefore low signal-to-noise as a result of the highly efficient distribution of magnetization through the spin system of dipolar coupled protons (known as 'spin diffusion'). Another efficient means of the dispersal of magnetization arises through the dipolar interaction of ^{13}C and ^{15}N with their directly bound proton spins, which

constitutes the main source of relaxation for ^{13}C- and ^{15}N-nuclei in isotope-labelled proteins (Cavanagh *et al*, 1996 and Sattler and Fesik, 1996).

Deuteration represents an attractive approach in dealing with both of the problems of excessively rapid relaxation of magnetization identified above. In respect to the rapid distribution of magnetization exhibited by competing relaxation pathways in a network of dipolar coupled protons, decreasing the proton density by introducing deuterons at positions that modify the efficiency of these relaxation pathways and thereby dramatically reduce the remaining ^{1}H line widths and improve the signal-to-noise ratio. Deuteration will also result in the reduction of dipolar interactions between ^{13}C and ^{15}N with their directly bound proton spin which results in increased relaxation times leading to smaller line-widths and a higher signal-to-noise. These effects arise by virtue of the significantly smaller gyromagnetic ratio (γ) of the deuterium spin when compared to the ^{1}H spin, roughly a factor of 6.5-times lower than that of ^{1}H ($\gamma_D \sim 1/6.5\ \gamma_H$), which results in relaxation rates that are scaled proportional to $(\gamma_D/\gamma_H)^2 \sim 0.024$. Therefore the dipolar contribution of a deuteron to the line-width of a ^{13}C spin attached to a molecule undergoing isotropic tumbling with a rotational correlation time of 15 ns is ~1.5 Hz, compared with ~20 Hz for the case of a ^{13}C-^{1}H pair (Yamazaki *et al*, 1994b).

The benefits to minimizing unwanted ^{1}H relaxation in protein NMR spectra are best realised for samples in which very high levels (> 80 %) of substitution of non-exchangeable protons has been obtained – so-called 'perdeuteration'; even in this circumstance the dipolar interactions between the remaining NH, OH and SH protons provides for some degree of 'spin-diffusional' broadening of the proton signals. In respect of the effects on ^{13}C relaxation in triple resonance pulse sequences, commonly used methods that provide this overall level of proton-deuteron substitution typically yield > 95 % substitution at the Cα atom, because of the metabolic pathways in *E. coli*, leading to very effective enhancements of the signal-to-noise performance for these samples.

Figure 2.2 provides a simplistic representation of the very significant reduction in proton density for a normal ('per-protonated') protein that is theoretically achievable with 100 % perdeuteration at the non-exchangeable positions.

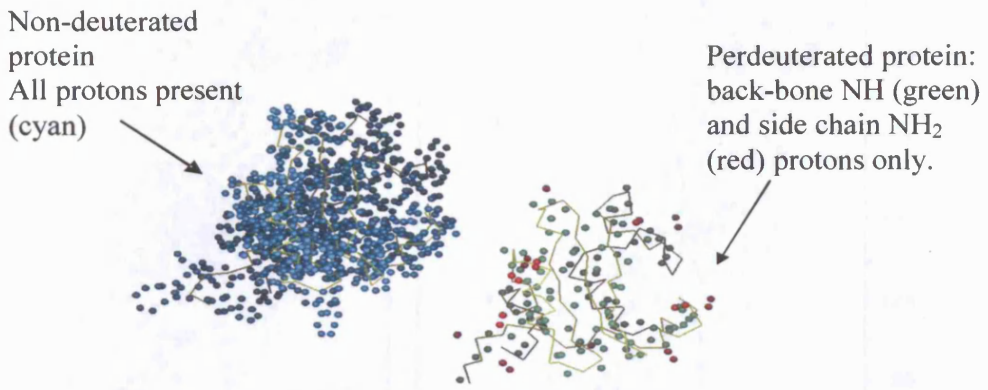


Figure 2.2: Depiction of the reduction in proton density that is achieved upon perdeuteration of small protein. The fewer protons present in solution due to ^{2}H-labelling results in 2D NMR spectral improvements that are not shown here.

The effects upon the ^{1}H line widths that result from perdeuteration of a protein are readily appreciated from a comparison of the [^{15}N]-HSQC NMR spectra of a [^{15}N]-labelled (protonated) protein and its perdeuterated equivalent. Figure 2.3 provides an example of such as comparison, in this case for the ~28 kDa amino terminal domain of enzyme I from *E. coli* (Garrett *et al*, 1997). Perdeuteration clearly results in enhanced spectral resolution as a result of the narrowing of the ^{1}H line widths.

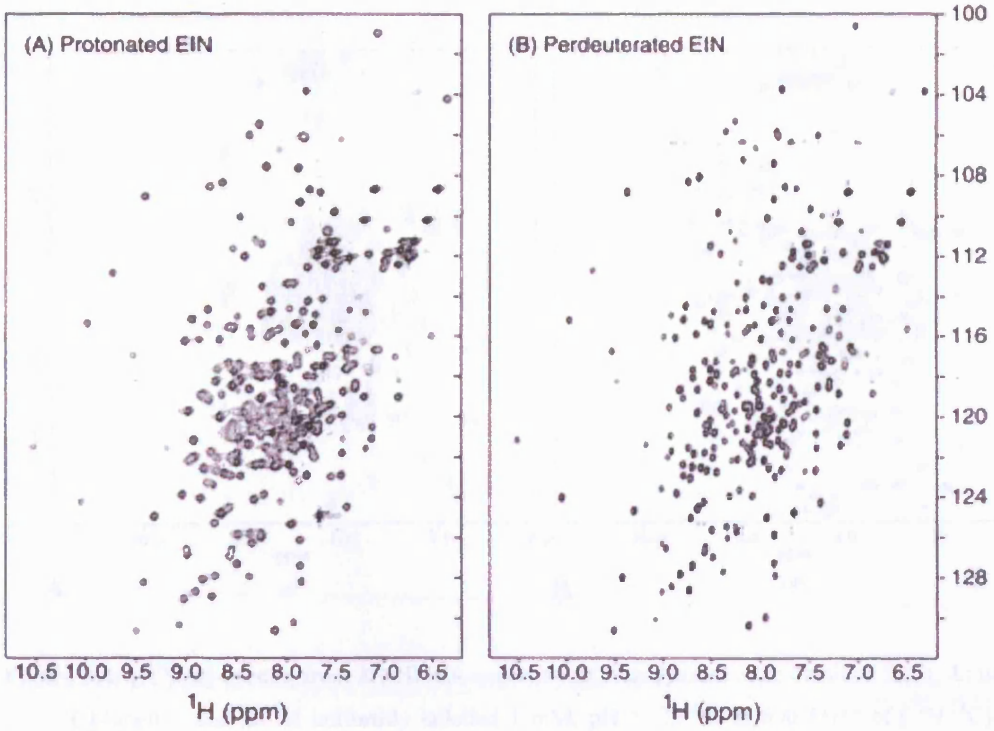


Figure 2.3: The difference in the 2D [^{15}N]-HSQC spectra between a 1 mM, pH 7, 25 °C, at 600 MHz ^{15}N-labelled (protonated) protein (**A**) and its perdeuterated equivalent (**B**) (Garrett *et al*, 1997). A significant improvement in spectral resolution of peaks and the corresponding reduction in cross peak overlap is seen as a result of [^{2}H]-labelling.

The effect of substitution of Cα-bonded protons with deuterons on the performance of triple resonance NMR experiments is illustrated in Figure 2.4 which shows 2D projection of 3D HNCA spectra recorded for [^{13}C,^{15}N]- and [^{2}H,^{13}C,^{15}N]-labelled samples of a phosphotyrosine-binding (PTB) domain (Sattler and Fesik, 1996). The dramatic improvement in achievable sensitivity and resolution is evident for the latter case.

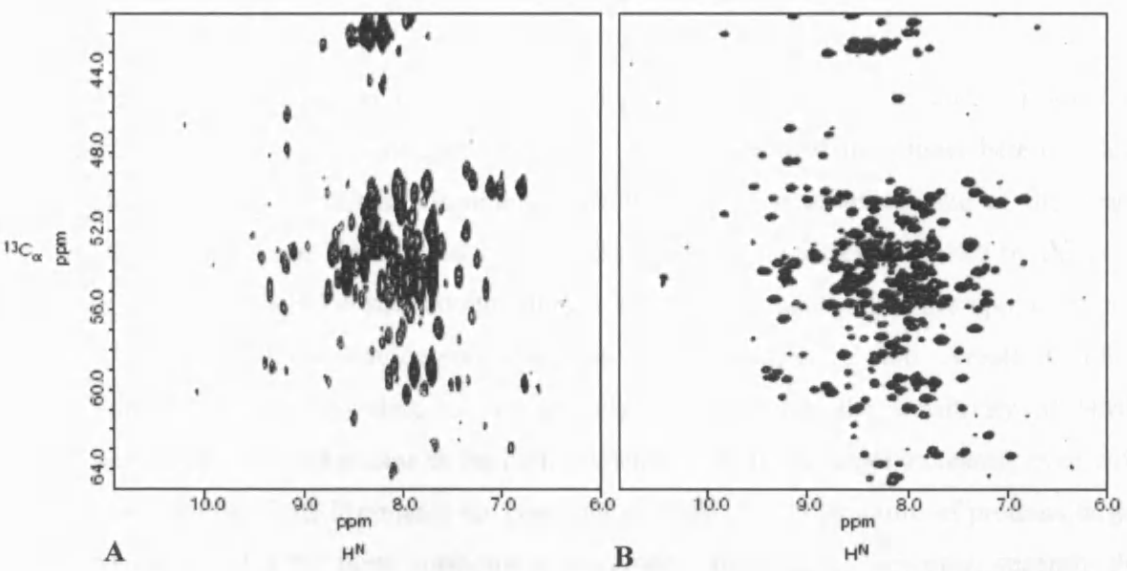


Figure 2.4: [H^N,Cα]-spectra from 3D-HNCA experiments. The spectra were obtained from, **A**: uniformly [^{15}N,^{13}C]-labelled and **B**: ^2H uniformly labelled 1 mM, pH 7, 25 °C, at 600 MHz of [^{15}N,^{13}C]-labelled PTB domain.

Experience shows that high level perdeuteration has a significant beneficial impact on the performance of many heteronuclear NMR experiments that are used to obtain the back-bone resonance assignments of medium-sized proteins (e.g. Venters *et al,* 1996). These assignments have a wide utility in a variety of contexts including chemical shift perturbation mapping (e.g. SAR-by-NMR), and the analysis of polypeptide back-bone dynamics through the measurement of ^{15}N relaxation rates. However it is worth remembering that, by definition, the substitution of the majority of non-labile protons with deuterons leads to difficulties in efforts to determine protein structure based upon the standard paradigm of collecting inter-proton NOE-based distance restraints. A number of alternative strategies to circumvent this problem have been developed (e.g. partial molecular alignment and residual dipolar coupling measurements): these techniques are beyond the scope of the work described in this thesis and are not discussed further.

2.11.5. Transverse relaxation optimized spectroscopy (TROSY).

Whilst [^{2}H]-labelling has and continues to prove beneficial to the study of small to medium-sized [^{13}C,^{15}N]-enriched proteins, the application of traditional heteronuclear NMR spectroscopy is still ultimately constrained by the effective size of the target molecule. In general the application of NMR to macromolecules is limited by the rates of transverse relaxation that operate during the pulse sequences that are applied. Since the dominant factor that determines these rates is related to the overall tumbling (rotational correlation) time, it is generally the case that the sensitivity of NMR experiments becomes poorer as the hydrodynamic size of the target increases, even with perdeuteration. Thus it remains the case that probing the 3D structure of proteins larger than ca. 40 kDa by these methods is extremely challenging. However, recently the development of new NMR methods that can be broadly collected under the heading transverse relaxation optimized spectroscopy (TROSY) have opened up the potential to obtain useful spectra for very much larger molecules (Pervushin *et al*, 1997). The principle that the TROSY methods exploit involves the interference effects of cross-correlation of distinct relaxation pathways within J-coupled NMR multiplets, such that the different multiplet components exhibit differential transverse relaxation rates (Riek *et al*, 2000). For example for an N-H group, the interference between dipole-dipole coupling and chemical shift anisotropy (CSA) leads to one of the (un-decoupled) ^{15}N (and ^{1}H) doublet components having a narrower line, and the other a broader line. Thus for 1/4 of the total signal contributing to a cross peak in an un-decoupled 2D HSQC spectrum, line broadening due to transverse relaxation is significantly reduced and a sharper, more intense signal results. Under conditions that optimise the interference effects (e.g. at the optimal magnetic field strength) pulse sequences that specifically select the slowly-relaxing multiplet component (e.g. TROSY-HSQC) can provide superior sensitivity and resolution than the more traditional (decoupled) HSQC spectrum, particularly for large molecules. TROSY performance is degraded when alternative [^{1}H]-relaxation networks exist, so the full potential of TROSY-based methods is only obtained when the protons are appropriately 'diluted out' to deal with ^{1}H-^{15}N spin-pair relaxation with remote protons due to ^{1}H dipolar coupling.

TROSY-based spectroscopy has provided for a plethora of applications to different types of protein targets that perhaps would previously have been considered beyond the scope of NMR otherwise have been refractory to NMR analysis. For example, the 3D

solution structure of bacterial malate synthase G (82 kDa) was successfully solved (Tugarinov *et al*, 2005) as was the complete back-bone amino acid assignment of a 44 kDa uniformly [^{15}N, ^{13}C]-labelled and fractionally (35 %) deuterated trimeric *B. Subtilis* chorismate mutase protein (Hu *et al*, 2003). The implementation of the TROSY principle has been shown to increase the sensitivity of key ^{1}H, ^{13}C, ^{15}N triple resonance correlation experiments (Eletsky *et al,* 2001). Perhaps more impressively, TROSY NMR methods has aided in studies of proteins in complex with the 900 kDa chaperone GroES-GroEL (Fiaux *et al*, 2002). Much current interest in NMR methods development is centred upon extending the concepts underpinning TROSY to chemical moieties other than the back-bone amide NH group, such as $^{13}CH_3$ in a $^{12}C,^{2}H$ background (Sprangers *et al*, 2007).

2.12. Discussion/Concluding remarks.

Multidimensional heteronuclear NMR spectroscopy is a powerful, rapidly developing field with an ever-expanding arsenal of analytical tools applicable in protein science. This project will make use of the practical aspects of protein based NMR spectroscopy, using optimized isotope protein labeling techniques, to gain an understanding, with the support of other areas of biochemical-based research, of the physiochemical properties of a variety of protein targets involved in aspects of bacterial persistence associated with *Mycobacterium tuberculosis,* as found in human and animal models.

Chapter 3

Peroxiredoxins and the investigation into *Mtb* TPx

3.1. Chapter summary.

A member of the peroxiredoxin protein family, *Mtb* thiol peroxidase (TPx) has recently been shown to catalyse the reduction of a range of peroxide species *in vitro*. These include: hydrogen peroxide, phosphatidylcholin-A hydroperoxide (a lipid hydroperoxide), cumene hydroperoxide (a phenyl hydroperoxide), t-butyl hydroperoxide (an organic hydroperoxide) and peroxynitrite (Zang *et al,* 1992). The ability of *Mtb* TPx to reduce these reactive oxygen species suggests that it may have a role in antioxidant defence in tuberculosis, similar to that of alkyl hydroperoxide reductase (AhpC) which is also believed to protect against organic peroxides and peroxynitrites in *Mtb* (Chae *et al,* 1994 and Jaeger *et al,* 2004). Both wild-type *Mtb* TPx (TPxWt) and the *Mtb* TPx Cys60 → Ser60 mutational variant (TPxC60S) exhibit quasi-identical homodimeric structures with an overall ellipsoidal shape. Although homodimeric in nature, *Mtb* TPxWt is functionally monomeric with two active reducing sites per dimeric unit. *Mtb* TPxWt is an atypical two-Cys peroxiredoxin with reactive cysteines Cys60 and Cys93 acting as the peroxidatic (C_P) and resolving (C_R) residues, respectively, within the active site. The crystal structures of TPxWt and the TPxC60S mutant, solved by independent groups, and apparently mimicking the oxidised and reduced states respectively, suggest substantial local rearrangement of secondary structure elements proximal to the active site, such as has been suggested previously for a homologous peroxiredoxin.

3.2. Peroxiredoxins.

Organisms living under aerobic conditions have developed an array of antioxidative systems to protect cells from oxidative damage caused by reactive oxygen and reactive nitrogen species. An emerging family of non-heme peroxidases, called peroxiredoxins, has been described in all kingdoms during the past few years (Chae *et al*, 1994 and Dietz *et al*, 2002). Anti-oxidant enzymes such as tryparedoxin peroxidase (TryP), alkyl hydroperoxide reductase (AhpC), and bacterioferritin-comigratory protein (BCP) belong to the peroxiredoxin protein family, which also includes thiol peroxidases (TPxs). AhpC has been shown to be pivotal for the survival of many pathogens in the oxidative environment of the host's phagocytes (Storz *et al*, 1987). These ubiquitous proteins are able to reduce organic and inorganic hydroperoxides with electrons provided mainly by NADPH or NADH and different proteins such as thioredoxin (Trx), glutaredoxin (Grx), cyclophilin-A, and subunit-F of alkyl hydroperoxide reductase, to produce either water (from H_2O_2) or corresponding alcohols (ROH) from ROOH species (Chae *et al*, 1994; Niimura *et al*, 1995; Lee *et al*, 2001 and Rouhier *et al*, 2002). In addition to protecting against reactive oxygen species, some peroxiredoxins can reduce and thereby detoxify reactive nitrogen (Bryk *et al*, 2000) and sulphur (Yim *et al*, 1994) species. Peroxiredoxins are also known to participate in signal transduction and cell proliferation (Wood *et al*, 2003a).

The defining characteristic of peroxiredoxins (Prxs) is the presence of conserved cysteine residues that are implicated in the catalytic cycle involved in reducing organic and inorganic hydroperoxides. As a family, these proteins are classified according to substrate specificity, structural features and oligomerisation state (Zang *et al,* 1992, Jaeger *et al*, 2004, Rho *et al*, 2006 and Trujillo *et al*, 2006). The peroxiredoxins were first classified into two sub-classes, the 1-Cys and the 2-Cys peroxiredoxins, according to the number of absolutely conserved cysteine residues and dependent on whether they form, as part of the reaction mechanism, intermolecular (Rhee *et al*, 2001) or intramolecular (Wood *et al*, 2003a) disulfide bridges, respectively. The 2-Cys peroxiredoxins were further sub-divided into typical and atypical 2-Cys peroxiredoxins. The basis of the distinction between 1-Cys, typical 2-Cys and atypical 2-Cys Prxs relates to the second step of the

peroxidase reaction mechanism - namely the resolution of the cysteine sulfenic acid intermediate (see below). The increasing number of peroxiredoxins found in all organisms has prompted more complex categorisation schemes. Thus, in mammals the classification of peroxiredoxins is sub-divided into six sub-groups (I to VI) on the basis of cellular localisation, protein size, and electron donor capability. Plant peroxiredoxins are classified into four subgroups (Dietz *et al*, 2002) based upon function and location: 1-Cys, 2-Cys, type II peroxiredoxin and peroxiredoxin Q. Here the focus is on the bacterial 2-Cys Prx group.

The basic reaction mechanism for typical and atypical 2-Cys Prxs is illustrated schematically in Figure 3.1 In each case, the reaction is initiated by the nucleophilic attack of one Cys side-chain thiol (the peroxidatic Cys, C_P) on the ROOH peroxide substrate leading to a cysteine sulfenic acid intermediate (C_P-SOH) and the corresponding alcohol product, ROH. The sulfenic acid is reduced by the action of a second, resolving Cys (C_R) thiol group leading to the generation of a C_P-S-S-C_R disulfide and loss of a water molecule. The reduced oxidized Prx is then recycled to the dithiol state by the action of an upstream reductase (PrxR) – usually a flavoprotein.

Typical 2-Cys Prxs are the largest class of Prxs and are identified by the conservation of their two redox-active cysteines, C_P (generally near residue 50) and the C_R (near residue 170) (Hofmann *et al*, 2002). Typical 2-Cys Prxs are obligate homodimers containing two identical active sites (Wood *et al*, 2002). In the second step of the peroxidase reaction (Figure 3.1A), the intermediate peroxidatic cysteine sulfenic acid (C_P–SOH) from one subunit is attacked by the resolving cysteine (C_R–SH) located in the C-terminal part of the other subunit. This condensation reaction results in the formation of a stable inter-subunit disulfide bond, which is then reduced by one of several cell-specific disulfide oxidoreductases (e.g. thioredoxin, AhpF, tryparedoxin or AhpD) (Nogoceke *et al*, 1997; Poole *et al*, 2000 and Wood *et al*, 2002), completing the catalytic cycle.

The second class of Prxs are the atypical 2-Cys Prxs, which have the same basic mechanism as typical 2-Cys Prxs but are functionally monomeric (Declercq *et al*, 2001). In these Prxs, both the C_P and the corresponding C_R residues are contained within the same polypeptide, with the condensation reaction resulting in the formation of an intramolecular disulfide bond (Figure 3.1B).

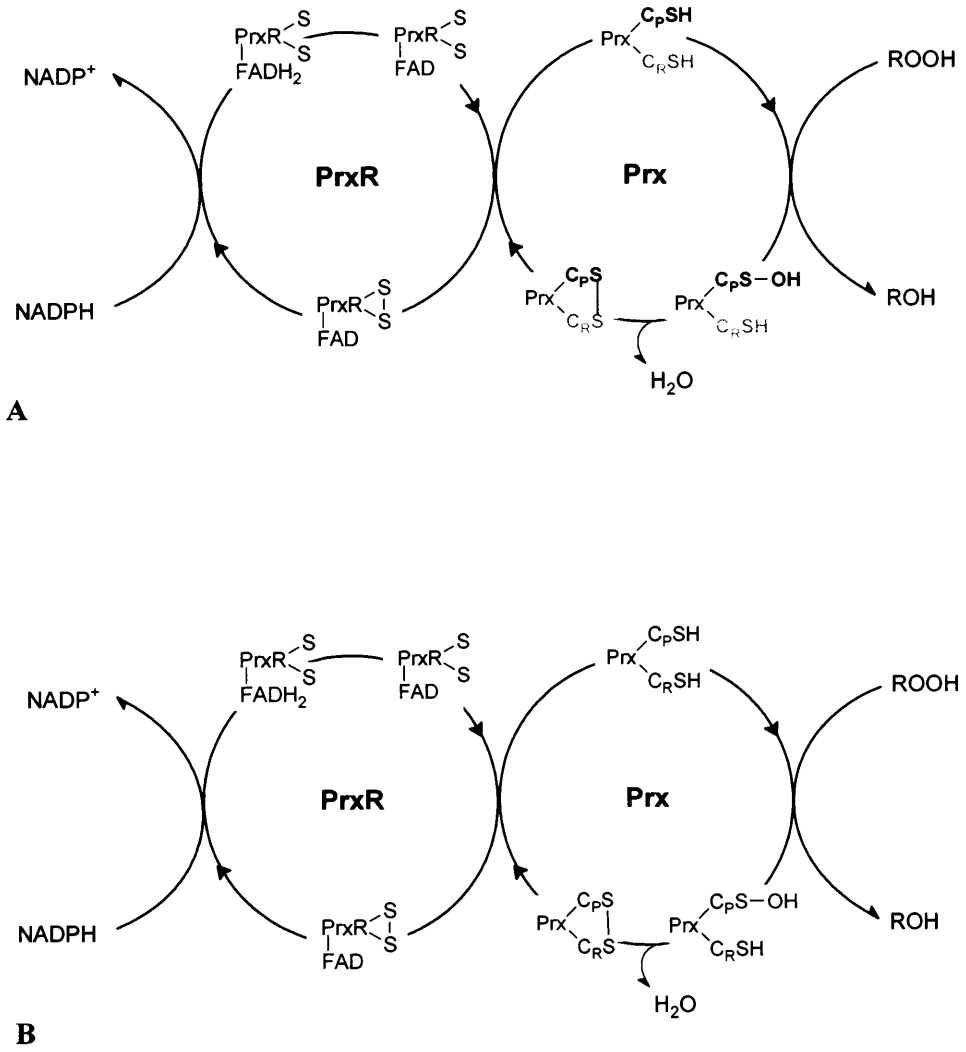


Figure 3.1: A: The typical 2-Cys Prx and **B:** atypical 2-Cys Prx mechanisms, with peroxidatic cysteines and resolving cysteines in the reduced (C_PSH and C_RSH, respectively), sulfenic-acid (C_PSOH) or disulfide (C_PS and C_RS connected) state. In the case of typical 2-Cys Prxs, which are functionally dimeric, the peroxidatic cysteine (black C_P) and resolving cysteine (pink C_R) originate from different subunits and condense to form an inter-subunit disulfide bond (blue). Atypical 2-Cys Prxs are functionally monomeric. Reduction of typical and atypical 2-Cys Prxs usually involves a flavoprotein disulfide reductase, which is oxidized from a dithiol (PrxR-FAD-SH_2) to a disulfide (PrxR-FAD-SS) state during Prx reduction.

Although within typical and atypical 2-Cys Prxs the C_R residue is conserved, its position within the polypeptide chain is different. Atypical 2-Cys Prxs appear to couple to thioredoxin and thioredoxin reductase as upstream electron donors, in the recycling of the intermolecular disulfide (Declercq *et al*, 2001).

3.3. *Mycobacterium tuberculosis* (*Mtb*) thiol peroxidase (TPx).

In the *Mtb* bacterial genome, several open reading frames have been identified encoding peroxiredoxins: two bacterioferritin-comigratory protein (BCP)-type peroxiredoxins [Rv1608c (BCPB) and Rv2521 (BCP)], a 1-Cys peroxiredoxin [Rv2238c (AhpE)] and two 2-Cys peroxiredoxins [Rv2428 (AhpC) and Rv1932 (TPx)]. The atypical 2-Cys peroxidase *Mtb* TPx was first described as culture filtrate protein 20 (CFP-20) in a search for vaccine candidates (Weldingh *et al*, 1998). It was characterized as one of the antigens strongly recognized in *Mtb* infected animals and induces a strong proliferative response in both humans and mice (identified by increasing numbers of IFN-γ releasing T-cells, as part of a delayed type hypersensitivity response). Proteomic studies of *Mtb* have indicated that TPx is found in the cell wall as well as CFP fractions (Rosenkrands *et al*, 2000a and 2000b).

Homologues of *Mtb* TPx are distributed throughout most or all eubacterial species, both Gram-negative and Gram-positive, and are found in pathogenic strains such as *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Helicobacter pylori* (Wan *et al,* 1997). Of these, biochemical and genetic analyses have been limited primarily to *E. coli* TPx (*Ec*TPx) and I will discuss aspects of *Mtb* TPx structure and function largely with respect to the reported characteristics of this protein. A degree of functional equivalence between *Mtb* TPx and *Ec*TPx is illustrated by the increased susceptibility of *E. coli* to oxidative stress when the TPx gene is deleted (Master *et al*, 2002).

3.4. Antioxidant defences of *Mtb*.

The action of the first-line tuberculostatic isoniazid (INH) depends upon the presence of the *Mtb* heme-containing catalase peroxidase KatG. The constitutive function of KatG is for the detoxification of H_2O_2, and laboratory strains lacking KatG are avirulent. However, clinical isolates lacking the catalase are virulent and proved to be resistant to INH treatment. This is because KatG is evidently required to chemically activate INH (Zhang *et al*, 1992). Loss of catalase should in principle make the pathogenic bacteria more susceptible to the host's oxidative immunological response, unless it is compensated for by other components of the bacterial antioxidant defence *in vivo*. It has been suggested that there are two candidate peroxidases that participate in this defence. One of these peroxidases is the typical 2-Cys peroxiredoxin alkyl hydroperoxide reductase-C (AhpC) which is often found at elevated levels in catalase-deficient clinical isolates (Sherman *et al*, 1996). The other peroxidase is TPx which belongs to the atypical 2-Cys peroxiredoxin class. Both TPx and AhpC exist as part of what could be termed as the 'second line' of antioxidant defence for *Mtb*. TPx has recently been demonstrated as being the most efficient peroxidase of the two peroxiredoxins in *in vitro* assays, as determined by rate constants for the reaction with hydroperoxides and peroxynitrite *in vitro* (Trujillo *et al*, 2006). Notably, the scope of TPx reducing ability includes a broad spectrum of hydroperoxides, including H_2O_2, as well as being involved in oxidative and nitrosative stress defence *in vitro* (Jaeger *et al*, 2003). TPx is specifically regenerated, in its reduced state, by two of the mycobacterial thioredoxins, *Mtb* TrxB and *Mtb* TrxC. However, *Mtb* maintains NADPH-driven detoxification predominantly through the TrxC-mediated pathway (Zang *et al,* 1992, Jaeger *et al*, 2004, Rho *et al*, 2006 and Trujillo *et al*, 2006). It has been speculated that TPx also meets special antioxidant requirements in $KatG^+$ strains by reducing hydroperoxides of complex lipids and peroxynitrites. Thus, both TPx and AhpC appear to complement KatG catalase in antioxidant defence in *Mtb*. These characteristics have led to the speculation that TPx, alongside AhpC, could be considered as a potential drug target. The targeted inactivation of AhpC and TPx could be especially relevant in cases of antibiotic-resistant tuberculosis, where $KatG^-$ strains may increasingly

emerge thus making INH treatment ineffective (Sherman *et al*, 1996). The importance of the apparently more efficient peroxidase, TPx, as a potential drug target is underscored by its low level of sequence similarity with mammalian PrxII, showing only 21 % identity, suggesting the potential for selective targeting of the bacterial TPx.

3.5. The mechanism of peroxidase activity of atypical 2-Cys peroxidase from *E. coli.*

The *Mtb* TPx protein exhibits 52 % sequence identity with *E. coli* TPx (*Ec*TPx) (PDB code: 1QXH), as determined by both the 3D structure alignment tool known as DALI (Holm *et al*, 1995) and the ClustalW sequence alignment tool (Higgins *et al*, 1994) showing conserved catalytic regions to be present (Figure 3.2). The catalytic regions are represented by the peroxidatic cysteine residues: C60 (*Mtb*) and C61 (*E. coli*), and the resolving cysteine residues: C93 (*Mtb*) and C95 (*E. coli*).

To date, one of the most documented and well understood peroxidase mechanisms is that of the atypical 2-Cys *Ec*TPx peroxidase. The following section will describe the peroxidatic mechanism of *Ec*TPx and the resulting chemical changes occurring to the catalytic cysteines.

Baker and Poole (2003) studied the mechanism of action of *Ec*TPx by applying rapid reaction kinetic experiments, as a coupled NADPH-linked peroxidase assay, used to measure the steady-state turnover of cyclic intermediates and products. Therefore examining the thiol group chemistry, and assessing the peroxide detoxifying properties of wild-type *Ec*TPx and four Cys → Ser mutants (C61S, C82S, C95S, C82S/C95S).

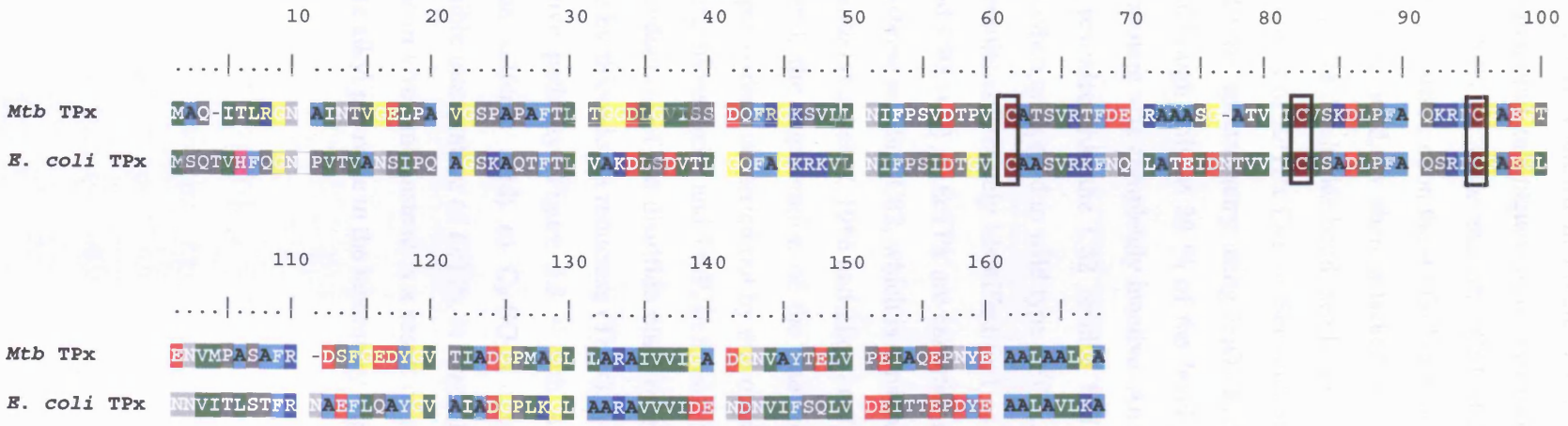


Figure 3.2: ClustalW sequence alignment of *Mtb* TPx with its closest structural homologue; *Ec*TPx. Conserved catalytic cysteine residues; Cys60 (*Mtb*)/Cys61 (*E. coli*), Cys83 (*Mtb*)/Cys82 (*E. coli*), and Cys93 (*Mtb*)/Cys95 (*E. coli*) are shown coloured brown and boxed. Amino acid numbering is for *Mtb* TPx as used as the primary source alignment sequence. Highly hydrophobic residues are coloured green, weakly hydrophobic are light blue, polar residues are grey, positively charged residues are dark blue, negatively charged residues are red, and glycines are yellow. Displayed in Rich Text format using BioEdit v4.0.6 (Hall, 1997).

The proposed mechanism of *Ec*TPx reduction of alkyl peroxide (ROOH) to an alcohol (ROH) is shown in Figure 3.3. A series of titration assays and tryptic peptide mapping techniques were executed using 5,5-dithiobis (2-nitrobenzoic acid) (DTNB) to show that the C61 (C_P) thiol group, which carries out a nucleophilic attack upon the -O-O- bond of the alkyl hydroperoxide resulting in a C61 sulfenic acid, is then attacked by the C95 residue (C_R) to form an intramolecular disulphide bond (see Figure 3.3 – path B) (Baker and Poole, 2003). The activities of *Ec*TPx Cys → Ser mutants C61S and C95S were measured by stopped-flow spectrometry using H_2O_2 as the substrate. The C95S mutant was active although only at 20 % of the level of wild-type *Ec*TPx while the C61S *Ec*TPx mutant was completely inactive. An equivalent reaction assay using C82S *Ec*TPx revealed that the C82 residue had no overall influence on peroxidase activity when compared to wild-type *Ec*TPx.

These results collectively identified C61 as the catalytic cysteine in *Ec*TPx. C61 (C_P) and C95 (C_R) in *Ec*TPx are essential for maximal catalytic activity but the third cysteine residue, C82, which is conserved in most atypical peroxiredoxins, is not essential (Cha *et al,* 1996 and Zhou *et al*, 1997).

In general, the regeneration of the intermolecular disulfide bonded Prx to the dithiol peroxidase is carried out by one of either cell-specific oxidoreductase such as: AhpF, thioredoxin, and TryP. In *E. coli*, the *Ec*TPx catalytic site is regenerated by the reduction of the disulfide bond by thioredoxin-1 (Trx1), which in turn is reduced by thioredoxin reductase (TrxR), which uses NADPH as a substrate. An alternative pathway (Figure 3.3 – path A) shows the conversion of C_P-SOH (cysteine sulfenic acid) to C_P-SO_2H (cysteine sulfinic acid), which causes irreversible inactivation of *Ec*TPx *in vitro*. However, this pathway is believed not to occur *in vivo*, and instead is a result of non-physiologically high levels of the substrate alkyl peroxide in the laboratory experiments.

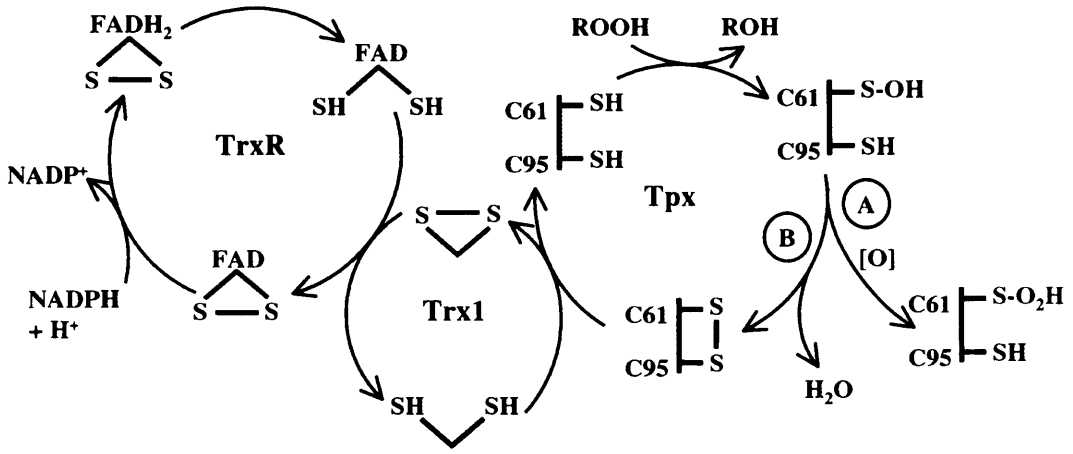


Figure 3.3: Proposed reaction scheme for reduction of alkyl peroxide by *Ec*TPx (Baker and Poole, 2003)

Recent identification of a C60 residue oxidised to a sulfenic acid form in *Mtb* (Rho *et al*, 2006), comparable to the Cys61-SH → Cys61-SOH transition identified for the C61 residue in *Ec*TPx during its catalytic peroxidatic mechanism, was made using 7-chloro-4-nitrobenzo-2-oxa-1, 3-diazole (NBD-chloride) labelling. The sulfenic acid form of C60 was detected by observing changes to the visible absorbance properties of the respective adducts of Cys61-SH and Cys61-SOH upon reacting with the sulfhydryl-reactive reagent, NBD. The following section will therefore highlight the significance of this *Mtb* TPx C60 residue in providing detoxifying ability for *Mtb* TPx and describe the NADPH-linked peroxidase assay systems used to assess reactivity.

3.6. NADPH-linked peroxidase assay for *Mtb*.

Rho *et al.* and Trujillo *et al.* each established a coupled activity assay for *Mtb* TPx to assess TPxWt (Trujillo *et al*, 2006 and Rho *et al*, 2006) and the mutated variant TPxC60S (Rho *et al*, 2006). In the assay system, TPx is subjected to H_2O_2 as a substrate, in the presence of an *E. coli* thioredoxin-thioredoxin reductase system which transfers electrons from NADPH to TPx. The assay is performed by monitoring the decrease in absorbance at 340 nm as NADPH becomes oxidised. As expected, of the two *Mtb* TPx variants the only one which demonstrates

NADPH oxidation is TPxWt; no activity was observed with the C60S mutant. An essentially identical result was obtained by Jaeger *et al.* where they combined a mycobacterial thioredoxin reductase and *Mtb* TrxC thioredoxin, with TPxWt in the presence of H_2O_2 and observed the thioredoxin reductase-dependent oxidation of NADPH (Jaeger *et al*, 2004). *Mtb* TPxWt also accepted *Mtb* TrxB, another mycobacterial thioredoxin as intermediate electron donor, although with lower reactivity when compared to TrxC. Oxidation of NADPH required all components of the thioredoxin-thioredoxin reductase system (thioredoxin reductase, thioredoxin and a peroxidase). This research group also examined the reaction rate differences that *Mtb* TPxWt and AhpC exhibit in this peroxidase assay (Jaeger *et al*, 2004). Their observations revealed that TPxWt exhibits an order of magnitude greater catalytic efficiency than *Mtb* AhpC. The TPx dependent reduction of peroxynitrite was also tested via stopped-flow analysis for wild-type *Mtb* TPxWt and two Cys → Ser mutants, C80S and C93S. In the presence of TPxWt, peroxynitrite decayed rapidly, with a markedly increased reaction with the *Mtb* TPx mutant forms (Trujillo *et al*, 2006 and Jaeger *et al*, 2004). As cysteine residues are involved in the catalytic activity of peroxiredoxins, Rho *et al.* examined the effect of sulfhydryl reactive reagents (NEM, IAA and NBD-chloride) on *Mtb* TPx and individual Cys → Ser mutants in the presence of dithiothreitol (DTT). The results revealed that the TPxWt activity to be completely inhibited when in the presence of H_2O_2, in much the same way as that achieved with *Ec*TPx under the same conditions (Rho *et al*, 2006). The observation of complete inhibition of *Mtb* TPxWt peroxynitrite reduction using N-ethyl-maleimide (NEM), a known reagent in blocking peroxidatic Prx (Nogocheke *et al*, 1997), was also made by Jaeger *et al,* (2003). Both Rho *et al.* and Jaeger *et al.* showed that *Mtb* TPx mutants C60S, C93S were inactive and C80S (Rho *et al*, 2006) was found to be active along with the wild-type TPx (Jaeger *et al*, 2003). When C93 was replaced with a Ser residue through site-directed mutagenesis, a decrease in NADPH-linked peroxidase activity was measured using the stopped-flow technique, as observed by the decline in tryptophan fluorescence of *Mtb* TRxB upon oxidation with TPxWt (Trujillo *et al*, 2006). These results collectively highlight the importance of two specific cysteine residues, C60 and C93, in *Mtb* TPx peroxidatic activity, but more importantly indicate the catalytically essential C60 residue for optimal peroxide detoxification.

The importance of the C60 residue to TPx activity was further tested by establishing its importance in protecting glutamine synthatase activity in the presence of reactive oxygen species (Rho *et al*, 2006, and Jaeger *et al*, 2004) and in the protection of supercoiled DNA from oxidative damage (Rho *et al*, 2006) using a thiol-dependent metal-catalysed oxidation (MCO) system (Kim *et al*, 1985). In the glutamine synthase measurements, *Mtb* TPx exhibiting two Cys → Ser point mutations at residues C80 and C93 was able to protect glutamine synthase from reactive oxygen species, whilst in the presence of DTT. Conversely, the C60S mutant was unable to protect glutamine synthase from the same reactive oxygen species, thereby indicating that C60 is the essential peroxidatic cysteine. For the supercoiled DNA protection assay the negative control in which the *Mtb* TPx was omitted from the reaction mixture resulted in nearly all the DNA being converted to a nicked form, whereas the positive control showed about 60 % protection from nicking in the reaction mixture. The mutant C60S failed to protect the supercoiled DNA from degradation. In contrast, the wild-type enzyme and variants C80S and C93S showed almost complete protection activity, indicating that C80 and C93 are not strongly involved in supercoiled DNA protection activity.
From the various activity measurements used to examine the peroxidase activity of *Mtb* TPx, the results show that Cys60 is the peroxidatic cysteine and Cys93 is the resolving cysteine in *Mtb* TPx.

In summary, the similarities observed between *Ec*TPx and *Mtb* TPx in the homologous cysteine positions at the C-terminal end (C95 in *Ec*TPx and C93 in *Mtb* TPx), the presence of a peroxidatic cysteine (C_P), the formation of a disulfide bridge between C_P and the conserved cysteine within the same subunit and the replacement of the distal cysteine (C95 or C93, respectively) by serine showing changes in activity (Rho *et al*, 2006 and Trujillo *et al*, 2006) reveals the cysteine conserved close to the N-terminal end at position 60 as the peroxidatic (C_P) cysteine in *Mtb*. The distal cysteine (C93) is suggested to be the functional equivalent of the resolving cysteine C_R found in typical 2-Cys-Prxs (Choi *et al*, 2003 and Baker and Poole, 2003), primarily due to its supportive role alongside the essential C60 residue in peroxidatic activity (Trujillo *et al*, 2006).

The formation of an internal disulfide bridge between both the peroxidatic and resolving cysteine residues in *Mtb* TPx appears to be an essential and indispensable catalytic requirement in the peroxidase-thioredoxin-thioredoxin reductase catalytic system.

3.7. Peroxiredoxin structure.

Seventeen 3D x-ray crystal structures of peroxiredoxins originating from various organisms have been reported (see Table 3.1). Often in order to obtain tractable crystals mutation of the active site C_p residue has been required. To date, only the *S. typhimurium* AhpC and human peroxiredoxin-V have been crystallized in both reduced and oxidized forms. Most peroxiredoxins form dimers but some typical two-Cys peroxiredoxins also form decamers consisting of a pentameric arrangement of dimers.

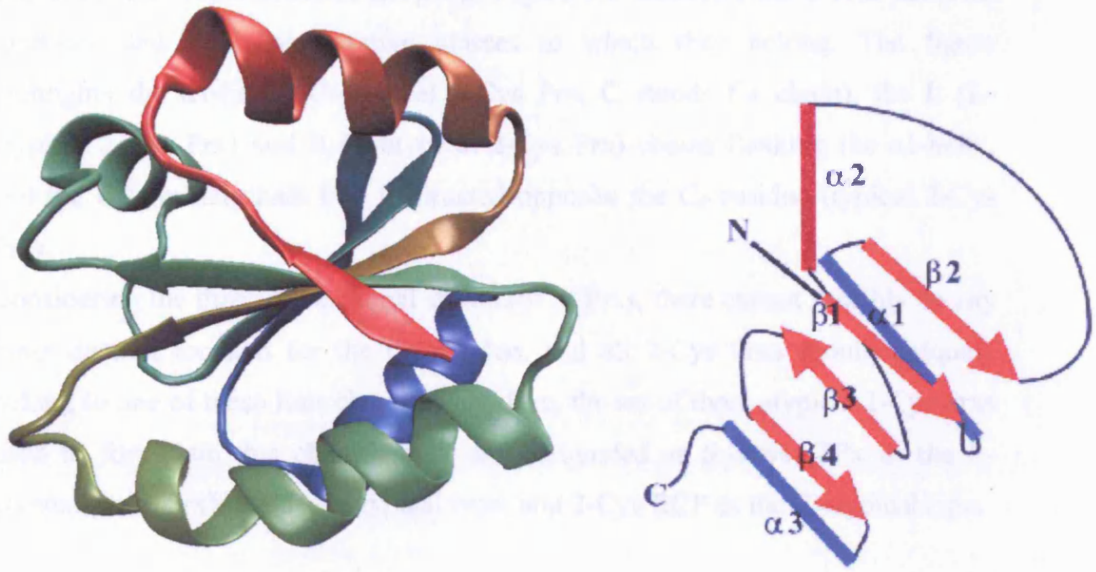
	Name:	Species:	X-ray structure resolution:	Oligomeric state:	Redox state:	Note (wild-type unless otherwise stated):	Reference:	PDB code:
Typical two-Cys Prxs	PrxI	Rat	2.6 Å	Dimeric	Oxidised	-	Hirotsu *et al*, 1999.	1QQ2
	PrxII	Human	1.7 Å	Decameric	Oxidised	-	Schroder *et al*, 2000.	1QMV
	TRyP	*Crithidia fasciculate*	3.2 Å	Decameric	Reduced	-	Alphey *et al*, 2000.	1E2Y
	AhpC	*Salmonella typhimurium*	2.5 Å	Decameric	Reduced	C46S mutation	Wood *et al*, 2003b.	1N8J
	AhpC	*Salmonella typhimurium*	2.5 Å	Decameric	Oxidised	-	Wood *et al*, 2002.	1KYG
	AhpC	*Amphibacillus xylanus*	2.9 Å	Decameric	Oxidised	-	Kitano *et al*, 2005.	1WE0
Atypical two-Cys Prxs	PrxV	Human	2.0 Å	Dimeric	Oxidised	-	Evrard *et al*, 2004.	1OC3
	PrxV	Human	1.5 Å	Monomeric	Oxidised	-	Declercq *et al*, 2001.	1HD2
	*Ec*TPx	*Escherichia coli*	2.2 Å	Dimeric	Oxidised	-	Choi *et al*, 2003.	1QXH
	*Sp*TPx	*Streptococcus pneumoniae*	2.3 Å	Dimeric	Reduced	-	Coordinates deposited/no publication.	1PSQ
	*Hi*TPx	*Haemophilus influenzae*	1.9 Å	Dimeric	Oxidised	-	Coordinates deposited/no publication.	1Q98
	TPx	*Mycobacterium tuberculosis*	1.75 Å	Dimeric	Oxidised	-	Rho *et al*, 2006.	1XVQ
	TPxC60S	*Mycobacterium tuberculosis*	2.1 Å	Dimeric	Reduced	C60S mutation	Stehr *et al*, 2006.	1Y25
One-Cys Prxs	AhpE	*Mycobacterium*	1.87 Å	Dimeric	Oxidised	-	Li *et al*, 2005.	1XXU

		tuberculosis						
	AhpE	*Mycobacterium tuberculosis*	1.9 Å	Dimeric	Reduced	-	Li *et al*, 2005.	1XVW
	PfAOP	*Plasmodium falciparum*	1.8 Å	Dimeric	Oxidised	-	Sarma *et al*, 2005.	1XIY
	PrxVI	Human	2.0 Å	Dimeric	Reduced	C91S mutation	Choi *et al*, 1998.	1PRX
	AhpC	*Mycobacterium tuberculosis*	2.4 Å	Dodecameric	Oxidised	-	Guimaraes *et al*, 2005.	2BMX

Table 3.1: Structural survey of deposited 3D x-ray crystal structures of peroxiredoxins (Prxs).

3.8. The tertiary structure - the thioredoxin fold.

Although sequence identity is low (typically 20-30 %) among the peroxiredoxin family members, all these structures share a thioredoxin-like fold, as indicated by PFAM analysis (accession number: PF08534). The thioredoxin fold is fairly structurally conserved in nature (Creighton, 2000). It is a protein fold common to enzymes that catalyze disulfide bond formation and isomerization through a conserved Cys residue situated in an active site region. The fold is named for the canonical example thioredoxin and is found in both prokaryotic and eukaryotic proteins. The thioredoxin fold is composed of an alpha/beta configuration illustrated in Figure 3.4. The fold's spatial topology consists of a four-stranded β-sheet directly sandwiched between two α-helices, α1 and α3 (colored blue in Figure 3.4B) with an additional offset α-helix, α2, orientated towards the N-terminal end of the polypeptide (colored dark red in Figure 3.4B).



A B

Figure 3.4: The thioredoxin fold. **A** Human thioredoxin, reduced form, (PDB: 1TRX), 1.7 Å resolution (Weischel *et al*, 1996). Figure obtained using PyMOL (DeLano, 2006). **B** A schematic diagram representing the thioredoxin fold comprising a 4β stranded sheet with 3α helices. The dark red α helix is one of three in the fold that does not sandwich the β-strands, shown in blue.

The thioredoxin structure shown in Figure 3.4 represents the 'simplest' example of 3D structures in this class. Other examples from the family, such as the TPxs, have additional, often small secondary structure elements that elaborate the fold (Wood *et al*, 2003a).

A paper by Choi *et al.* describes a generalised scheme showing that 2-Cys Prxs can be classified into four molecular clades or branches (Choi *et al*, 2003). They begin their description by selecting three atypical 2-Cys Prxs, *Mtb* TPx, *E. coli* BCP and human Prx that differ from each other in the location of their resolving C_R residues and a typical 2-Cys Prx, represented by AhpC. Based on structural observations using the four 2-Cys Prxs, the research group identify four different locations of the C_R residue relative to the location of the conserved C_P residue (See Figure 3.5). The proposition is made that Prxs should uniquely belong to one of the four classes based on the location of the C_R residue and also upon the disulfide bridge formed within the polypeptide as intersubunit, L-intrasubunit, R-intrasubunit and intrachain bridges.

The schematic representation shown in Figure 3.5 identifies these four different locations and the representative classes to which they belong. The figure highlights the α1-helix (C-atypical 2-Cys Prx, C stands for chain), the L (L-atypical 2-Cys Prx) and R (R-atypical 2-Cys Prx) chains flanking the α1-helix, and the C-terminal chain that is situated opposite the C_P residue (typical 2-Cys Prx).

Considering the three-dimensional structures of Prxs, there cannot sensibly be any other distinct location for the C_R residue, and all 2-Cys Prxs should uniquely belong to one of these four classes. Therefore, the set of three atypical 2-Cys Prxs used to formulate this classification are designated as follows: TPx as the L-atypical type; PrxV as the R-atypical type; and 2-Cys BCP as the C-atypical type.

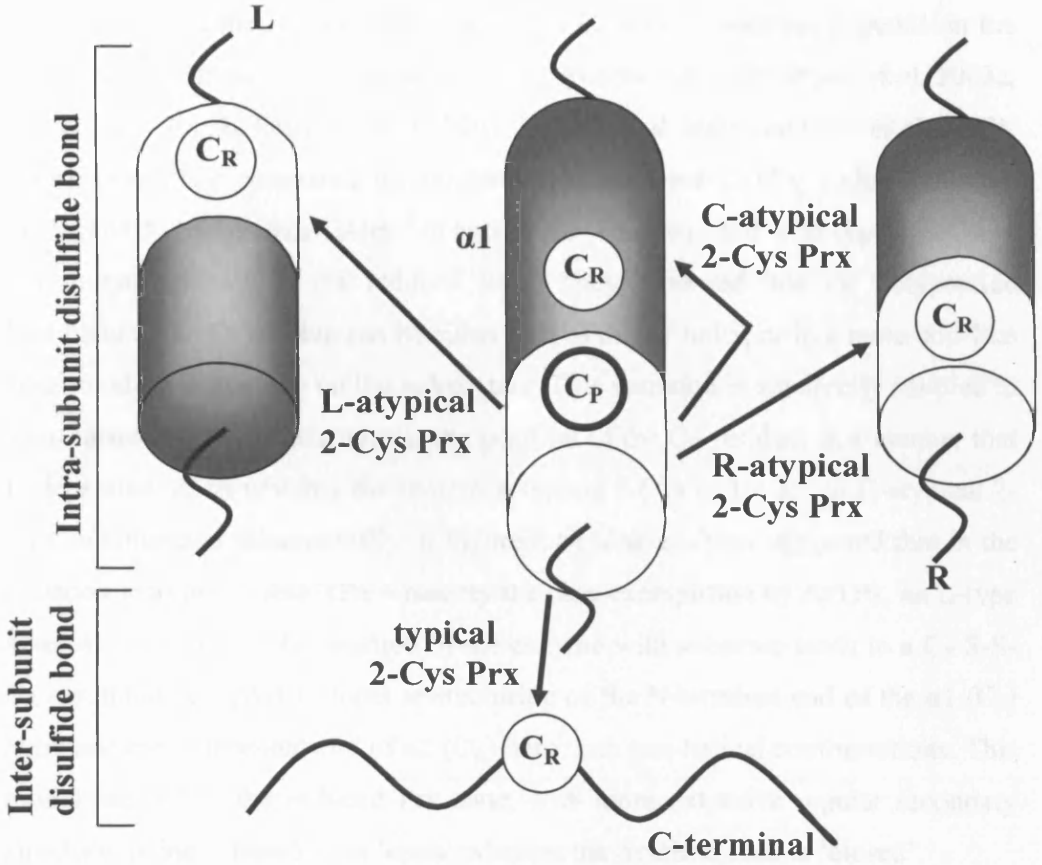


Figure 3.5. A schematic diagram showing that the 2-Cys Prxs can be classified into four molecular clades. Arrows indicate the directions of the disulfide formation. Shaded cylinders denote the α-helices and white cylinders denote the region that contains the C_R residue that undergoes a loop-to-helix or a helix-to-loop transition depending on redox states. L and R labels for the helices flanking the Cp-containing segment on α-helix α1. These labels follow the scheme described by Choi *et al*, 2003. For *Mtb* TPx helix L corresponds to α4. Figure adapted from Choi *et al*, 2003.

3.9. Oxidised *Ec*TPx structure and implications for local refolding as part of the reactive mechanism.

The structural comparison by Choi *et al.* led to the suggestion that the local conformation of the regions surrounding Prx C_P and C_R residues depends on the redox states of these two cysteine residues (Wood *et al,* 2002, Wood *et al,* 2003a, Wood *et al,* 2003b, Declercq *et al*, 2001, Evrard *et al*, 2004 and Choi *et al*, 2003). Specifically they compared the structures of oxidised *Ec*TPx, reduced human PrxV, and *S. typhimurium* AhpC in two states – the oxidised wild-type and $C_P \rightarrow$ Ser mutant mimicking the reduced state. They observed that the polypeptide backbone at the C_P residue can be either part of the α1 helix, or in a more coil-like conformation depending on the redox state. This variation is apparently coupled to local unfolding of the structure at the position of the C_R residue, in a manner that is dependent upon whether the protein is typical 2-Cys or L-, R-, or C-atypical 2-Cys (summarised schematically in Figure 3.5). The analysis suggested that in the situation relevant to *Mtb* TPx – namely the case exemplified by *Ec*TPx, an L-type atypical 2-Cys Prx – the reaction of the enzyme with substrate leads to a C_P-S-S-C_R disulphide coupled to local restructuring of the N-terminal end of the α1 (C_P) helix and the C-terminal end of α2 (C_R) helix into non-helical conformations. This model has led to the reduced Prx state, with more extensive regular secondary structure, being referred to as 'open' whereas the oxidised state is 'closed'.

3.10. The quaternary structure – *Mtb* TPx dimer.

Two research groups have independently solved X-ray structures of *Mtb* TPx (Rho *et al*, 2006 and Stehr *et al*, 2006) which form quasi-identical homodimers with an overall ellipsoidal shape of similar dimensions. A structure of wild-type *Mtb* TPx (TPxWt), from the *Mtb* Structure Genomics Consortium in Los Alamos, USA, resembles more the structures of *Ec*TPx and *S. pneumoniae* TPx (*Sp*TPx) in their oxidised forms, with characteristics consistent with the presence of a C_P-S-S-C_R disulphide bond, and thereby presumably mimics the closed, oxidised state of the enzyme. The *Mtb* TPxWt crystal structure resembles 2-fold related monomers, one molecule per asymmetric unit, whereas in all other Prx structures the dimer is

in the asymmetric unit. An intriguing aspect of this structure is that a major segment of the polypeptide chain (residues 94-99) is not clearly represented in the electron density. In addition there is no interpretable electron density for the side chain of the C_P residue Cys60 (see below). Polar/charged interactions at the dimer interface of TPxWt involve residues in loop regions, as opposed to interacting β-strands from opposing interface monomers as is the case in typical two-Cys peroxiredoxins and one-Cys peroxiredoxins (Hirotsu *et al*, 1999 and Schroder *et al*, 2000).

The group of Hecht and co-workers, together with our collaborators at the GBF Braunschweig laboratory, solved the structure of the $C_P \rightarrow$ Ser mutant of *Mtb* TPx (TPxC60S). The structure of this variant differs from that of TPxWt in a manner broadly consistent with the local restructuring model of Choi *et al*. The TPxC60S structure has the extra regular secondary structure and C_R position that corresponds to an 'open' conformation, and thus probably mimics the reduced state of the wild-type protein. The two molecules in the asymmetric unit of TPxC60S have nearly identical structures, with a back-bone atom root mean square deviation (R.M.S.D.) of 0.3 Å (Stehr *et al*, 2006).

3.11. Structural comparisons: Similarities and differences between *Mtb* TPxC60S and *Mtb* TPxWt.

This section discusses the structural comparisons of the two *Mtb* TPx variants with each other, and with a selection of PDB deposited 3D Prx structures.

At present there are two crystal forms of *Mtb* TPx currently deposited in the PDB: a mutated variant of *Mtb* TPx designated as *Mtb* TPxC60S (PDB code: 1Y25) and an oxidised wild-type form of *Mtb* TPx (TPxWt) (PDB code: 1XVQ). Both structures exhibit an extended thioredoxin-like fold composed of a central β-sheet core (strands β3-β7) with flanking α-helices (α1-α4), β-strand segments (β1 and β2) and one $\alpha 3_{10}$ helix.

TPxWt and TPxC60S, although structurally thioredoxin-like, are somewhat different in the detailed secondary structure composition and organisation. The automatically generated (PDBSUM) secondary structure topology maps for

TPxWt and TPxC60S are presented in Figure 3.6. The two structures have in common β1, β2, the 3_{10} α-helix turn in a loop connecting β-strands β3, β4, β6 and β7 in the central β-sheet, and flanking helices α1-α4. Compared to TPxWt, TPxC60S has an extra β-strand (β5) in the central sheet. In addition, the length and residue boundaries of α1 and α2 helices are significantly different. In TPxWt, α1 starts at residue 61 and ends at residue 74, a total of three residues longer than the equivalent α1 helix in TPxC60S. In helix α2, the reverse is true. In TPxC60S, α2 (starting at residue 85 and ending at residue 97) is longer by seven residues when compared to α2 in TPxWt (starting at residue 85 and ending at residue 90).

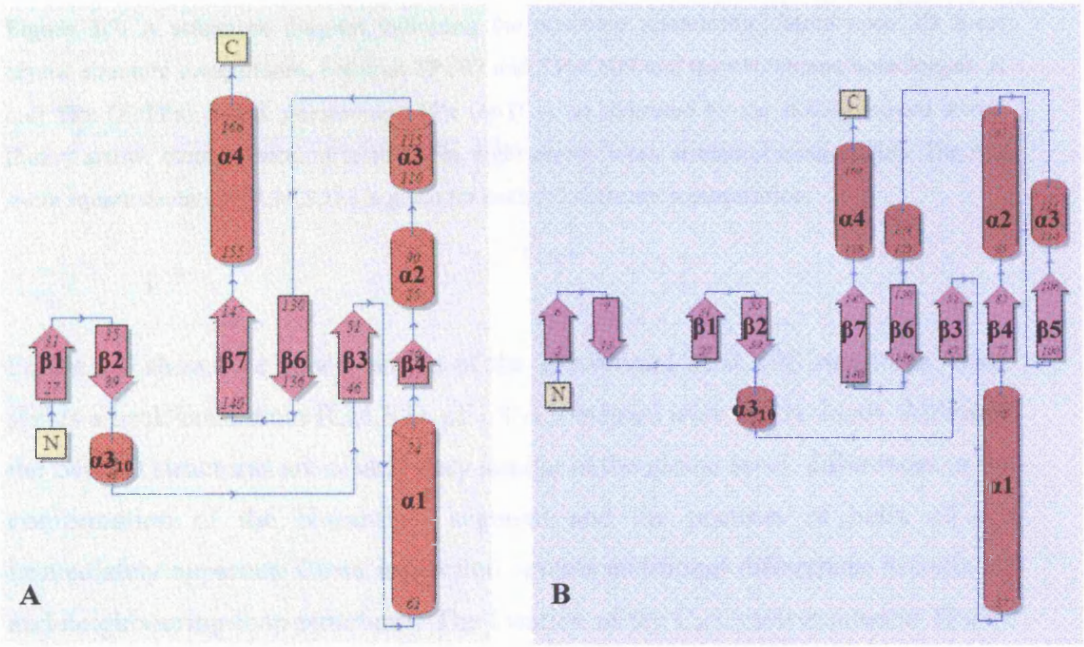


Figure 3.6: Secondary structure maps of **A:** oxidised TPxWt (PDB: 1Y25) and **B:** reduced TPxC60S (PDB: 1XVQ). β-strands are indicated as purple arrows and α-helices are represented as red cylinders. Figure adapted from PDBSUM.

The following schematic describes the structural relationship that TPxWt and TPxC60S have with each other and with their two closest homologues, *Ec*TPx and *Sp*TPx (Figure 3.7). The figure provides the best-fit superposition R.M.S.D. values, of all back-bone atoms, for each pair of structures.

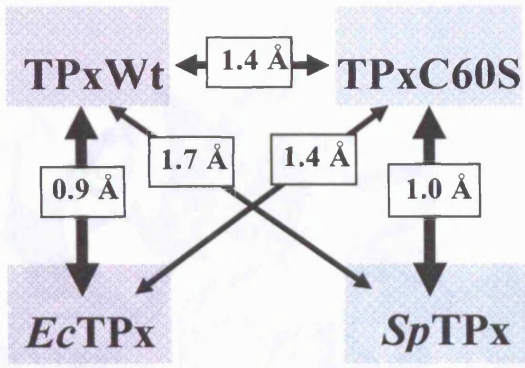


Figure 3.7: A schematic diagram indicating the structural relationship, based upon 3D X-ray crystal structure comparisons, between TPxWt and TPxC60S and the two closest homologues, *E. coli* TPx (*Ec*TPx) and *S. pneumoniae* TPx (*Sp*TPx), as indicated by the double-headed arrows (heavy arrow: strong structural relationship, light arrow: weak structural relationship). The root mean square deviation (R.M.S.D.) is given for each 3D structure superposition.

Figure 3.8 shows the superposition of the TPxWt and TPxC60S structures which yields a back-bone atom R.M.S.D. of 1.4 Å, obtained over 164 residues. Although the two 3D structures are clearly very similar at the global level, differences in the conformation of the N-terminal segment and the position of helix α4 are immediately apparent. Close inspection reveals additional differences in helix α2 and neighbouring loop structures. The location of the C_P Cys60 residue in TPxWt and the equivalent Ser60 (mutated) residue in TPxC60S are found to differ substantially between the two structures. In TPxWt the Cys60 residue is part of the N-terminus of helix α1, whereas in the TPxC60S mutant Ser60 is outside the helix α1, which starts at position 61, in a loop region.

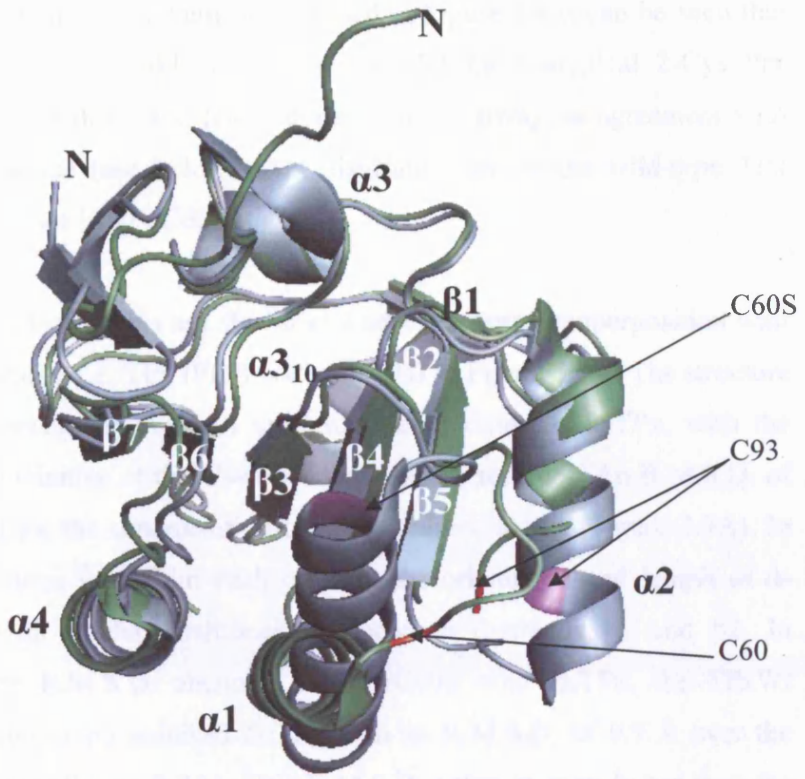


Figure 3.8: Superposition of reduced and oxidised *Mtb* TPx as represented by TPxC60S (blue) and *Mtb* TPxWt (green), respectively, with the positions of the C_P (TPxWt) and C_P-equivalent (TPxC60S), including C_R residues coloured magenta for TPxC60S and red for TPxWt. Figure obtained using PyMOL (DeLano, 2006)

With respect to the N-terminal region, the distal orientation of the N-terminal extremity found with TPxWt (compared to TPxC60S and other thiol peroxidases) is suggested by Rho *et al.* to be due to intermolecular interactions resulting from a different crystal packing arrangement (Rho *et al*, 2006). Indeed, TPxC60S adopts an N-terminal configuration consistent with other TPxs. In the TPxC60S crystal structure residues Ala2, Gln3 and Asn9 in one monomer make hydrogen bond contacts with Thr28, Thr30 and Ser110 in the other monomer resulting in a compact N-terminus.

In respect of the other differences between the two 3D structures, concerning helices α1 and α2, the variation occurs in a manner consistent with the Prx local restructuring model of Choi *et al.* Examination of the position and orientation of helix α1 in the two structures reveals a tilt and shift in its position by approximately 4 Å inducing a similar shift in helix α2 (Figure 3.8). Combined

with the differences in helix length summarised in Figure 3.6 it can be seen that TPxWt and TPxC60S resemble the general model for L-atypical 2-Cys Prx polypeptides in the oxidised and reduced states, respectively, in agreement with the presumed presence (see below) of a disulfide bond in the wild-type TPx enzyme and its absence in TPxC60S.

The two TPx crystal structures are shown as a separate best-fit superposition with the structure of oxidised *Ec*TPx (PDB code: 1QXH) in Figure 3.9A. The structure of TPxC60S superimposes well on the structure of oxidised *Ec*TPx, with the exception of the vicinities of the C_P-equivalent and C_R residues. An R.M.S.D. of 1.4 Å is obtained for the superposition of all back-bone atoms (Figure 3.9A). In general, the structures vary from each other in the orientation and length of α-helices α1–α4, with milder positional variances in β-strands β1 and β2. In comparison to the R.M.S.D. obtained for TPxC60S with *Ec*TPx, the TPxWt structure superimposes on oxidised *Ec*TPx with an R.M.S.D. of 0.9 Å over the entire C^{α} back-bone (Figure 3.9A). This R.M.S.D. value is even better than the superposition of TPxWt with TPxC60S (1.4 Å). The lower R.M.S.D. value obtained for TPxWt with oxidised *Ec*TPx, over that achieved by TPxC60S, is representative of the oxidised state of TPxWt in solution and is therefore consistent with the 2-Cys Prx model proposed by Choi *et al.*

When comparing the structures of *Mtb* TPxC60S with a reduced form of *S. pneumoniae* TPx (*Sp*TPx) (PDB code: 1PSQ), another close Prx family homologue, a back-bone R.M.S.D. of 1.0 Å for all the residues in the crystal structures is obtained. This fit includes the binding pocket in TPxC60S. The only regions of the TPxC60S structure exhibiting any significant differences to the *Sp*TPx structure are found in residues 12–23, which includes a loop region proximal to β-strand β1, and residues 71–77 which includes a region of the polypeptide linking helix α1 to β-strand β4 (Figure 3.9B).

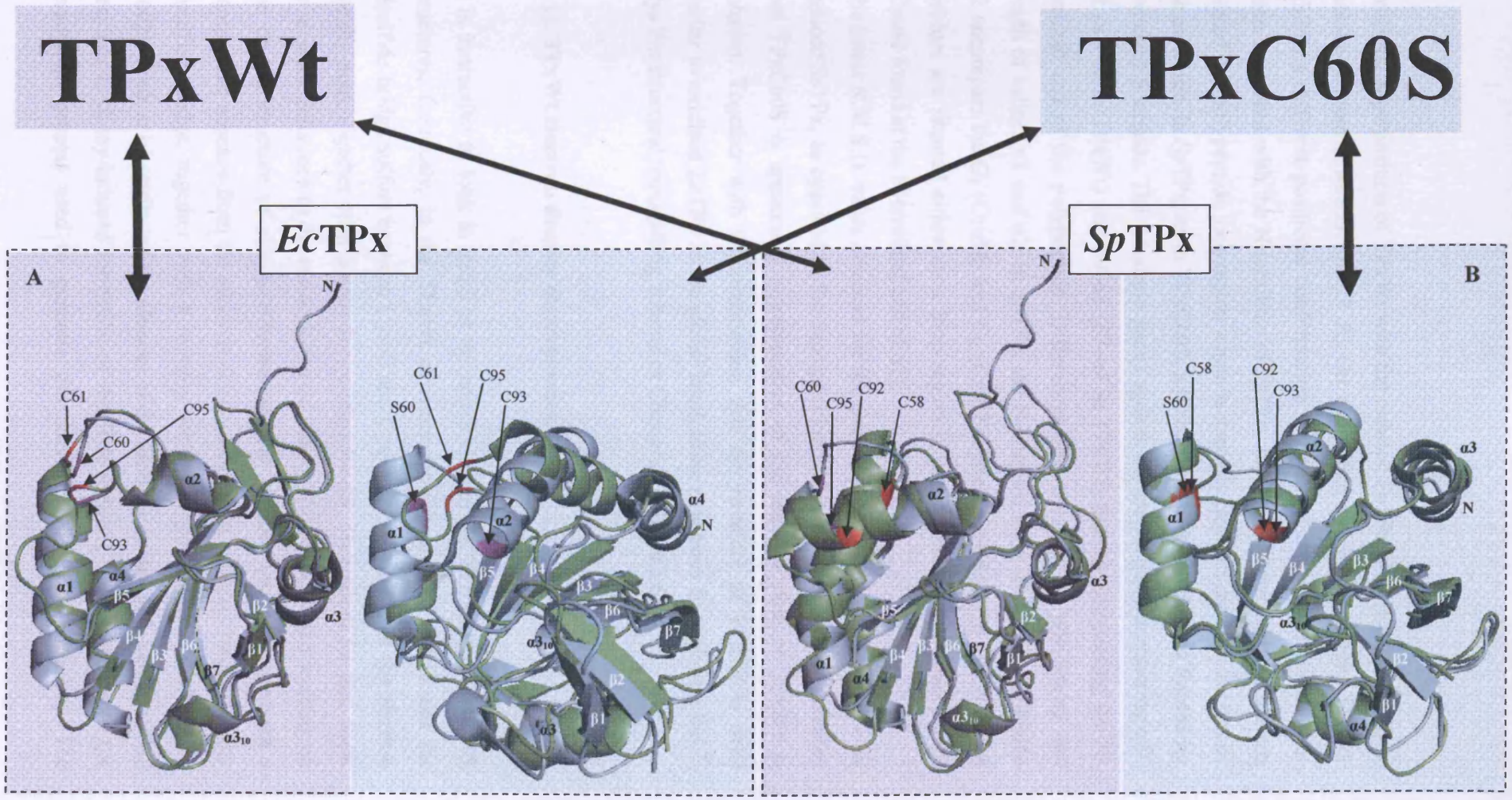


Figure 3.9 A: 3D structural superposition of TPxWt (magenta background (BG)) and TPxC60S (blue BG) with *Ec*TPx. **B:** 3D structural superposition of TPxWt (magenta background (BG)) and TPxC60S (blue BG) with *Sp*TPx. The arrows indicate the 'strength' of structural similarity as indicated in Figure 3.7.

Structural superposition of TPxWt with the reduced *Sp*TPx (Figure 3.9B) yields a back-bone atom R.M.S.D. of 1.7 Å. On examining the superposition, TPxWt exhibits a different positional conformation for the first six N-terminal residues when compared with the N-terminus of *Sp*TPx. In TPxWt, these six residues are located in a polypeptide loop region which is spatially isolated, distal to the rest of the molecule. In *Sp*TPx, as in TPxC60S, a β-hairpin is present which is formed by two short β-strands. The β-hairpin packs against the loop regions connecting α3–β8 (α3–β6 in TPxWt) and β9–α4 (β7–α4 in TPxWt) thereby compacting the N-terminal end of the polypeptide. Differences also occur in the positioning and length of helices α1 and α2. In *Sp*TPx, the N-terminal helix α1 and the extended α2 encompass the C_P (Cys58) and C_R (Cys92) cysteine residues. In TPxWt these residues are situated either on a loop region (Cys93) or close to a loop region (Cys60 found at the N-terminal end of α1).

The lower R.M.S.D. value obtained for the global superposition of TPxC60S with reduced *Sp*TPx, as opposed to the superposition of TPxWt with *Sp*TPx, indicates that TPxC60S is apparently representative of the reduced state of TPxWt in solution. Together with the observation that the structure of TPxWt is most similar to oxidised *Ec*TPx at the global level, these structures fit well with the 2-Cys Prx structural remodelling scheme of Choi *et al.* (Figure 3.6).

3.12. TPxWt electron density discrepancies.

It is instructive to look in detail at the active sites of the TPxWt and TPxC60S structures. Curiously, in the TPxWt structure the Cβ atom of Cys60 and the disulfide bridge sulfurs between Cys60 and Cys93 are missing from the electron density map, together with density for residues 94–99. To account for this, Stehr *et al.* made reference to the research carried out by Weik *et al.* When working on the crystal structure of acetylcholinesterase, this research group encountered a comparable absence from the electron density map of an expected intermolecular disulfide bridge, together with a missing section of polypeptide (Weik *et al*, 2002). Weik *et al.* attributed the absence of the disulfide bridge and corresponding residues to X-ray-induced formation of disulfide radicals, a consequence of the irradiation process used to generate the data for the crystal structure. The

additional missing residues were presumed to exhibit weak electron densities due to radiation damage.

3.13. Active site structure.

In all Prx structures, the active site is located proximal to the dimer interface. In the TPxWt structure (Rho *et al,* 2006), the C_P and C_R cysteines, as well as the conserved Thr57 and Arg130 residues, are located in a well-defined T-shaped cleft with surface dimensions 19 Å x 17 Å x 8 Å with a hole of ~8 Å in diameter on the right side of the cleft (Figure 3.10).

When the TPxWt active site region is compared to the corresponding region in TPxC60S, there is a characteristic variation in the organisation of the local secondary structural elements - as previously highlighted (see above). The C_P-equivalent Ser60 residue in TPxC60S is part of the α1 helix, as opposed to being located on a loop region located to the N-terminal side of α1 as in TPxWt (Figures 3.10 and 3.11). The C_R residue in TPxWt is located on a loop region extending from the C-terminal end of the α2 helix, while in TPxC60S the C_R residue is located on the α2 helix itself.

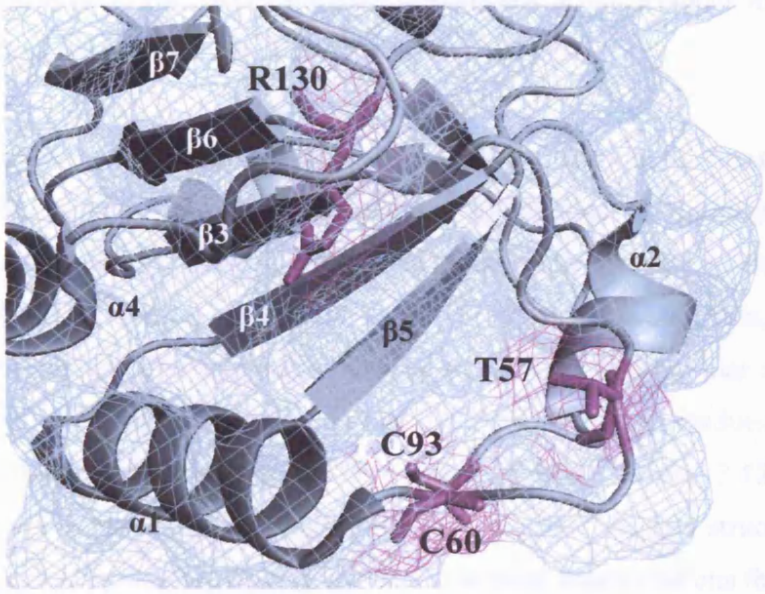


Figure 3.10: TPxWt active site comprised of 2 catalytic cysteines, C_P and C_R, and including Arg130 and Thr57, all shown coloured in magenta. Figure composed using PyMOL (DeLano, 2006).

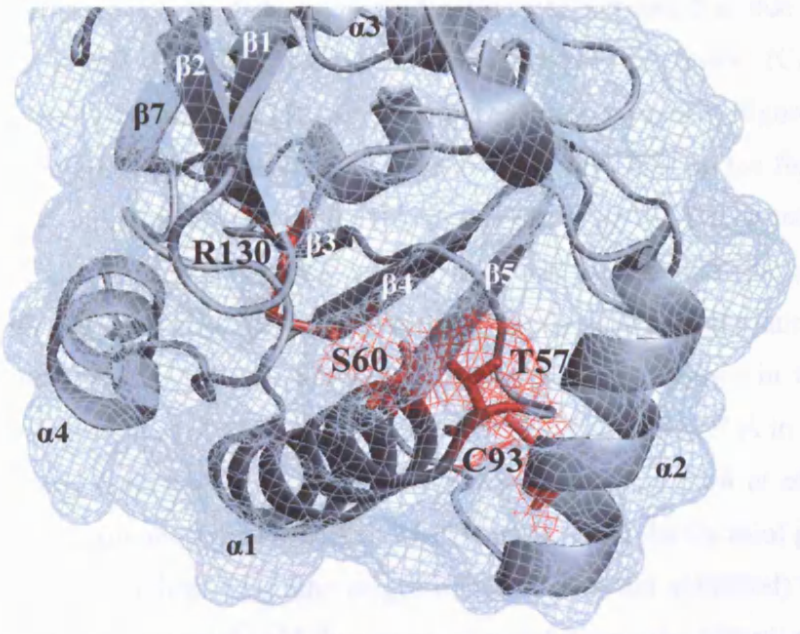


Figure 3.11: TPxC60S active site comprised of the catalytic cysteine, Cys93 and Ser60, plus two C_P cleft residues Arg130 and Thr57, all shown coloured in red. Figure composed using PyMOL (DeLano, 2006).

A total of 19 out of the 27 amino acid residues found on the surface of the TPxC60S binding pocket cleft are conserved in *Ec*TPx (the closest homologue), with 13 conserved in the *Sp*TPx enzyme. Two-thirds of the residues that line the bottom of the cleft are hydrophobic.

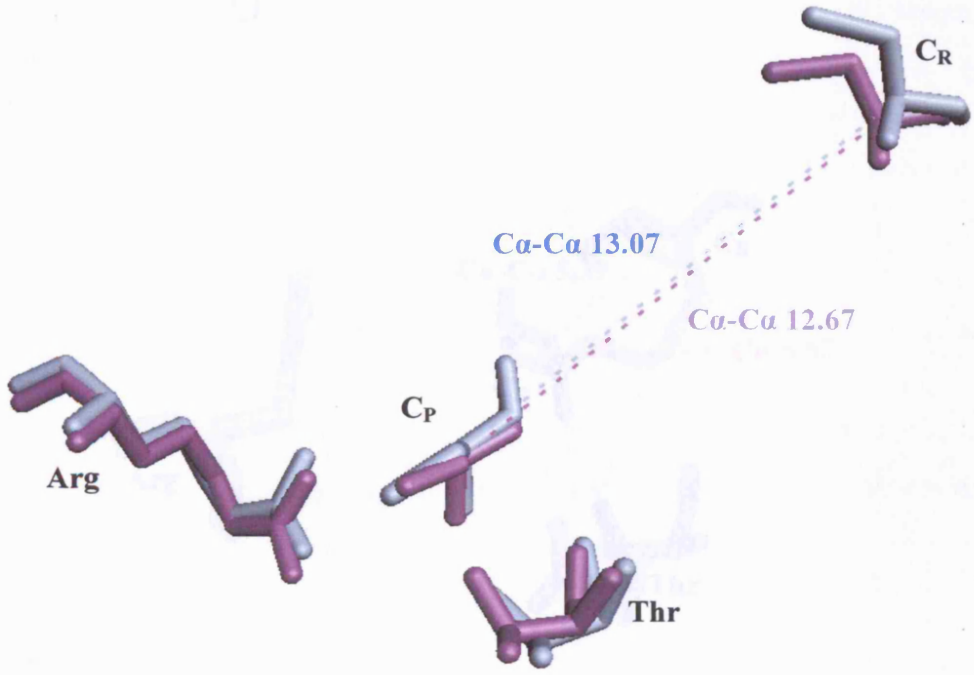
On the basis of multiple Prx sequence alignments and early indications of the 3D structure Hofmann *et al.* suggested that the C_P residue is 'activated' by the proximity of a nearby positive charge of a conserved basic residue and by hydrogen bonding with a conserved serine or threonine side-chain, which is presumed to promote the active site cysteine to form the thiolate that can readily attack the ROOH substrate (Hofmann *et al*, 2002). These three residues constitute what is colloquially known as the Prx catalytic triad. Figure 3.12 shows a schematic representation of the binding pocket regions of the TPx structures from *Mtb*, *E. coli* and *S. pneumoniae* displayed as structural superpositions that indicate the inter-C_α between the C_P (or C_P-equivalent) and C_R residues.

In each case, the position of the conserved active site Arg and Thr side chains is highlighted. In *Mtb* TPx the catalytic triad is constituted by Cys60 (C_P), Thr57 (the conserved Thr) and Arg130 (the conserved basic residue). Figures 3.12A shows the superposition of the active site residues of *Mtb* TPx (in the form of the C60S mutant) superposed on the equivalent residues of the *Sp*TPx structure. The residues superpose almost exactly, suggesting that the interactions of the C_P residue with the other two side chains of the catalytic triad are maintained in the two structures. Notably the separation of the C_P and C_R residues in these two structures is substantial (Cα-Cα 12.67 Å in TPxC60S; Cα-Cα 13.07 Å in *Sp*TPx). These two structures appear to fit well with the proposal of Choi *et al.* that the activation of C_P involves the conserved Thr Oγ positioning the C_P thiol proton for abstraction by a catalytic base (the origin of which was not specified) while the conserved Arg residue might aid this process by stabilizing the effective negative charge on the C_P sulphur atom. Consequently, the p*Ka* value of the C_P residue is expected to be decreased and the nucleophilic attack on the hydroperoxide substrate to be facilitated (Choi *et al*, 2003). Interestingly the detail of the H-bonding pattern of the Ser60 residue in the TPxC60S mutant is not completely consistent with this model (Figure 3.13). Perhaps because of the substitution at the C_P residue, or because of the presence of the acetate ion in the active site, the structure does not predict an H-bond interaction between the side chains of Thr57 and Ser60, despite their evident proximity (2.89 Å). Arguably the precise arrangement of the side chains in this structure would be different in the reduced state of TPxWt.

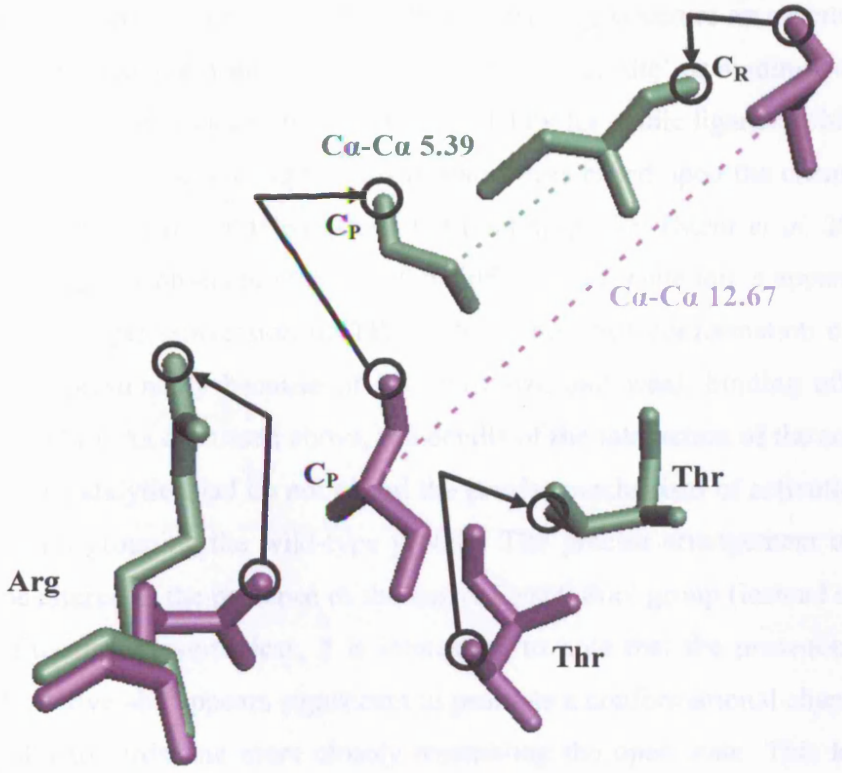
Figure 3.12C shows the superposition of the active site residues of TPxWt and *Ec*TPx, structures that are both presumed to contain a C_P-C_R disulphide bond. Again the superposition of these atoms appears to be very close. The presence of the disulphide bond is consistent with the relatively small C_P-C_R residue separations (Cα-Cα 5.39 Å in TPxWt; Cα-Cα 5.58 Å in *Ec*TPx) (Figure 3.12C).

However, when the TPxC60S binding pocket region is compared to that of oxidised TPxWt, the position of the residues involved in the catalytic triad are substantially different (Figure 3.12B). The most notable difference is in the position of the C_P and C_R residues. In TPxWt, the C_P and C_R residues are presumed (despite the absence of electron density for some atoms) to be bonded by the disulphide bridge linkage; while in TPxC60S the distance between the two

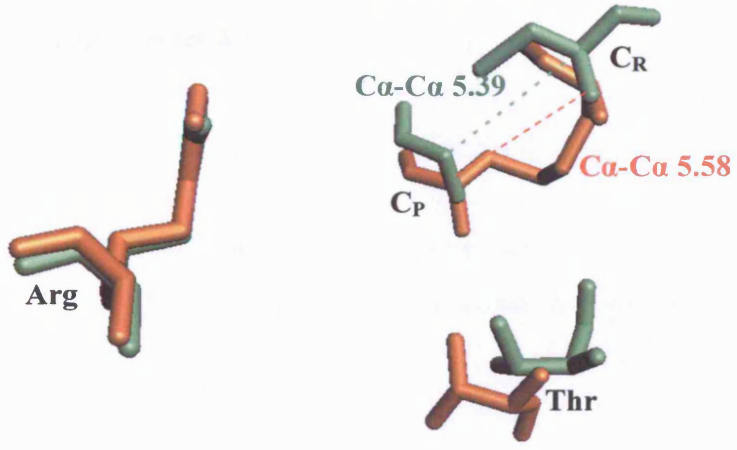
residues is substantially greater (Cα-Cα 12.67 Å). If the structures of *Mtb* TPxWt and TPxC60S represent the conformations of oxidised and reduced TPx pertinent to the reaction mechanism of Figure 3.3, then it appears that the general features are conserved with TPxs from *E. coli* and *S. pneumoniae*. Moreover, the overall nature of the structures combined with a detailed analysis of the catalytic triad side chains appears to highlight that a significant remodelling of the Prx structure is required to bring the C_P and C_R side chains together to form the oxidised disulphide bonded reaction intermediate from an initial state in which the side chains are much further apart.



A



B



C

Figure 3.12: A stick-representation of the catalytic triad (C_P, Arg and Thr), together with the C_R residue, found in the binding pocket of most Prxs. A superposition of the catalytic triad of **A**: TPxC60S (magenta) and *Sp*TPx (cyan), **B**: TPxC60S (magenta) and TPxWt (green). Arrows and circles indicate the movement of the distal N-terminal end of each residue, from the reduced state position to the oxidised-equivalent state position. **C**: TPxWt (green) and *Ec*TPx (orange). C_P and C_R residue distances (Å) are shown as doted line in each figure-set.

In the magnesium acetate buffer used to generate the TPxC60S crystal structure, free acetate ions were present in solution. In the resulting structure an acetate ion was observed to bind at a position identified as the 'active site' or binding pocket (Figure 3.13). This suggests a potential affinity of TPx for acidic ligands, which is understood to be common among various types of Prxs based upon the chemistry of the residues commonly situated within the binding-pocket (Stehr *et al,* 2006). The crystallographers observed that the TPxC60S-bound acetate ion is apparently insufficient to trigger conversion of TPxC60S to the open conformation of the oxidised state presumably because of its small size and weak binding affinity (Stehr *et al,* 2006). As discussed above, the details of the interaction of the acetate group with the catalytic triad do not reveal the precise mechanism of activation of the Cys60 thiol grouping the wild-type protein. The precise arrangement of the triad may be altered in the presence of the native Cys60 thiol group (instead of the Ser60 substitution). Nevertheless, it is interesting to note that the presence of a ligand in the active site appears significant to promote a conformational change of *Mtb* TPxC60S towards one more closely resembling the open state. This leaves open the question of how the protein's C_R residue samples the reactivity of the C_P residue; does this require a substrate-induced conformational change, or a sulfenic acid modulated allosteric transition, or any other mechanism? This matter is considered further in Chapter 4.

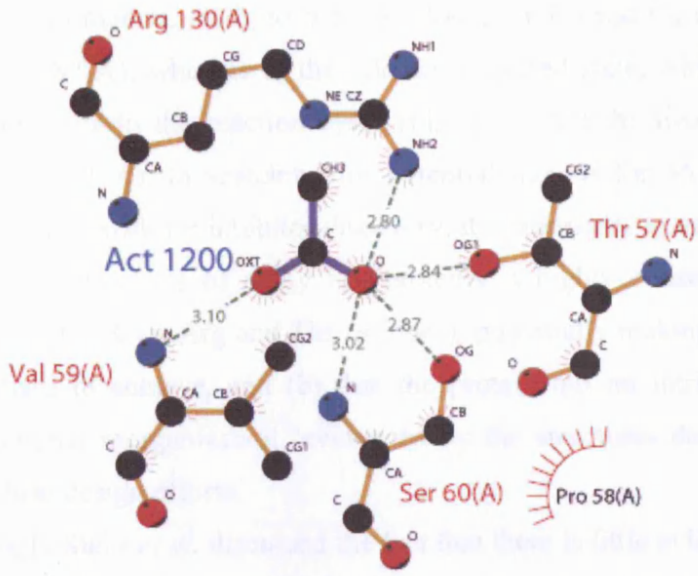


Figure 3.13: A ball and stick representation of the TPxC60S binding pocket with associated acetate ion. Black dots are the carbon atoms, blue dots are nitrogen and red dots are oxygen. The letter (A) that accompanies each residue shown is in reference to all residues present in the same protomer.

3.14. Discussion.

Stehr *et al.* revealed that the acetate ion, found to bind to the TPxC60S binding-pocket, forms hydrogen bonds from a carboxylate oxygen atom to a number of closely orientated side chain residues that include Ser60, Thr57, and Arg130 as well as the back-bone amide NH of Ser60 (Stehr *et al*, 2006). They postulate that similar contacts may be possible for the typical Prx substrates, e.g. hydrogen peroxide, *t*-butyl hydroperoxide and cumene hydroperoxide (Jaeger *et al*, 2004 and Stehr *et al*, 2006).

In the context of the broad aims of the work described in this thesis the availability of the crystal structures of TPxWt and TPxC60S is extremely useful. The comparison of the structures and their homologues clearly indicates that the L-atypical 2-Cys Prx reaction scheme of Choi *et al.* should be relevant to the mechanism of *Mtb* TPx. Thus one can anticipate that the structure around the active site is dependent upon the redox state of the enzyme. For example in the

'open' reduced state, which presumably is a representation of the catalytically primed conformation, ready to bind substrate, the C_P and C_R residues are some distance (9.87 Å), whereas in the 'closed' oxidised state, which must occur at some later stage in the reaction cycle (Figure 3.12C) the two are more tightly associated (4.61 Å). In searching for potential ligands for *Mtb* TPx that might provide a framework for inhibitor discovery, it is necessary to confront (a) the fact that the C_P active site of many Prx proteins is highly conserved (e.g. by the catalytic triad of Cys, Arg and Thr residues), potentially making the *Mtb*-specific activity hard to achieve, and (b) that the protein has an intrinsic potential for conformational reorganisation (evidenced by the structures described here) that may confuse design efforts.

Interestingly Stehr *et al.* discussed the fact that there is little evidence for dynamic properties of *Mtb* TPx in the TPxC60S structure. Thus they found no pattern in the electron density, or in the X-ray temperature factors, that suggested the potential for dynamic behaviour in this region. They concluded that "the [structural changes] are either induced by ligand binding or oxidation of the peroxidatic cysteine. The acetate molecule ...is evidently not sufficient to trigger conversion to the open conformation of the oxidized state" (Stehr *et al*, 2006). One can argue with the validity of these comments, but it remains a fact that along the reaction coordinate the TPx protein appears to interconvert between substantially different sub-states, and by definition this can only be achieved by a dynamic process.

The NMR investigation of *Mtb* TPx discussed in subsequent chapters was performed at the same time that the crystallographic results were emerging, and it is interesting to analyse these results, obtained in solution, in the context of the foregoing descriptions from X-ray diffraction.

Chapter 4

TPxWt and TPxC60S – Results: secondary structure determination and preliminary NMR spectroscopic data

4.1. Chapter summary.

E. coli strains transformed with plasmids for the expression of *Mtb* TPxWt (wild-type) and a (Cys60 → Ser60) mutant TPxC60S were obtained from the German Research Centre for Biotechnology (GBF) supplied on antibiotic-prepared LB agar plates. Following confirmation of amino acid sequence by the amino acid sequencing company MWG-Biotech™, both TPx proteins were subsequently expressed in large quantities, and purified from the soluble fraction using the methods described in Sections 2.2-2.3. Sufficient pure protein was obtained from 3 L culture volumes to yield samples of 2.0–4.5 mM concentration. The oligomerisation state and secondary structure content was assessed using analytical size exclusion chromatography (ASEC) and circular dichroism (CD) respectively. For NMR, the expression protocols for proteins was modified to include uniform multi-isotope labelling to produce [^{15}N, ^{2}H, ^{13}C]-enriched TPxC60S for triple resonance spectroscopic experiments. 80 % of the resonances corresponding to the back-bone atoms of TPxC60S were assigned. The corresponding "missing" cross peaks from the back-bone assignment process has been hypothesized as being due to dynamic motion between alternate states, resulting in chemical exchange between nuclei leading to line broadening. Potential scenarios to account for the pattern of line broadening are discussed.

4.2. TPxWt and TPxC60S – expression and purification.

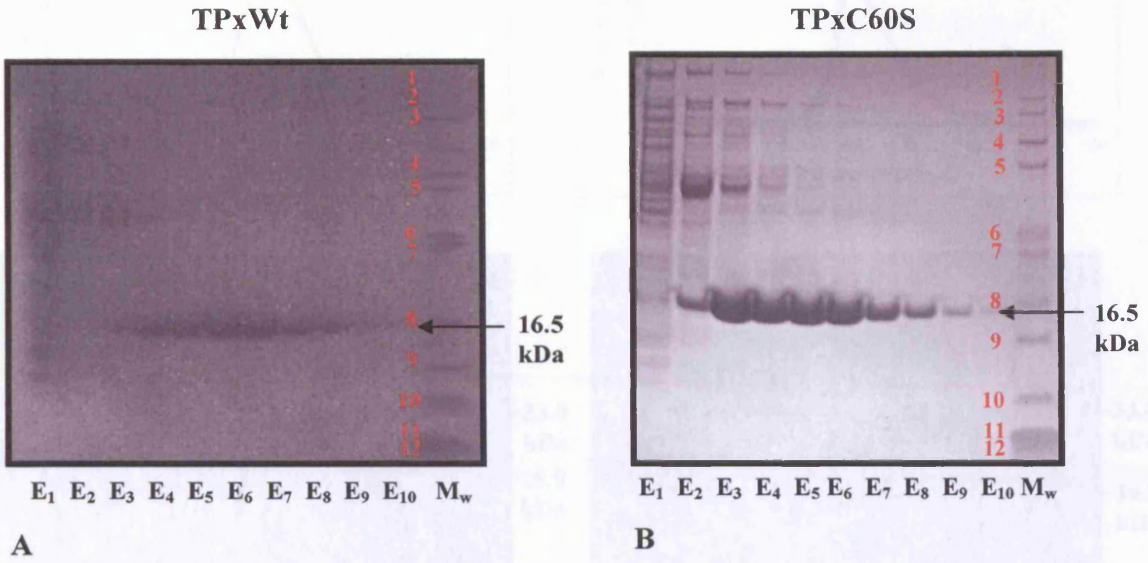
Studying proteins using NMR techniques typically requires proteins in multi-milligram quantities at millimolar concentrations in aqueous buffers. In many cases, technical and/or economic limitations can become apparent in the production of proteins.

In order to carry out a biologically relevant study of both TPxWt and TPxC60S (the amino acid sequences of each protein was confirmed through amino acid sequencing carried out by MWG-Biotech™) the buffer conditions must be kept as close to physiological pH (pH 7.4) and the salt concentration kept as *in vivo*, optimally below 200 mM for suitable NMR spectroscopic study. Before attempting a large scale preparation of either of the TPx proteins, a small scale preparation was performed to assess the "foldedness" (stability and spectral characteristics (discussed in Section 4.5)) of the TPx proteins in typical conditions commonly applied in protein NMR experiments. A 1 L recombinant *E. coli* cell culture was grown for each protein independently (see Section 2.3). Following overnight induction using isopropyl β-D-1-thiogalactopyranoside (IPTG), each culture was then assessed for protein expression levels in both the cell pellet and soluble fractions (as described in Section 2.2). Following rupture by French press, SDS-PAGE gels were run (as described in Section 2.2) revealing that each *Mtb* TPx clone did express substantial yields in the soluble fraction at the expected molecular weight (16.5 kDa). In fact, there was little evidence of TPx in the insoluble fraction.

It was straightforward to demonstrate that the expressed TPx proteins could be isolated by immobilised metal ion affinity chromatography (IMAC) using Ni-NTA resin (see Section 2.3.3). The proteins were retained on the resin during buffer washes with low concentrations of imidazole, and could be recovered in essentially pure form by elution with 150-300 mM imidazole.

To obtain sufficient material for biophysical investigations, I performed a larger scale preparation of each TPx protein: 3 L culture volume followed by cell rupture, and protein purification by Ni-NTA IMAC and preparative size exclusion chromatography (SEC) (see Sections 2.2-2.3).

Figures 4.1A and 4.1B show the SDS-PAGE gels obtained after the Ni-NTA chromatography purification for both TPxWt and TPxC60S, respectively.

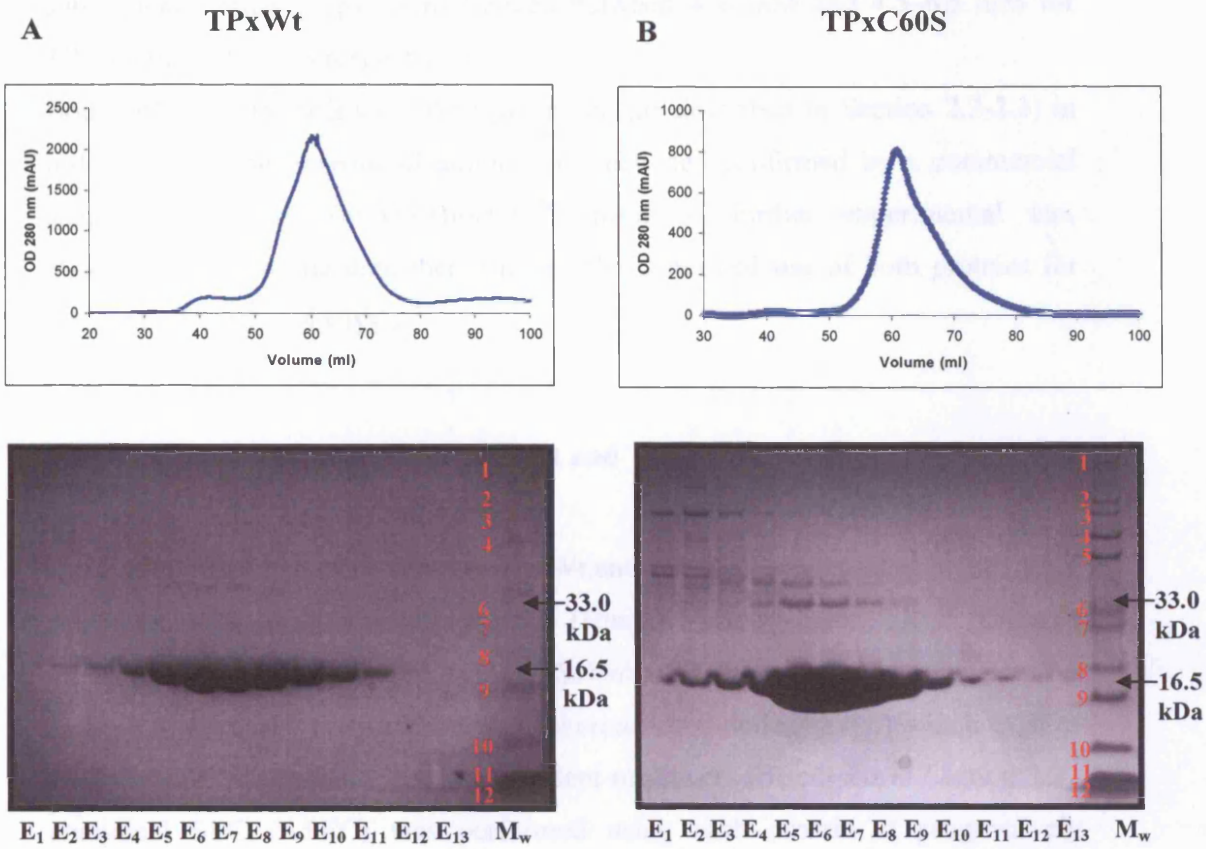


Key: E_x = Elution fraction, M_w = Molecular weight marker

Figure 4.1: A and **B** SDS-PAGE gels of [^{1}H]-TPxWt and TPxC60S purified by Ni-NTA IMAC Lanes labelled E_1-E_{10} represent the fraction obtained with increasing imidazole concentration in the elution buffer: E_1 (5 mM imidazole) – E_{10} (500 mM imidazole). Red numbering indicates the M_w protein marker sizes for A and B gels, 1: 200 kDa, 2: 116.3 kDa, 3: 97.4 kDa, 4: 66.3 kDa, 5: 55.4 kDa, 6: 36.5, 7: 31 kDa, 8: 21.5 kDa, 9: 14.4 kDa, 10: 6 kDa, 11: 3.5 kDa, and 12: 2.5 kDa.

Based solely on the qualitative interpretation of the SDS-PAGE protein gels produced, the Ni-NTA purification process reveals that both TPxWt and TPxC60S are obtained at high levels. As reducing state SDS-PAGE conditions were used, both proteins appear on the gel at the monomeric molecular weight (16.5 kDa). IMAC fractions (E_3-E_9) were then pooled and concentrated (as described in Section 2.2.) and the TPx proteins further purified by being passed down a Superdex-75 SEC column (as described in Section 2.3.4.). This step separates out any large aggregate impurities and additionally removes the bulk of the imidazole buffer from the IMAC step.

Figures 4.2A and 4.2B show the SEC chromatograms and corresponding protein gels obtained for the SEC of TPxWt and TPxC60S, respectively.



Key: E_1 (35 mL) – E_{13} (95 mL) = Fraction number, M_w = Molecular weight marker

Figure 4.2: A and **B:** SDS-PAGE of [^{1}H]-TPxWt and TPxC60S purification by Superdex-75 SEC respectively. Each lane corresponds to an eluted fraction of 5 mL. Red numbering indicates the M_w protein marker sizes for A and B gels, 1: 200 kDa, 2: 116.3 kDa, 3: 97.4 kDa, 4: 66.3 kDa, 5: 55.4 kDa, 6: 36.5, 7: 31 kDa, 8: 21.5 kDa, 9: 14.4 kDa, 10: 6 kDa, 11: 3.5 kDa, and 12: 2.5 kDa.

The results demonstrate that in each case the TPx protein is eluted as the dominant species with only relatively minor contamination by other (larger) proteins.

These additional species could either represent oxidised TPx oligomers (e.g. the band at ~33 kDa; Figure 4.2) or contaminant proteins that persist through the SEC process. Subsequent attempts to separate TPxWt and TPxC60S from the identified higher order oligomeric species using ion exchange chromatography (IEC) and varying salt concentrations (200-650 mM) proved unsuccessful (as described in Section 2.3.5).

The elution fractions corresponding to the protein peaks in both chromatograms (volume 50-80 ml; elution fractions E_4-E_{11}) were then pooled and subsequently concentrated to 1 mL volume. The final concentration, averaged over a number of

subsequent similar preparations, ranged between 4–6 mM and 4.5–6.5 mM for TPxWt and TPxC60S respectively.

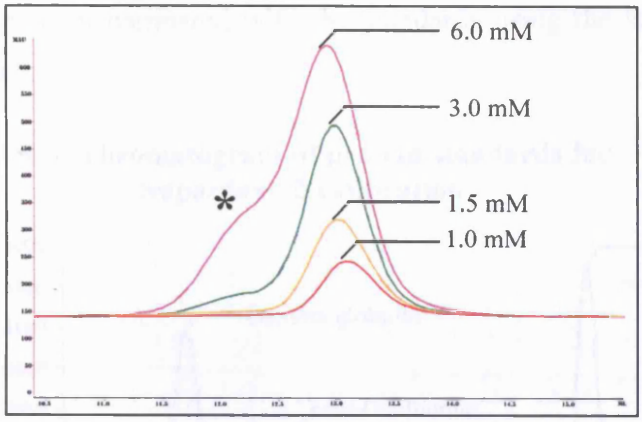
Both *Mtb* TPx proteins were then processed (as described in Section 2.2-2.3) in order to have the N-terminal amino acid sequence confirmed by a commercial sequencing service (MWG-Biotech™), prior to further experimental use. Successful sequencing data then allowed the continued use of both proteins for further biochemical analysis.

4.3. Analytical SEC analysis of TPxWt and TPxC60S.

In an attempt to determine whether TPxWt and TPxC60S can exist as higher order oligomers like some other Prxs, PrxII (Human – decameric), TRyP (*Crithidia fasciculate* – decameric) and AhpC (*Salmonella typhimurium*, *Amphibacillus xylanus* – decameric, *Mycobacterium tuberculosis* – dodecameric) which exist as various state oligomeric redox-independent multimers (discussed in Section 3.7), analytical SEC (ASEC) was performed using each protein at progressively decreasing protein-concentrations, ranging from 6 mM to 1 mM (as described in Section 2.3.5.). Figures 4.3A and 4.3B display the ASEC data obtained for *Mtb* TPxWt and TPxC60S, respectively, using a Superdex-75 matrix and a buffer containing 1 mM DTT to maintain the reduced state.

ASEC: TPxWt

A



ASEC: TPxC60S

B

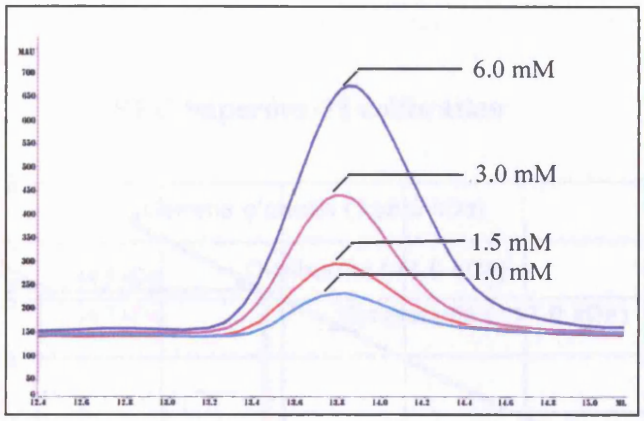


Figure 4.3: **A** and **B** Superdex-75 ASEC chromatograms of *Mtb* TPxWt and TPxC60S, respectively, at different protein concentrations. The asterisk in Figure **A** indicates a shoulder feature for the wild-type protein, attributed to a concentration dependent higher molecular weight species.

ASEC allows the assessment of the hydrodynamic properties of the target protein by comparison of the elution volumes with a range of standards of known properties. To this end the Superdex-75 column was calibrated with four molecular weight standards (Figure 4.4A). The elution volumes of TPxWt and TPxC60S where then compared with the standards using the standard semi-log plot (Figure 4.4B).

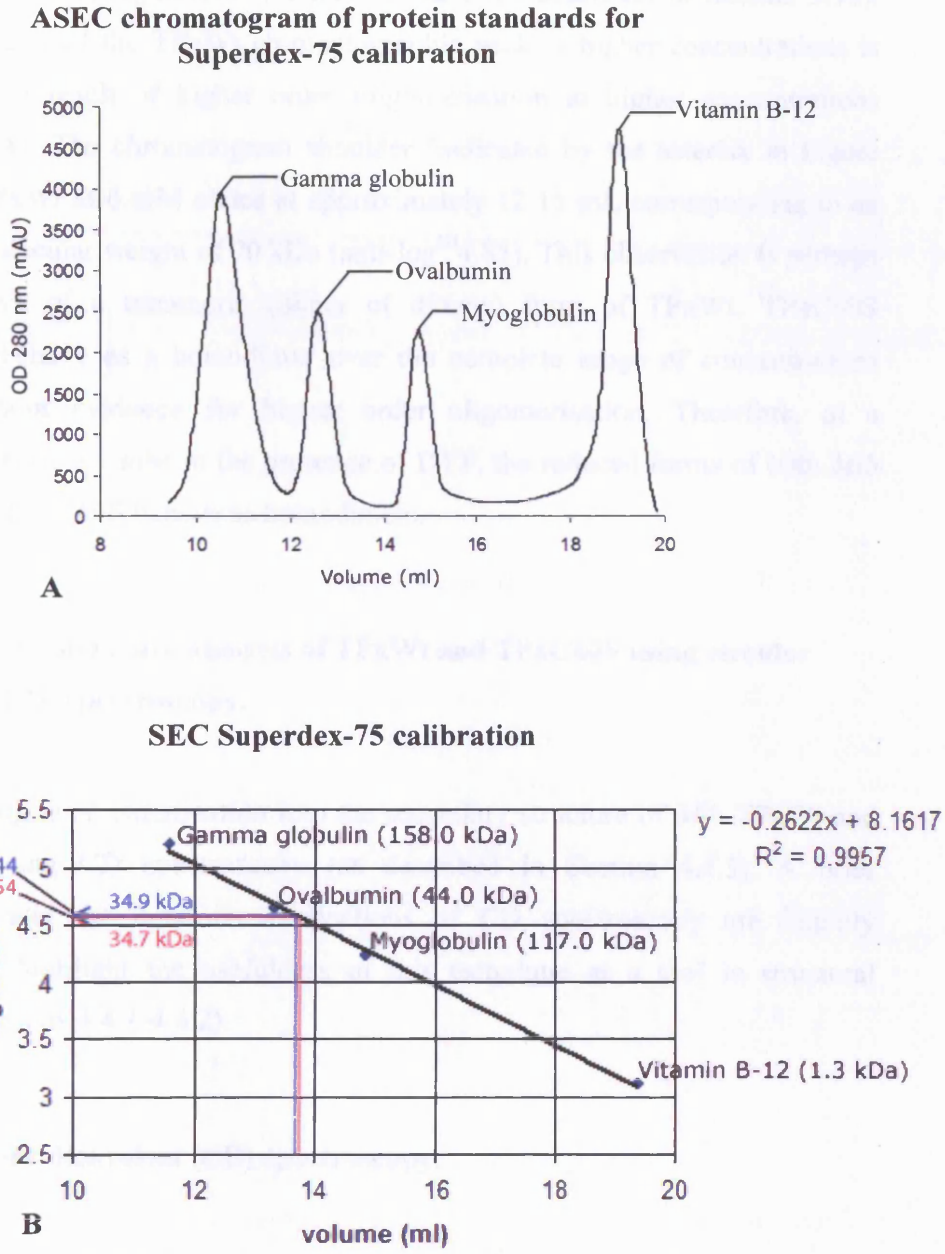


Figure 4.4: A: ASEC of protein standards obtained for the Superdex-75 column. **B** Semi-log calibration plot generated from the protein standards elution data in **A**, showing the experimental elution volumes of TPxWt (blue) and TPxC60S (red)

The most dilute protein samples of TPxWt and TPxC60S (1 mM) eluted with an apparent molecular weight of 34.9 kDa and 34.7 kDa respectively. TPxWt has a predicted monomer molecular weight of 16.9 kDa and therefore the ASEC result suggests that the protein is present in solution as a homo-dimer. The result is consistent with analytical ultracentrifugation studies of *E. coli* TPx (*Ec*TPx) and *S. pneumoniae* TPx (*Sp*TPx), which have been shown to be non-covalent homo-dimers in solution, regardless of oxidation state (as described in Section 3.10). The asymmetry of the TPxWt chromatographic peak at higher concentrations is presumably a result of higher order oligomerisation at higher concentrations (Figure 4.3A). The chromatogram shoulder (indicated by the asterisk in Figure 4.3A) of TPxWt at 6 mM elutes at approximately 12.15 mL corresponding to an apparent molecular weight of 70 kDa (anti-$\log^{10}$4.85). This observation is perhaps representative of a tetrameric (dimer of dimers) form of TPxWt. TPxC60S appears to behave as a homodimer over the complete range of concentrations tested, without evidence for higher order oligomerisation. Therefore, at a concentration of ~ 1 mM in the presence of DTT, the reduced forms of both *Mtb* TPxWt and TPxC60S behave as homodimers.

4.4. Secondary structure analysis of TPxWt and TPxC60S using circular dichroism (CD) spectroscopy.

The following is an examination into the secondary structure of *Mtb* TPxWt and TPxC60S using CD spectroscopy (as described in Section 4.4.3). A brief description and the common applications of CD spectroscopy are initially provided to highlight the usefulness of this technique as a tool in structural analysis (sections 4.4.1-4.4.2).

4.4.1. Circular dichroism (CD) spectroscopy.

CD is a low resolution technique when compared to more complex methods of structural analysis of proteins through NMR spectroscopy and X-ray crystallography.

Circular dichroism (CD) spectroscopy measures differences (as a function of wavelength) in the absorption of left-handed polarized light versus right-handed polarized light which arise due to structural asymmetry. CD data is often provided as the mean residue elipticity (MRE), in degrees cm^2 $dmol^{-1}$ $residue^{-1}$, for each amino acid residue measured and is calculated using the following equation:

$$\frac{CD\ signal}{(10x\ number\ of\ amino\ acids)\ x\ protein\ concentration\ (Molar)\ x\ path-length\ (cm)}$$

Chiral or asymmetric molecules produce a CD spectrum because they absorb left and right handed polarized light to different extents and thus are considered to be 'optically active'. Biological macromolecules such as proteins and DNA are composed of optically active elements and because they can adopt different types of three-dimensional structures, each type of molecule produces distinct CD spectrum (Whitmore, 2004 and Kalodimos, 2004).

4.4.2. Applications of circular dichroism spectroscopy.

Despite its low resolution, CD spectroscopy is useful in a variety of contexts including:

- Determining whether a protein is folded, and if so characterizing its secondary structure content and thereby infer characteristics of the overall tertiary structure, and the structural family to which it belongs.
- Comparing the structures of a protein obtained from different sources (e.g. species or expression systems) or comparing structures for different mutants of the same protein.
- Demonstrating comparability of solution conformation after changes in manufacturing processes or formulation.
- Studying the conformational stability of a protein under physical "stress" - thermal stability, pH stability, and stability to denaturants - or how this stability is altered by buffer composition or addition of stabilizers and excipients.
- Determining whether protein-protein interactions alter the conformation of protein.

For proteins, the secondary structure composition can be determined by CD spectroscopy in the "far-UV" spectral region (190–250 nm). At these wavelengths the chromophore is the peptide bond, and the signal observed is found to be dependent on the local structure of the polypeptide chain.

Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and amplitude of CD spectrum. This concept is illustrated by the graph in Figure 4.5, which shows spectra for identical peptides exhibiting different secondary structure conformations. Note: Delta ε (y-axis below) is a simple conversion of a proteins MRE data into a more manageable, clearly displayable form with the units (mdeg $M^{-1}cm^{-1}$).

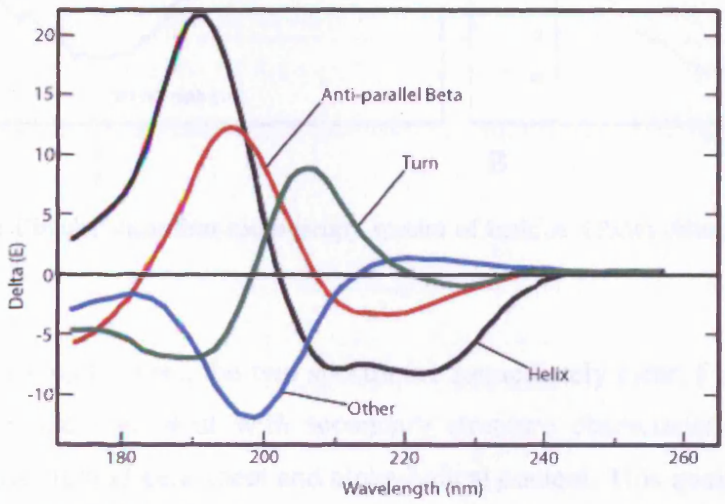


Figure 4.5: Characteristic far-UV CD signatures representing different secondary structures (adapted from Kalodimos, 2004)

The approximate fraction of each secondary structure type that is present in any protein can thus be determined by analyzing its far-UV CD spectrum as a sum of fractional multiples of such reference spectra for each structural type.

Like all spectroscopic techniques, the CD signal reflects an average of the entire molecular population in the sample. Thus, while CD can determine that a protein contains about 50 % alpha-helix, it cannot determine which residues are specifically involved in the alpha-helical portion, or if the sample is in exchange between equimolar amounts of helical and non-helical structures.

Far-UV CD spectral measurements require 20-200 µL of solution and containing 50 µg/mL to 1 mg/mL protein in a buffer which does not have a high absorbance in this region of the spectrum. High concentrations of DTT (>2 mM), histidine, or imidazole, for example, cannot be used in the far-UV region.

4.4.3. Circular dichroism spectroscopy of TPxWt and TPxC60S.

Figures 4.6A and 4.6B show the far-UV CD spectra obtained for TPxWt and TPxC60S respectively, using the method as described in Section 2.3.6.

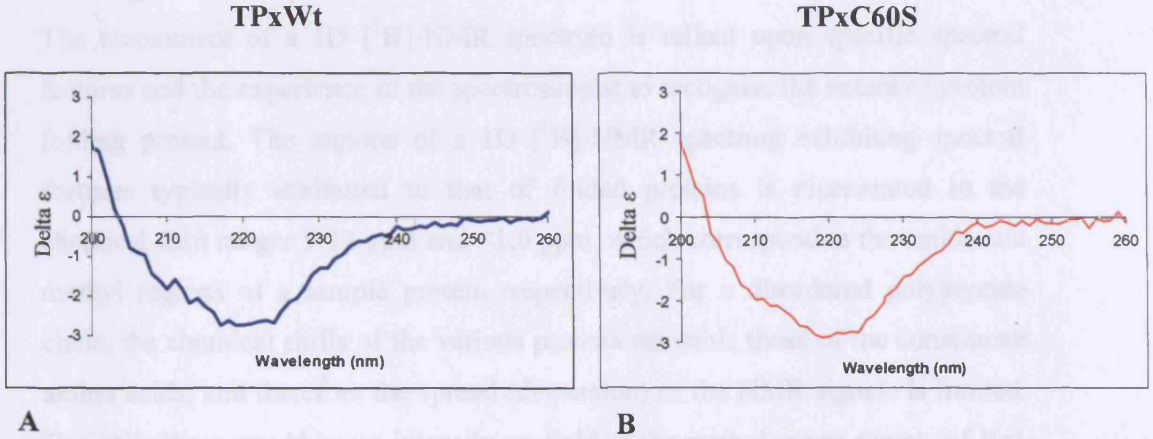


Figure 4.6: Circular dichroism spectroscopy spectra of both; **A** TPxWt (blue) and **B** TPxC60S (red)

The similarities between the two spectra are immediately clear. For both proteins the spectra are consistent with secondary structure characteristics common to proteins with mixed beta-sheet and alpha-helical content. This qualitative measure is based on the fact that both spectra fail to match exactly with either of the characteristic far-UV CD signatures illustrated in Figure 4.5. Both spectra have a minimum at ~220 nm. For a completely alpha-helical content protein, two minima are common, 222 and 208 nm, while a protein composed entirely of anti-parallel β-sheets has a single minimum at ~217 nm. Both TPx proteins lack a minima at 217 nm and as a result of lacking the complete alpha-helical minima complement it would be essentially reasonable to propose that both proteins are composed of an alpha-helical/beta-sheet mixture. The shifted minimum at ~220 nm, as opposed to an expected alpha-helical minimum of 222 nm, may be accounted for by the intermediate spectral line-shape that would exist for such mixed alpha-helical/beta-sheet secondary structure content proteins. This result is consistent with the crystal structures that exist for both TPxWt (PDB code: 1XVQ) and TPxC60S (PDB code: 1Y25) and is typical of the thioredoxin-like fold that both proteins resemble (as detailed in Sections 3.8. and 3.11.).

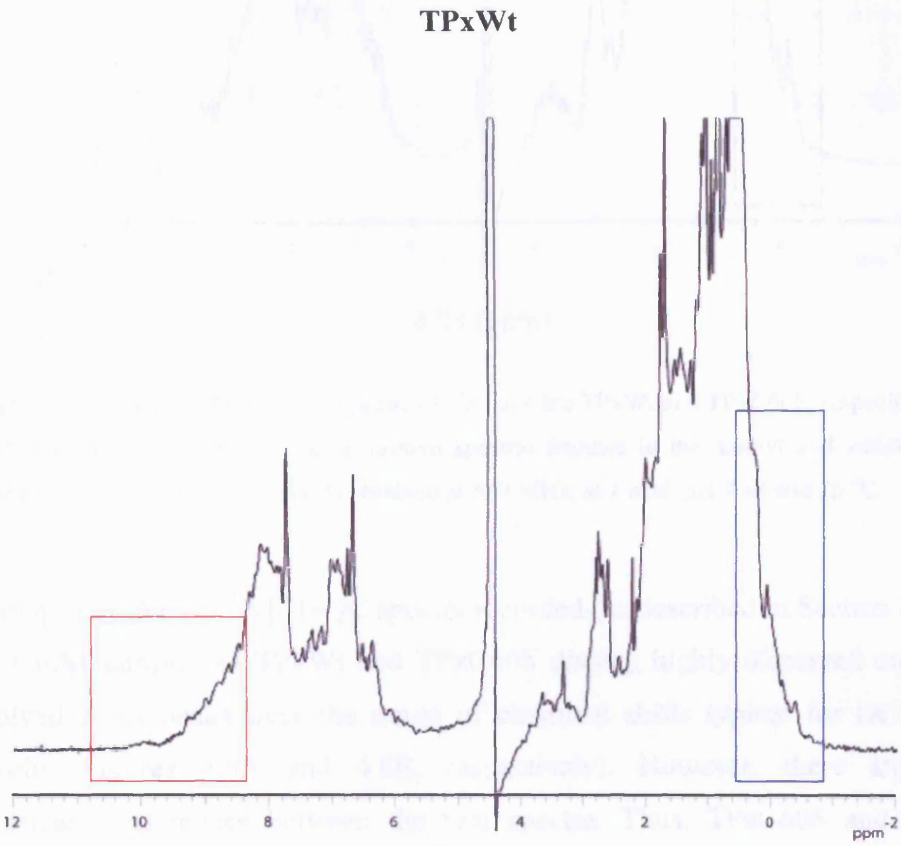
4.5. Heteronuclear NMR spectrosopy of TPxWt and TPxC60S.

Analysis by 1D [^{1}H]-NMR is often performed to assess whether a protein is folded, and is usually carried out on unlabelled protein samples prior to isotope labelling for more sophisticated 2D/3D NMR measurements.

The assessment of a 1D [^{1}H]-NMR spectrum is reliant upon specific spectral features and the experience of the spectroscopist to recognise the extent of protein folding present. The regions of a 1D [^{1}H]-NMR spectrum exhibiting spectral features typically attributed to that of folded proteins is represented in the chemical shift ranges 9-11 ppm and <1.0 ppm, which correspond to the amide and methyl regions of a sample protein respectively. For a disordered polypeptide chain, the chemical shifts of the various protons resemble those of the constituent amino acids, and therefore the spread (dispersion) of the NMR signals is limited. Typically there would be no intensity up-field of the methyl group signals of Val, Ile and Leu at 0.8 ppm, or downfield of the back-bone amide NH groups at 8.3-8.7 ppm (excepting those few side chain signals that might appear in the region below ~10 ppm, e.g. the Trp ring indole NH signal). In contrast, for an ordered folded polypeptide the chemical shift dispersion is typically substantially increased so that some aliphatic protons may appear at chemical shifts <0.8 ppm, and some amide NH groups appear in the region 8.6-10.5 ppm. The presence of signals in these regions of the 1D spectrum is usually taken as a good indication that the protein contains a folded component. The relative intensity of the spectral envelope between ~0.8 ppm and the signals up-field of 0.8 ppm (<0.8 ppm) is sometimes used to qualitatively assess the degree of folding (though there are risks to this approach). The unlabelled [^{1}H]-TPxWt and TPxC60S preparations were examined by 1D NMR which yielded spectra apparently consistent with a high degree of folding (data not shown).

This result paved the way for the use of isotope labelling of the two proteins for detailed analysis.

Figures 4.7A and 4.7B show the 1D ^{1}H NMR spectra obtained for single isotope (^{15}N) labelled preparations of both proteins. The spectra of TPxWt and TPxC60S show both proteins to be folded, with the mutant exhibiting slightly sharper resolved spectral features compared to the wild-type protein. These spectral characteristics suggest that TPxWt and TPxC60S are globularly folded molecules. The fact that the resonances of the TPxWt spectrum appear slightly broader than those of the mutant may be consistent with a higher order self association as witnessed in the corresponding SEC experiments (Section 4.3).



A δ ^{1}H (ppm)

TPxC60S

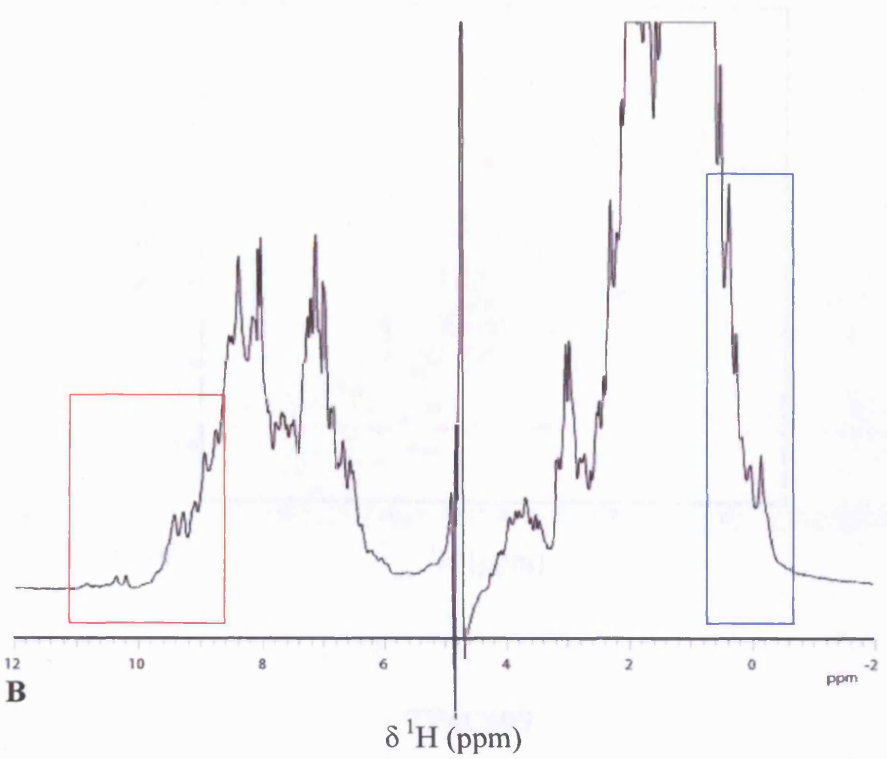
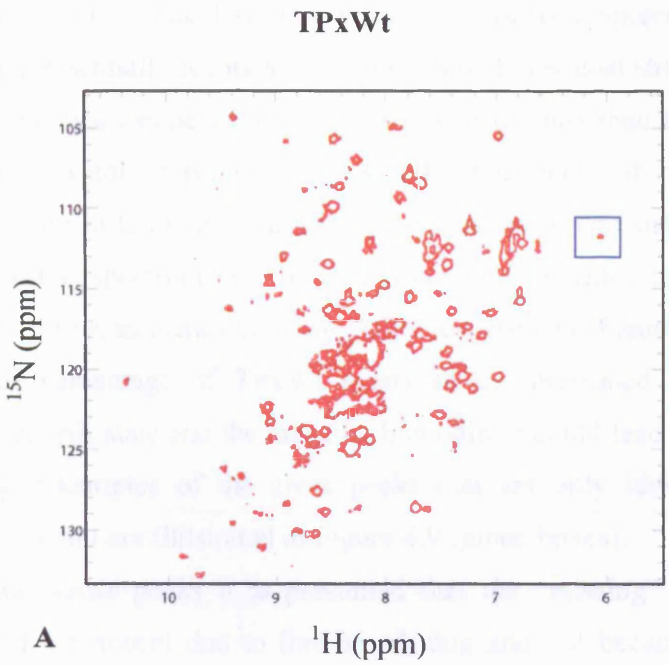


Figure 4.7: A and **B** 1D ^{1}H NMR spectra of ^{15}N-labelled TPxWt and TPxC60S, respectively. Red and blue boxes indicate the folded protein spectral features in the methyl and amide regions, respectively. Both A and B spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

Subsequent 2D [^{1}H, ^{15}N]-HSQC spectra recorded (as described in Section 2.3.7.3) for 1 mM samples of TPxWt and TPxC60S display highly dispersed and well-resolved cross peaks over the range of chemical shifts typical for an ordered protein (Figures 4.8A and 4.8B, respectively). However, there are some significant differences between the two spectra. Thus, TPxC60S and TPxWt exhibit 149 and 130 cross peaks respectively. The cross peaks in the TPxC60S spectrum are better defined than those in the wild-type spectrum by virtue of exhibiting higher intensities and a more uniform appearance.



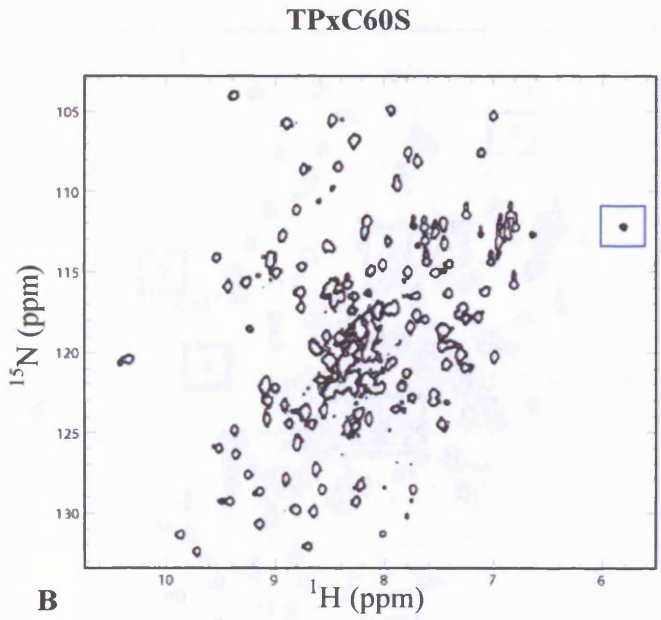


Figure 4.8: **A** and **B** display the 2D [^{15}N]-HSQC spectra of TPxWt and TPxC60S, respectively. A folded Arg side chain signals are highlighted in blue boxes. Both A and B spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

Figure 4.9 shows a superposition of [^{15}N]-HSQC spectra of TPxWt and TPxC60S. Overall, the pattern of chemical shifts are highly conserved suggesting that the two proteins have essentially identical structures. Small chemical shift differences are apparent for many cross peaks, but these are typically less than the width of a cross peak and probably represent very small adjustments of the TPxC60S structure in response to the mutation. The broad cross peak line shapes observed in the TPxWt HSQC spectrum may be as a result of a tendency to form higher order oligomeric forms, as identified in the ASEC experiment (Figure 4.3A). Even if only a small percentage of TPxWt exists as the presumed tetramer, the exchange between this state and the majority homodimer could lead to substantial line broadening. Examples of the cross peaks that are only identified in the spectrum of the mutant are illustrated in Figure 4.9 (green boxes).

For these "extra" cross peaks it is presumed that the "missing" peaks in the TPxWt spectrum are absent due to line-broadening and not because of a large chemical shift difference.

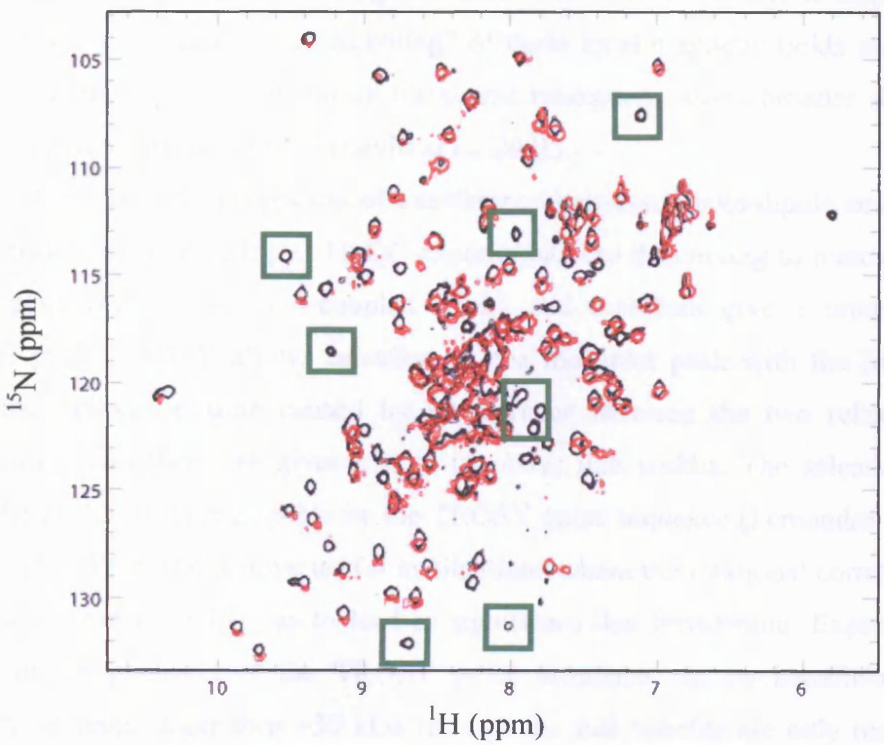


Figure 4.9: Superposition of [^{15}N]-HSQC spectra obtained for TPxWt (red) and TPxC60S (black). Green boxes indicate the cross-peaks found in the mutant-type variant that are not found in the wild-type *Mtb* TPx variant Spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

4.5.1. Transverse relaxation optimised spectroscopy (TROSY) NMR using [^{1}H, ^{15}N]-labelled TPxWt.

In an attempt to increase the sensitivity of the cross peaks obtained for TPxWt using [^{1}H, ^{15}N]-HSQC NMR experimentation, transverse relaxation optimised spectroscopy (TROSY) NMR was implemented (as described in Section 2.3.7.5.).

4.5.2. Transverse relaxation optimised spectroscopy (TROSY) NMR.

^{15}N transverse relaxation is predominantly caused by two mechanisms: dipole-dipole interaction and chemical shift anisotropy (as well as applied field in-homogeneity). Dipole-dipole coupling causes a local magnetic field between ^{15}N and ^{1}H nuclei; the rotation of a protein molecule causes this field to fluctuate in relation to the large applied magnetic field. Chemical shift anisotropy refers to the directional local field exerted by surrounding nuclei; rotation of a protein molecule relative to the applied magnetic field causes this local field to change in direction and magnitude. The "wobbling" of these local magnetic fields leads to phase coherence loss and therefore transverse relaxation causes broader signals with a weaker signal/noise ratio (Levitt *et al*, 2001).

The TROSY principle makes use of interference between dipole-dipole coupling and chemical shift anisotropy. HSQC experiments use decoupling to remove the splitting of peaks from spin-coupled nuclei and therefore give a broadened average peak. TROSY allows selection of the multiplet peak with the longest transverse relaxation time caused by interference between the two relaxation mechanisms and therefore gives rise to narrower line widths. The selection of multiplet peaks is made possible by the TROSY pulse sequence (Fernandez *et al*, 2003). TROSY NMR is most useful in situations where the rotational correlation (tumbling) time is so long as to lead to significant line broadening. Experience shows that application of the TROSY pulse sequence can be beneficial for globular proteins larger then ~30 kDa (though the full benefits are only realised when the non-exchangeable C-H protons are replaced with deuterium atoms).

Figure 4.10 shows the [^{1}H, ^{15}N]-TROSY spectrum obtained for TPxWt and a comparison with the HSQC spectrum of the same protein.

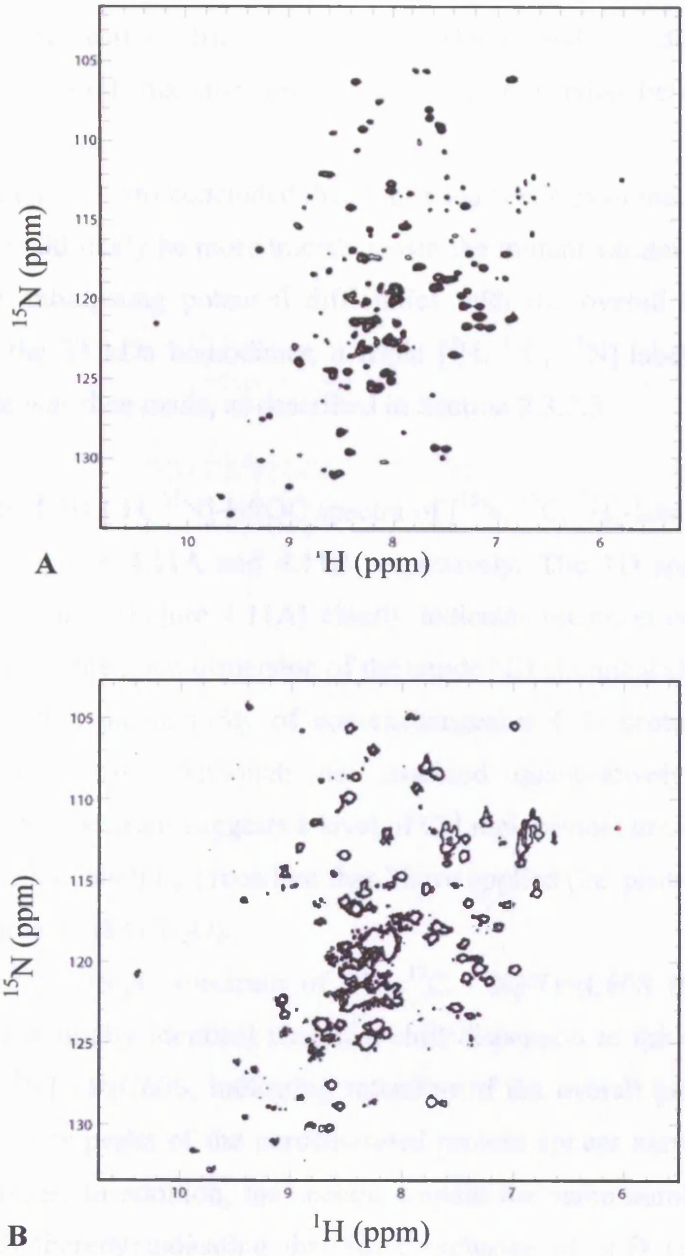


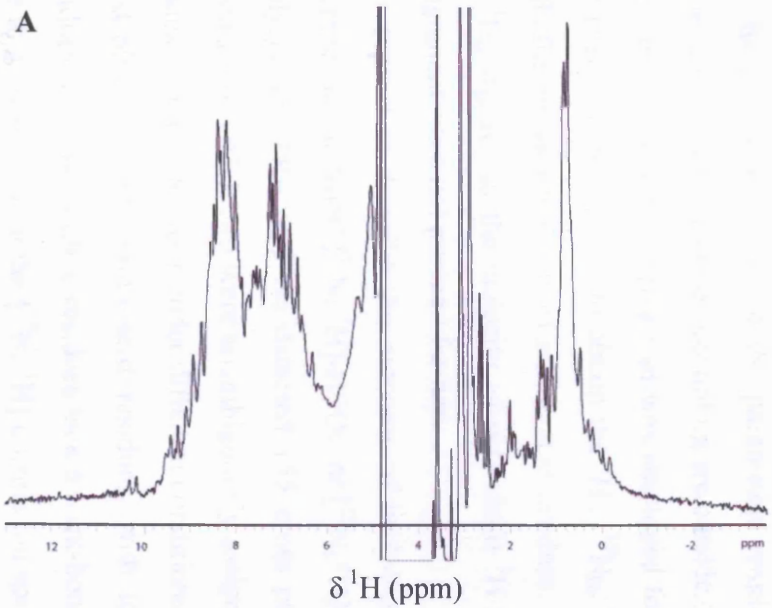
Figure 4.10: **A**: [^{1}H, ^{15}N]-Transverse relaxation optimised spectroscopy (TROSY) NMR experiment of *Mtb* [^{15}N]-labelled TPxWt at 1 mM protein concentration, measured at 25°C and performed according to the experimental parameters described in Section 2.3.7.5 on the 600 MHz spectrometer. **B**: [^{1}H, ^{15}N]-HSQC spectrum of 1 mM [^{15}N]-labelled TPxWt.

Unfortunately, no new cross peaks were identified in the use of TPxWt in the TROSY method spectrum, despite evident improvements in the lindwidths of individual cross peaks (compared to the [^{1}H, ^{15}N]-HSQC spectrum). This result perhaps indicates that the origin of the cross peak intensity variation might be a reflection of chemical exchange processes, which leads to differential line broadening. Such exchange may result from inter-conversion between alternate oligomerisation states.

From this evidence, it was concluded that continued NMR examination of the *Mtb* TPx protein would likely be more tractable with the mutant variant TPxC60S. To this end, and anticipating potential difficulties with the overall hydrodynamic behaviour of the 33 kDa homodimer, a triple [^{2}H, ^{13}C, ^{15}N]-labelled TPxC60S protein sample was then made, as described in Section 2.3.2.3.

The 1D [^{1}H] and 2D [^{1}H, ^{15}N]-HSQC spectra of [^{15}N, ^{13}C, ^{2}H]-labelled TPxC60S are shown in Figures 4.11A and 4.11B respectively. The 1D spectrum of the perdeuterated sample (Figure 4.11A) clearly indicates retention of the globular fold (evidenced by the good dispersion of the amide NH chemical shifts ~6.0-10.5 ppm) and also that the majority of non-exchangeable C-H protons have been replaced by deuterons. Although not assessed quantitatively the overall appearance of the spectrum suggests a level of CH replacement greater than 85 %, as is typical for the labelling procedure that I have applied (i.e. protonated glucose carbon-13 source in 98 % D_2O).

The 2D [^{1}H, ^{15}N]-HSQC spectrum of [^{2}H, ^{13}C, ^{15}N]-TPxC60S (Figure 4.11B) demonstrates essentially identical chemical shift dispersion to the corresponding spectrum of [^{15}N]-TPxC60S, indicating retention of the overall protein fold. As expected, the cross peaks of the perdeuterated protein appear narrower than the protonated sample. In addition, the spectra contain the same number of signals (Figure 4.11C) thereby indicating that back exchange of N-D groups to N-H groups is complete. This observation is important because artificially incomplete back exchange of the N-D groups would limit the completeness of the subsequent analysis of the spectrum.



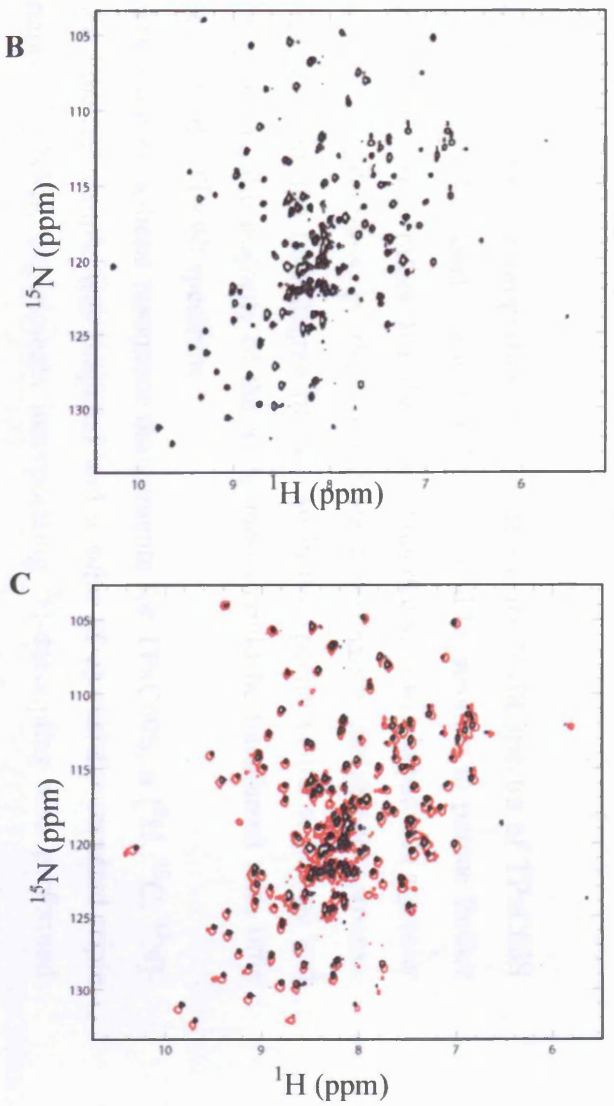


Figure 4.11: **A:** 1D [^{1}H] NMR spectrum of [^{2}H, ^{13}C, ^{15}N]-TPxC60S and **B:** [^{1}H, ^{15}N]-HSQC spectrum of [^{2}H, ^{13}C, ^{15}N]-TPxC60S. C: spectral superposition of [^{1}H, ^{15}N]-HSQC spectra of [^{2}H, ^{13}C, ^{15}N]-TPxC60S with [^{15}N]-TPxC60S, showing that both spectra match exactly in cross peak number. All spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

4.6. TPxC60S – amino acid back-bone assignment.

The apparently superior properties of the preliminary NMR spectra of TPxC60S compared to TPxWt clearly indicated that it would be sensible to pursue further analysis of the former rather than the latter. This decision would provide a greater chance of completeness in the all-important resonance assignment process. Thereafter, given the high degree of overlap in the spectra of the wild-type and mutant proteins, the majority of the assignments could be transferred with little ambiguity to the TPxWt spectrum.

To obtain the back-bone resonance assignments for TPxC60S, a [^{2}H, ^{13}C, ^{15}N]-isotope labelled sample was prepared and a suite of essentially standard triple resonance 3D NMR experiments incorporating ^{2}H-decoupling was performed. Full details of the experimental parameters employed for each experiment are listed in Section 2.3.7.6, along with a brief description of the methodology employed in the analysis of the recorded datasets (Section 2.10.3). Suffice to say here that the pulse sequences used are well-established and have been used multiple times in our laboratory. A full description of the pulse sequence design and the particular choice of the parameters employed for solvent suppression, shaped pulses and selective decoupling are outside of the scope of this thesis.

With the assignment scheme that was employed for TPxC60S, it was anticipated that is should be feasible to obtain the $^{1}H_{N}$, $^{15}N_{H}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}CO chemical shifts for the vast majority of amino acid residues. Because of the [^{2}H]-labelling, the $^{1}H_{\alpha}$ signal and the majority of side chain ^{1}H signals were absent and these assignment were not pursued further.

It is typical to describe the outcome of back-bone resonance assignment with reference to the 'root' [^{15}N, ^{1}H]-HSQC or [^{15}N, ^{1}H]-TROSY spectrum. In the final analysis, for TPxC60S we detected 155 cross peaks in the [^{15}N, ^{1}H]-TROSY spectrum of which two were unambiguously assigned to arginine side chain N_{ε}-H groups by the behaviour under different conditions of 'aliasing'. The construct of TPxC60S has 164 amino acid residues, with ten proline and no tryptophan residues. Because proline residues lack a back-bone H-N proton, they do not give rise to a cross peak in the [^{15}N, ^{1}H] correlation spectrum, though one can hope to identify the Pro C_{α} and C_{β} shifts in the 3D HNCA and HN(CA)CB experiments

through cross peaks that connect the N-H group of the succeeding residue in the chain. Taking into account that the amino terminal α–NH_3^+ group is invariably not detected in protein NMR spectra under physiological conditions, we would in principle ideally expect to observe 154 back-bone amide N-H cross peaks in the [^{15}N, ^{1}H]-HSQC spectrum of TPxC60S.

Figures 4.12 and 4.13 show an example of the cross peak connectivities that could be unambiguously established using the commonly adopted 'strip plot' representation. Figure 4.12 shows the connections observed in the HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB spectra for residues Thr144-Val147. Figure 4.13 illustrates the corresponding connections that yield the CO shifts via the HN(CO)CA and HN(COCA)CB spectra.

The complete set of connectivities that were obtained in the various triple resonance datasets are schematically illustrated in Figure 4.14A and 4.14B. Note that because of redundancy the dipeptide assigned here to Val81-Ser82 could alternatively be attributed to Val63-Ser64.

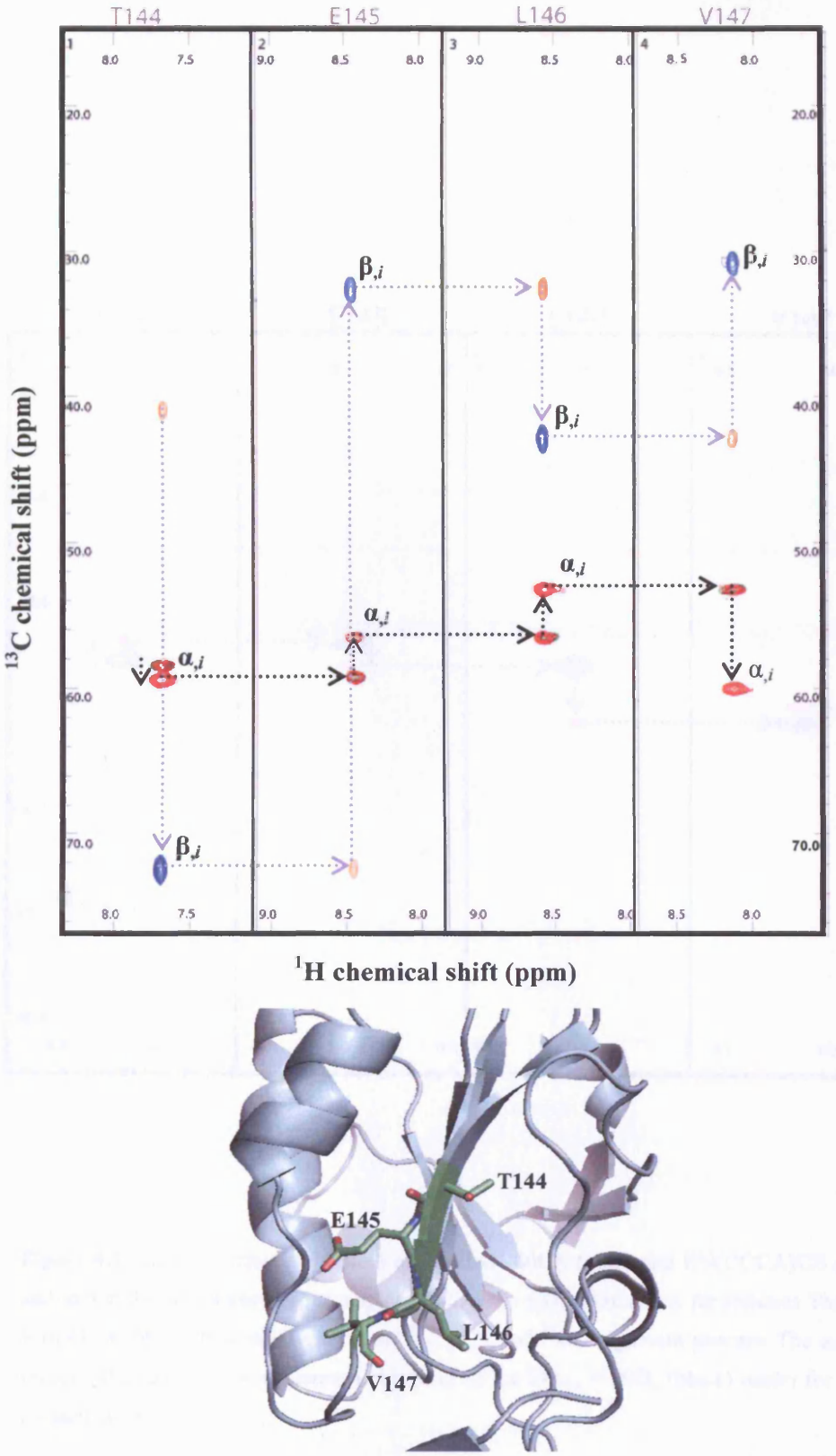


Figure 4.12: A: Four overlaid strip plots of the HNCA (red cross peaks), HN(CO)CA (red and green cross peaks), HN(CA)CB (orange cross peaks) and HN(COCA)CB (blue cross peaks) spectra showing examples of $C_{\alpha,i-1} \rightarrow C_{\alpha,i}$ and corresponding $C_{\beta,i-1} \rightarrow C_{\beta,i}$ correlations for residues T144-V147 of TPxC60S applied during the $^{1}H^{N}$, ^{15}N and ^{13}C assignment process. The arrows indicate the consecutive sequence assignment of the $C_{\alpha,i-1} \rightarrow C_{\alpha,i}$ (black) and $C_{\beta,i-1} \rightarrow C_{\beta,i}$ (purple) nuclei for each respective strip plot. **B:** A 3D structural view of TPxC60S highlighting the assigned T144-V147 residues shown as ball-and-stick models with CPK colouring.

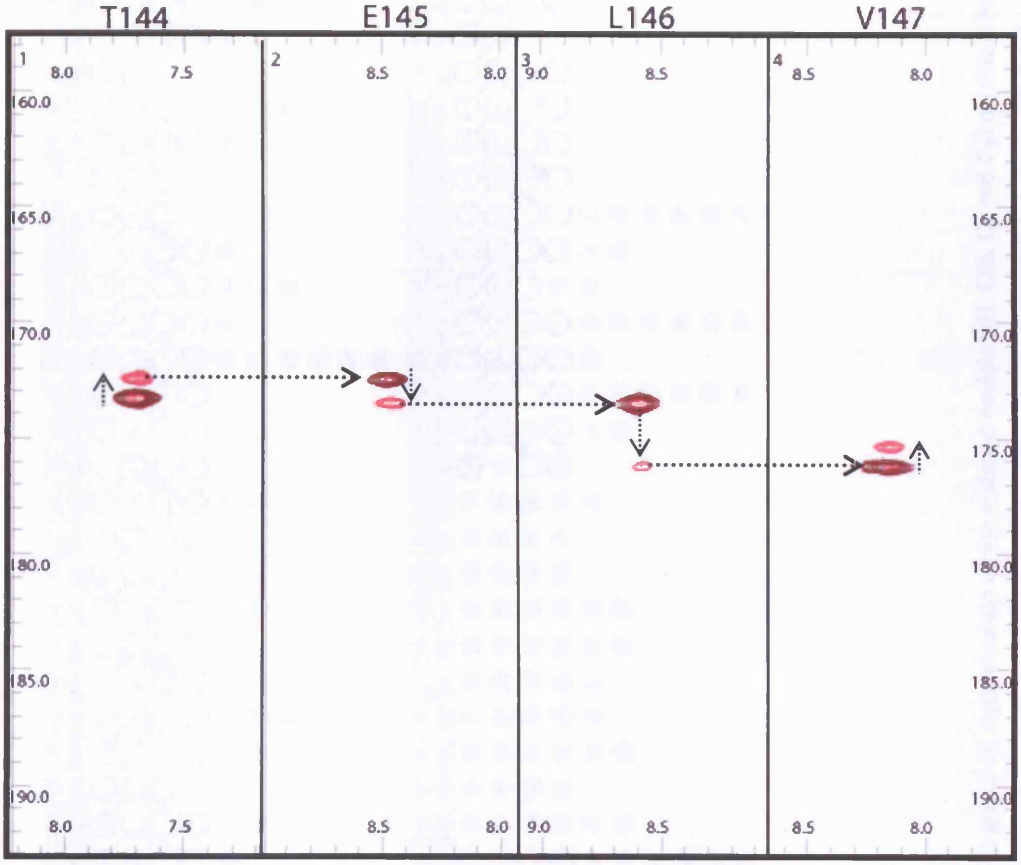
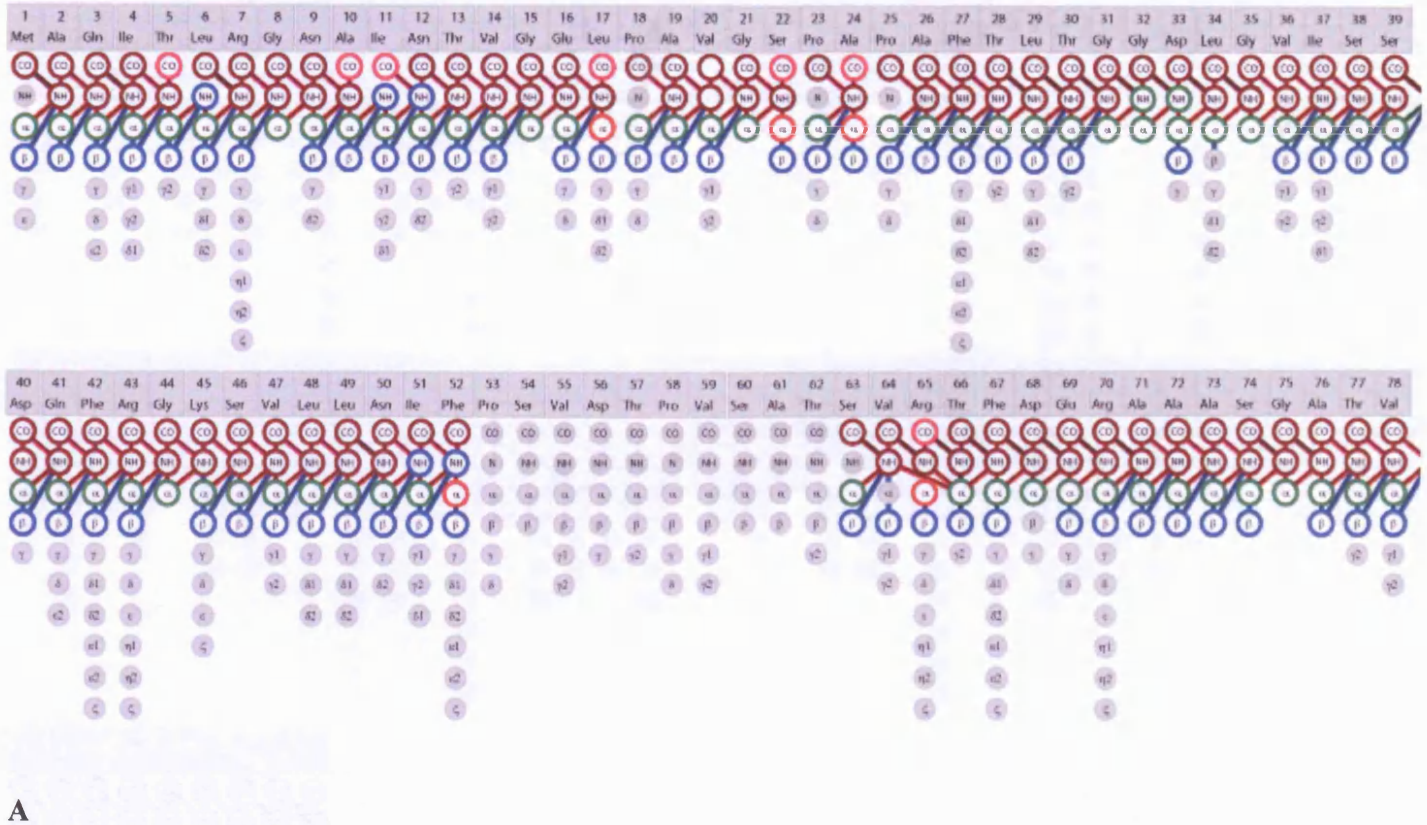
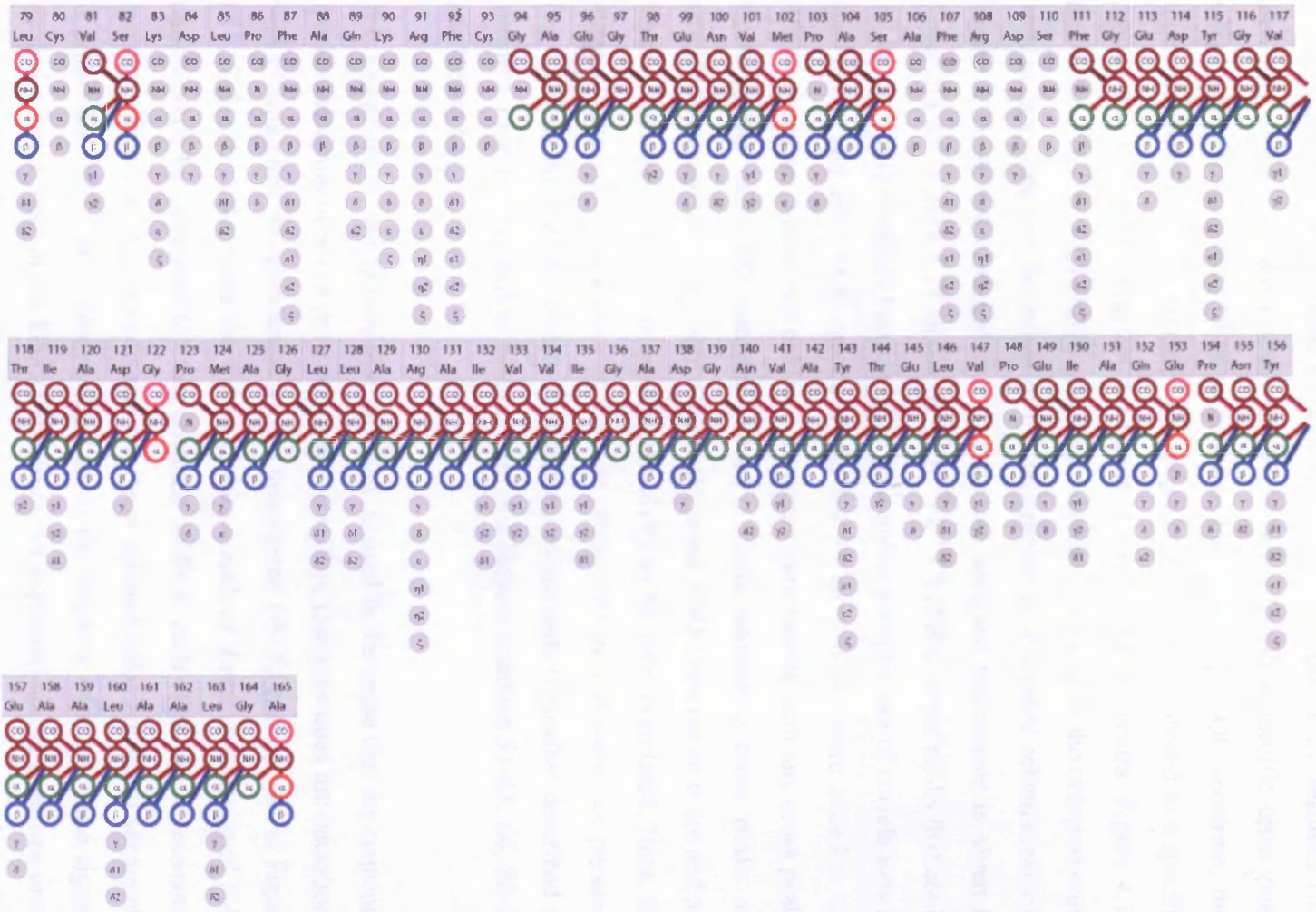


Figure 4.13: Four overlaid strip plots of the HN(CO)CA (pink) and HN(COCA)CB (pink and dark red) spectra showing examples of $CO_{,i-1} \rightarrow CO_{,i}$ correlations for residues Thr144-Val147 of TPxC60S applied during the $^{1}H^{N}$, ^{15}N and ^{13}C assignment process. The arrows indicate the consecutive sequence assignment of the $CO_{,i-1} \rightarrow CO_{,i}$ (black) nuclei for each respective strip plot.



A

Figure 14.4: A and **B** (page 89): Assignment graph – a graphical representation of the assigned nuclei (NH, CO, C_α and C_β) obtained from TPxC60S using the triple resonance experiments performed as described in section 4.6. Colour coding: HNCA (red), HNCB (blue), HN(CO)CA (green), HNCO (brown) and HN(CA)CO (pink). The colours in the assignment graph indicate which spectrum was used to establish the assignment or connection. The other nuclei, below C_β, represent the carbon and nitrogen amino acid back-bone of individual residue nuclei that were not assigned using the resonance experiments previously mentioned. Note: the chemical shift values for residues 81-82 can be equally mapped to residues 64-65. Therefore the assignment presented here is based upon the author's judgement.


79 Leu 80 Cys 81 Val 82 Ser 83 Lys 84 Asp 85 Leu 86 Pro 87 Phe 88 Ala 89 Gln 90 Lys 91 Arg 92 Phe 93 Cys 94 Gly 95 Ala 96 Glu 97 Gly 98 Thr 99 Glu 100 Asn 101 Val 102 Met 103 Pro 104 Ala 105 Ser 106 Ala 107 Phe 108 Arg 109 Asp 110 Ser 111 Phe 112 Gly 113 Glu 114 Asp 115 Tyr 116 Gly 117 Val
118 Thr 119 Ile 120 Ala 121 Asp 122 Gly 123 Pro 124 Met 125 Ala 126 Gly 127 Leu 128 Leu 129 Ala 130 Arg 131 Ala 132 Ile 133 Val 134 Val 135 Ile 136 Gly 137 Ala 138 Asp 139 Gly 140 Asn 141 Val 142 Ala 143 Tyr 144 Thr 145 Glu 146 Leu 147 Val 148 Pro 149 Glu 150 Ile 151 Ala 152 Gln 153 Glu 154 Pro 155 Asn 156 Tyr
157 Glu 158 Ala 159 Ala 160 Leu 161 Ala 162 Ala 163 Leu 164 Gly 165 Ala

B

During the execution of the TPxC60S assignment, a number of gaps in the cross peak connectivity were identified and in the end a total of 27 residues (not including proline residues) could not be associated with a specific cross peak assignment. Against that, 18 cross peaks in the [^{15}N, ^{1}H]-HSQC spectrum, that appear to derive from backbone N-H groups, could not be attributed to a specific residue due to absent connectivities in the triple resonance spectra. Figure 4.15 shows the [^{2}H, ^{15}N, ^{13}C]-HSQC spectrum labelled according to the obtained cross peak assignments and the extent of the assignment is illustrated schematically in Figure 4.16. A full chemical shift list for the assigned resonances is given in Appendix A6. Figure 4.17 shows the [^{2}H, ^{15}N, ^{13}C]-HSQC cross peaks that could not be explicitly assigned because of an incomplete complement of correlations in the HNCA and HNCACB spectra. A further 55 cross peaks were added to the HSQC spectrum that do not exhibit convincing correlations with any cross peaks found in the other 3D resonance spectra. These interesting cross peaks are provided because they exist above the background HSQC resonance noise and are features that may lack the necessary sensitivity to be fully correlated. Note, the sequence specific assignments of the side chain NH_2 groups were not pursued. The residues without back-bone N-H group assignments (hereafter described as 'missing') are highlighted in Figure 4.16, and include residues 53-63, 66, 80-93 and 105-111.

It is important to clarify here that 'missing' is used in the sense that the sequence specific assignment is not (yet) possible – for 18 of these residues the back-bone N-H groups must be represented by the unassigned HSQC cross peaks in Figure 4.17. Table 4.1 documents the extent to which each of these 18 unassigned N-H signals could be correlated to other chemical shifts in each of the triple resonance spectra. (Chemical shift coordinates for these connectivities are listed separately in Appendix A7.1.) It is noteworthy that in the majority of cases these signals yielded a cross peak in the HNCO spectrum (as expected) and at least one cross peak in the HNCA spectrum (presumably corresponding to the intra-residue C_α).

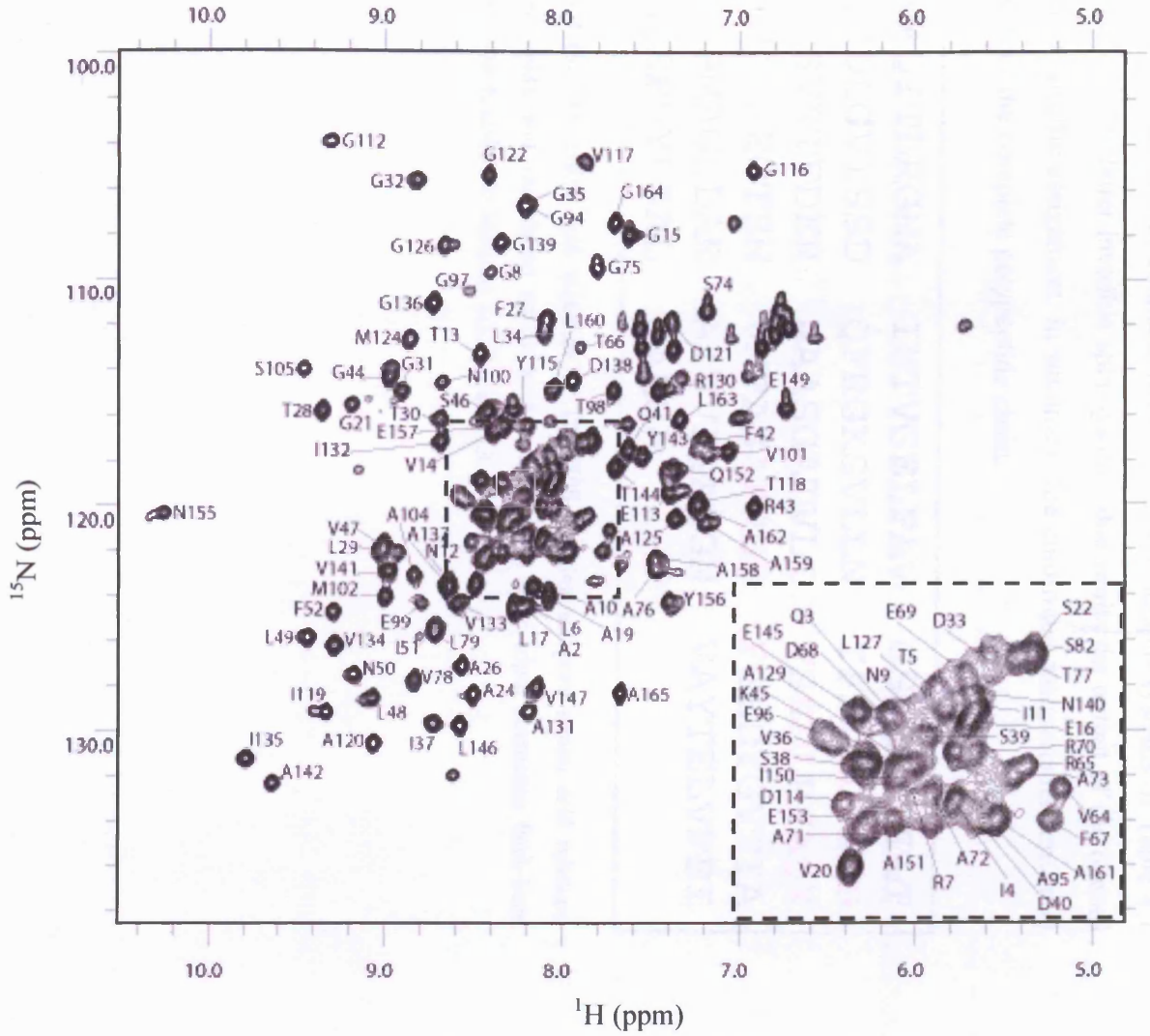


Figure 4.15: Assigned 2D ^{13}C-decoupled [^{1}H, ^{15}N]-HSQC spectrum of 1 mM (pH 7.48) [^{2}H, ^{13}C, ^{15}N]-labelled TPxC60S recorded on a 500 MHz Varian UNITYplus spectrometer at 298 K.

However in many instances correlations to a second C_α chemical shift or any C_β shift were apparently absent. Database-derived predictions of the amino acid type, based on the limited chemical shifts available, yield only low confidence results (Table 4.1). However even accounting for these visible yet unassigned cross peaks it is apparent then that the inventory of backbone N-H signals is incomplete; apparently nine residues do not yield an identifiable HSQC cross peak. Taken together, it is the loss of connectivities for the visible spin systems of Table 4.1 coupled with the latter invisible spin systems that limits the extent of the overall sequence specific assignment. In summary, the confirmed assignments constitute ca. 80 % of the complete polypeptide chain.

MAQITLRGNA	INTVGEL**P**AV	GS**P**A**P**AFTLT
GGDLGVISSD	QFRGKSVLLN	IF**P**SVDT**P**VS
ATSVRTFDER	AAASGATVLC	VSKDL**P**FAQK
RFCGAEGTEN	VM**P**ASAFRDS	FGEDYGVTIA
DG**P**MAGLLAR	AIVVIGADGN	VAYTELV**P**EI
AQE**P**NYEAAL	AALGA	

Figure 4.16: The amino acid sequence of TPxC60S showing assigned amino acid residues, prolines (bold), and unassigned amino acid residues (red) after triple resonance back-bone assignment as described in Sections 2.2.13.6 and 2.3.

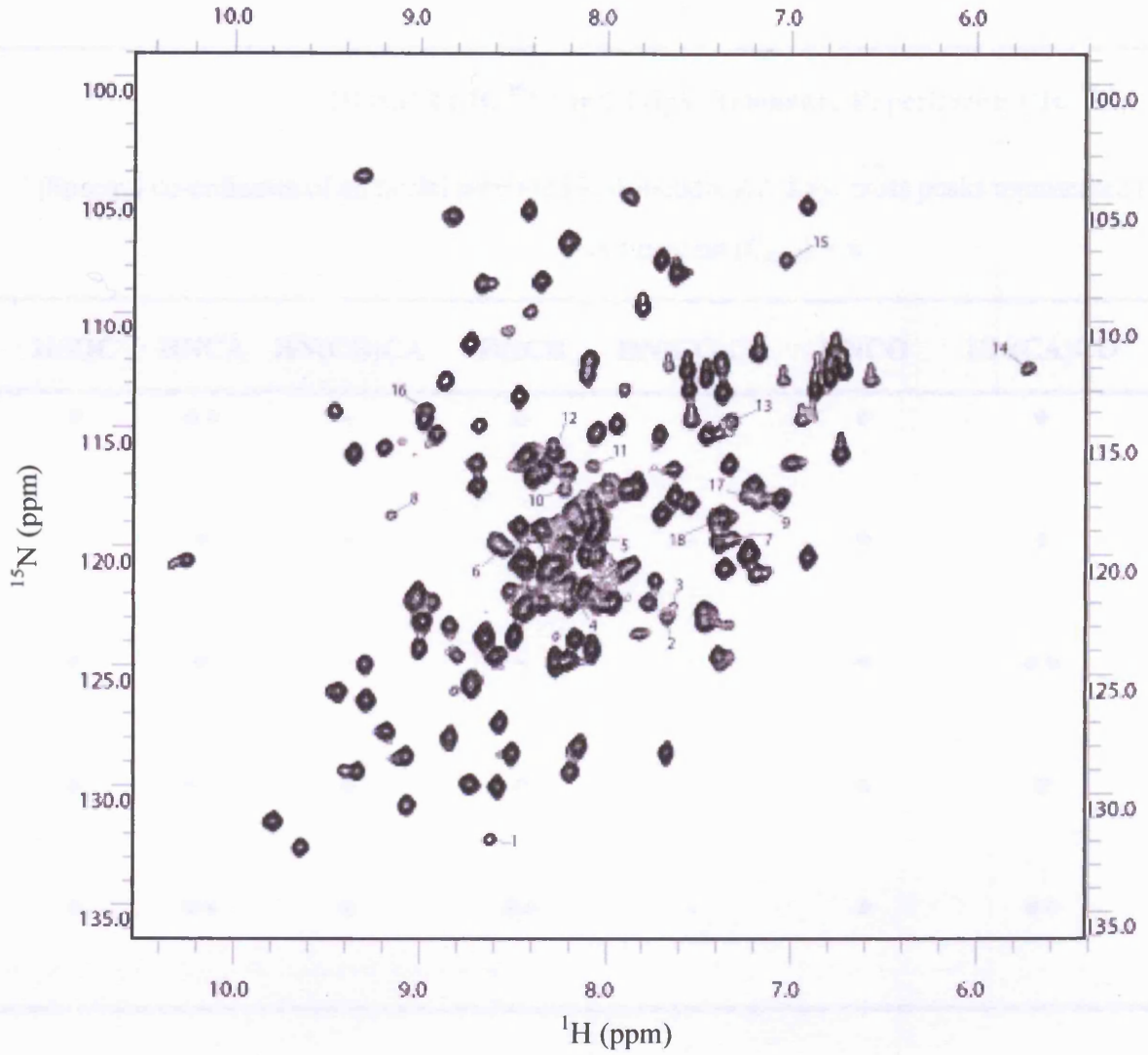


Figure 4.17: 2D ^{13}C-decoupled [^{1}H, ^{15}N]-HSQC spectrum of 1 mM (pH 7.48) [^{2}H, ^{13}C, ^{15}N]-labelled TPxC60S recorded on a 500 MHz Varian UNITYplus spectrometer at 298 K showing 18 cross peaks (denoted 1-18) that exhibit chemical shift correlations in two or more of the triple resonance experiments performed, but could not be assigned.

2D NMR (^{1}H, ^{15}N) and Triple Resonance Experiments (^{1}H, ^{13}C)

(Spectral co-ordinates of all nuclei presented in Appendix A7. Key: cross peaks represented as ($C_{\alpha,i}$) = ● second co-ordinate set ($C_{\alpha,i-1}$) = ●

Spin System	HSQC	HNCA	HN(CO)CA	HNCB	HN(CO)CB	HNCO	HN(CA)CO	Predicted residue
1	●	●●	-	●	-	●	●	Met 9.4%
2	●	●	-	●	-	●	●	Asp 15.7%
3	●	●	-	-	-	●	●●	Tyr 5.2%
4	●	●●	●	●	-	●	●	Leu 11.2%
5	●	●●	●	●●	-	●	●●	Glu 12.9%

6	●	●●	●	-	-	●	●●	**Glu 12.0%**
7	●	●	-	-	-	●	-	**Ala 12.6%**
8	●	●	-	-	-	●	-	**Ser 10.9%**
9	●	●●	-	●	-	●	●	-
10	●	●	-	-	-	-	●	-
11	●	●	-	-	-	●	-	-
12	●	●	-	●	-	●	-	-
13	●	●●	●	-	-	●	●	-

14	●	-	-	-	-	●	-	-
15	●	●	-	-	-	●	●	-
16	●	●●	●	●	-	●	●	-
17	●	●	-	-	-	●	-	-
18	●	●	-	-	-	-	-	-

Table 4.1: Spin systems identified from unassigned cross peaks found in the TPxC60S HSQC spectrum (Figure 4.17) and displayed to indicate the presence of $C_{\alpha,i}$ (black) and $\mathbf{C}_{\alpha,i-1}$ (purple) cross peaks as coloured dots for each respective 2D NMR and triple resonance experiment. The co-ordinates for all available cross peaks are provided in Appendix X.

4.7. Prediction of secondary structure elements from ^{13}C chemical shift data.

The deviation of chemical shift values from putative random coil values is often used in prediction and assessment of the secondary structural elements within the folded protein (Wishart and Sykes, 1994; Wishart and Bigam, 1995a and 1995b). Thus it has been shown that there is a statistical bias of the 'secondary' shift of C_α and C_β resonances that correlates with the local secondary structure (Spera and Bax, 1991). In β-strands and extended structures the C_α nuclei experience an up-field chemical shift, whilst in α-helices they experience a down-field chemical shift. Conversely, the deviation of C_β nuclei chemical shifts from random coil values displays the opposite correlation with local secondary structure.

It is of interest to examine the predicted secondary structure composition of TPxC60S from the chemical shift assignments that have been obtained (displayed in Appendix A6). To do this it is useful to compute the quantity $\Delta C_\alpha - \Delta C_\beta$, where ΔC_α represents the difference between the observed chemical shift and the random coil value. The quantity $\Delta C_\alpha - \Delta C_\beta$ combines the influences of the secondary structure on the chemical shifts and makes this a more sensitive predictor of the secondary structure than either of ΔC_α and ΔC_β alone. Typically one would plot the value of $(\Delta C_\alpha - \Delta C_\beta)$ smoothed $((\Delta C_\alpha - \Delta C_\beta)_{smoothed,i})$ over a window of i neighbouring residues, usually $i = 3$ (Salzmann *et al*, 2000):

$$(\Delta C_\alpha - \Delta C_\beta)_{smoothed,i} = (\Delta C_{\alpha,i+1} + \Delta C_{\alpha,i} + \Delta C_{\alpha,i-1} - \Delta C_{\beta,i+1} - \Delta C_{\beta,i} - \Delta C_{\beta,i-1})/3$$

The forgoing holds for $^{13}C_\alpha$ and $^{13}C_\beta$ shifts obtained from proteins that have not been subject to replacement of non-exchangeable protons with deuterons. It is recognised that in the case of a perdeuterated protein the ^{13}C shifts obtained are systematically perturbed because of substituent [^2H]-isotope effects. Standard correction terms for the [^2H]-isotope effects have been suggested and are listed in Table 4.2, and can be quiet large compared to the influence of secondary structure of the ^{13}C shift. Thus application of these corrections to the observed ^{13}C shifts is required prior to the computation of $\Delta C_\alpha - \Delta C_\beta$.

Residue[a]	$C_\alpha$corr[b]	$C_\beta$corr[b]
G	-0.39	-
N, D, S, H, F, W, Y, C	-0.55	-0.77
K, R, P	-0.69	-1.11
Q, E, M	-0.69	-0.97
A	-0.68	-1.00
I	-0.77	-1.28
L	-0.62	-1.26
T	-0.63	-0.81
V	-0.84	-1.20

Table 4.2: Correction values in ppm applied to calculate for the ^{2}H-isotope affect upon C_α and C_β chemical shift values for a random coil protein structure. [a]A one letter code is used to describe the residue in question. [b]$C_{\alpha/\beta}$corr stands for the chemical shift correction value used to account for deuteration of random coil residue nuclei.

The [^{2}H]-isotope corrected values of ΔC_α - ΔC_β have been obtained for the assigned TPxC60S residues and a plot of $(\Delta C_\alpha - \Delta C_\beta)_{smoothed,i}$ is shown in Figure 4.18C.

Salzmann *et al.* define the specific cut-off point for identifying secondary structure elements as being above and below a range of +/-1.4 ppm (Salzmann *et al*, 2000). Therefore, when ΔC_α - ΔC_β > 1.49 ppm an α-helix is expected and when ΔC_α - ΔC_β < 1.49 the likelihood of a β-strand being present in the secondary structure is enhanced. The β-strand content is identified as being negative in the correlated plot as a result of highly localised hydrogen bonding.

Figure 4.18A shows the consensus secondary structure composition of TPxC60S as defined by the chemical shift data obtained from the triple resonance experiments. The comparison of this $(\Delta C_\alpha - \Delta C_\beta)_{smoothed,i}$ data to the same secondary structure composition and location obtained from the X-ray crystal structure of TPxC60S (PDB: 1Y25) (Figure 4.18B) is well correlated.

In general the content and location of α-helices in TPxC60S is well preserved, following the triple resonance experiments, but with some major α-helical content discrepancies proving vitally important in completing the back-bone assignment. The most obvious irregularity obtained is that of α-helix α2 (containing the C93,

residues 85-97), which is completely missing resonance data. Other α-helices have incomplete resonance assignments, such as α-helix α3 (residues 110-115), which is missing N-terminal residues 110-112 inclusive. The TPxC60S α-helix α1 (residues 63-75) is also shorter, by six N-terminal residues, when compared to the X-ray crystal structure version (residues 57-75). To account for the complete α-helical complement of TPxC60S; α-helix α4 exhibits a complete resonance assignment (residues 155-164), correlating well with the X-ray crystal structure α4.

The β-strand content and location in the chemical shift data obtained for TPxC60S is very well correlated to the X-ray crystal structure, with all β-strands present and situated in virtually identical locations baring small deviations. These minor shift deviations, together with the missing α-helical regions are presumably associated with the distinct absence of resonance data, as shown by the bracketed regions in Figure 4.18A.

In summary, the quality of the back-bone resonance assignment carried out using TPxC60S has been confirmed. This conclusion is based upon the comparative agreement reached between the predictions of the secondary structure elements, by the NMR ^{13}C chemical shift analysis, with the secondary structure elements in the X-ray crystal structure of TPxC60S. The deviations observed in some α-helical and β-strand elements are presumed to be a consequence of the three defined regions of missing resonances and would be improved once the complete assignment is achieved.

As a comparative to the manual approach in obtaining a predicted structure of TPxC60S, the automated structural prediction program, designed by Berjanskii, Neal and Wishart, 2006, known as PREDITOR (PREDIction of TORsion angles from chemical shift and homology) was used. PREDITOR is able to accurately predict a large number of protein torsion angles (φ, ψ, ω, χ) using either ^{1}H, ^{13}C or ^{15}N chemical shift assignments or protein sequence (alone) as input. The data output obtained (not shown) using the TPxC60S chemical shift resonances from PREDITOR proved to be highly comparable, with all α-helical and β-strand elements being comparable in both number and location.

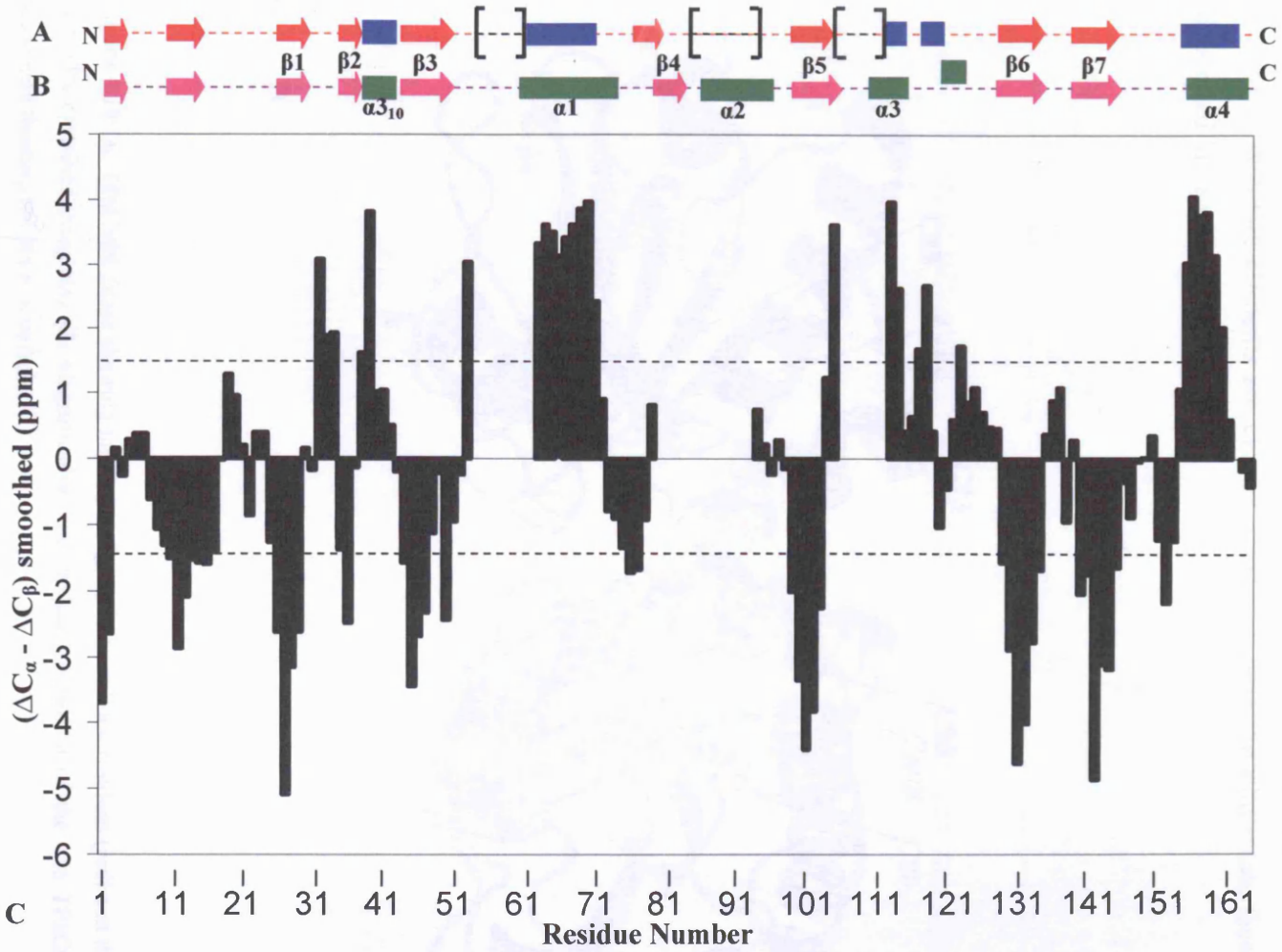
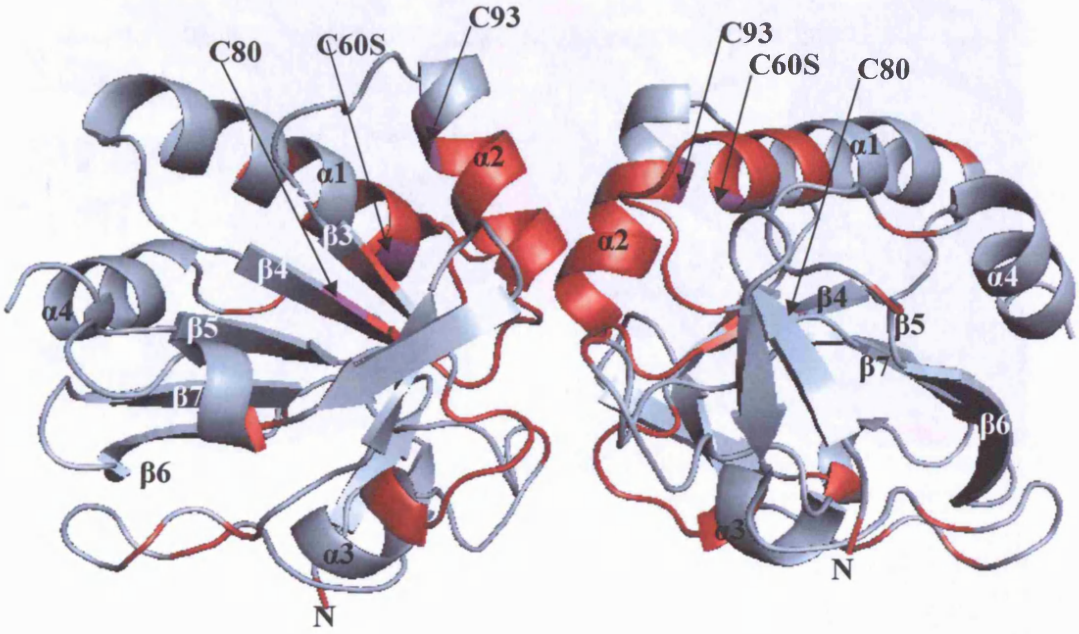


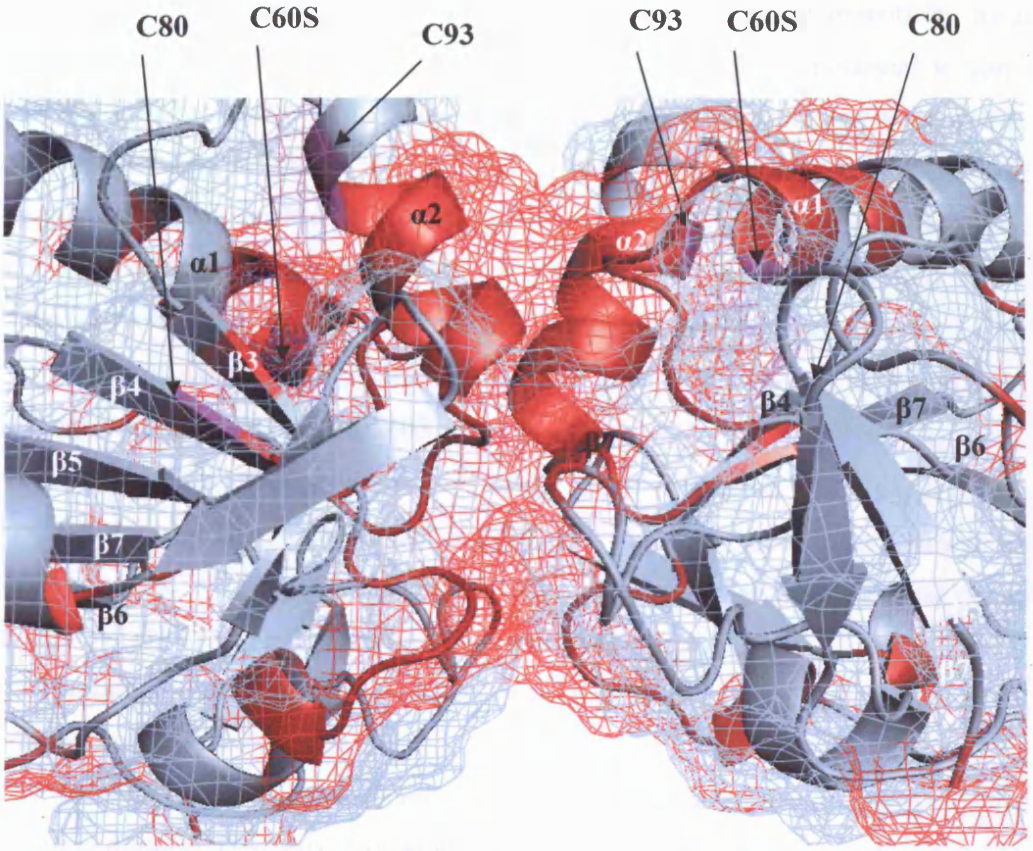
Figure 4.18: Prediction of secondary structure of TPxC60S from ^{13}C chemical shift data. **A:** Consensus schematic representation of predicted TPxC60S α-helical (blue rectangle) and β-strand (red block arrow) location using NMR-derived chemical shift data. Bracketed regions indicate sections of the amino acid sequence of TPxC60S that did not produce chemical shift values. **B:** Consensus schematic representation of TPxC60S (PDB: 1Y25) α-helical (green rectangle) and β-strand (violet block arrow) location obtained from X-ray crystallographic data. **C:** Predicted secondary structure histogram displaying chemical shift data as a function of residue number. The dashed line indicates the 1.4 ppm threshold value for the identification of secondary structure elements. Values higher or lower than 1.4 ppm or -1.4 ppm indicate the presence of α-helical or β-strand regions respectively.

By mapping the missing residues to the PDB deposited crystal structure of TPxC60S (PDB: 1Y25) it can be seen that missing assignments cluster predominantly around the TPxC60S C_P-equivalent Ser60 residue on alpha helix α1 and Cys93 also on alpha helix α1 (C_R) as well as the non-catalytically essential Cys80 on beta-strand β4 (Figure 4.19A) and on several loop regions. These secondary structure elements are closely associated with the dimer interface, as illustrated in Figure 4.19B.



A

Figure 4.19: **A**: TPxC60S dimer showing missing assignment back-bone residues (red) and the 3 *Mtb* TPx cysteines (magenta). **B**: Magnification and surface interaction of the two TPxC60S monomers flanking the dimer interface.



B

4.8. Discussion.

The comparison of ASEC profiles and [^{15}N, ^{1}H]-HSQC NMR spectra of TPxWt and TPxC60S proteins suggest that whilst the latter behaves essentially as a mono-disperse homodimer, the former shows evidence of the potential to form higher order oligomeric forms. Thus the wild-type protein yields an asymmetric, concentration dependent ASEC elution profile, whereas the mutant gives a symmetric elution peak at a volume consistent with a homodimer. In addition, TPxWt gives poorer quality NMR spectra (fewer cross peaks and more heterogeneous intensity distribution) when compared to the mutant, even at 1 mM concentration. Given these characteristics, we have focussed our chemical shift assignment efforts on the TPxC60S variant.

Application of an essentially standard set of triple resonance NMR experiments provided high quality datasets that allowed the sequence specific assignment of the majority of N-H cross peaks. However incompleteness in both the total count of HSQC/TROSY N-H correlations, and absences of H-N-C_α and H-N-C_β correlations in the triple resonance data mean that 27 residues from a total of 165 (~20 %) could not be assigned.

The "missing" cross peak resonances appear to be associated with the region of the homodimeric TPxC60S crystal structure that encompasses the active site and the inter-protomer contact surface (Figure 4.19).

Given that the N-H cross peak absences cannot be attributed to incomplete back-exchange of N-D deuterons for protons (see above), explanation for the non-uniform completeness of the inter-nuclear correlations in the triple resonance spectra must appeal to other factor(s). The intensity of a given cross peak in these spectra is a complex function of the relaxation properties of the nuclei involved in the coherence transfer pathway, and ultimately upon the local dynamics of the polypeptide chain. Line broadening that could lead to loss of intensity is commonly associated with exchange between two (or more) molecular substates on a timescale that is of the order of the difference between the relevant substate chemical shifts, and is usually on the micro- to milli-second timescale. Figure 4.20 illustrates how this might arise from the classic case of two-site exchange and the effect on the observe NMR resonances.

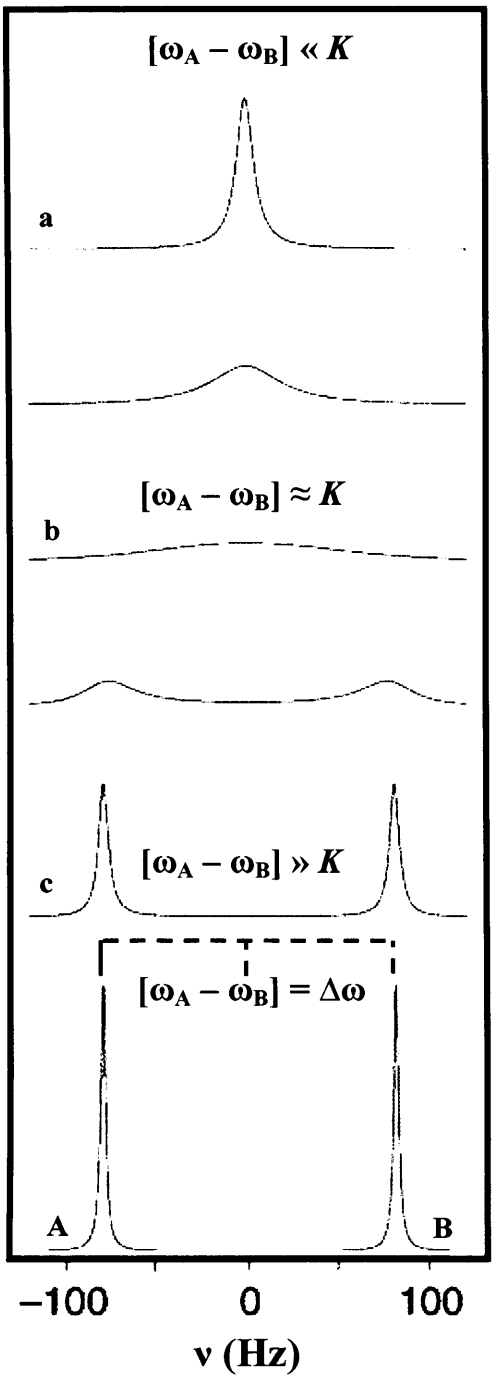


Figure 4.20: An illustration of two species **A** and **B** and the corresponding resonance frequencies obtained during chemical exchange. In the absence of chemical exchange these species produce two resonance lines at the characteristic Lamor precession frequencies ω_A and ω_B. If the equilibrium concentrations of A and B are equal $K_{A\rightarrow B} = K_{B\rightarrow A} = K$. Fast, intermediate and slow exchange between the two species is illustrated with lineshapes a, b and c respectively together with equations indicating the relationship between the chemical difference $\Delta\omega$ and the exchange rate constant K.

Thus the expectation is that for exchange between two environments on the intermediate timescale line broadening leads to potentially 'vanishing' peak intensities.

The non-uniform intensities of TPxC60S cross-peaks therefore presumably reflect the differential effects of internal motions along the polypeptide chain on this intermediate timescale. In multidimensional spectra the broadening of a cross peak may result from one of more of the correlated resonances, but unravelling the precise origin of 'absent' NMR intensity is, perhaps by definition, intrinsically difficult.

The distribution of the "missing" assignment residues in the crystal structure of TPxC60S points to the effects of these putative motions being clustered around the active site and inter-domain interface of the homodimer. One can speculate that the implied reorganisation of the structure may be related to exchange between substates that are either involved in the catalytic mechanism, remodelling of the dimer interface, or both. It is intriguing to consider that the presumed mechanism of *Mtb* TPx, based upon the known structures of oxidised and reduced homologues (see Section 3.11) directly predicts that the structure of the protein must change significantly along the reaction coordinate. Thus it seems clear that in the reduced catalytically competent state, the C_P and C_R residues are located at a distance from one another, but in the oxidized state the same two residues form an intra-monomer disulphide bond, having been brought into proximity by local helix-to-coil rearrangements. As Figure 4.21 suggests conversion between these open and closed states could either occur as a result of an intrinsic property of the isolated reduced polypeptide (i.e. in the absence of a substrate); be induced upon formation of the protein-peroxysubstrate complex that is formed at an early stage of the reaction mechanism; or be induced by the formation of the sulfenic acid intermediate following contact with the substrate.

Presumably resolution of the sulfenic acid intermediate proceeds by nucleophilic attack of the C_R thiol upon the C_P S=O group and by definition requires close approach of the C_R and C_P residues.

The pattern of the spectral characteristics encountered in the analysis of the NMR spectrum of TPxC60S implies a segmental dynamic profile, where one or more regions may well be undergoing conformational exchange leading to line broadening. Given that the line broadening effects appear to cluster around the

active site and the interprotomer interface, it is tempting to speculate that the spectrum may reveal constitutive internal motion on the intermediate timescale that could be consistent with the scenario depicted in panel A of Figure 4.20. In particular since these motions are suggested to involve helix-to-coil transitions, and it is well established that $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ shifts are sensitive to the local secondary structure, it is not impossible that such changes would result in selective broadening of the $^{13}C_{\alpha}$ or $^{13}C_{\beta}$ resonances and thereby lead to absences of correlations in HNCA and HNCACB spectra, even when there is little evidence of line broadening for the corresponding N-H moieties. Confirmation of this hypothesis would require more in-depth characterisation of the relevant relaxation rate constants, which presently is beyond the scope of this thesis.

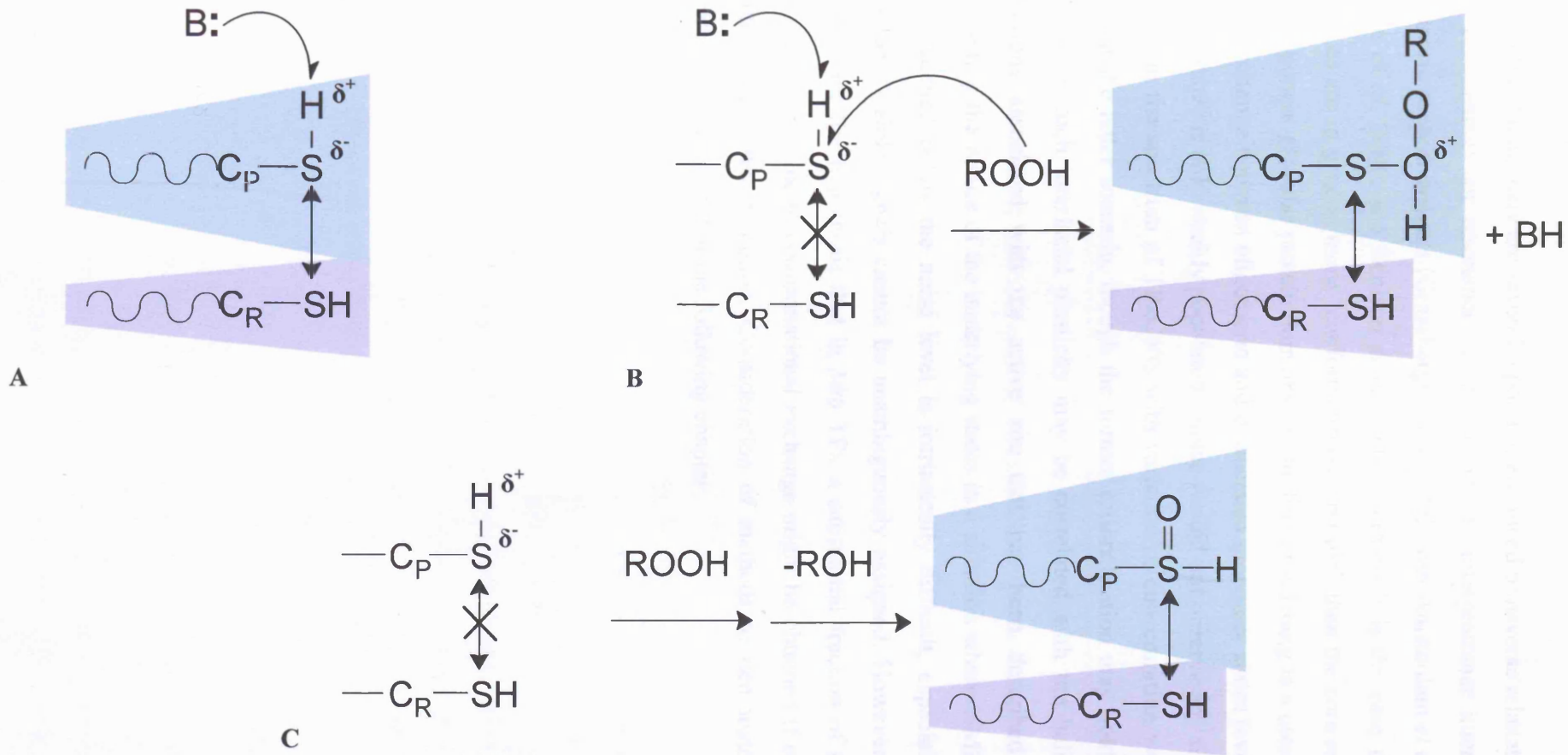


Figure 4.21: The mechanism of TPx proteins implies a conformation change that requires large changes in the distance between C_P and C_R side chains. Motion of both C_R and C_P residues accompanied by helix-to-coil back-bone conformational changes seems likely based upon the available crystal structures. Oxidized TPx, wherein the C_P-C_R distance is constrained by disulphide bond formation, occurs only after the formation of the sulfenic acid intermediate. The diagram illustrates three scenarios: **A**, motion of the C_R and C_P residues that allows close approach spontaneously in the absence of substrate; **B**, motion of C_R and C_P occurs only after nucleophilic attack of the C_P thiol group on the substrate peroxide to form a covalent protein:substrate complex; and **C**, motion of C_R and C_P occurs only after decomposition of the protein:substrate complex to form the C_P sulfenic acid intermediate. Squiggly lines, magenta and blue colours highlight states in which the protein internal motion allows close approach of C_R and C_P. The relative extent of predicted motion is suggested by the area of the colour. The double headed arrow emphasizes that this motion is to- and fro- (in exchange) between conformational substates. 'B:' indicates the base that putatively polarises the C_P thiol group prior to attack upon the peroxy substrate ROOH.

On the other hand, there are several reports of enhanced transverse relaxation rates (line broadening) for resonances corresponding to interprotomer interfaces in oligomeric protein systems (Grassberger *et al*, 1992; van Amsterdam *et al*, 2001; Mohan *et al*, 2006; and Benison *et al*, 2006). Perhaps it is the case that such interfaces are in general more "conformationally plastic" than the core regions of the component globular protein domains, or the line broadening is a consequence of the exchange between oligomeric and dissociated monomer states (even if one of these states is only weakly populated). Since we did not observe any significant changes in the spectrum of TPxC60S with variation of concentration we tend to discount the latter scenario, though the former rationalisation may hold true (in which case such interfacial plasticity may be correlated with the helix-to-coil transitions associated with the active site that has been described above). Unravelling the nature of the underlying states in a situation where NMR signals are broadened below the noise level is intrinsically difficult, especially when associated visible signals cannot be unambiguously assigned. However indirect evidence for the hypothesis that in *Mtb* TPx a substantial fraction of the cross peaks are absent due to conformational exchange might be obtained if one could "quench" the internal motion. Consideration of methods to demonstrate such quenching is the subject of the following chapter.

Chapter 5

Results: Unraveling the mystery of the missing TPxC60S HSQC backbone assignments and exploring the ligand binding capacity of *Mtb* TPx

5.1. Chapter summary.

In recording a series of NMR spectra, for example in obtaining resonance assignments, it is imperative to maintain the sample conditions and other external parameters within very narrow ranges of variation. This is because the spectral parameters (chemical shifts, resonance linewidths and peak intensities) can be very sensitive to such variation. On the other hand systematic variation of the sample conditions can be used to deliberately perturb the spectrum, thereby exploiting or exploring this dependence. For example one can straightforwardly monitor the dependence of the protein tumbling time (hydrodynamic status), foldedness (local structural stability), and activity (ligand binding) using the NMR spectrum as a function of the pH, protein concentration, experimental running temperature or sample buffer components.

Initially, the back-bone assignment process applied using TPxC60S yielded assignments for 128 amino acid backbone N-H groups (out of a total of a potemtial 155 non-proline residues). However for 27 residues assignment proved impossible, in spite of the presence of 18 cross peaks in the HSQC spectrum that

could not be connected (Chapter 4). In an attempt to find these "missing" residues, a series of experimental parameters was carefully varied, combined with monitoring of the TPxC60S NMR spectrum in the hope of finding new or different features. The parameters varied include: sample pH, experimental running temperature and the application of ligands implicated in interactions with *Mtb* TPx. TPxC60S was shown to be particularly pH sensitive, with the identification of several 'new' cross peaks in the HSQC spectra exhibiting one or more correlations in the triple resonance experiments. However, with seven 'new' cross peaks identified, none could be unambiguously linked to the currently unassigned sections of the TPxC60S back-bone. The temperature variations revealed *Mtb* TPx to be particularly thermally stable upto temperatures of 95 ^{o}C but with little success in revealing 'new' cross peaks in the concurrent HSQC spectra taken. The use of ligands provided mixed results, with some having greater effects than others, with neither proving substantially bulky enough to ellicit quenching of the fluctuating altered conformation. TPxWt was also probed to determine the reason for the poor intensity and line-broadened cross peaks in the HSQC NMR spectra produced by examining the possibility of intra (intrinsic) or inter chain (external multimer) association by using excess DTT. The conclusion of which describes TPxWt as being governed by mechanistically driven internal motion, closely modulated by residue Cys60.

5.2. TPxC60S melting temperature (T_m) assessed using circular dichroism (CD) spectroscopy.

As a prelude to NMR investigations of the effect of temperature on the NMR spectrum of TPxC60S we set out to characterise the dependence of the overall folding status of TPxC60S upon the temperature using the convenient means of circular dichroism spectropolarimetry (CD). From the perspective of the NMR experiment, which potentially risks relatively large amounts of sample to irreversible unfolding and aggregation, it is useful to know the limiting temperature range within which the protein maintains its natively folded state.

The method applied during the CD data collection and subsequent processing is described in Section 2.3.6 and the results are shown in Figure 5.1. The CD data obtained reveals TPxC60S as being particularly thermo-stable, apparently still giving rise to a measurable elipticity in the range 205-240 nm at the highest temperature sampled, 95 °C (Figure 5.1A). Upon reversing the temperature back down to 20 °C, TPxC60S apparently regains its original spectral form. The interchange of conformations upon heating up to 95 °C and back down to 20 °C appears essentially completely reversible. In between these two temperatures the magnitude of the CD spectrum clearly follows a classical sigmoidal trajectory characteristic of a cooperative two state transition with a mid-point melting temperature (T_m) around 63 +/- 1 °C (average of fitted heating and cooling curve fits; Figure 5.1B). Taken together the results are consistent with a conversion of natively folded TPxC60S to a higher energy, partially folded state, possibly a molten globule-like conformation.

The prediction from these measurements is that one can anticipate that the NMR spectrum of TPxC60S should be dominated by the natively folded state at temperatures up to 55 °C where the fraction of partially folded protein should be less than 5 %. Above this temperature there is a risk that the NMR spectrum will be complicated by the presence of a substantial fraction of the partially folded state.

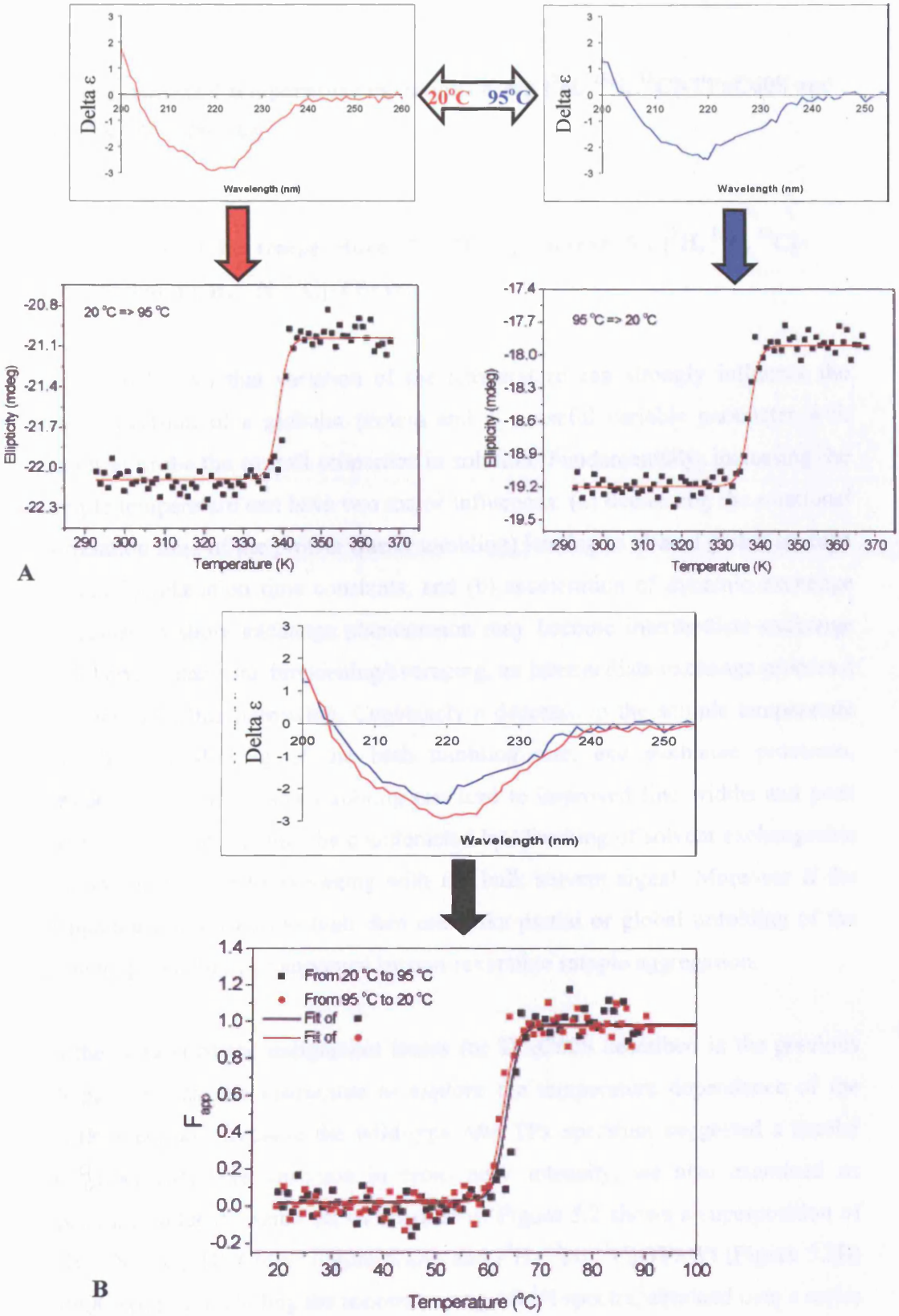


Figure 5.1: A: CD spectra of TPxC60S, before (red – 20°C) and after (blue – 95°C) temperature melt, with the corresponding data-plot fits for the heating and cooling down process. **B**: CD spectral overlay of TPxC60S spectra from **A**. The data-plot fit is the overlaid normalised data-plots from **A**, presented as the apparent fraction (F_{app}) following the subtraction of the baseline.

5.3. Experimental temperature variations using [^{2}H, ^{15}N, ^{13}C]-TPxC60S and [^{2}H, ^{15}N, ^{13}C]-TPxWt.

5.3.1. Increasing the temperature of NMR experiments for [^{2}H, ^{15}N, ^{13}C]-TPxC60S and [^{2}H, ^{15}N, ^{13}C]-TPxWt.

It is well known that variation of the temperature can strongly influence the NMR spectrum of a globular protein and is a useful variable parameter with which to probe the overall properties in solution. Fundamentally, increasing the sample temperature can have two major influences: (a) decreasing the rotational correlation time of the protein (faster tumbling) leading to altered global average T_1 and T_2 relaxation time constants; and (b) acceleration of dynamic exchange processes: a show exchange phenomenon may become intermediate exchange with concomitant line broadening/averaging, an intermediate exchange process a fast one with line narrowing. Conversely a decrease in the sample temperature will lead to slowing of the both tumbling rate, and exchange processes. Importantly, whilst faster tumbling can lead to improved line widths and peak intensities, this effect may be counteracted by bleaching of solvent exchangeable proton signals due to averaging with the bulk solvent signal. Moreover if the temperature is raised too high then one risks partial or global unfolding of the protein, possibility accompanied by non-reversible sample aggregation.

In the context of the assignment issues for TPxC60S described in the previous chapter, it was felt instructive to explore the temperature dependence of the NMR spectrum. Because the wild-type *Mtb* TPx spectrum suggested a similar but more extensive variation in cross peak intensity, we also examined its spectrum under the same set of conditions. Figure 5.2 shows a superposition of [^{2}H, ^{15}N, ^{13}C]-TPxC60S (Figure 5.2A) and [^{2}H, ^{15}N, ^{13}C]-TPxWt (Figure 5.2B) HSQC spectra, including the accompanying 1D ^{1}H spectra, obtained over a series of increased temperatures. Close inspection of these spectra reveal that for both of *Mtb* TPx proteins there is no increase in cross peak numbers over those observed in the room temperature spectrum. At the final temperature point, 50 °C (blue spectra in Figures 5.2A and 5.2B), both the *Mtb* TPx protein spectra

experienced losses in signal intensity and a disruption to the homogenous line shapes for the majority of cross peaks. For experiments performed at 50 °C and above the spectra contained increasing numbers of signals with ^{1}H chemical shifts at 8.4 +/- 0.2 ppm, though this was clearly more evident for the mutant than the wild-type protein. We presume that these changes reflect partial unfolding or (more likely) degradation of the sample: upon lowering the temperature again the spectra retained this features consistent with disordered peptides. At the highest temperature tested 58.8 °C the HSQC experiment indicated that both protein variants are undergoing rapid protein degradation, resulting in increasingly disrupted and un-interpretable spectra (data not shown).

Despite the apparent slow degradation of a population of protein molecules in these samples, it is straightforward to inspect the more highly resolved regions of the spectra across the temperature range. Thus, for [^{2}H, ^{15}N, ^{13}C]-TPxC60S there was no loss in the number of resolved cross peaks present in the HSQC spectra at higher temperature. However for [^{2}H, ^{15}N, ^{13}C]-TPxWt the number of resolved cross peaks began to decrease as higher temperature spectra are obtained.
Taken together these data indicate that both proteins exhibit a degree of sensitivity to signal loss or degradation at higher temperatures, without any evidence of spectral resolution of 'extra' cross peaks that might derive from regions of the protein suspected to display exchange-derived line broadening in the spectra recorded at 25 °C, either by selection of a single conformer or by acceleration of interconversion of multiple conformers.

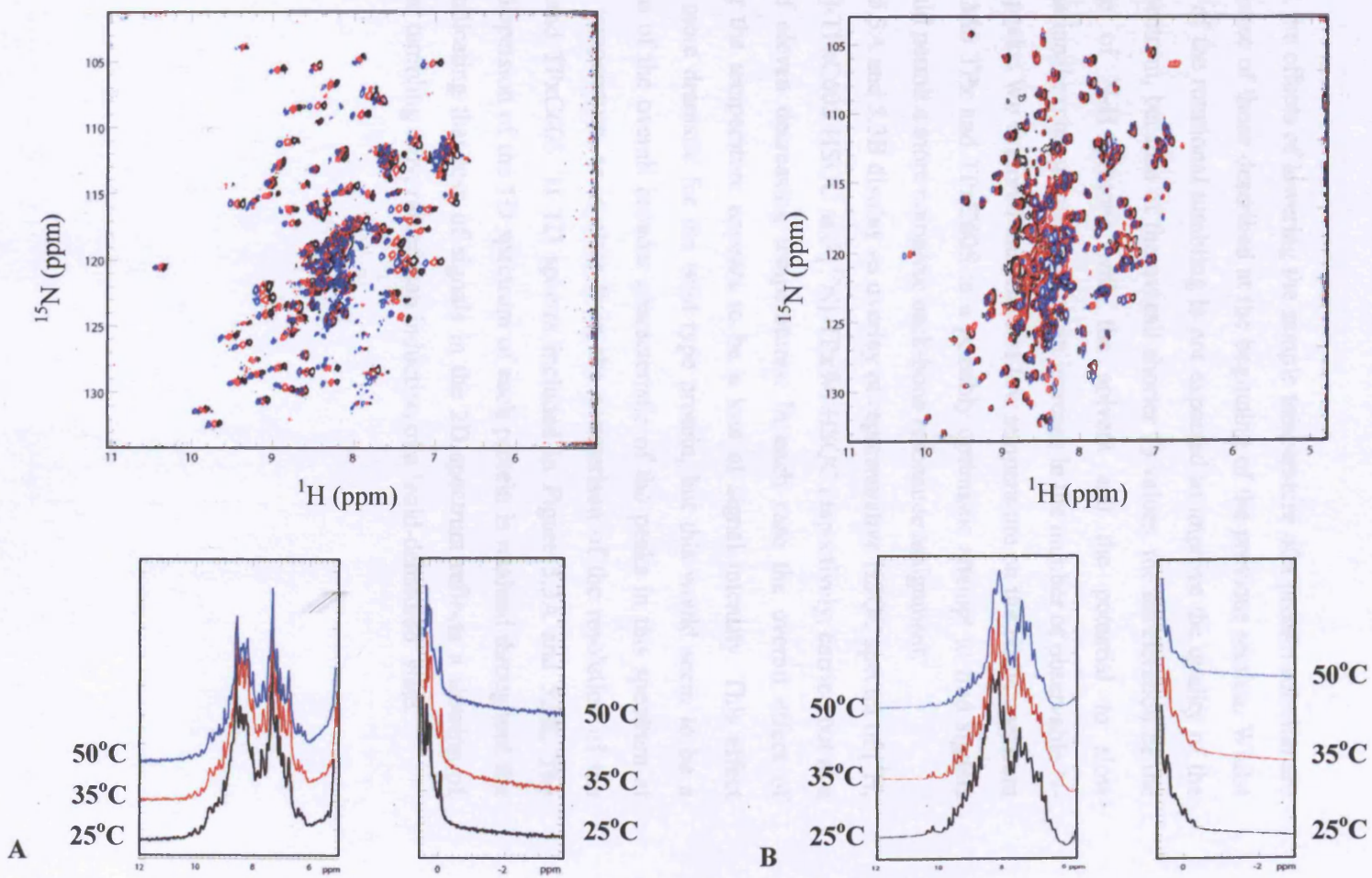


Figure 5.2: HSQC and 1D overlays of 1 mM, pH 7.48 (at 500 MHz); **A**: [^{2}H, ^{15}N, ^{13}C]-TPxC60S and, **B**: [^{2}H, ^{15}N, ^{13}C]-TPxWt HSQC spectra, at various temperatures; 25 °C (black), 35 °C (red) and 50 °C (blue).

5.3.2. Decreasing the temperature of NMR experiments for [^{2}H, ^{15}N, ^{13}C]-TPxC60S and [^{2}H, ^{15}N, ^{13}C]-TPxWt.

In NMR the effects of lowering the sample temperature of a protein solution are the converse of those described at the beginning of the previous section. Whilst slowing of the rotational tumbling is not expected to improve the quality of the NMR spectrum, because of the overall shorter T_2 values, the deceleration of the exchange of N-H protons with the solvent and the potential to slow conformational exchange may lead to an increase in the number of observable N-H cross peaks. We explored the effect of low temperature on the NMR spectrum of both *Mtb* TPx and TPxC60S in a possibly optimistic attempt to find signals that would permit a more complete back-bone resonance assignment.

Figures 5.3A and 5.3B display an overlay of representative HSQC spectra of [^{2}H, ^{15}N, ^{13}C]-TPxC60S HSQC and [^{15}N]-TPxWt HSQC respectively, carried out at a series of eleven decreasing temperatures. In each case the overall effect of lowering the temperature appears to be a loss of signal intensity. This effect appears more dramatic for the wild type protein, but this would seem to be a reflection of the overall broader characteristic of the peaks in this spectrum at ambient temperature, as is clear from the comparison of the resolution of the TPxWt and TPxC60S ^{1}H 1D spectra included in Figure 5.3A and 5.3B. The overall dispersion of the 1D spectrum of each protein is retained throughout the series, indicating that loss of signals in the 2D spectrum reflects a slowing of molecular tumbling only, rather than induction of a 'cold-denatured' state.

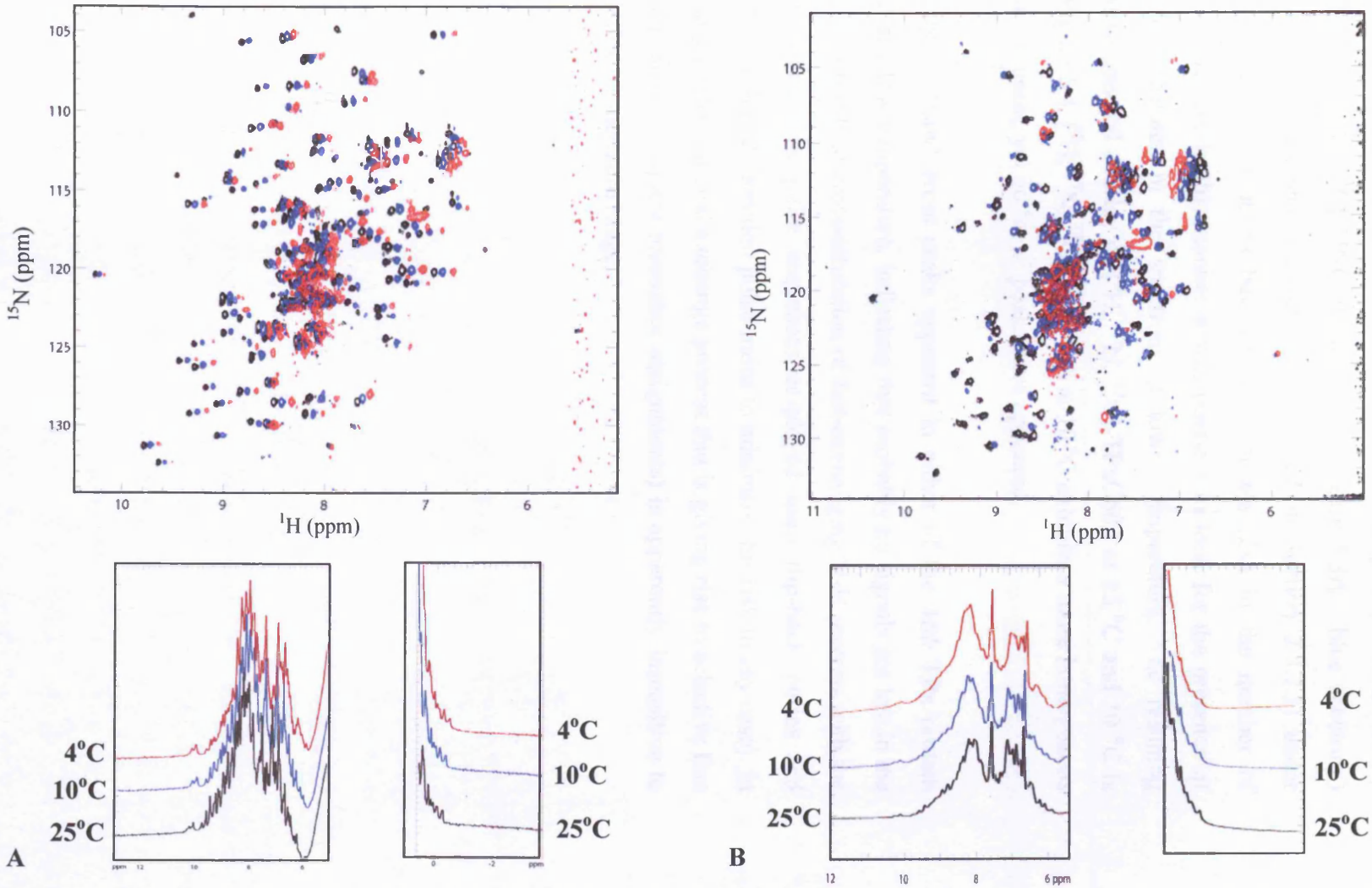


Figure 5.3: HSQC and 1D overlays of 1 mM, pH 7.48 (at 500 MHz); **A**: [^{2}H, ^{15}N, ^{13}C]-TPxC60S and, **B**: [^{15}N]-TPxWt HSQC spectra carried out at various temperatures; 25 °C (black), 10 °C (blue) and 4 °C (red).

For the spectra recorded in Figure 5.3 the number of transients recorded per t_1 increment was held constant, so the signal-to-noise ratio is normalised to the total recording time. In an attempt to improve the spectral quality of the HSQC spectrum of [^{2}H, ^{15}N, ^{13}C]-TPxC60S at 10 °C (Figure 5.3A - blue spectrum) another HSQC experiment was run (as described in Section 2.3.7.3) under otherwise identical conditions but with an increase (2x) in the number of transients collected. In this manner it was possible to look for the presence of new weaker features in the spectrum at lower temperature. The resulting comparative spectral overlay of [^{2}H, ^{15}N, ^{13}C]-TPxC60S at 25 °C and 10 °C is seen in Figure 5.4. [^{2}H, ^{15}N, ^{13}C]-TPxC60S at 10 °C exhibits a more homogenous cross peaks present, yet no 'new' peaks have appeared.

Importantly no 'new' cross peaks appeared in either of the *Mtb* TPx protein spectra at any low temperature, indicating that probably no signals are lost in the spectrum as a result of cross-saturation of fast-exchanging N-H protons with the bulk H_2O signal (the pulse sequences employed used flip-back pulses and WATERGATE-based detection pulse trains to minimize this risk in any case). In addition any conformational exchange process that is giving rise to selective line broadening (and incomplete resonance assignments) is apparently insensitive to the accessible temperature range.

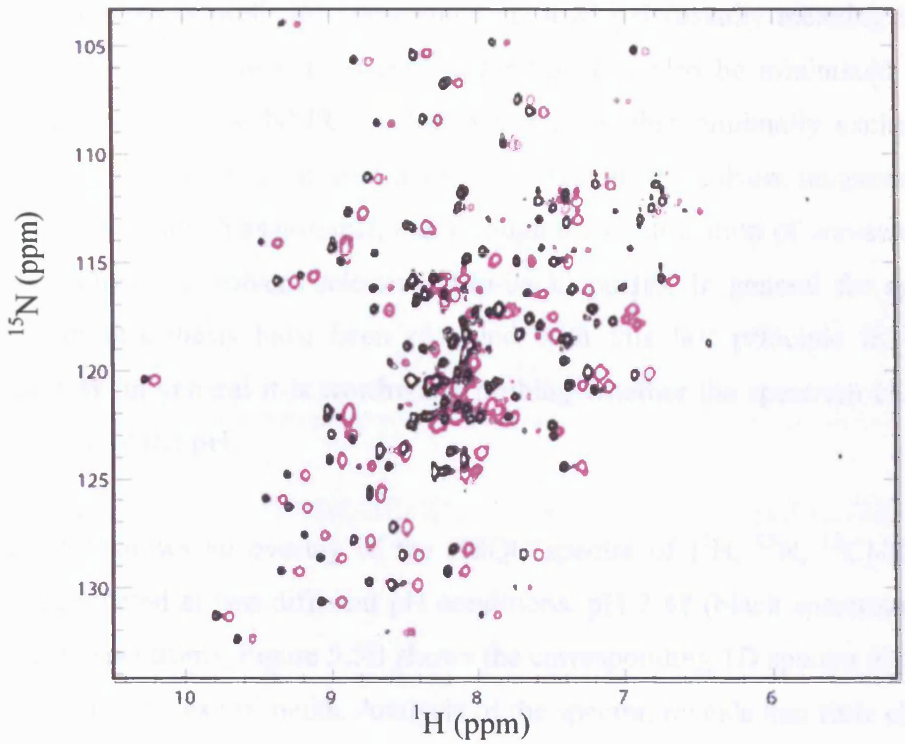


Figure 5.4: HSQC overlay of 1 mM, pH 7.48, at 500 MHz of [^{2}H, ^{15}N, ^{13}C]-TPxC60S HSQC spectra carried out at two different temperatures and data-point increments: 25 °C (black) – 64 points and 10 °C (magenta) – 128 points.

5.4. 1D [^{1}H] and [^{2}H, ^{15}N, ^{13}C]-labelled-HSQC: TPxC60S and TPxWt pH variations.

Another facile means to perturb the NMR spectrum of a protein is via variation of the solution pH. In general, whilst the H^{+} ion concentration is not expected to have an influence upon line broadening due to conformational exchange, it is well known to have a strong impact upon the rate of exchange of N-H, O-H and S-H protons ('exchangeable' protons) with the bulk solvent H_2O. Thus when NMR experiments are recorded in a way that leads to saturation of all or part of the solvent resonance there is a risk that signals corresponding to these protons lose intensity or are saturated themselves (by a process termed cross-saturation), in a manner that depends quantitatively upon the pH. In general the lower the pH the slower the rate of exchange of protons with the bulk solvent and the lower the

risk of loss of signal intensity due to cross-saturation. Thus it is often said that NMR of proteins is preferably performed at 'low' pH (usually meaning below neutral pH). In fact, the risk of cross-saturation can also be minimized to an extent by recording the NMR spectra in a manner that minimally excites the solvent H_2O resonance, or at least attempts to restore the solvent magnetisation to the +Z axis as much as possible, e.g. through the combination of non-selective hard RF pulses and solvent selective 'flip-back' pulses. In general the spectra reported in this thesis have been recorded with this last principle in mind. Nevertheless, in general it is worthwhile probing whether the spectrum changes as a function of the pH.

Figure 5.5A shows an overlay of the HSQC spectra of [^{2}H, ^{15}N, ^{13}C]-TPxWt spectra conducted at two different pH conditions, pH 7.48 (black spectrum) and pH 6.0 (red spectrum). Figure 5.5B shows the corresponding 1D spectra obtained prior to the HSQC experiments. Analysis of the spectra, reveals that little change is observed. The protein clearly remains in solution and displays the same degree of resonance dispersion indicating that the overall fold is maintained. Upon lowering the pH yet further, from pH 6.0 to pH 5.0, TPxWt precipitates into a stringy deposit that clings to the side of the NMR tube, without settling. The deposit was not fully re-solubilized upon raising the pH again to 7.5. Thus it appears that much below pH 6.0 TPxWt undergoes partial or complete irreversible denaturation. As a result of this investigation ~pH 6.0 is regarded as the pH minima recommended prior to irrevocable structural change occurring to [^{2}H, ^{15}N, ^{13}C]-TPxWt. Importantly, no 'new' HSQC cross peaks are observed with a [^{2}H, ^{15}N, ^{13}C]-TPxWt protein sample at pH 6.0; the only discernible changes were small shifts (less than, or of the order of the resonance cross-section) in a few cross peak positions.

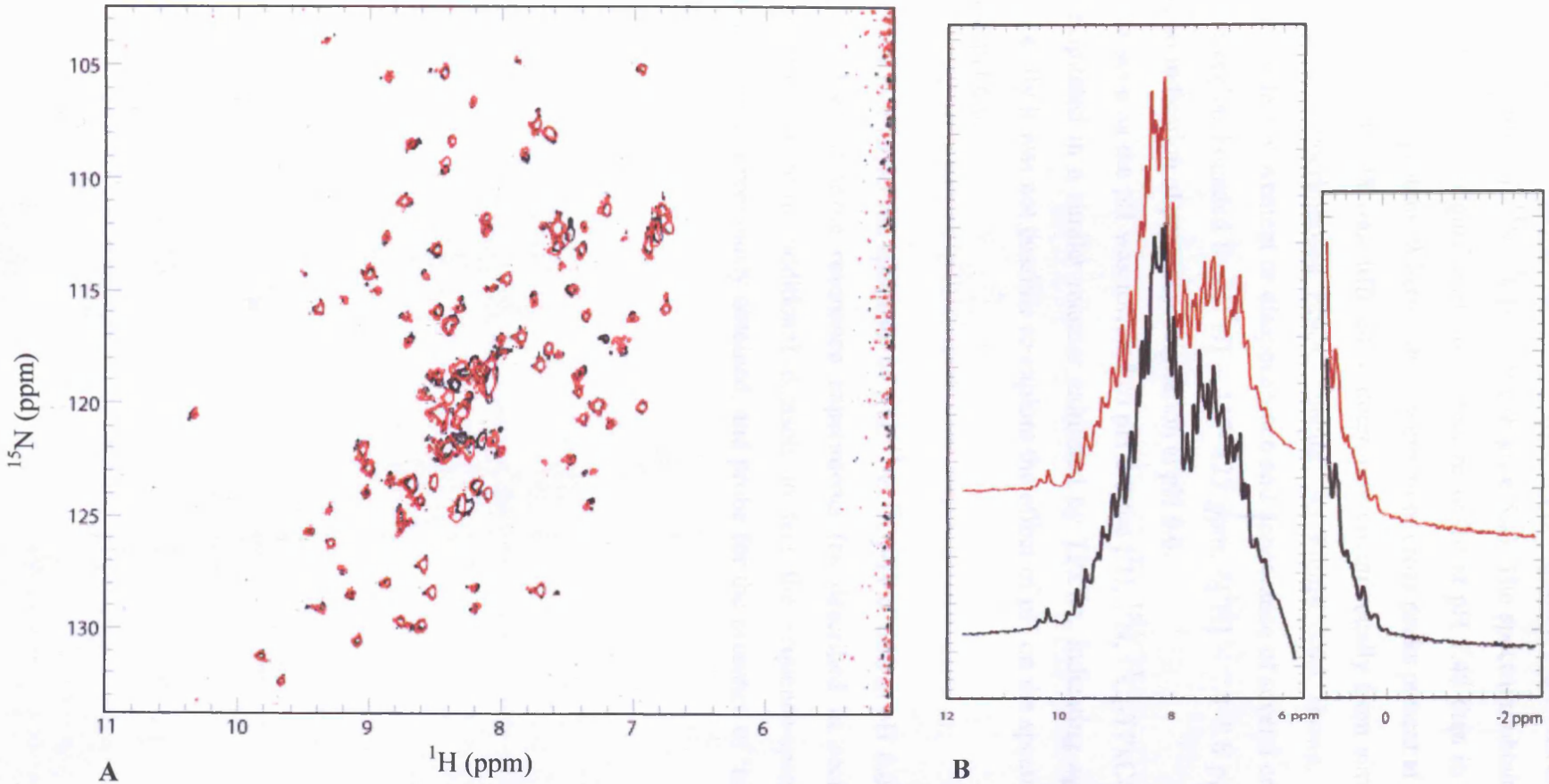


Figure 5.5: Comparative overlays of 1 mM, 25 °C, at 500 MHz of [^{2}H, ^{15}N, ^{13}C]-TPxWt HSQC (**A**) and 1D (**B**) NMR spectra, respectively, carried out at two different pH values; pH 7.48 (black spectrum) and pH 6.0 (red spectrum).

[^{2}H, ^{15}N, ^{13}C]-TPxC60S was similarly examined in pH variation experiments as those conducted with [^{2}H, ^{15}N, ^{13}C]-TPxWt (Figure 5.6). The spectrum obtained at pH 6.0 differed more significantly from that recorded at pH 7.48 than in the case of the wild-type protein. Whether there were more cross peaks present at pH 6.0 compared to pH 7.48 was difficult to determine unequivocally from simple inspection of the superimposed HSQC spectra. As Figure 5.6A shows, the changes may reflect movement or disappearance and appearance of several cross peaks in the region bounded by $\delta[^{15}N] \sim 117–125$ ppm; $\delta[^{1}H] \sim 7.8–8.8$ ppm. There was no indication of protein precipitation at pH 6.0.
However, as soon as the pH was lowered to pH 5.0 the [^{2}H, ^{15}N, ^{13}C]-TPxC60S protein precipitated in a similar manner exhibited by TPxWt, indicating again that operationally it was not possible to explore the effect of pH on the spectrum much below pH 6.0.

It was decided to analyse the spectrum of [^{2}H, ^{15}N, ^{13}C]-TPxC60S at pH 6.0 by recording a subset of triple resonance experiments (as described in section 2.3.7.6), thereby providing additional datasets to test the sequence-specific resonance assignments previously obtained, and probe for the presence of 'new' cross peaks.

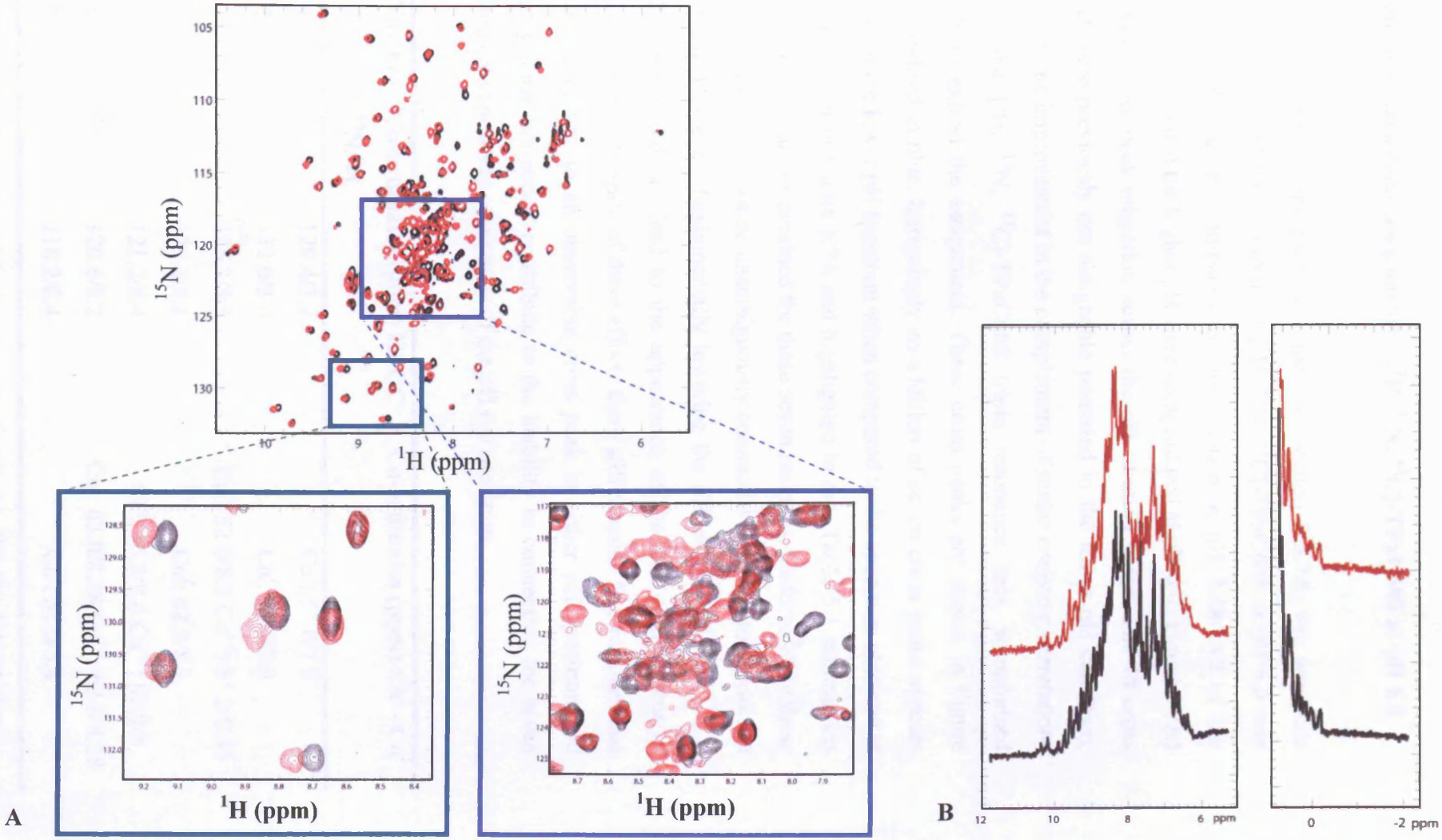


Figure 5.6: Spectral overlays of 1 mM, 25 °C, at 500 MHz of [^{2}H, ^{15}N, ^{13}C]-TPxC60S HSQC (**A**) and 1D (**B**) NMR spectra, respectively, carried out at two different pH values; pH 7.48 (black) and pH 6.0 (red spectrum).

5.5. Amino acid back-bone assignment – [^{2}H, ^{15}N, ^{13}C]-TPxC60S at pH 6.0.

Using the experimental approach described in Section 2.3.7.6, the sequence specific assignment of the spectrum of [^{2}H, ^{15}N, ^{13}C]-TPxC60S at pH 6.0 was attempted and the results compared to those obtain at pH 7.48. All of the assignments obtained at the higher pH were recapitulated at pH 6.0 (Figure 5.7A) by observing cross peak migration across the pH titration range. The 18 cross peaks that were previously not assignable persisted in the lower pH condition, but despite some improvement in the complement of scalar coupling correlations present in the [^{2}H, ^{15}N, ^{13}C]-TPxC60S triple resonance data, it remained impossible to extend the assignment. These cross peaks are shown in Figure 5.7A, highlighted in blue. Intriguingly an addition of seven cross peaks appears exclusively at the lower pH spectrum when compared to the spectrum obtained at pH 7.48, as shown in Figure 5.7A and highlighted in red. Table 5.1 summarises the detectable correlations obtained for these seven residues. Unfortunately these 'new' cross peaks could not be unambiguously connected to other cross peaks of the spectrum. However, fundamentally lowering the pH of the [^{2}H, ^{15}N, ^{13}C]-TPxC60S sample did not lead to the appearance of the full complement of missing cross peaks. In spite of these efforts there still remains five residues that still do not give rise to an observable cross peak in either pH spectrum and presumably these 'absences' contribute to the inability in connecting the seven unassigned cross peaks to other parts of the pH 6.0 spectrum.

Peak Number	Co-ordinates (ppm) HSQC ^{15}N/^{1}H	Co-ordinates (ppm) $C\alpha^{i}$ - $C\alpha^{i-1}$
1	129.4/7.9	$C\alpha^{i}$: 57.0/7.9
2	132.0/8.0	$C\alpha^{i}$: 53.5/7.9
3	125.3/8.3	$C\alpha^{i}$: 52.0/8.3 $C\alpha^{i-1}$: 61.2/8.35
4	122.2/8.1	$C\alpha^{i}$: 62.2/8.1
5	121.2/8.4	$C\alpha^{i}$: 56.3/8.4 $C\alpha^{i-1}$: 62.8/8
6	120.6/8.2	$C\alpha^{i}$: 62.0/8.28 $C\alpha^{i-1}$: 62.8/8.28
7	118.2/8.4	$C\alpha^{i}$: 56.0/8.4

Table 5.1: HSQC coordinates for the seven cross peaks found using [^{2}H, ^{15}N, ^{13}C]-TPxC60S at pH 6.0.

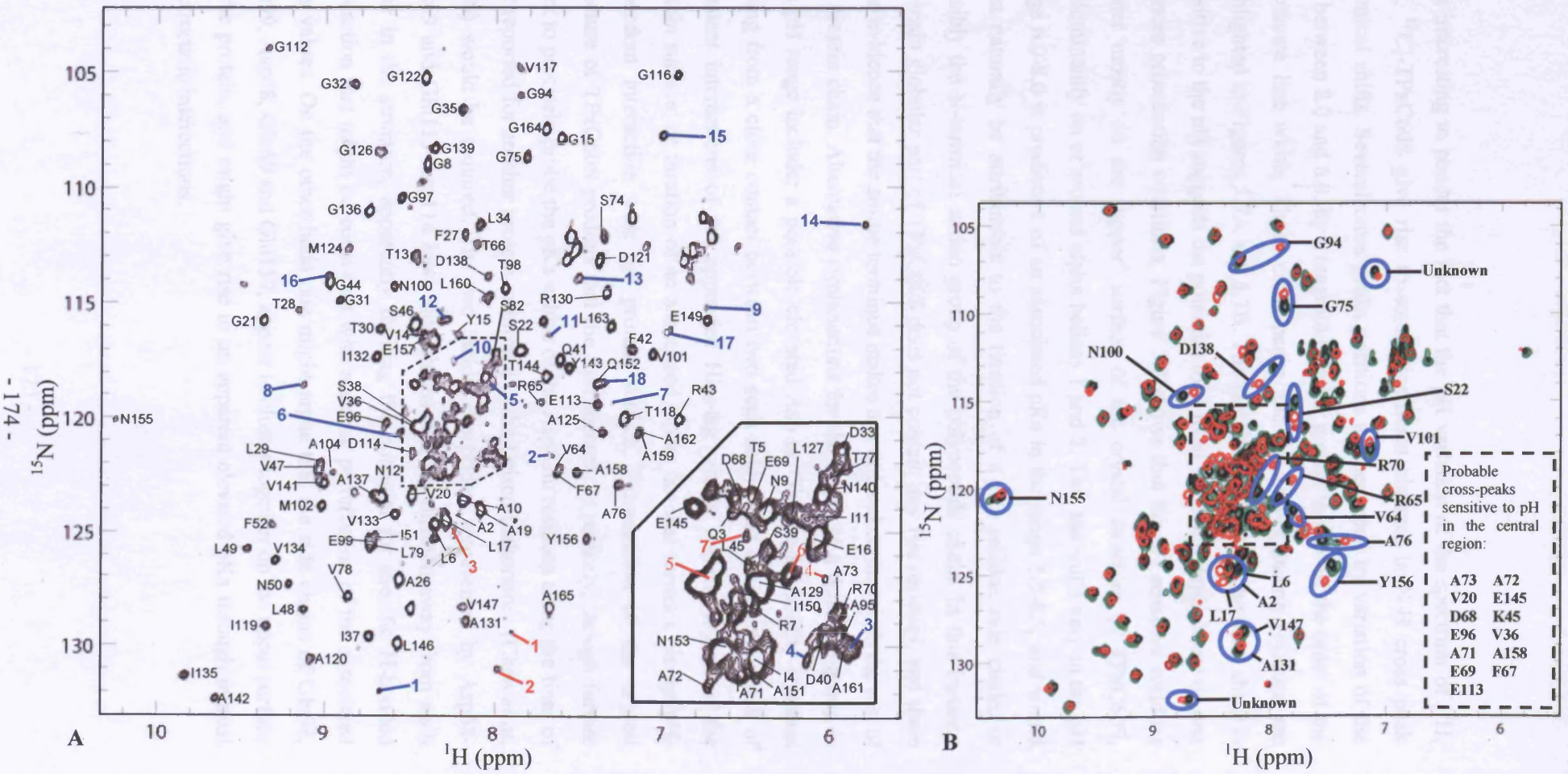


Figure 5.7 A: The assigned spectrum of 1 mM, 25 °C, at 500 MHz of [^{2}H, ^{15}N, ^{13}C]-TPxC60S at pH 6.0. The red numbering denotes the 'new' cross-peaks observed and are directly related to **Table 5.1**. The blue numbering denotes the previous 18 unassigned cross peaks from the pH 7.48 spectrum. **B**: Superposition of TPxC60S spectra at differing pH conditions, 7.5 (black), 6.5 (green) and 6.0 (red). Probable cross peaks sensitive to changes in pH are indicated.

It is interesting to ponder the fact that the pH variation of the spectrum of [^{2}H, ^{15}N, ^{13}C]-TPxC60S give rise to some significant changes in N-H cross peak chemical shifts. Several cross peaks positions are perturbed by variation of the pH between 8.0 and 6.0, by a magnitude either greater then or of the order of the resonance line width. These cross peaks and the corresponding residues are highlighted in Figures 5.7A and 5.7B. The fact that a subset of chemical shifts is sensitive to the pH suggests the potential for a local structural dependence on one or more protonation equilibria. Figure 5.8 shows that the pH sensitive residues cluster mostly on the 'upper' surface of the crystal structure of TPxC60S, predominantly on or around alpha helices 1 and 2. That the shifts vary in the pH range 6.0-8.0 is predictive of an associated pKa in the range 5.5-8.5, and would most naturally be attributable to the titration of a His residue side chain, or possibly the N-terminal amino group of the polypeptide chain. In this instance the main globular part of TPxC60S does not contain any His residues, and there is no evidence that the amino terminus makes any ordered contact with the rest of the protein chain. Alternative explanations for the origin of a titratable group in this pH range include: a possible elevated Asp or Glu residue acidity constant arising from a close contact between two such acidic side chains; the result of transient interactions of the appended His_6-tag with a specific region of the protein surface; or titration of an associated buffer ion that shows a charge/pH-dependent interaction with the protein surface. Examination of the crystal structure of TPxC60S predicts that the first scenario is unlikely, though further work to precisely probe the pKa values of the Asp/Glu residues along the lines of that reported for another protein examined in the Driscoll laboratory (Chen *et al*, 2000) would be required. The two 'acidic dipeptides' represented by Asp68-Glu69 and Glu113-Asp114 have side chains that are directed away from each other in the structure, apparently limiting the potential for specific H-bonded interaction that might be consistent with a strong perturbation of the associated pKa values. On the other hand one might argue that the side chains of Glu98, Glu99, Asp68, Glu69 and Glu157, appear to cluster together on the upper surface of the protein, and might give rise to an apparent elevated pKa through mutual electrostatic interactions.

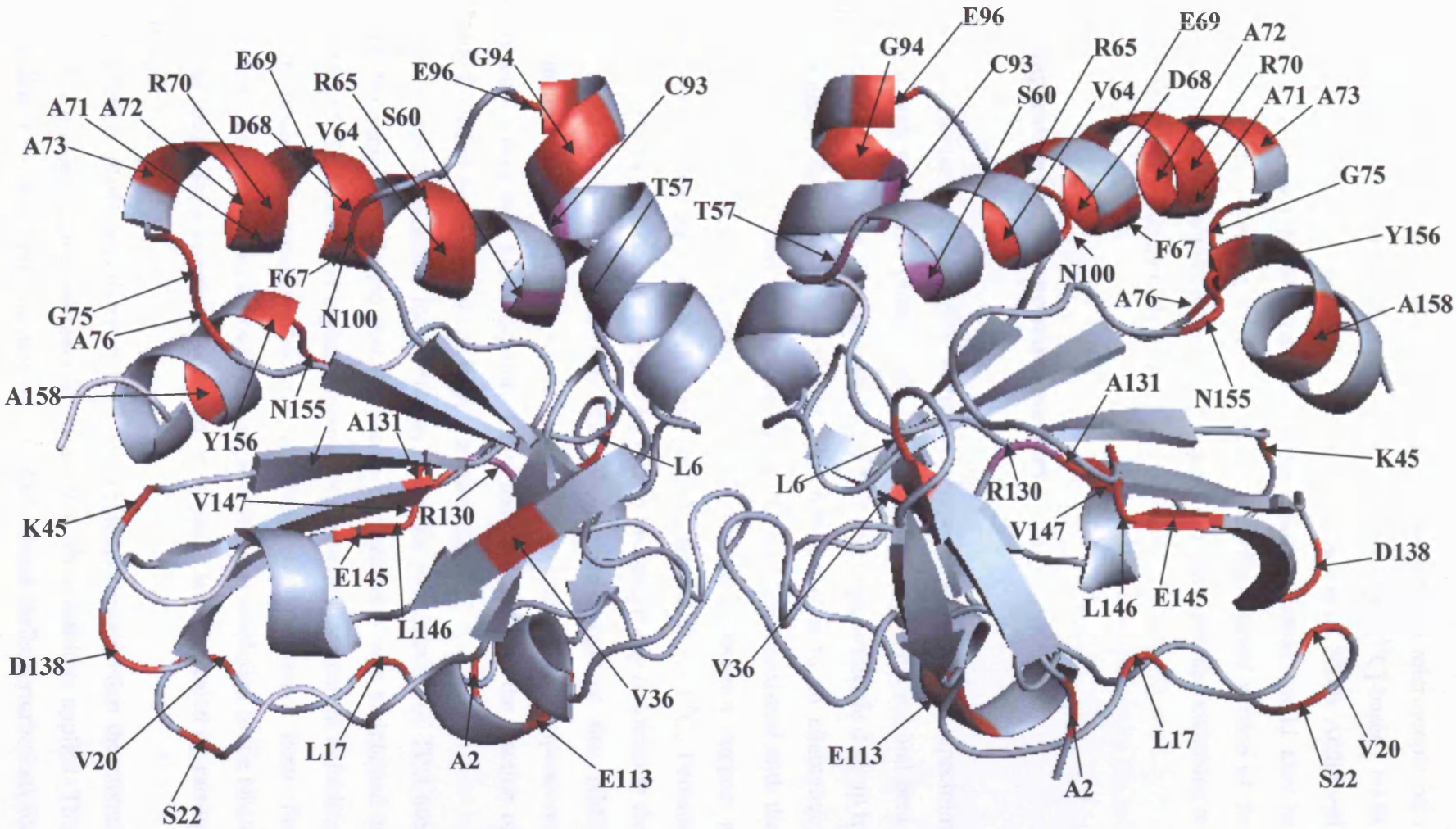


Figure 5.8: Crystal structure of *Mtb* TPxC60S (PDB: 1Y25) indicating all the amino acid residues (red) perturbed upon a change of pH from 7.48-6.0. The residues coloured magenta indicate the C60S mutated residue and other constituents of the catalytic triad.

Resolution of these ideas requires direct measurement of the microscopic pKa values for these residues that can be obtained by [^{13}C]-based NMR measurements of the type described by Chen *et al.* (Chen *et al*, 2000). Additional further work to resolve the potential for the other scenarios would also be required, including repeating the experiments for a His_6-cleaved version of the protein, which would require a new DNA expression construct, and switching to a cationic buffer, e.g. Tris, or Bis-Tris.

5.6. Temperature and pH variation summary.

Circular dichroism of TPxC60S at room temperature yielded a spectrum consistent with a globular protein fold containing both alpha-helical and beta-sheet secondary structure. The analysis of the CD spectropolarimetric data up to 95 °C suggests that the natively folded TPxC60S converts to an alternative, presumably partially folded state that retains a CD signature consistent with the persistence of significant secondary structure. The data obtained suggest a cooperative two-state transition with a mid-point T_m around 63 +/- 1 °C. Perhaps surprisingly the protein does not appear by CD to be completely unfolded at the highest temperature tested. However this observation suggested that NMR measurements of the protein would sensibly be limited to temperatures significantly lower than the mid-point of this transition, so that the fraction of partially folded molecules would be kept to a minimum.

Variation of the temperature for NMR spectroscopic examination of TPxC60S and TPxWt samples indicated that the native conformation was maintained at least up to 55 °C. At this and higher temperatures the effects of partial unfolding was irreversible degradation became apparent. Unfortunately, from the perspective of attempting to uncover new signals for the completion of the NMR back-bone resonance assignment, no extra cross peaks were revealed by raising the temperature.

The inspection of pH sensitivity of both *Mtb* proteins revealed that the mutant variant had a more aggressive reaction to the low pH conditions applied. This was exhibited by observing characteristic HSQC-related shifts/perturbations for several cross peaks, together with the disappearance/appearance of others. A

subsequent recapitulation of the TPxC60S triple resonance assignment obtained at pH 7.48 was reproduced at pH 6.0. By comparing TPxC60S spectra at the two different pH conditions, 32 cross peaks were identified as being especially sensitive to the low pH condition applied, including two that are as yet unassigned due to a variety of cross peak "absences" in the related triple resonance experiments. Most intriguing of all was the identification of a further seven cross peaks (seemingly exclusive to pH 6.0 TPxC60S spectra) exhibiting either complete or partial C_α i and i^{-1} correlations in a partially-acidic environment. Although a significant observation, these seven "new" cross peaks could not be unambiguously connected to the currently unassigned sections of the TPxC60S sequence.

The next step is in the use of ligands – as a sequestering tool (if scenario three (as described in Section 4.8) is to be believed as the principal reason for internal dynamic motion), to further explore the dynamic fluctuations exhibited by TPxC60S and in so doing aid in completing the back-bone assignment. The following section will deal with the ligands used in this project to resolve the issue described above.

5.7. Ligand binding using TPxC60S and TPxWt – prelude.

In the following sections, a series of experiments are described that attempt to probe the interaction of *Mtb* TPx with chemical entities that are supposed or suspected to interact with the active site of the protein (as detailed in Section 2.3.7.4). Specifically the interaction of *Mtb* TPx is tested for magnesium acetate, an organic compound known as Diclofenac, and the thiol reactive compound DTDP is discussed. For the first two cases, crystallographic data suggest that the compound has the potential to interact directly with the protein active site. In the last case it was anticipated that DTDP might react with and 'trap' the side chain of the resolving cysteine Cys93, which might become exposed during the overall reaction mechanism of the protein. In each case it was hoped that a productive interaction with the compound might perturb the presumed underlying dynamics that gives rise to substantial cross peak absences in the NMR spectrum.

5.7.1 Magnesium acetate ligand binding.

The crystals of TPxC60S reported by Stehr *et al.* were obtained using a buffer that had the following composition: 20 % PEG 8000, 100 mM sodium cacodylate pH 6.5 and 200 mM magnesium acetate. Electron density at the active site of the final model was interpreted as involving an acetate anion (Figure 5.9A) with H-bonds between the ligand carboxylate group and the side chains of C60S (including the back-bone nitrogen), Thr57, and Arg130 (Figure 5.9B). The ligand methyl group makes van der Waals contacts with hydrophobic residues Leu6, Arg7, Pro53, Met124, and Ile150. The overall binding mode is similar to the interaction of a benzoate ion with human peroxiredoxin 5 (Declerq *et al*, 2001).

The bulk of the NMR data for TPxC60S described so far in this thesis have been obtained using a buffer with composition 50 mM NaH_2PO_4, 100 mM NaCl, 1 mM DTT and 0.1 mM EDTA. Whilst it is was not expected that introduction of acetate ions to the buffer would have a dramatic effect, nevertheless we formally titrated a sample of TPxC60S with a solution of magnesium acetate up to a final acetate concentration of 400 mM, with monitoring by 1D ^{1}H and 2D [^{15}N,^{1}H]-HSQC spectroscopy (Figure 5.9C). As largely anticipated, changes to the spectrum in the presence of acetate where slight. However, a small number of assigned cross peaks, listed in Table 5.2, are perhaps perturbed in their chemical shift(s). Importantly in this regard the direction of the chemical shift changes for these signals (see Figure 5.10) is characteristically different from those obtained in the pH titration series described in Section 5.4, suggesting that the shift observed may well be due to specific interactions of the protein with the acetate ions.

Figure 8 corresponding boxes	Co-ordinates (ppm) HSQC ($^{15}N/^{1}H$)	Amino acid assignment
1	107.8/7.65	G94
	108.2/7.61	G75
2	120.8/10.51	N155
3	120.2/7.74	R65
4	124.8/8.89	A104

Table 5.2: Chemical shifts co-ordinates and cross peak assignments relating to perturbed residues observed in Figure 5.10B, [^{2}H, ^{15}N, ^{13}C]-TPxC60S in 200 mM magnesium acetate.

The corresponding residues cluster around the active site (Figure 5.11B) which might be consistent with binding of the acetate ion in the active site, as in the crystal structure, with three of the four alpha helices harbouring four of the five perturbed residues: α1:R65, G75; α2:G94; and α4:N155. The remaining perturbed amino acid residue, A104, is found on beta strand β5. Of course it is difficult to know whether phosphate ions in the 'standard' buffer used for the other NMR experiments also interact with the active site region. Importantly no 'new' cross peaks have appeared in the HSQC spectrum as a result of the presence of acetate ions. [Similarly, experiments in which TPxWt was titrated with magnesium acetate did not lead to any significant improvement in the spectrum (data not shown).] Thus any interaction of the acetate ion with the protein is presumably rather weak, and the presence of acetate within the active site of the TPxC60S crystal structure does not represent a highly stable interaction. As described by Stehr *et al*, (2006) acetate certainly does not appear to promote the formation of conformations consistent with the formation of a C_R-C_P disulphide bond, though perhaps this should not be expected in any event.

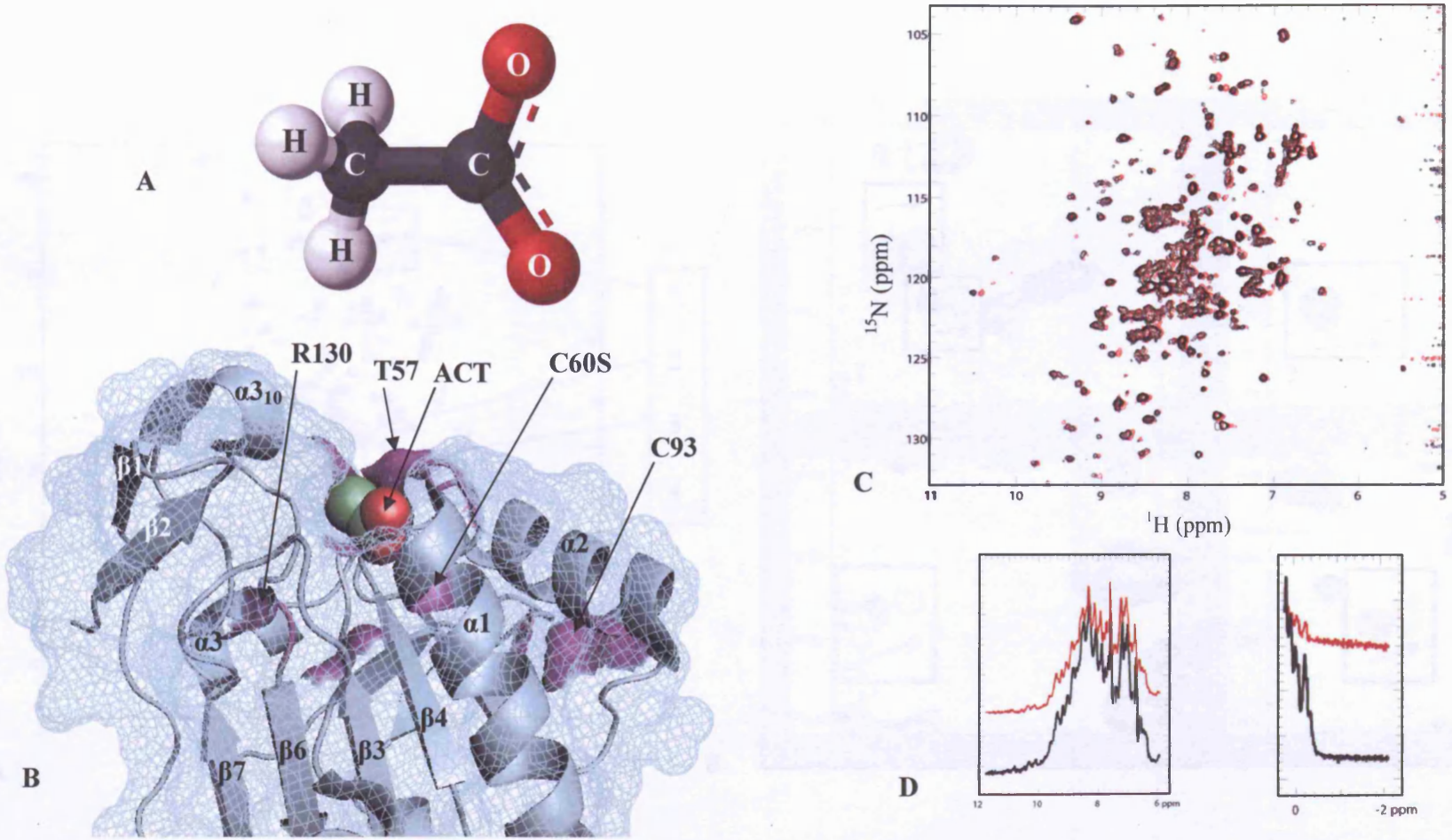


Figure 5.9: A: The chemical structure of acetate anion (CH3COO) found in magnesium acetate, (drawn using BioRad's KnowItAll™) using CPK colouring. **B**: *Mtb* TPxC60S active site; comprised of two catalytic cysteines, Cys60S and Cys93, plus the two important C_P activating cleft residues; Arg130 and Thr57, all shown coloured in magenta. The acetate (ACT) molecule is as indicated by the coloured balls Picture drawn using PyMOL v0.99 (DeLano, 2006). [^{2}H, ^{15}N, ^{13}C]-TPxC60S spectral overlays of; HSQC (**C**) and 1D (**D**) NMR spectra, respectively, carried out at two different magnesium acetate concentrations; 0 mM magnesium acetate (black) and 200 mM magnesium acetate (red). Both C and D spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

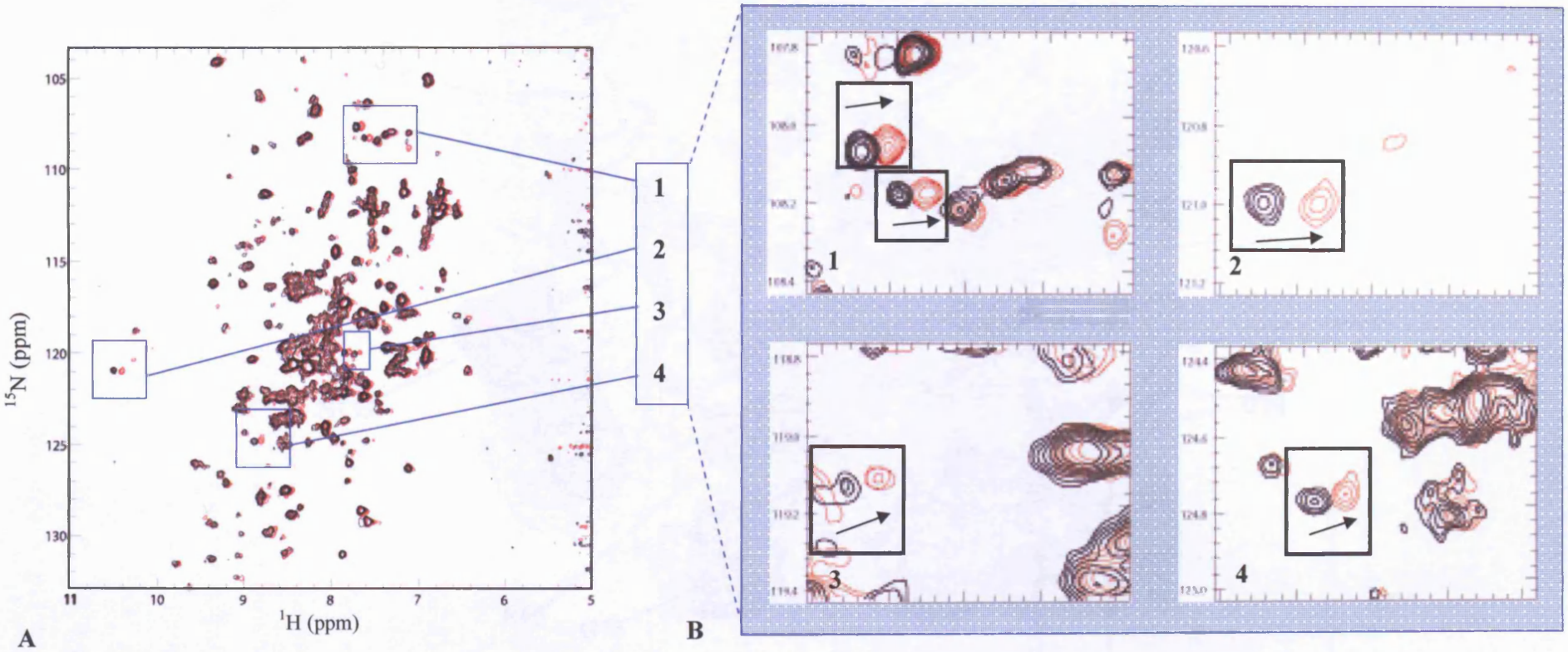


Figure 5.10: **A**: [^{2}H, ^{15}N, ^{13}C]-TPxC60S HSQC spectral overlays at two different magnesium acetate concentrations; 0 mM magnesium acetate (black) and 200 mM magnesium acetate (red). **B**: Close-up of perturbed residues from A, with movement indicated by boxed arrow. Spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

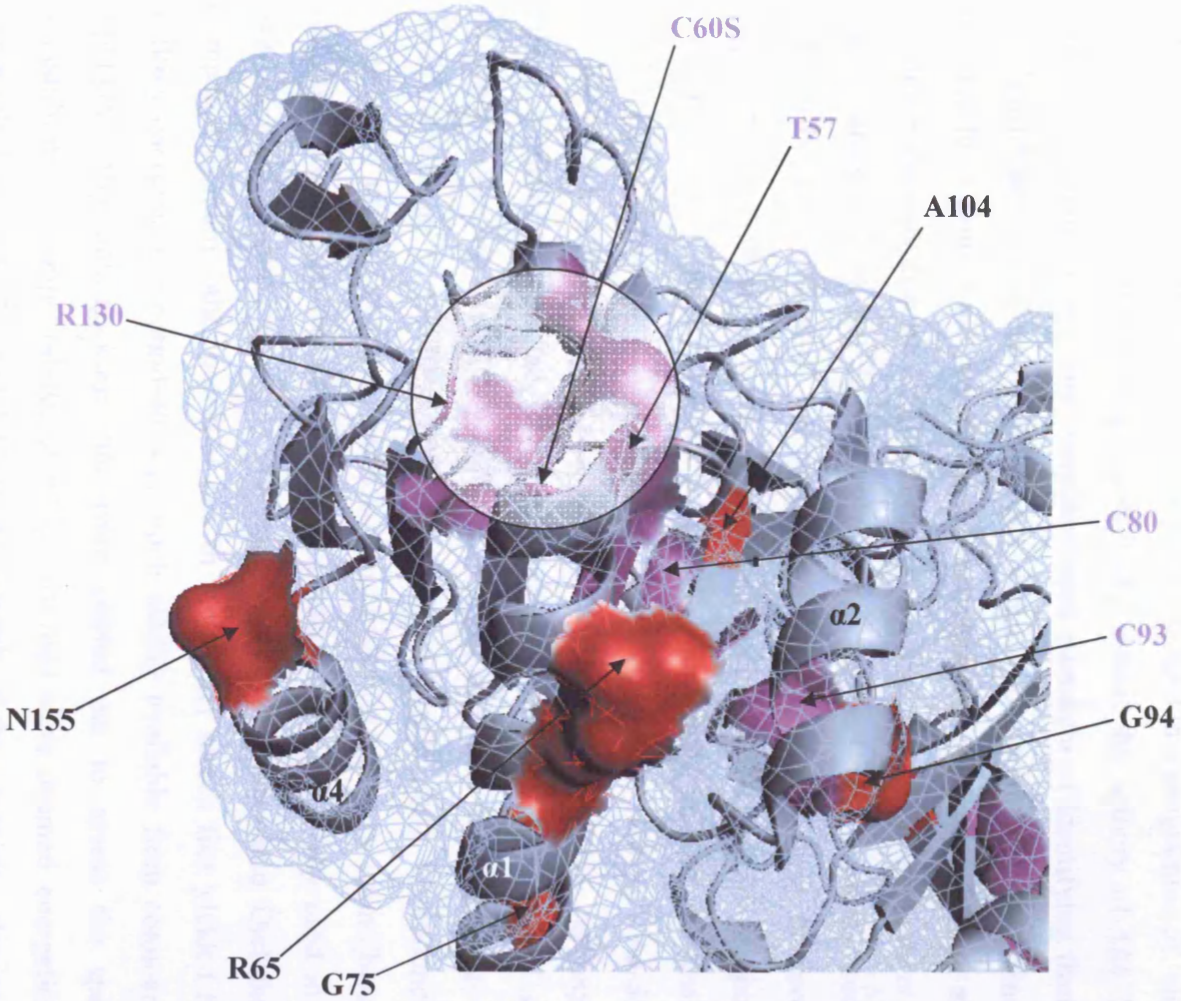


Figure 5.11: Perturbed residues mapped onto the TPxC60S crystal structure (PDB: 1Y25) shown in red and the catalytic, activating and mutated C60S amino acid residues shown in magenta. The binding site is circled.

5.7.2 Diclofenac sodium salt – Compound-40.

Our collaborators (GBF Braunschweig) in the EU Framework Programme 7 project that funded the work presented here, and that also determined the X-ray crystal structure of TPxC60S have also been engaged in a programme of 'virtual screening' for ligands that might potentially inhibit the activity of *Mtb* TPx. Virtual screening refers to the computer-based paradigm of identifying from an *in silico* library of low molecular weight chemical compounds putative binding partners for a protein surface. Different incarnations of virtual screening exist, varying in the way that the docking solutions are searched and in the manner that the hits are scored. A full description of the virtual screening methods used by our collaborators is beyond the scope of this thesis. In summary however a ligand-protein docking program (Autodock3) combined with two compound diversity-sets that provide molecular spatial coordinates and atomic charges (NCI, ~1990 compounds; ZINK, ~49000 compounds) was used to 'auto-dock' flexible ligands to the crystal structure of TPxC60S. From a ZINK carboxylate subset, a molecule denoted as compound-40 (Figure 5.12A) emerged as a substance exhibiting a relatively high (i.e. strongly negative) docking energy and predicted binding to the TPxC60S active site (Figure 5.12B). Compound-40, identified by the full chemical name as 2-[(2,6-dichlorophenyl)- amino]benzoic acid, is a known non-steroidal anti-inflammatory drug commonly used in the field of medical pharmacology, marketed under the trade name Diclofenac. Compared to many other potential hits in the virtual screen that yielded high docking energies, compound-40 was much readily available from commercial suppliers. Fifty auto-docking runs were carried out to assess the spatial compatibility of compound-40, of which six runs were deemed energetically favourable (Figure 5.12C and 5.12D). As a result of this promising theoretical binding potential for compound-40, co-crystallisation of the compound with TPxC60S was then performed. This effort yielded a 2.4 Å resolution model of the complex between TPxC60S and compound-40, electron density consistent with a molecule of compound-40 occupying the active site (Figure 5.13A). The overall conformation of the TPx protein is unchanged, with only a slight displacement of the location of the side chain of Arg7 at the edge of the binding

pocket that makes room for the bulky dichlorophenyl ring. The compound-40 carboxylate group makes H-bonds to the side chains of Ser60 and Thr57 as for acetate in the previously reported structure (Figure 5.13B). [There is also evidence for a second molecule of compound-40, sitting atop the first molecule in the active site (Figure 5.13B) which is in a more peripheral location with few contacts to the protein itself. This second molecule of compound-40 makes a pi-stacking interaction with a symmetry-related neighbour in the lattice (not shown).]

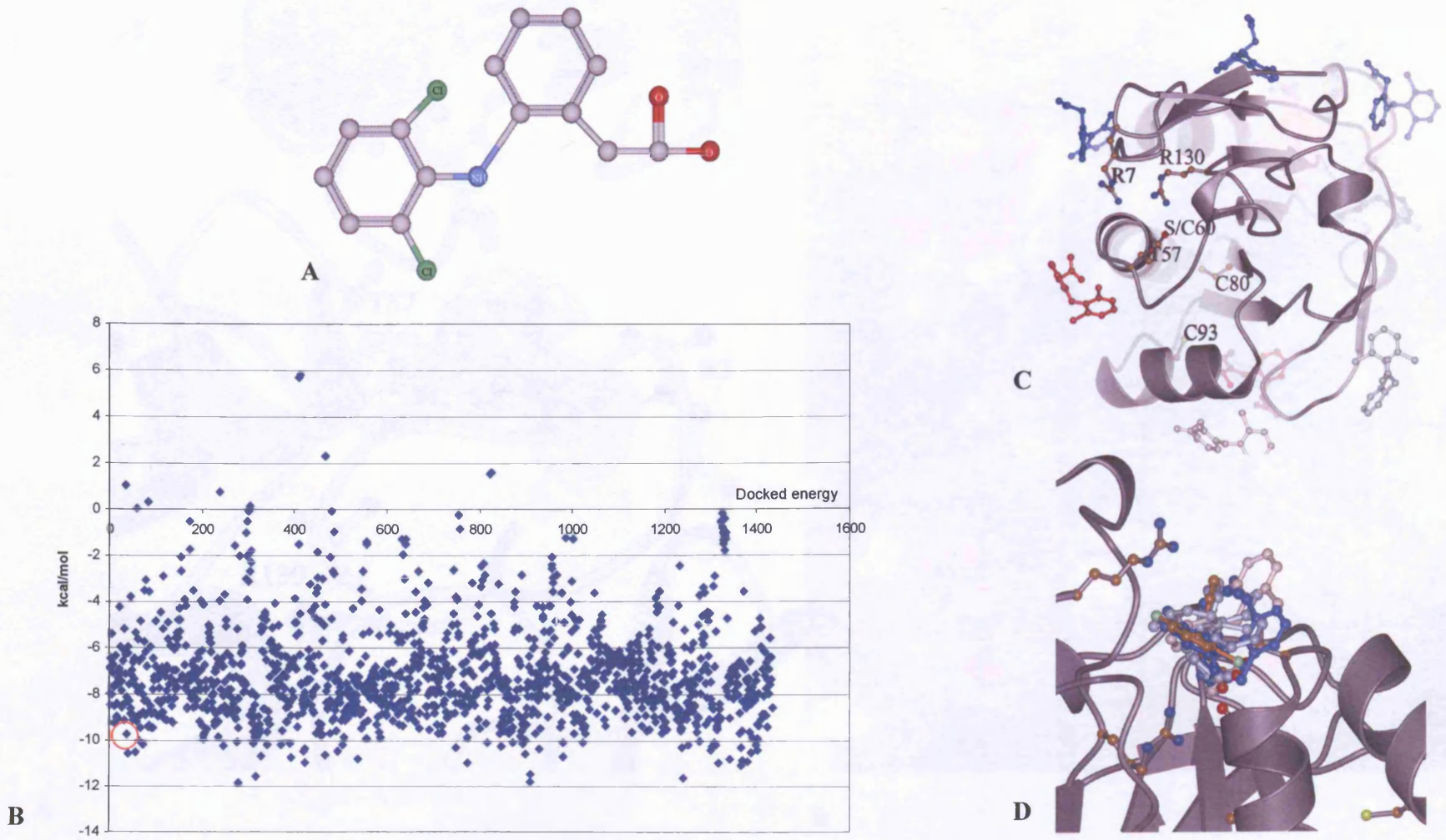
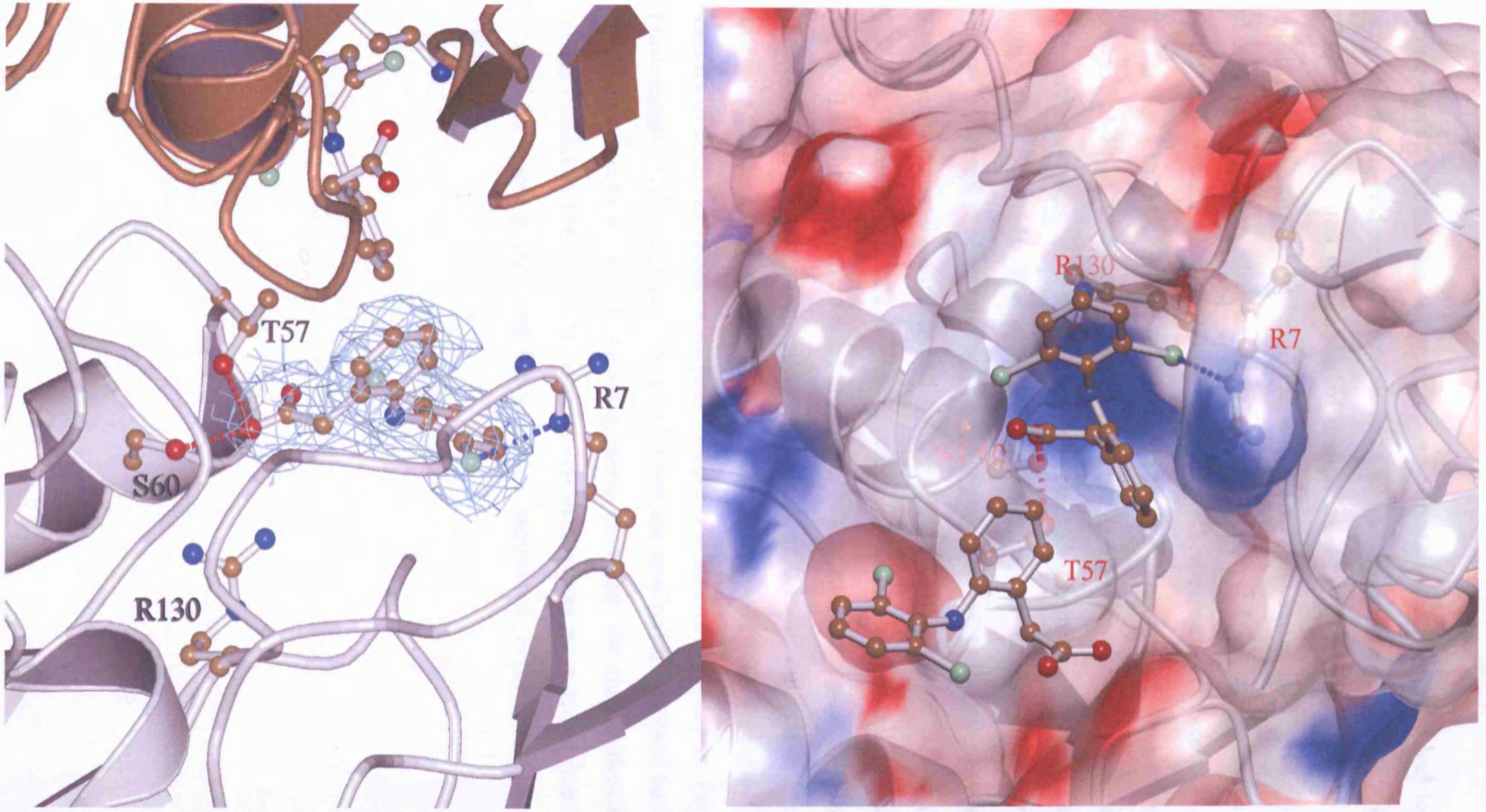


Figure 5.12: **A**: The chemical structure of compound-40 (drawn using BioRad's KnowItAll™) using CPK colouring. **B**: Theoretical TPxC60S ligand binding using the ZINK carboxylate compound group. Subset (1450 compounds) map, indicating compound-40 circled in red. **C**: TPxC60S structure indicating compound-40 as the high energy molecule at the active site, high energy compounds are shown in red, low energy compounds in blue. **D**: 6 out of the 50 attempted of auto-docking conformations of compound-40 with TPxC60S.



A B

Figure 5.13: **A**: The binding of compound-40 with the active site of TPxC60S using CPK colouring (drawn using PyMOL™). **B:** Crystal structure of TPxC60S indicating the association of two compound-40 molecules to the active site per monomeric unit. Colouring indicates the hydrophobic nature of the surrounding surface residues close to the active site of TPxC60S (blue > hydrophobic; red < hydrophobic). All hydrogen bond contacts are illustrated by dashed lines.

Given that the virtual screening and X-ray crystallography efforts had together provided some evidence for a productive interaction of compound-40 with *Mtb* TPx, we investigated the impact of compound-40 on the nature of the NMR spectrum of isotope-labelled TPxC60S. Thus, a series of NMR titration experiments was recorded, as described in Section 2.3.7.4, to probe the effect compound-40 had on the 2D [^{15}N-^{1}H]-HSQC spectrum. The results of the titration of 1 mM [^{15}N]-labelled TPxC60S with a concentrated stock solution of compound-40 up to a final concentration of 50 mM are shown in Figure 5.14. (No further changes to the spectrum were detected at compound-40 concentration higher than 50 mM). The appearance of the spectra indicate that TPxC60S retains its overall structure throughout the titration series, as the majority of cross peaks are completely unperturbed by the presence of putative ligand (Figure 5.14A). However a small subset of cross peaks show a small but definite chemical shift change that varies monotonically as more compound-40 ligand is added. The chemical shift changes are rather small, in all but one or two cases amounting to less then the apparent cross peak line width (Figure 5.14C). The identity of each perturbed cross peak (where it is known) is listed in Table 5.3. Frustratingly, of these effects, the two largest are ascribed to resolved back-bone amide cross peaks for which the assignments have thus far eluded identification (peaks 1 and 3 in Table 5.3). By elimination these peaks must derive from the set of residues whose assignment has not been uncovered (Figures 4.16 and 4.17, Table 4.1).

Figure 8 corresponding cross peak numbering	Co-ordinates (ppm) HSQC (^{15}N-^{1}H)	Amino acid assignment	Location on *Mtb* TPxC60S crystal structure
1	107.80ppm – 7.10ppm	**UNKNOWN**	**N/A**
2	105.59ppm – 8.82ppm	Gly32	Loop connecting β1 and β2
3	112.40ppm – 5.77ppm	**UNKNOWN**	**N/A**
4	116.14ppm – 7.00ppm	**Glu149**	**Loop connecting α4 and β7**
5	116.35ppm – 7.63ppm	**Leu128**	**Loop connecting α1 and β6**
6	118.35ppm – 7.37ppm	**Gln152**	**Loop connecting α4 and β7**
7	114.51ppm – 8.68ppm	Asn100	Loop – connecting α2 and β5
8	122.07ppm – 7.77ppm	**Phe67**	**α1**
9	124.35ppm – 7.39ppm	Tyr156	α4
10	128.13ppm – 8.31ppm	Val147	Loop connecting α4 and β7
11	120.72ppm – 8.31ppm	**Ile150**	**Loop connecting α4 and β7**
12	122.05ppm – 8.33ppm	**Ala151**	**Loop connecting α4 and β7**
13	122.11ppm – 8.42ppm	Glu153	Loop – connecting α4 and β7

Table 5.3: The co-ordinates and amino acid assignments of the perturbed amino acid residues identified on addition of compound-40 to TPxC60S, as displayed in Figure 5.14A–C. More highly shifted cross peaks are highlighted in bold.

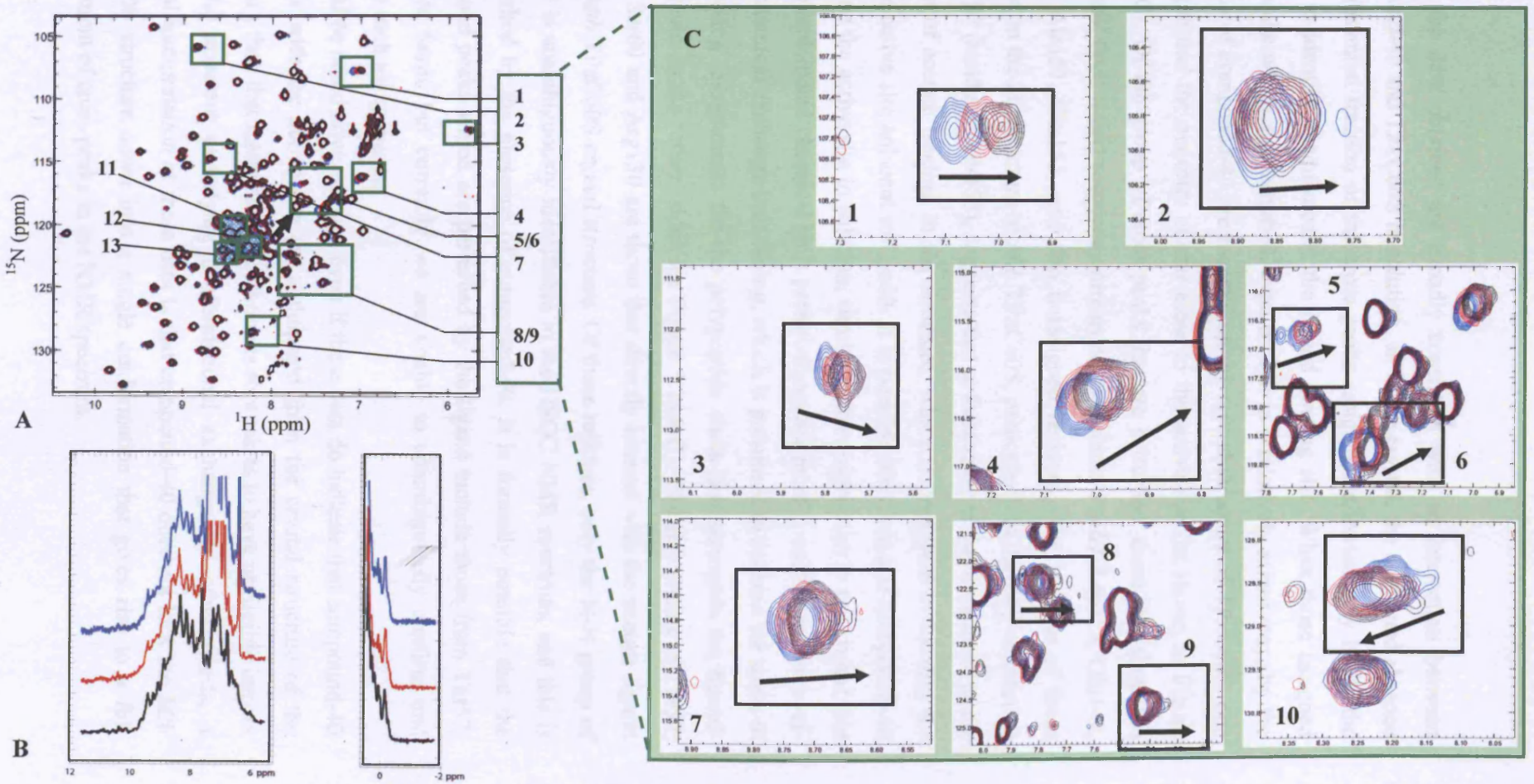
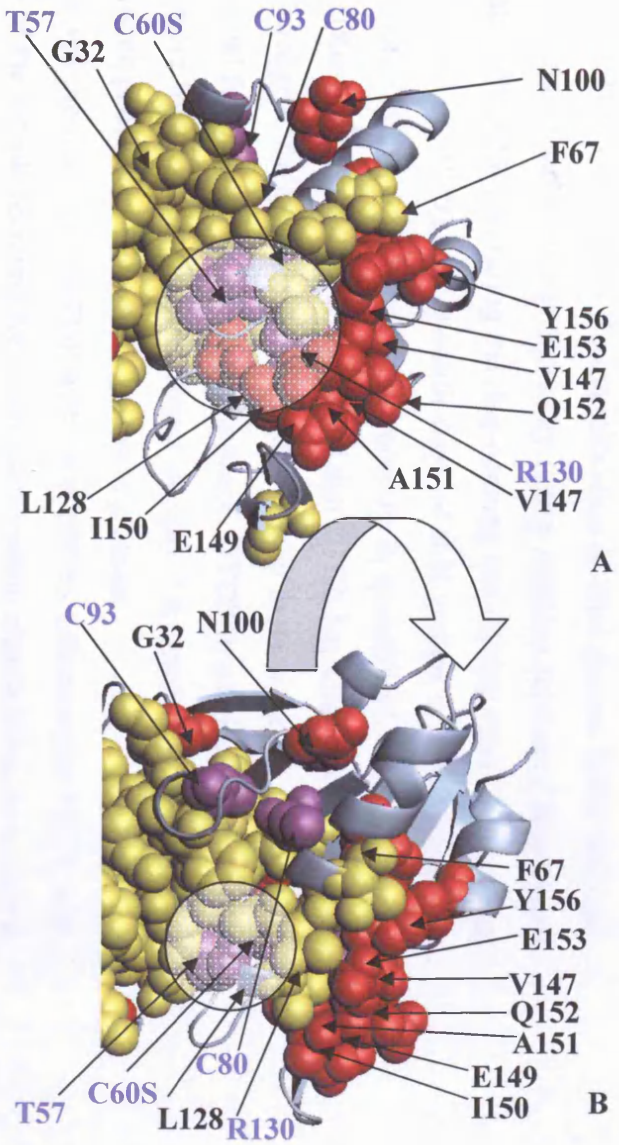


Figure 5.14: **A**: Titration HSQC and **B**: 1D NMR spectral overlay of TPxC60S with compound-40. Green boxes and green boxes with pale blue masking highlight cross peaks that are selectively perturbed in the presence of the ligand - blue spectrum = 0 mM, red spectrum = 30 mM, black spectrum 50 mM of compound-40. **C**: Green boxed sections only from Figure 5.14A showing the chemical shift perturbation on increasing compound-40 addition (direction indicated by the arrows). All spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

Whilst the data obtained are broadly consistent with an interaction between compound-40 and TPxC60S in solution, as in the crystal, the continued absence of a substantial fraction of the cross peaks assignments intrinsically limits the ability to identify unequivocally the ligand binding site. When those assigned cross peaks which are perturbed, even though in a relatively minor way, by the presence of compound-40 are mapped on the crystal structure of TPxC60S, one can argue that the majority cluster close to the active site (as shown in Figure 5.15A-C). Of the group of cross peaks that are perturbed, those that show the 'strongest' (still weak) response correspond to residues Phe67, Leu128, Glu149, Ile150, Ala151, Gln152, and two unassigned residues. The location of these residues on the 3D structure of *Mtb* TPxC60S, particularly in the loop connecting α4 and β7 (residues 149-153), suggests that at least some of these are close to the location of acetate binding in the structure, which is presumed to identify the putative active site adjacent to Cys60. It is possible therefore that compound-40 does bind the active site in solution, similar to the observation in the crystal, but that the anticipated chemical shift perturbations are mostly masked by the over-riding chemical exchange broadening, which is presumed to account for some of the 'missing' assignments for the polypeptide chain that surrounds the ligand-binding site on the 'other' side (see Figure 5.15A-C). The side chains of Thr57, Val59, Ser60 and Arg130 are those that directly interact with the acetate ligand in the *Mtb* TPxC60S crystal structure. Of these residues, only the N-H group of Arg130 is unambiguously identifiable in the HSQC NMR spectrum, and this is unperturbed by the presence of compound-40. It is formally possible that the unassigned peaks which are perturbed by the ligand include those from Thr57, Val59 or Ser60 but currently we are unable to unambiguously confirm and validate such assignments.

It should be noted however that, even if these data do indicate that compound-40 interacts with the active site (as anticipated from the crystal structure of the complex), then that interaction *in solution* is so weak as to have negligible impact upon the apparent underlying conformational exchange. In other words, a potential interpretation of these data is that compound-40 does not lock the *Mtb* TPxC60S structure down into a single conformation that gives rise to a full complement of cross peaks in the NMR spectrum.



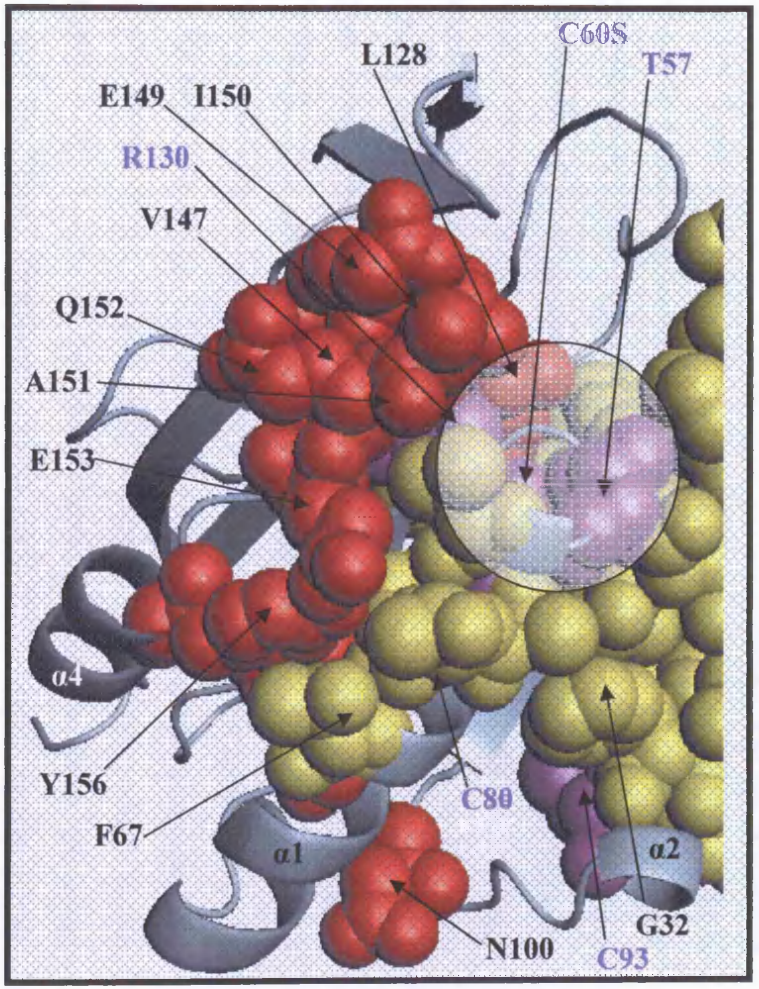


Figure 5.15: **A - C**: Crystal structure of *Mtb* TPxC60S (PDB: 1Y25) with compound-40, perturbed residues coloured red and catalytic residues coloured magenta. Missing residues, from back-bone assignment, are coloured yellow. Active binding site highlighted in semi-transparent circled region. **B**: 15° rotation of **A**. **C**: close-up of a single monomer of TPxC60S. Drawn using PyMOL™.

5.7.3. 4,4'-dithiodipyridine (DTDP).

Thiol chemistry is a rapidly expanding field in basic and applied bioscience (Packer, 1995 and Mrksich and Whitesides, 1996). The quantitative measurement of protein S-H groups is a task achieved using one of several differing methods which include electrochemical (Nishiyama and Kuninori, 1992), fluorimeteric (Kirley, 1989), and spectropotometric assays (Reiner *et al*, 2002). Of the three different methods, the spectrophotometric assay system is both rapid and easy to implement at the expense of a decrease in sensitivity. One of the most widely used spectrophotometric methods in the quantification of thiols is Ellman's assay. As is shown in Figure 5.16A, the assay involves the reaction of a mercaptan thiol (R-SH) with Ellman's reagent 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB). The reaction splits the symmetric DTNB into the mixed disulphide (R-S-TNB) releasing one equivalent of free thionitrobenzoate anion (TNB^{2-}) for each S-H group. The reaction can be followed spectrophotmetrically by the monitoring of the intense light absoption at 412 nm, which is pH independent at pH > 7.3, of the thionitrobenzoate product. Both the educt (DTNB) and the mixed disulphide product (R-S-TNB) absorb poorly at this wavelength.

In a similar manner to the use of Ellman's reaction, 2,2'-dithiodipyridine (DTDP) (Figure 5.16B) can be used for the quantification of thiol groups. DTDP has been used in characterisation of enzymatically active cysteine residues (Brockehurst and Little, 1973), for measuring the ring-opening and closing rates of thiosugars (Grimshaw *et al*, 1979), the quantification of S-H groups in small molecules (Egwim and Gruber, 2001), and most recently in quantification of cysteine in proteins (Riener *et al*, 2002). It is reported that DTDP has some advantages over Ellman's reagent, for example, in being intrinsically more sensitive, though with the potential for higher background absorbance. DTDP is relatively small and it is also more hydrophobic and uncharged at ~pH 7.0, consistent with a faster reaction with poorly accessible thiol groups in proteins.

We followed the reaction of DTDP with TPxC60S by heteronuclear NMR, with aliquots of the sample removed for spectrophotometric examination. As a control the reaction of DTDP with bovine serum albumin (BSA) was concurrently

monitored by the optical absorbance at 324 nm (Figure 5.17). The full methodological details of these experiments are described in Section 2.3.9. Increasing concentrations of DTDP were incubated with the protein substrate for 50 minute periods during which the 324 nm absorbance reading was allowed to come to settle. Curiously the reaction of DTDP with TPxC60S, monitored by absorbance as a function of time (but including increasing DTDP loadings as indicated in Figure 5.17) follows a pseudo-sigmoidal pattern (Figure 5.18). With both BSA (Figure 5.17A) and TPxC60S (Figure 5.17B) prepared in the absence of DTT the UV absorbance at 324 nm was found to increase in the presence of excess DTDP indicating that the thiol reactive agent is being consumed. In each case the total absorbance appears to approach a limiting value suggesting saturation of the thiol binding sites. Quantitatively the data suggest that BSA contains < 1 mole of reactive thiol groups per mole of protein, similar to measurements for BSA reported in the literature (Riener *et al*, 2002). For TPxC60S the data suggest the presence of potentially two reactive thiol groups (Cys80 and Cys93) per mole of protein homodimer, suggestive of the presence of a single reactive cysteine in each protomer. [An absorbance reading taken five days later indicated no further reactivity of TPxC60S. The non-integral number of reactive thiol groups obtained presumably suggests some error in the estimation of the protein concentration based upon the predicted extinction coefficient.]

Concurrent with the spectrophotometric analysis we followed the reaction of DTDP with TPxC60S by 2D [^{15}N,^{1}H]-HSQC NMR spectrum (Figure 5.18). A spectrum was recorded for each addition of DTDP. Figure 19 displays a subset of the spectra recorded without addition of DTDP, with DTDP concentrations of 1 mM, 2.6 mM, and 4.6mM, and finally a spectrum recorded after prolonged exposure to excess DTDP.

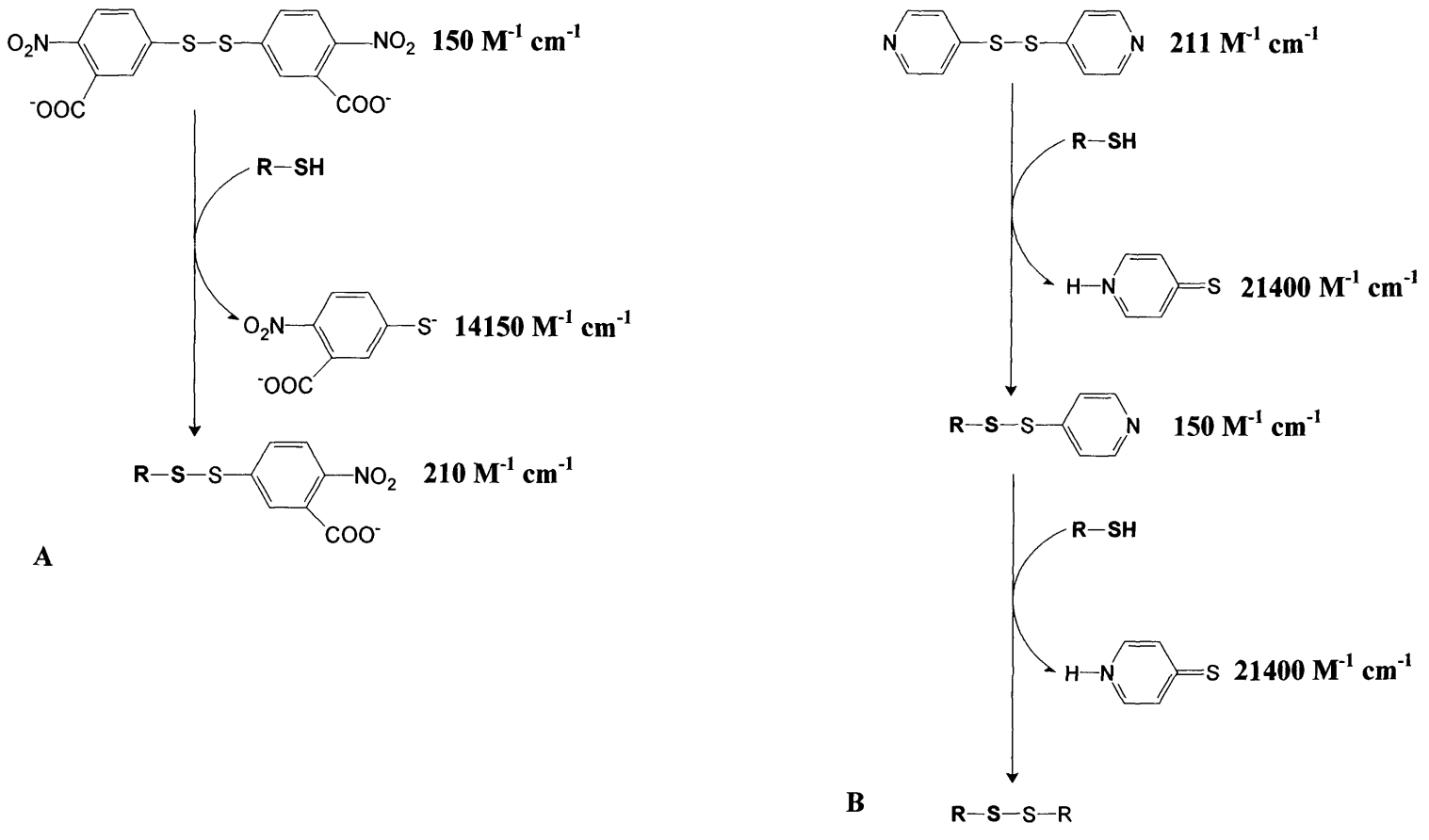
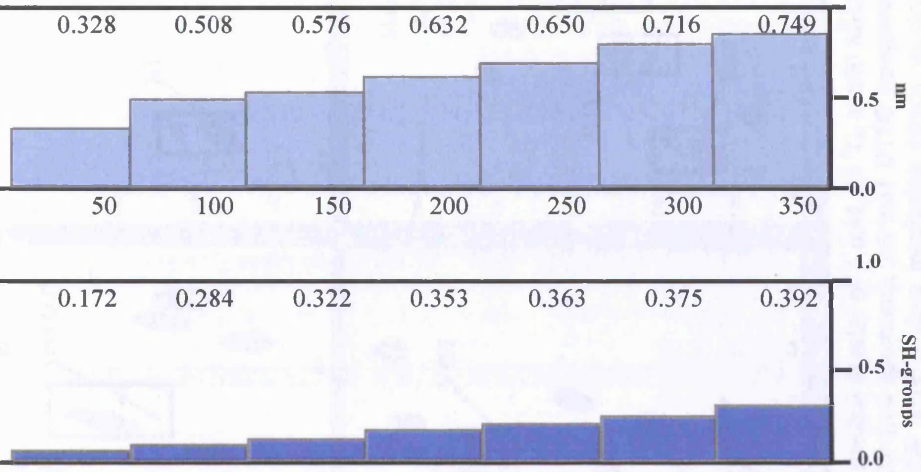


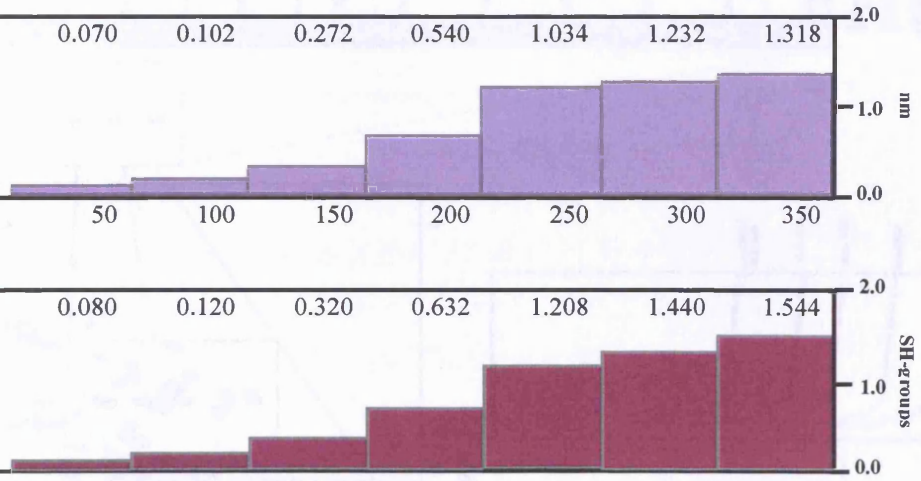
Figure 5.16: **A**: Reaction of a mercaptan with Ellman's reagent, yielding a mixed disulfide and one equivalent of TNB^{2-}. The molar absorptivities (ε_{412}) all refer to pH > 7.29. **B**: Two-step reaction of a mercaptan with DTDP, yielding one equivalent of 4-thiopyridone in each step. The molar absorptivities (ε) all refer to λ = 324 nm. Adapted from Reiner *et al*, 2002 and drawn using BioRad's KnowItAll™.

Experiments:	1	2	3	4	5	6	7
Reagent (DTDP) concentration (µM):	200	400	1000	1600	2600	3600	4600
Absorbance at 324nm:	0.328	0.508	0.576	0.632	0.650	0.716	0.749
Reagent/ protein incubation time (min):	50	100	150	200	250	300	350
Free –SH group reacting*:	0.172	0.284	0.322	0.353	0.363	0.375	0.392



A

Experiments:	8	9	10	11	12	13	14
Reagent (DTDP) concentration (µM):	200	400	1000	1600	2600	3600	4600
Absorbance at 324nm:	0.070	0.102	0.272	0.540	1.034	1.232	1.318
Reagent/ protein incubation time (min):	50	100	150	200	250	300	350
Free –SH group reacting*:	0.080	0.120	0.320	0.632	1.208	1.440	1.544



B

Figure 5.17: A: Measurement of BSA sulfhydryls by the DTDP method results. **B**: Measurement of *Mtb* TPxC60S sulfhydryls by the DTDP method results. * denotes value obtained from the equation in Section 2.3.9.

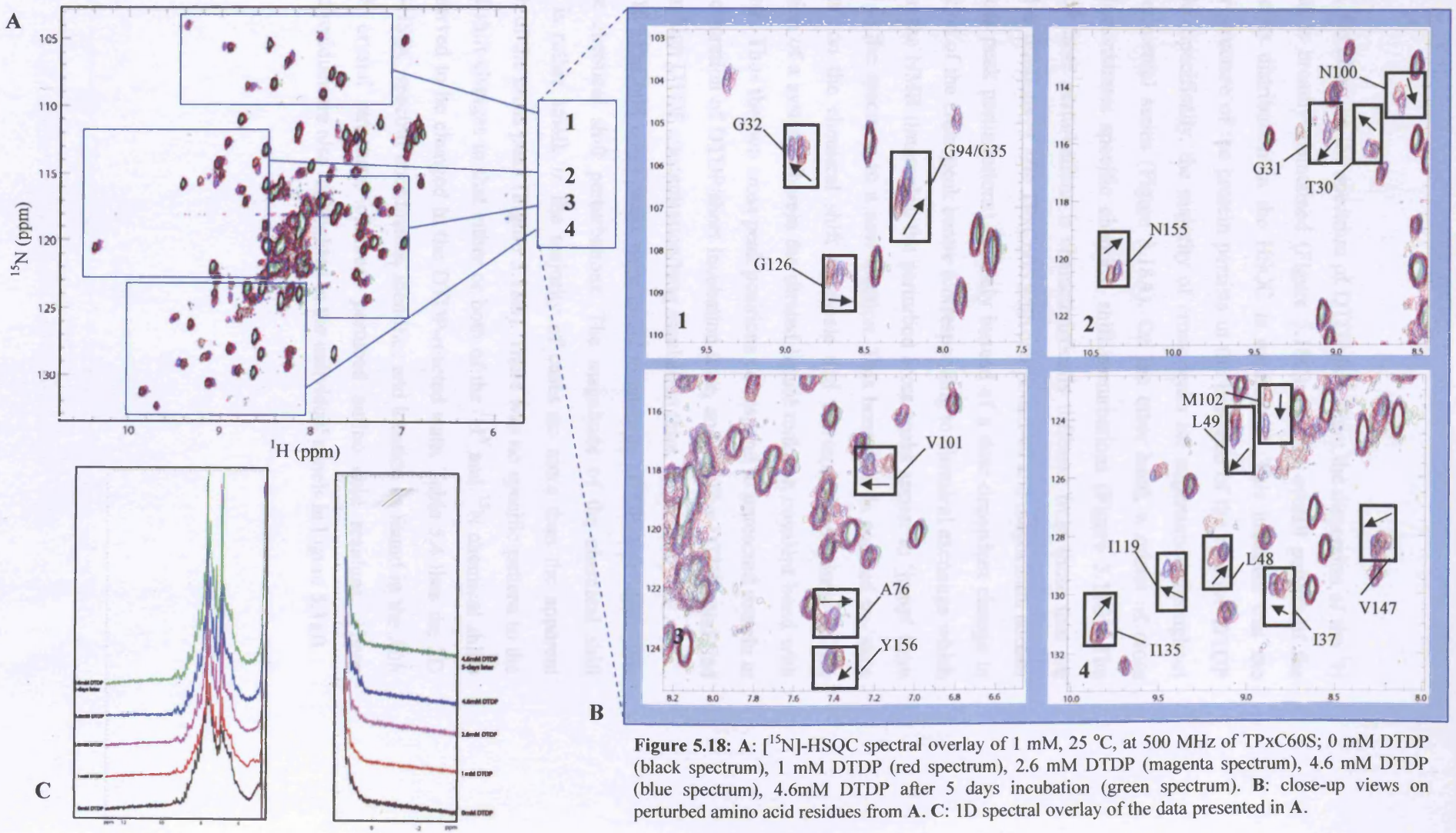


Figure 5.18: A: [^{15}N]-HSQC spectral overlay of 1 mM, 25 °C, at 500 MHz of TPxC60S; 0 mM DTDP (black spectrum), 1 mM DTDP (red spectrum), 2.6 mM DTDP (magenta spectrum), 4.6 mM DTDP (blue spectrum), 4.6mM DTDP after 5 days incubation (green spectrum). **B**: close-up views on perturbed amino acid residues from **A**. **C**: 1D spectral overlay of the data presented in **A**.

Overall, the effect on the spectrum of DTDP is benign: the dispersion of the ^{1}H spectrum is broadly maintained (Figure 5.18C) and the overall pattern of the cross peaks distribution in the HSQC is maintained. This indicates that the globular structure of the protein persists in the presence of the reactive DTDP molecule. Specifically, the majority of cross peaks are unperturbed throughout the experimental series (Figure 5.18A). On the other hand, a subset of cross peaks demonstrates specific chemical shift perturbations (Figure 5.18B). The nature of these perturbations is characteristically different from those that are detected in titrations of *Mtb* TPxC60S with compound-40 and magneisum acetate (four cross peak perturbations). Namely instead of a dose-dependent change in the location of the cross peak centre (corresponding to chemical exchange which is fast on the NMR timescale), the perturbed cross peaks appear to 'jump' from location in the spectrum to a new location. This behaviour is equated to 'slow exchange' on the chemical shift timescale and corresponds precisely to the expectation of a system wherein the titrated ligand makes a covalent bond with the protein. Thus the two cross peak positions correspond to unreacted protein at low concentration of DTDP/short incubation time, and the Cys-DTDP modified protein at high DTDP concentration/long incubation time, respectively.

In total 18 TPxC60S cross peaks were found to undergo DTDP-dependent slow exchange chemical shift perturbations. The magnitude of the chemical shift changes is rather small, in the majority of cases no more than the apparent linewidth of the cross peak (Figure 5.18B). There was no specific pattern to the chemical shift changes in that either or both of the $^{1}H^{N}$ and ^{15}N chemical shifts were observed to be changed in the DTDP-reacted state. Table 5.4 lists the 2D [^{15}N,^{1}H]-HSQC spectral coordinates, identities and location as found in the *Mtb* TPxC60S crystal structure, of these pertubed amino acid residues. These perturbed residues are also highlighted in the individual panels in Figure 5.18B.

Perturbed residue:	^{15}N, ^{1}H chemical shift coordinates:	Location in *Mtb* TPxC60S crystal structure
T30	116/8.8	Loop connecting β1 and β2
G31	115/9.0	Loop connecting β1 and β2
G32	106/8.9	Loop connecting β1 and β2
G35	107/8.3	β2
I37	130/8.8	β2
L48	129/9.2	β3
L49	126/9.5	β3
A76	123/7.5	Loop connecting α1 and β4
G94	107/8.3	α3
N100	114/8.7	Loop connecting α3 and β6
V101	118/7.1	Loop connecting α3 and β6
M102	124/9.0	Loop connecting α3 and β6
I119	129/9.4	Loop connecting α3 and β6
G126	108.5/8.7	Loop connecting α3 and β6
I135	131.5/9.8	β6
V147	128.5/8.2	Loop connecting α4 and β7
N155	120.5/10.4	Loop connecting α4 and β7
Y156	124/7.4	α4

Table 5.4: A list of perturbed amino acid residues on incubation of DTDP with TPxC60S.

The perturbed amino acid residues have been mapped onto the *Mtb* TPxC60S crystal structure (PDB: 1Y25), in order to observe their location relative to the potentially DTDP-reactive cysteine side chains in the protein (Figure 5.19). It is noteworthy that Cys80 is essentially completely buried in the core of the protein structure, which its side chains forming van der Waals contacts with the side chains of Val78, Ser82, Val101, and Ala104, based upon their proximities. A calculation with program NACCESS indicates that only 2 % of the side chain is exposed to a solvent probe with 1.4 Å radius. On the other hand the side chain of Cys93 is more exposed, ~18 % according to NACCESS, and could be envisaged to become more exposed to the protein surface with relatively minimal changes in a small number of torsion angles.

All 18 cross peaks that are perturbed in the DTDP titration, correspond to residues that cluster predominantly in a region that could more sensibly be described as surrounding Cys93, rather than Cys80. As Table 5.4 describes several residues are found on alpha helices α2 and α4, with others located on beta-strands β1-β4 and β7. However, a significant proportion is located on various interconnecting loop regions. In both protomers, the majority of perturbed residues cluster close to the dimer interface – a region believed to encompass mobile secondary structural elements hindering the backbone resonance assignment of particular sections of the TPxC60S amino acid sequence.

The results of the DTDP titration are as follows. Clearly the slow exchange characteristic of the spectral perturbations is consistent with a covalent modification of the protein attributable to reaction of DTDP with an exposed thiol group. Given the apparent stoichiometry of the reaction which suggests that not more than 1 Cys residue per TPxC60S protomer reacts with DTDP, and the pattern of residues in the structure of the protein that are sensitive to the chemical modification, it is presumed that DTDP reacts with the more exposed Cys93 side chain. It seems extremely unlikely that Cys80 could react with DTDP with global remodelling of the protein structure – a circumstance that is inconsistent with the rather limited chemical shift effects in the titration. Interestingly since from the NMR perspective the reaction appears to go essentially to completion, i.e. that

the cross peaks corresponding to the un-reacted protein diminish to vanishing intensity by the end of the experiment, it appears that the Cys93 residues in each half of the TPxC60S homodimer are equally reactive. Notably there is no evidence in the spectra for a Cys93-SH/Cys93-SS-TP heterodimer, an asymmetric species that might give rise to an increase in the numbers of NMR cross peaks due to symmetry breaking. Assuming that this species does not represent the end point of the reaction, the fact that we do not detect this species at intermediate points in the titration indicates that the chemical shift effects of the DTDP-modification are 'local' to the site of reaction. That is, any structural perturbation that accompanies the modification of the Cys93 residue is not transmitted across the dimer interface (a conclusion which again would be consistent with the relatively small magnitude of the DTDP-induced chemical shift changes).

Yet again this experiment demonstrates that it is possible to perturb the spectrum of TPxC60S, albeit in a relatively minor way, without inducing the appearance of 'new' cross peaks. Thus even the covalent attachment of a chemical moiety to the side chain of Cys93 appears insufficient to quench the intermediate timescale polypeptide dynamics that has been invoked in this thesis to account for the 'missing' cross peaks.

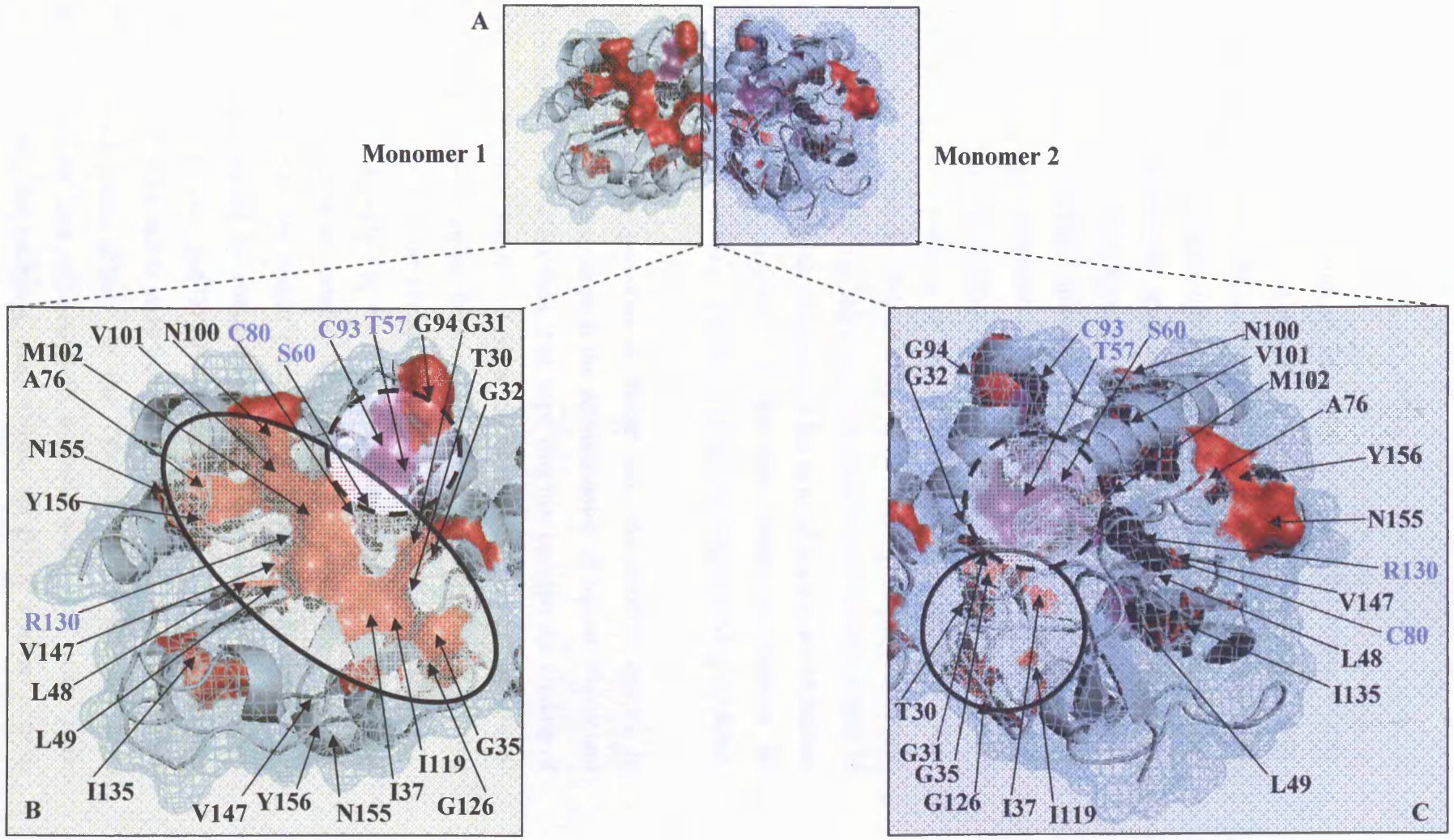


Figure 5.19: **A**: TPxC60S crystal structure (PDB: 1Y25) exhibiting perturbed residues on addition of DTDP. **B**: Monomer 1 and **C**: Monomer 2: perturbed residues situated close to dimer interface. Both monomers show perturbed residues to cluster (heavy outline) close to the recognised TPxC60S binding pocket (dashed outline).

5.8. Ligand binding summary.

The analysis of the NMR spectrum of TPxC60S throughout the course of this work has been impacted by the inability to complete the back-bone resonance assignment, in large part due to the apparent absence of 30 N-H cross peaks, and leaving 18 observable cross peaks unidentified (see Chapter 4). The underlying assumption that has pervaded our approach to this aspect is that the cross peak absences are the result of dynamical exchange broadening, presumed to arise from interconversion between multiple conformational sub-states on the intermediate timescale. As discussed already, it is attractive to imagine that the conformational exchange is related, or directly attributable to the observed differences between the structures of oxidized and reduced L-atypical 2-Cys peroxiredoxins (see Chapter 3): perhaps even the TPxC60S protein is constantly sampling the conformational trajectory between these two structures, which is predicted to include local folding/unfolding of the ends of the a1 and a4 helices and permitting the close approach of the side chains of residues 60 (corresponding to the peroxidatic cysteine, C_P) and 93 (the resolving cysteine, C_R).

We sought to explore the potential to 'freeze out' the putative motion, to 'improve' the NMR spectrum through the administration of ligands that might interact with the active site residues. The hope was that perhaps the binding of the ligand would either inhibit or slow down the rate of conformational exchange, thereby either trapping the structure in one conformation, or limiting the line broadening effects of the exchange respectively.

Specifically we tested the effect on the TPxC60S spectrum of acetate ions (found to occupy the active site in the original TPxC60S crystal structure); compound-40 (a known low molecular weight pharmaceutical substance identified as a potential active site ligand by computational screening and subsequently co-crystallised with the protein); and DTDP (a material that is reactive with cysteine side chains would be predicted to add a bulky chemical moiety to the thio-S atom of Cys93). In each case, effects on the TPxC60S NMR spectrum were observable: for acetate and compound-40 a small number of cross peaks migrated a little with a fast exchange characteristic; for DTDP a subset of cross

peaks underwent a step-wise (as if 'slow exchange') shift consistent with a covalent modification. Importantly however in each case the chemical shift perturbations were few in number and small in magnitude, often no more than the width of the cross peak, and unfortunately no 'extra' cross peaks were ever detected. The conclusion is that the NMR data suggest that whilst each ligand may well interact with the active site as expected, in no instance is that interaction so strong as to influence the proposed conformational exchange so as to permit the completion of the back-bone resonance assignment. Table 5.5 summarizes the results obtained from the three compounds used during these experiments.

Compound used in ligand binding:	**Protein used:**	**The number of cross peak perturbations:**	**Any new cross peaks in the resulting HSQC spectra:**
Magnesium	TPxC60S	4	NO
acetate	TPxWt	0	NO
Compound-40	TPxC60S	13	NO
DTDP	TPxC60S	18	NO

Table 5.5: Summary of the ligand binding experiments conducted during the course of this project using the two different *Mtb* TPx variants.

5.9. Oxidised vs. reduced TPxWt and the significance of the Cys60–Cys93 disulfide bond.

As has been shown elsewhere in this thesis (Chapter 4), the NMR spectra obtained for the wild-type *Mtb* TPx, whilst resembling the spectra for the C60S mutant, are always of a poorer quality. For example, whilst the chemical shift dispersion of the ^{1}H 1D spectra of both TPxWt and TPxC60S forms are typical of folded globular proteins, the 2D [^{15}N,^{1}H]-HSQC spectrum of the wild-type contains many fewer cross peaks than for the mutant. (Importantly the subset of peaks that are visible for TPxWt has a similar distribution of chemical shifts to the peaks also detected in the mutant, again indicating an overall similar globular

fold.) We have been intrigued in how the simple substitution of a single sulphur atom in the thiol group of Cys60 by an oxygen atom in each protomer chain could have such a dramatic influence on the overall quality of the NMR spectrum. We thought it important to check whether the chemical reactivity represented by the Cys60 side chain (which is after all the peroxidatic active site group), where it could potentially covalently bond with other parts of the same or other protein chains, was responsible for the overall appearance of the spectrum. To this end we recorded NMR spectra of *Mtb* TPxWt in various relatively elevated concentrations of DTT.

Figure 5.20 shows a series of 1D ^{1}H and 2D [^{15}N,^{1}H]-HSQC spectra of isotope-labelled wild-type TPx recorded with DTT concentrations up to 10 mM. Whilst there were hints from the 1D spectra that the line width of the resonances was slightly improved in the presence of excess DTT there is relatively little difference between the corresponding 2D spectra. This result suggests that in each case the protein maintains a consistent oligomeric state and that it remains essentially completely reduced. There is a consistent difference in the 'quality' of these NMR spectra compared to that of the TPxC60S mutant (Figure 5.20C), particularly in terms of the visible cross peak count.

5.10. Oxidised vs. reduced TPxWt: summary.

Clearly the spectrum of the wild type protein is impacted by line broadening effects that are more extensive even than those that appear to operate in the spectrum of the C60S mutant. Formally it is possible that the extra broadening could result from higher order self-association of the protein (e.g. by the transient formation of dimers of homodimers) leading to an overall increase in the average rotational correlation time. NMR spectra recorded at 0.5 and 1.0 mM protein concentration are essentially the same (data not shown), and this observation and the non-uniform nature of the comparative peak intensities between the wild-type and C60S TPx spectra argue against this interpretation. Rather we suspect that the additional line broadening observed for the wild-type protein may reflect enhanced amplitude of exchange broadening associated with the putative conformational exchange that has been described above for the mutant protein.

Consistent with this interpretation we note that there is no cross peak in the spectrum of the wild-type that does not occur in that of the mutant. Most likely there is a common mechanistic origin for the line-broadening in each case, with the rate or amplitude of the putative internal motion modulated by the chemistry of the side chain of residue 60. In this respect it is intriguing to speculate that the rate determining step of the mechanism of exchange depends upon the relative strength of the non-bonded (perhaps H-bonding) contacts for the Cys60 thiol –SH and Ser60 hydroxyl –OH moieties respectively.

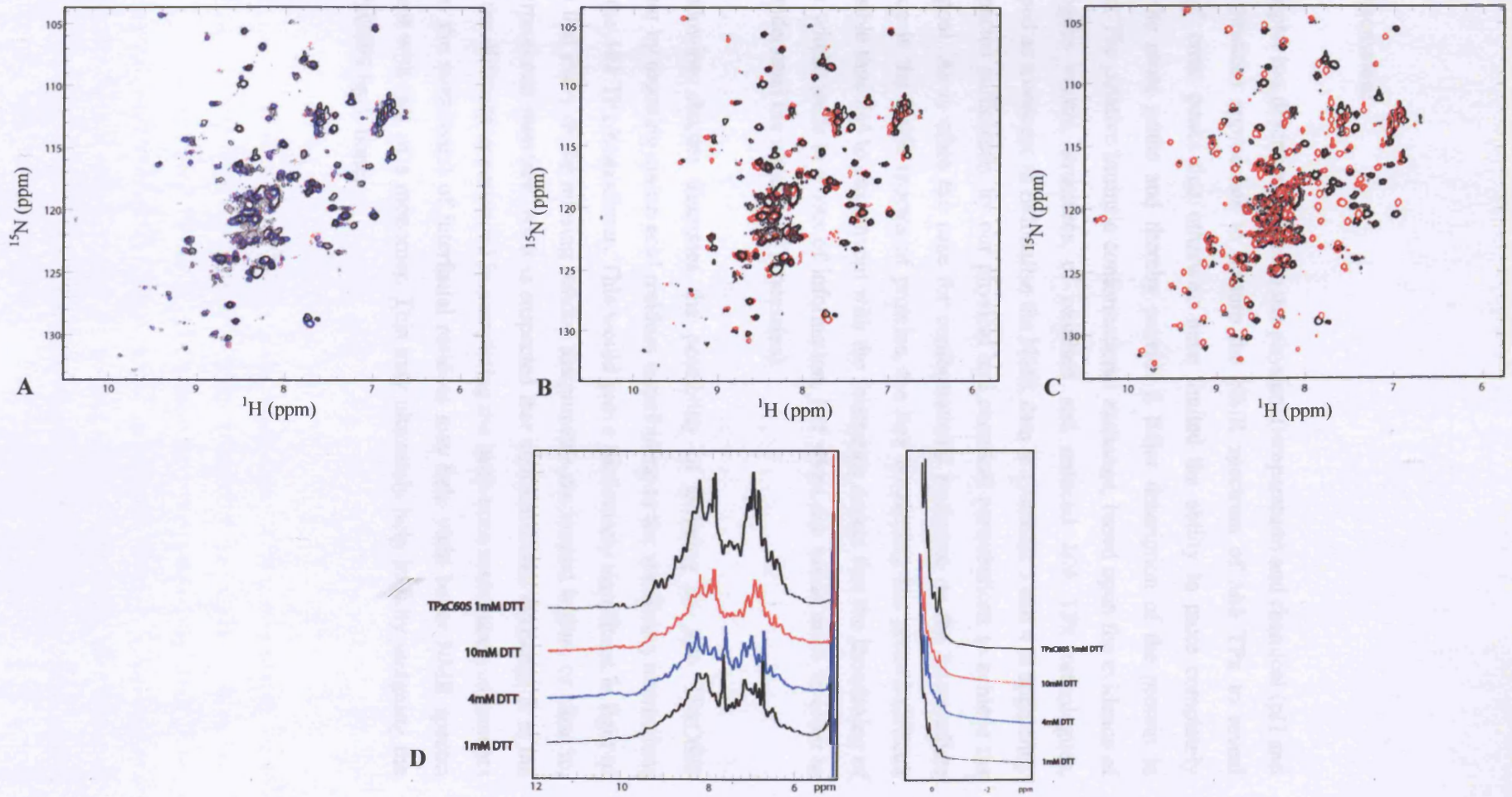


Figure 5.20: **A**: HSQC spectral overlay of TPxWt at various DTT concentrations: 1 mM (red), 4 mM (magenta) and 10 mM (black). **B**: Two isolated spectral HSQC overlays of TPxWt at two different DTT concentration extremes: 1 mM (red) and 10 mM (black). **C**: Two HSQC spectral overlays. One of TPxWt at 10 mM DTT (black) and the other of TPxC60S at 1 mM DTT (red). **D**: 1D spectral overlay of TPxWt at differing DTT concentrations, as seen in **A**, and TPxC60S at 1 mM DTT (black spectrum – top). All spectra obtained at 500 MHz, at 1 mM, pH 7.48 and 25 °C.

5.11. Discussion:

This chapter has described efforts to use physical (temperature) and chemical (pH and ligand titration) approaches to perturb the NMR spectrum of *Mtb* TPx to reveal 'missing' cross peaks that otherwise have limited the ability to more completely assign the cross peaks and thereby provide a fuller description of the protein in solution. The putative intrinsic conformational exchange, based upon the evidence of homologous crystal structures of oxidised and reduced *Mtb* TPx orthologues, developed as a concept to rationalise the NMR data in Chapters 3 and 4 is apparently not impacted sufficiently by our physical and chemical perturbations to achieve the stated goal. As is often the case for conformational exchange on the intermediate timescale in the NMR spectra of proteins, the line broadening has proved difficult (impossible thus far) to circumvent with the frustrating aspect that the broadening of the line widths leads to a loss of information, just when one needs more in order to better understand the underlying process(es).

The following chapter discusses the possibility of isolating an *Mtb* TPxC60S monomer by mutating amino acid residues contributing to the stabilizing interactions within the *Mtb* TPx homodimer. This would prove particularly significant in light of the fact that many of the missing residue assignments are located within, or close to, the interprotomer interface. As it is suspected that conformational exchange is at the root of the difficulty experienced in completing the back-bone resonance assignments for TPx, the mutation(s) of interfacial residues may help yield better NMR spectra consistent with that of a monomer. This may ultimately help in fully assigning the *Mtb* TPxC60S back-bone.

Chapter 6

Project summary, conclusions and discussions

The general approach to the rationalisation of the appearance of the NMR spectrum of *Mtb* TPx in this thesis has been slanted towards an expectation that the protein is undergoing conformational exchange that is related in some respect to the observed differences in the X-ray structures of oxidised and reduced *Mtb* TPx and species homologues. Because of the exchange broadening, which leads to a loss of information because of diminished, often vanishing, cross peak intensity we are intrinsically inhibited from using the NMR spectrum to unravel the nature of the exchanging conformations in solution. There is another aspect to this scenario that has so far not been exposed in this description. Namely, it is formally possible that conformational exchange that is proposed may also involve reorganisation of the inter-protomer interface of the *Mtb* TPx homodimer. Whilst no such rearrangement of the interface is suggested by the available crystal structures, it remains possible that the interface in solution is conformationally 'plastic'. Examination of the crystal structure of *Mtb* TPx indicates that the buried surface area in the homodimer contact is ca. 725 $Å^2$ per protomer. Whilst not a particularly small contact surface area, it is not incomparable with instances where NMR data have implied or confirmed that a dimer in a crystal is able to disassociate in solution, albeit to a greater or lesser extent (Jones and Thornton, 1996 and Plevin *et al*, 2007).

In this context it is interesting to recount the recent example, from work in the same laboratory, where the bacterial enzyme Dimethylarginine dimethylaminohydrolase (DDAH), a homodimer of two 29 kDa promoters, was studied by both X-ray crystallography and NMR. The contact surface buried in the interface in this instance was ca. 860 $Å^2$ per protomer, and the protein behaved essentially as a dimer in solution at concentrations used for NMR experiments. When resonance assignment of

the [^{15}N, ^{13}C, ^{2}H]-labelled DDAH was attempted, several cross peaks could not be assigned or were not detected. Again it was suspected that conformational exchange was at the root of this difficulty, though in this instance it was much more the case that reorganisation of the protomer-protomer interface, rather than any more gross reorganisation of the protein structure, is the mechanism by which the selective exchange broadening arises. (This conclusion was drawn in part because of the five-bladed β/α-propeller nature of the DDAH fold – it seems rather unlikely that this structure can locally unravel without global effects on the NMR spectrum – plus there was no crystallographic data to support the invocation of a locally reorganised polypeptide topology). On the basis of a per-residue analysis of the inter-protomer contact surface, mutations of interfacial residues were introduced into the DDAH protein and these variants were screened by analytical size exclusion chromatography and sedimentation equilibrium analytical ultracentrifugation to assess the hydrodynamic status. Ultimately a double mutant DDAH was derived that behaves exclusively as a monomer, and retains a level of biological activity equivalent to the wild-type. The double mutant DDAH yielded essentially ideal NMR spectra, consistent with a monomeric species. Full resonance assignments were obtained in a straightforward manner for this form of the protein. Comparison of these results with the spectrum of the wild-type homodimer indicated that the 'missing' cross peaks in the latter case derived from residues either at or close to the interprotomer interface in the crystal structure, lending support to the proposition that the exchange broadening could arise from local reorganisation of that interface in solution.

Given this example, which in many respects mirrors concepts that have arisen in the examination of the *Mtb* TPx protein, it is instructive to consider whether a similar analysis of the homodimer interface observed in the crystal might provide a basis for decoupling the exchange broadening effects that could arise from the proposed local reorganisation of the α1 and α2 helices (including the C_P and C_R active site residues) and any conformational plasticity of the homodimer interface.

Thus, the sites that constitute the interaction surface between the *Mtb* TPx homodimer protomers in the homodimer crystal contact surface were assessed in a residue-by-residue fashion using NACCESS, a program used to calculate the solvent accessible surface area (*SASA*) of a molecule from a PDB (Protein Data Bank) format file. To

identify the residues involved that contribute stabilizing interactions in the *Mtb* TPx homodimer, we examined the predicted changes in side chain *SASA* for each residue *i* ($\Delta SASA_i$) upon self-association of two monomers into the symmetric homodimer. By this assessment, a total of 19 residues - spanning a total area of ca. 725 Å^2 per protomer – exhibit a fractional change in side chain *SASA* (i.e. $fSASA_i > 0$), as shown in Figure 6.1B. The value of $fSASA_i$ is calculated according the formula below:

$$f\mathrm{SASA}_i = \frac{SASA_{i\,monomer} - SASA_{i,\mathrm{dimer}}}{SASA_{i,\mathrm{dimer}}}$$

For a residue whose side chain is entirely enclosed in the interface $fSASA_i \rightarrow 1$; for a residue entirely outside the contact surface $fSASA_i \rightarrow 0$. The residues exhibiting the highest $fSASA_i$ (> 0.8) are shown in red in Figure 6.1B and include: Ser54 and Asp56 (both located on the loop region between β3→α1); Asp84 and Leu85 (both located on the loop region between β4→α2); Pro86, Phe87 and Ala87 (all three located on α2); and Leu127 (located on loop region linking β6→α3). Additional residues that show a smaller value of $fSASA_i$ are indicated by the graded colour scheme in Figure 6.1B. Judged simply on the number of residues implicated by this calculation, only ~10 % of the residues in the protein contribute to dimer formation to any degree and the rest of the surface appears white in the diagram.

Of the side-chain exposed residues located at the interacting surface, a sub-total of nine residues display a significant absolute change in *SASA* (i.e. $\Delta SASA_i > 25$ Å^2) accounting for 509 Å^2 of the interacting side chain regions per protomer (corresponding to 81 %), as shown in Figure 6.1C. Of these residues, two (Leu85 and Phe87) are exhibit the dominant contribution to the total side chain $\Delta SASA$ for the homodimer interface. Examination of the crystal structure of *Mtb* TPx indicates that the opposing Leu85 residues form a pair-wise symmetric hydrophobic side chain-side chain contact, buttressed on each side (as it were in *cis*) by the side chains of Phe87. Each Phe87 phenyl ring makes a hydrophobic interaction, in *trans*, with the methyl group of Thr57 of the opposite protomer.

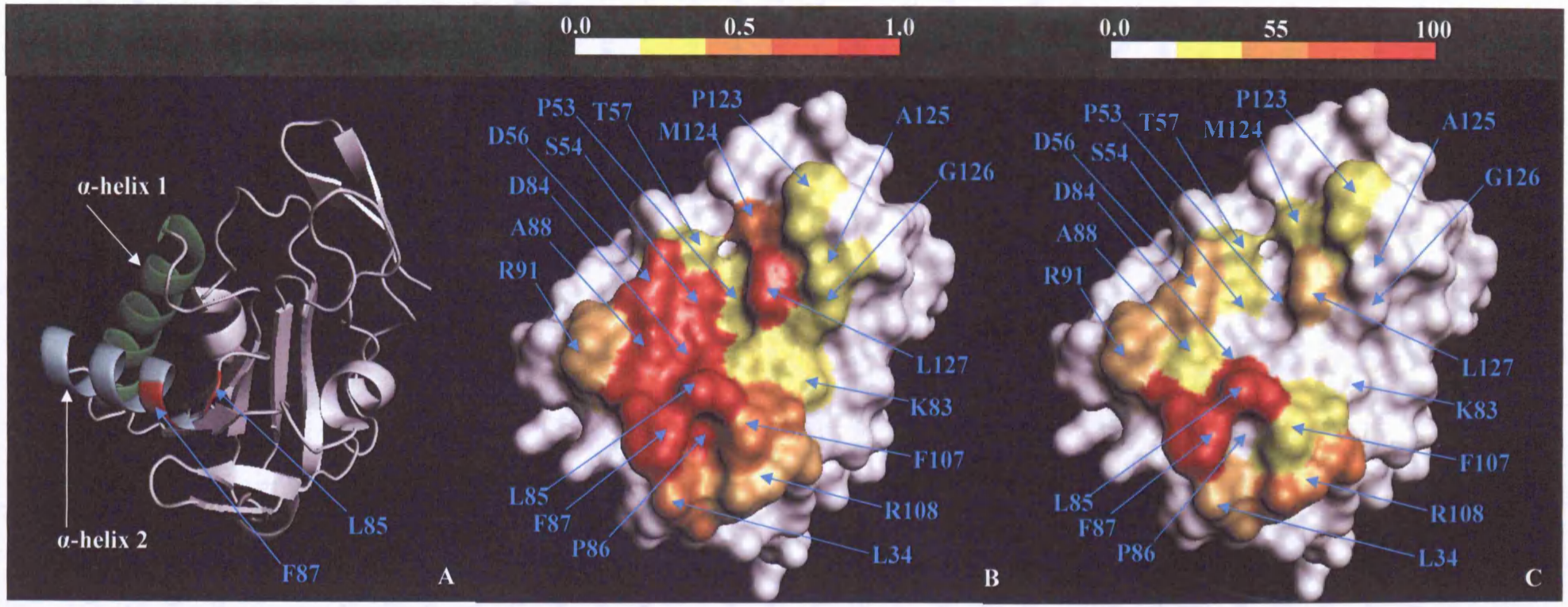


Figure 6.1: Analysis of solvent-accessible surface area (*ASA*) for TPxC60S. **A**: Cartoon representation of *Mtb* TPxC60S protomer indicating the location of the two alpha-helices responsible for the catalytic triad involving the C60S mutation. Also indicated is residue L85, the primary residue involved in the homodimer interaction; the binding face is to the front. **B** and **C**. Molecular surface representation of a TPxC60S protomer in the same orientation as **A**, coloured according to (**B**) fractional change in *ASA* per residue, $fASA_i$, and (**C**) absolute change in *ASA* per residue, $\Delta\ ASA_i$, in units of $Å^2$, upon formation of the homodimer. For a residue that is enclosed entirely in the interface $fASA_i$, $\rightarrow 1$; for a residue entirely outside the contact zone $fASA_i$, $\rightarrow 0$. All images were created using PyMOL.

Additional significant contributors to the interaction surface for *Mtb* TPx, assessed by the criteria $fSASA_i > 0.5$ and $\Delta SASA_i > 50$ Å^2 include Leu127, Asp84, Asp56, and Arg108. It is interesting to consider whether these residues are conserved in homologous proteins. A BLAST search with the amino acid sequence of *Mtb* TPx returns a large number of close homologues (CLUSTALW alignment in Figure 6.2). The alignment shows that in general the chemical characteristics of these positions is strongly conserved suggesting that perhaps all of these proteins exist in a homodimeric (or higher order) oligomerisation state. In fact in the top 100 BLAST hits residues at positions corresponding to positions 85 and 87 are absolutely conserved. Mindful however that these BLAST hits are dominated by TPx orthologues, and that TPx shares an overall structural architecture with a great many other peroxiredoxins and proteins in the thioredoxin-fold super-family – at least some of which are monomeric in solution – we sought to test whether these positions were maintained as hydrophobic residues in the wider class of structurally related proteins. Thus we performed a DALI search of the Protein Data Bank using the coordinates of an isolated *Mtb* TPx protomer. DALI is a programme that systematically compares a single protein coordinate set with the rest of the PDB to find examples of closely related structures, even in the absence of extensive sequence homology. In the DALI output the similarity between the query structure and any 'hit' is characterised by a 'Z-score', where $Z > 2.0$ implies a significant structural homology. For *Mtb* TPx DALI returned 150 coordinate sets with a Z-score greater then 2.0, of which we concentrated on the subset of 27 PDB entries with $Z > 5.0$, corresponding to sequence identities with *Mtb* TPx in the range 10-45 % and back-bone RMSD of the order of 1.5-3.8 Å. Of these proteins, ten are reported or predicted to be monomeric in solution. The amino acid sequences and the secondary structure content of these ten proteins are depicted in Figure 6.3. Interestingly when we analyse the sequences of these proteins at the positions corresponding to the homodimer interface of *Mtb* TPx, the chemical characteristics are strongly divergent (Figure 6.4). Thus for the position corresponding to *Mtb* TPx Leu85, these monomeric homologues can present the more polar Asp, Gln or Lys side chains. Likewise for the position corresponding to *Mtb* TPx Phe87 the homologues sometimes present Lys, Glu, or Gly. Whilst the absence of an absolutely conserved secondary structure in this region (Figure 6.3) makes it a potentially optimistically naïve concept, it seems that a potential route to obtain a monomeric variant of *Mtb* TPx would be to mutate either one or both of Leu85 or

Phe87 into alternative residue types, with the specific side chain substituents selected from amongst these examples. At the time of writing attempts at this objective are underway by others in the laboratory.

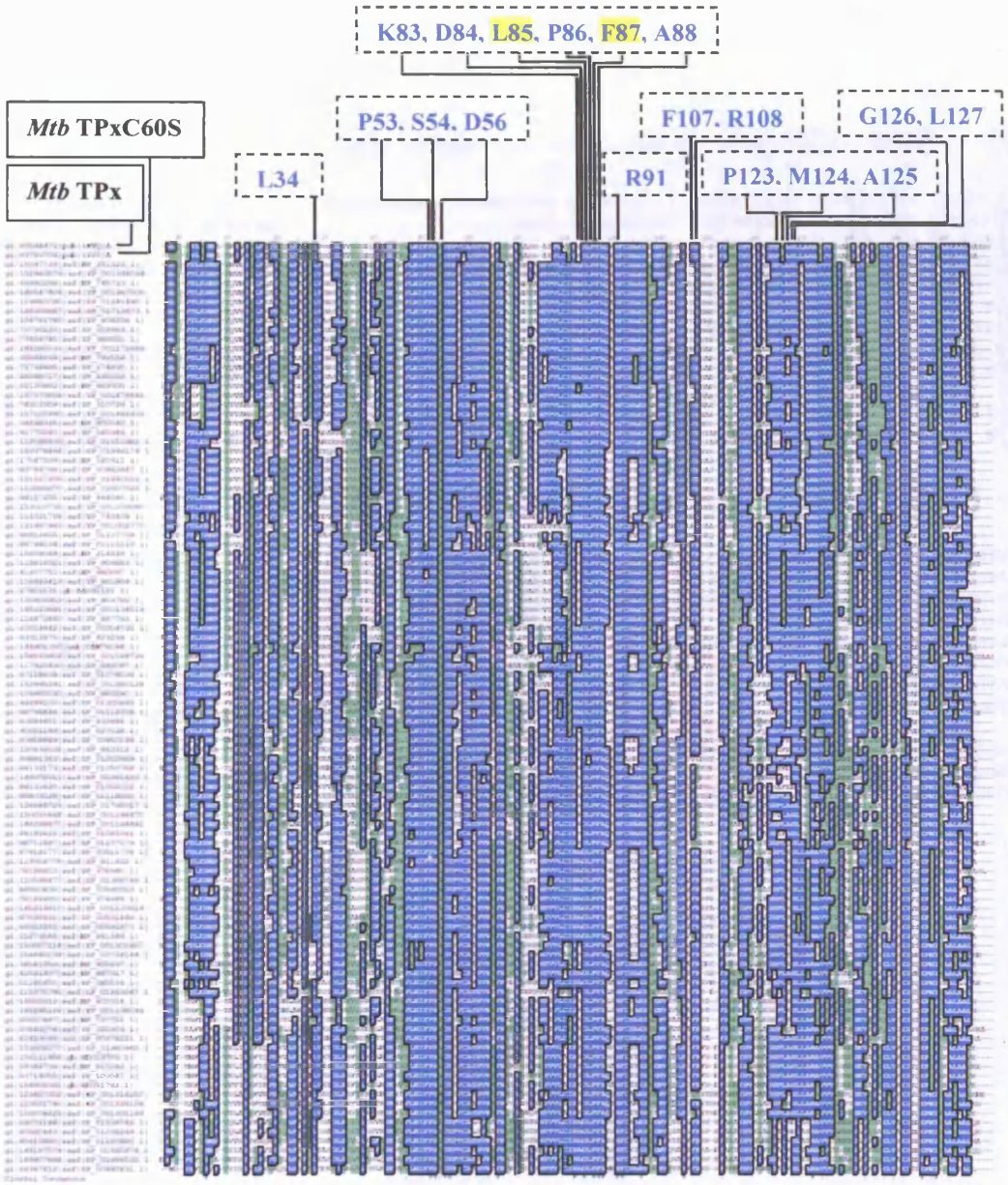


Figure 6.2: A ClustalW alignment of the top 100 sequence-related homologues for *Mtb* TPx (top of the list). By definition, blue segments indicate identical residues and green segments indicate similar residues. The amino acid residues identified as responsible for homodimer stability of *Mtb* TPx, in particular residues Lue85 and Phe87 (additionally highlighted in yellow), are shown above the alignment. All 18 residues are coloured blue and representative of high conservation across the alignment. Species and protein identifiers are listed on the far left of the alignment.

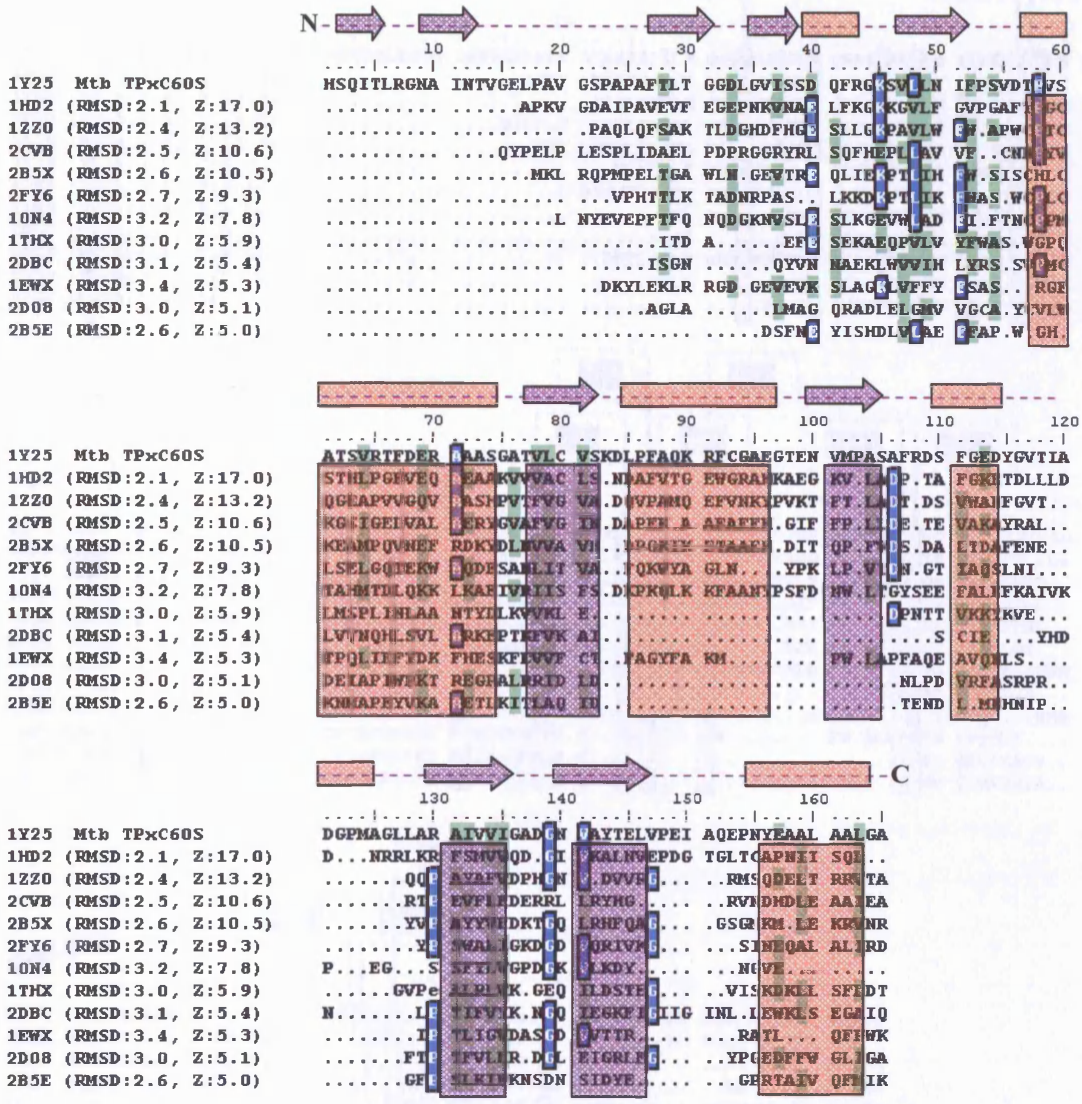


Figure 6.3: Structural alignments of *Mtb* TPxC60S (top of each alignment) with the top ten based Z-scored (numbers shown to the left) monomer PDB structures (PDB identifier and RMSD shown to left) as obtained from the Dali structural comparison server. The structural alignment highlights the α-helical (orange)/β-strand (purple) content of the aligned structures as compared to the secondary structure composition of *Mtb* TPxC60S (shown at the top of the alignment).

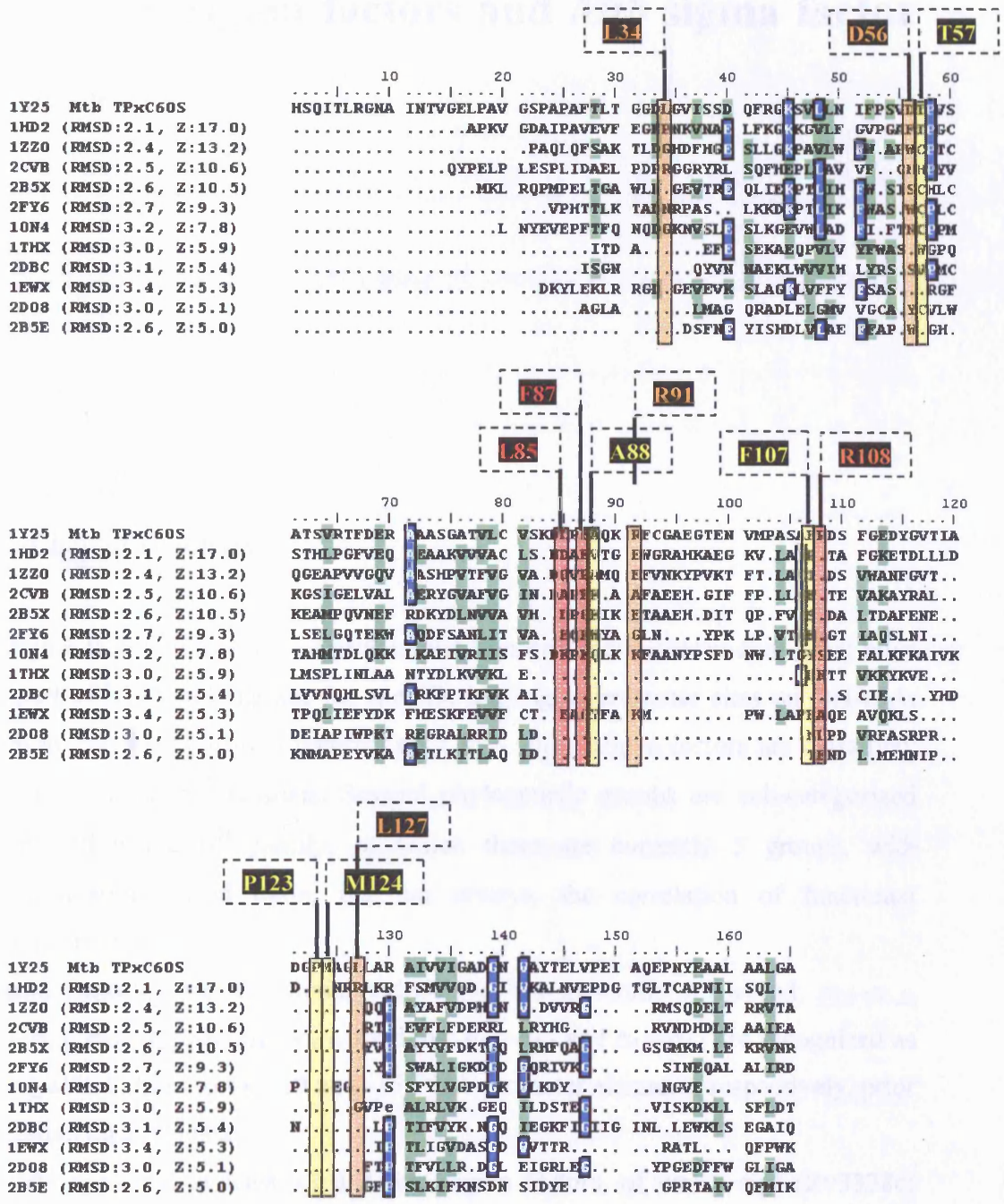


Figure 6.4: Structural alignments of *Mtb* TPxC60S (top of each alignment) with the top ten based Z-scored (numbers shown to the left) monomer PDB structures (PDB identifier and RMSD shown to left) as obtained from the Dali structural comparison server. The structural alignment is as shown in **Figure 6.3** but with the residues exhibiting the greatest extent of $\Delta SASA_i$ colour coded as shown in **Figure 6.1C**.

Chapter 7

Bacterial sigma factors and *Mtb* sigma factor J (sigJ)

7.1. Chapter summary.

A sigma factor (σ factor) is a prokaryotic transcription initiation factor that must be part of RNA polymerase for specific binding to promoter sites on DNA. At present, two large protein families exist into which sigma factors are classified; the σ^{70} and the σ^{54} families. Several phylogenetic groups are sub-categorized from within the σ^{70} family, of which there are currently 5 groups, with diversification based upon, but not always, the correlation of functional characteristics.

Sigma factors have four main regions that are generally conserved; $\sigma_{1.1}$–$\sigma_{1.2}$, $\sigma_{2.1}$–$\sigma_{2.4}$, $\sigma_{3.0}$–$\sigma_{3.2}$ and $\sigma_{4.0}$–$\sigma_{4.1}$. Regions; $\sigma_{2.1}$–$\sigma_{2.4}$ and $\sigma_{4.0}$–$\sigma_{4.1}$ are recognized as being able to bind to the -10 and -35 DNA promoter elements, respectively, prior to transcript synthesis.

In *Mtb* there are at least 13 different sigma factors, of which sigJ (Rv3328c) exhibited increased levels of mRNA synthesis in late stationary-phase cultures. SigJ proved to be strongly up-regulated and its importance deduced to be significant in helping *Mtb* to survive prolonged stationary phase, to withstand stress conditions and to tolerate antibiotics. This reasoning is the driving force by which sigJ was singled out for experimental analysis.

7.2. Introduction – SigJ.

Coates and Hu, (2001b), investigated the transcription of 82 genes from *Mtb* in the hope of identifying those genes involved in maintaining viability of 100-day stationary phase bacteria and persistent bacteria, both before and after antibiotic treatment using rifampicin. Rifampicin is used together with other antibiotics in treating *Mtb* infected individuals (Department of Health Services – Australia, 2002).

Coates and Hu, (2001b), showed using *in vitro* and *in vivo* studies that persistent bacilli which are tolerant to antibiotics are also transcriptionally active. Rifampicin treatment selectively isolated those bacteria still persisting post-antibiotic/antimicrobial treatment as high doses of rifampicin kills actively replicating bacteria.

Using a mini-DNA array they selected those genes recognised as being up-regulated or highly expressed in adaptation to stationary phase in other bacteria. Up-regulation of genes suggests an important role in stationary-phase survival in *Mtb*.

The mRNA levels from three target areas were monitored: exponential-phase cultures (4 days), late stationary-phase dormant cultures (100 days) and in persisters (the 100-day culture, after rifampicin treatment). The data collected showed that 12 % of the genes investigated exhibited an increase in transcription level (as opposed to the mRNA levels monitored during exponential growth) in the late stationary-phase cultures (100 days), 44 % of the genes investigated showed no change in transcription level, while another 44 % exhibited a two- or more fold decrease in mRNA level.

In the cultures treated with rifampicin, no up-regulation was observed. From the 82 genes monitored, 67 genes remained unchanged in mRNA levels monitored while 10 genes were down-regulated.

Of those genes whose mRNA levels increased in the late stationary-phase cultures, sigJ (Rv3328c), with a 17-fold increase, was found to have the highest change in mRNA level of all the genes monitored suggesting it may have a significant role in the non-replicating form of *Mtb*.

The results reveal that sigJ together with sigI, both being strongly up-regulated (two-fold mRNA level increase during stationary phase growth for sigI), may serve as important alternative sigma factors which control the expression of stationary phase-associated genes which help *Mtb* to survive prolonged stationary phase, to withstand stress conditions and to tolerate antibiotics.

7.3. Bacterial sigma factors and transcription.

7.3.1. Overview.

In bacteria, specific transcriptional initiation by RNA polymerase (RNAP) requires a single polypeptide (sigma (σ) factor), which binds to core RNAP to form the active holoenzyme (Darst *et al,* 1996) and is directly involved in promoter recognition, DNA melting and promoter escape and clearance (Borukhov *et al,* 2002). The following sections examine sigma factor reliance for RNAP for transcription initiation in bacteria and the various clades into which sigma factors are classified depending upon their discriminate properties.

7.3.2. Bacterial RNA polymerase.

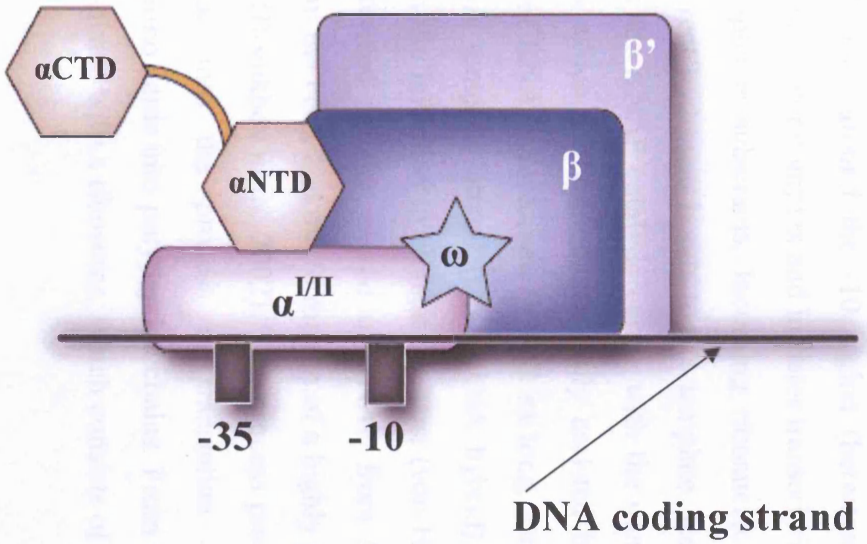
The DNA-dependent RNAP is the principal enzyme of the transcription process and is a final target in many regulatory pathways that control gene expression in all living organisms (Borukhov *et al,* 2002).

RNAP is a complex multisubunit enzyme that contains a dissociable sigma subunit responsible for promoter recognition. In bacteria, RNAP is responsible for the synthesis of all the different types of RNA in the cell such as messenger RNA, ribosomal RNA and transfer RNA. Typically, the vast majority of transcription in rapidly growing bacteria requires a single, primary σ factor subunit similar to σ^{70} of *Escherichia coli.* Control of genes transcribed by this dominant RNA polymerase often rests with DNA-binding repressor and activator proteins. In other cases, however, transcription is regulated by a σ factor

switching mechanism in which the primary σ subunit is replaced by an alternative σ factor with distinct promoter selectivity. Alternative (encompassing the classified σ factor grouping; Group 3 and 4) σ factors function in place of the primary σ factor by binding to core RNA polymerase to allow promoter recognition. In general, alternative σ factors control specialised regulons which are active during growth transitions, in stationary phase, in response to stress conditions and during morphological differentiation (Helmann, 2002).

The bacterial RNAP exists in two forms: the core enzyme or the holoenzyme. The core enzyme has a relative molecular mass of approximately 400 kDa and consists of four subunits (Figure 7.1).

RNAP is conserved in all living organisms. Thus, bacterial RNAP, archaeal RNAP, and eukaryotic RNAP I, RNAP II, and RNAP III, are members of a conserved protein family, termed the 'multisubunit RNAP family'. Members of this protein family contain a conserved subunit of ~160 kDa (β' in bacterial RNAP; A in archaeal RNAP; RPA1, RPB1, and RPC1 in eukaryotic RNAP I, II, and III), a conserved subunit of ~150 kDa (β in bacterial RNAP; B in archaeal RNAP; RPA2, RPB2 and RPC2 in eukaryotic RNAP I, II, and III), a conserved subunit of ~35 kDa (α^{I} in bacterial RNAP; D in archaeal RNAP; RPC5 in eukaryotic RNAP I and RNAP III; RPB3 in eukaryotic RNAP II), a conserved subunit of ~10 – 35 kDa (α^{II} in bacterial RNAP; L in archaeal RNAP; RPC9 in eukaryotic RNAP I and RNAP III; RPB11 in eukaryotic RNAP II), and a conserved subunit of ~6 kDa (ω in bacterial RNAP; K in archaeal RNAP; RPB6 in eukaryotic RNAP I, II, and III). Bacterial RNAP contains only these conserved subunits. Archaeal and eukaryotic RNAP contain these conserved subunits as well as additional subunits (Ebright, 2000).



- **β** – Subunit which includes the RNA polymerase catalytic site.
- **β'** – Subunit, includes DNA binding site.
- **α^{I}** and **α^{II}** – Two sequence identical subunits both possessing a C-terminal (**αCTD**) and N-terminal (**αNTD**) domains. Both domains are involved as scaffolding units that interacts with other proteins that regulate transcription.
- **ω** – Role not well characterized.

Figure 7.1: Diagrammatic representation of the RNAP five subunit assembly, with annotation.

7.3.3. Bacterial transcription.

The transcription cycle in bacterial cells can be divided into three main stages: initiation, elongation and termination. Although it is catalytically active, the core enzyme is incapable of initiating transcription efficiently and with specificity. For this, it must bind an initiation factor, σ, to form a holoenzyme that can recognize specific DNA sequences (promoters) (Gross *et al,* 1992). A promoter sequence is located upstream from the start site of transcription and contains the consensus sequence to which the polymerase binds. During initiation, the holoenzyme specifically binds to two conserved hexamers in the promoter at nucleotide (nt) positions -35 and -10 relative to the transcription start site (+1), to form a closed promoter complex (Borukhov *et al,* 2002). *E. coli* promoters include the so-called TATA box (-10 bp – rich in adenine and thymine base pairs) and the -35 region (-35 bp upstream) from start site. The holoenzyme specifically recognizes these promoter regions (Cummings and Klug, 1994; Bishai and Manabe, 2000 and Borukhov *et al,* 2002). The holoenzyme unwinds the double stranded DNA around the -10 region (between nt -12 and +2), resulting in the open promoter complex and initiates transcription in the presence of nucleotide triphosphates substrates. Incoming ribonucleotide triphosphates (RTPs) form the correct hydrogen bonds to the template. New phosphodiester bonds are formed at the RNAP catalytic centre with the concurrent release of pyrophosphate (PPi) which is thermodynamically assisted by PPi hydrolysis. After the synthesis of RNA approximately 9-12 nt long, 8nt or 9nt are base-paired with the DNA template strand (DNA–RNA hybrid), the transcription complex passes from the initiation to elongation stage (von Hippel, 1998). This transition is characterized by the escape of RNAP from the promoter, the dissociation of σ from the core, and the formation of a highly processive ternary elongation complex (Borukhov *et al,* 2002). This process produces mRNA that is then incorporated into the process of translation – the biological polymerization of amino acids into polypeptide chains. From the transcriptional process, mRNA associates with a ribosome, which consists of two subunits (one

large and one smaller – based upon their sedimentation behaviour in sucrose gradients) and several rRNA components.

Specialized, alternative σ factors direct transcription of specific regulons during unusual physiological or developmental conditions such as heat shock or the presence of extracellular stresses (Darst *et al,* 1996).

The core RNAPs from bacterial and eukaryotic cells, which are catalytically active in RNA chain elongation, are homologous in structure and function (Ahearn *et al,* 1987), showing that the structural and functional nature of core RNAPs is ubiquitous.

Bacterial σ factors are an essential component of prokaryotic RNA polymerase and determine promoter selectivity. The substitution of one σ factor for another can redirect some or all of the RNAP in a cell to activate the transcription of genes that would otherwise be silent.

7.4. The sigma-70 (σ^{70}) family – Introduction.

Bacterial σ factors belong to two large protein families, both apparently unrelated; the σ^{70} and the σ^{54} families (Helmann, 1994 and 2002). Within the σ^{70} family, there are several phylogenetic groups that often, but not always, correlate with function (Table 7.1). Thereby showing the variable structural nature relative to functional characteristics found in this family.

Group 1 or primary sigma factors have a highly conserved sequence identity through a variety of diverse organisms. The regions within a sigma sequence involved in promoter recognition are especially highly conserved. This is not surprising, as all known sigma factors appear to recognize closely related promoter sequences. There are three amino acid sequence segments that show almost complete conservation among primary sigma factors. They include two segments that overlap promoter recognition regions (RpoD box) that include the -10 DNA binding element, a conserved 20-mer that overlaps the -35 recognition region and a third conserved region that comprises a 14-mer just downstream of the RpoD box. These conserved sequences can be used to discriminate Group 1 from Group 3 and 4 sigma factors.

Group 2 sigma factors are regarded as non-essential for exponential cell growth. However, the Group 2 sigma factors are highly similar to the primary sigma factors in their DNA-binding regions, suggesting that the two groups of sigma factors may recognize similar promoter regions.

Group 3 sigma factors, of which Group 4 can be regarded as a subfamily, exhibit poor sequence identity to the sigma factors of Group 1. Group 3/4 sigma factors fall into functional subgroups that include members from diverse organisms. The members of a subgroup are highly similar in sequence identity to one another, regardless of the source organism from which they originate. This suggests that Group 3/4 sigma factors that regulate similar activities have common functional constraints.

Group 5 sigma factors are a fairly new addition to the sigma factor classification system and include functional characteristic found exclusively in this group. Group 5 sigma factors are integral to bacterial resistance against elevated toxic conditions by stimulating appropriate bacterial gene expression.

7.5. The σ^{70} protein family – structure and homology.

The family of related σ^{70} proteins share four regions of amino acid sequence homology designated σ_1–σ_4, which are further divided into several sub-regions (Figure 7.2). The regions $\sigma_{4.2}$, $\sigma_{2.3}$–$\sigma_{2.4}$ were shown to recognize the -35, -10 and the colloquially known 'extended -10' (approximately nucleotide residues -17 to -13) elements of a DNA promoter region, respectively (Barne *et al,* 1997 and Fenton *et al,* 2000). However, interactions between residues in region 3.0 and the upstream 'extended -10' motif substitute for region 4 interactions with the -35 element. Research by Campbell *et al*, (2002) showed that a crystallized $\sigma_{2\text{-}3}$ fragment is sufficient for transcription from an extended -10 promoter. In addition, regions $\sigma_{2.3}$–$\sigma_{2.4}$ are known to interact specifically with the -10 region of the non-template strand in the open promoter complex (Marr and Roberts, 1997 and Borukhov *et al,* 2002).

Group	Description
1	Responsible for regulating most transcription in rapidly growing bacterial cells. DNA promoter sequence elements -35 and -10 are recognized by Group 1 sigma factors when in complex with core RNAP. Contain four characteristic conserved regions: σ_1 to σ_4.
2	Nonessential paralogues to the Group 1 sigma factors. One of the differences between Group 1 and Group 2 sigma factors is the differing selectivity for promoters lacking a consensus -35 element or having a cytosine adjacent to the upstream thymidine of the -10 DNA element. Contain all 4 of the conserved sequence regions found in Group 1.
3	Also known as 'alternative sigma factors'. Under conditions of either stress or developmental processes the number of RNAPs associated with Group 3 sigma factors increases in relation to Group 1 sigma factors. Such RNAP-Group 3 sigma factor complexes function as global regulators allowing the coordinate activation of numerous unlinked operons. Region σ_1 and (often) region σ_3 are absent.
4	Also known as the extracytoplasmic function (ECF) subfamily but also an extended classification encompassing the 'alternative sigma factors' from Group 3. In many cases the ECF sigma factor is cotranscribed with a regulatory transmembrane anti-sigma factor. ECFs often control functions associated with some aspect of the cell surface or transport and generally involve cell envelope functions that include secretion and extracytoplasmic stress.
5	Also known as the TxeR subfamily. Sigma factors in this group have been identified as being positive regulators of toxin and bacteriocin gene expression.

Table 7.1: Five groups by which sigma factors are classified depending upon functional characteristics, as reviewed by Helmann 2002.

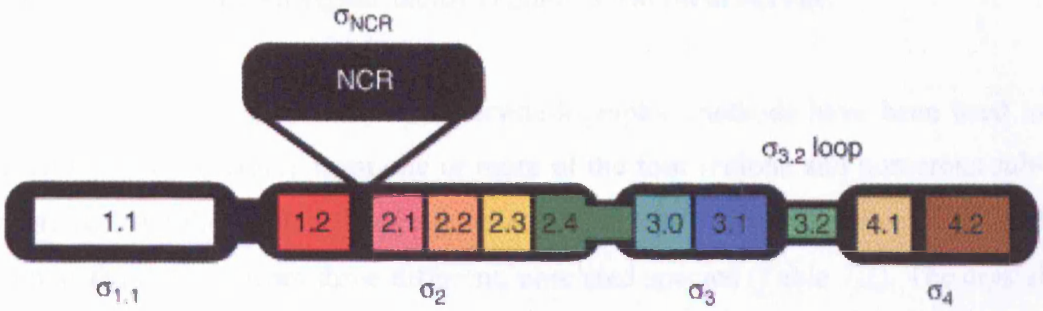


Figure 7.2: Diagrammatic representation of the σ^{70} family of sigma factors (Dove *et al*, 2003).

Darst *et al*, (1996) solved the crystal structure of σ_2 (as well as deducing the domain organization of *E. coli* σ^{70}), which is the most highly conserved region in the σ^{70} family. σ_2, a 39 kDa domain was implicated in core RNAP binding recognition with σ^{70} and binding of the –10 promoter element on the non-template DNA strand. Region σ_2 was also implicated in promoter melting and in the separation process of the RNAP complex from DNA at the end of the transcriptional process.

Region 3 is divided into two sub-regions that essentially extend from the end of region 2.4 to the beginning of region 4. The more conserved region 3.1 bears a weak resemblance to the helix-turn-helix (HtH) DNA binding motif, while conserved amino acid positions in region 3.2 are largely acidic. However, region 3.2 has been implicated in core-RNAP binding (Lonetto *et al*, 1992).

7.6. Sigma factor and sigma factor region - structural survey.

Several papers to date exist where crystallographic methods have been used to elucidate the structures from one or more of the four regions and numerous sub-regions, from the σ^{70} family of sigma factors. 2 papers currently depict 3 of the 4 connected regions from three different, unrelated species (Table 7.2). The crystal structure of region $\sigma_{1.1}$ has not yet been determined.

Region 1.1 is believed to lack a stable structure (Sorenson *et al*, 2004) as a result of being rapidly and completely degraded during limited proteolysis (Campbell *et al*, 2002). Region 1.1 is also a self-inhibitory domain, which is known to mask the DNA-binding regions of σ before it binds to the core RNAP (Dombroski *et al*, 1993). Its precise role in transcription is not understood.

From the survey conducted the structure of an intact σ factor containing σ_2, σ_3 and σ_4 has not been determined without either, core RNAP included, DNA or anti-σ factors used as co-crystallizing binding partners, or isolating individual domains using limited proteolysis. So the disposition of the domains in a free σ factor form remains elusive and thus the structural basis for auto-regulation remains unknown.

σ factor region(s)	Details:	PDB ID Number	Reference
$\sigma_{2.1} - \sigma_{2.4}$ $\sigma_{4.1} - \sigma_{4.2}$	2.00Å resolution crystal structure of *E. coli* σ factor, σ^E, in complex with the N-terminal domain of its anti- σ factor, RseA. Complex size: ~30kDa.	1OR7	Campbell *et al*, 2003.
$\sigma_{2.1} - \sigma_{2.4}$	2.40Å resolution crystal structure of N-terminal (obtained from limited proteolysis) σ^R factor from *S. coelicolor*, σ^{RN}, in complex with its anti- σ factor, RsrA. Complex size: ~22kDa. σ^R involved in thiol-disulphide redox homeostasis.	1H3L	Burton *et al*, 2002.
$\sigma_{1.2} - \sigma_{3.0}$ $\sigma_{4.1} - \sigma_{4.2}$	2.50Å – 3.00Å resolution crystal structure of 3 individual domains, obtained using limited proteolysis, from σ^A (σ^{70} homolog) factor from *T. aquaticus*. One crystal structure complex of σ_4 plus DNA. Molecular weights of the 2 domains modeled: 1. $\sigma_{2.0}$ (includes $\sigma_{1.2}$) – $\sigma_{2.4}$: ~22 kDa 2. $\sigma_{4.1} - \sigma_{4.2}$: ~6kDa	$\sigma_{1.2} - \sigma_{3.1}$: 1KU2 $\sigma_{4.1} - \sigma_{4.2}$: 1KU3 $\sigma_{4.1} - \sigma_{4.2}$/DNA: 1KU7	Campbell *et al*, 2002.
$\sigma_{1.2} - \sigma_{2.4}$ $\sigma_{3.0} - \sigma_{3.2}$ $\sigma_{4.1} - \sigma_{4.2}$	2.60Å resolution crystal structure of *T. thermophilus* σ factor, σ^{70}, in complex with RNAP core. Complex size: ~420kDa.	1IW7	Borukhov *et al*, 2002.
$\sigma_{2.1} - \sigma_{2.4}$	2.60Å resolution crystal structure of *E. coli* σ factor, σ^{70}, containing region 2 (obtained from limited proteolysis), including part of region 1.2. Complex size: ~39kDa.	1SIG	Malhotra *et al*, 1996.
$\sigma_{2.1} - \sigma_{2.4}$ $\sigma_{3.0} - \sigma_{3.2}$ $\sigma_{4.1} - \sigma_{4.2}$	3.25Å and 2.30Å resolution crystal structure of *A. aeolicus* σ factor, σ^{28} in complex with its anti- σ factor, FlgM. Complex size: ~38kDa.	1RP3	Sorenson *et al*, 2004.
$\sigma_{4.1} - \sigma_{4.2}$	2.30Å resolution crystal structure of: λcI (bacterial activator from bacteriophage which binds to specific DNA sequences, operators, just upstream of the promoter -35 element)/σ_4 (σ factor from *T. aquaticus*) /DNA. Complex size: ~18kDa.	1RIO	Jain *et al*, 2004.

Table 7.2: Structural survey for all available σ factors currently deposited in the PDB database.

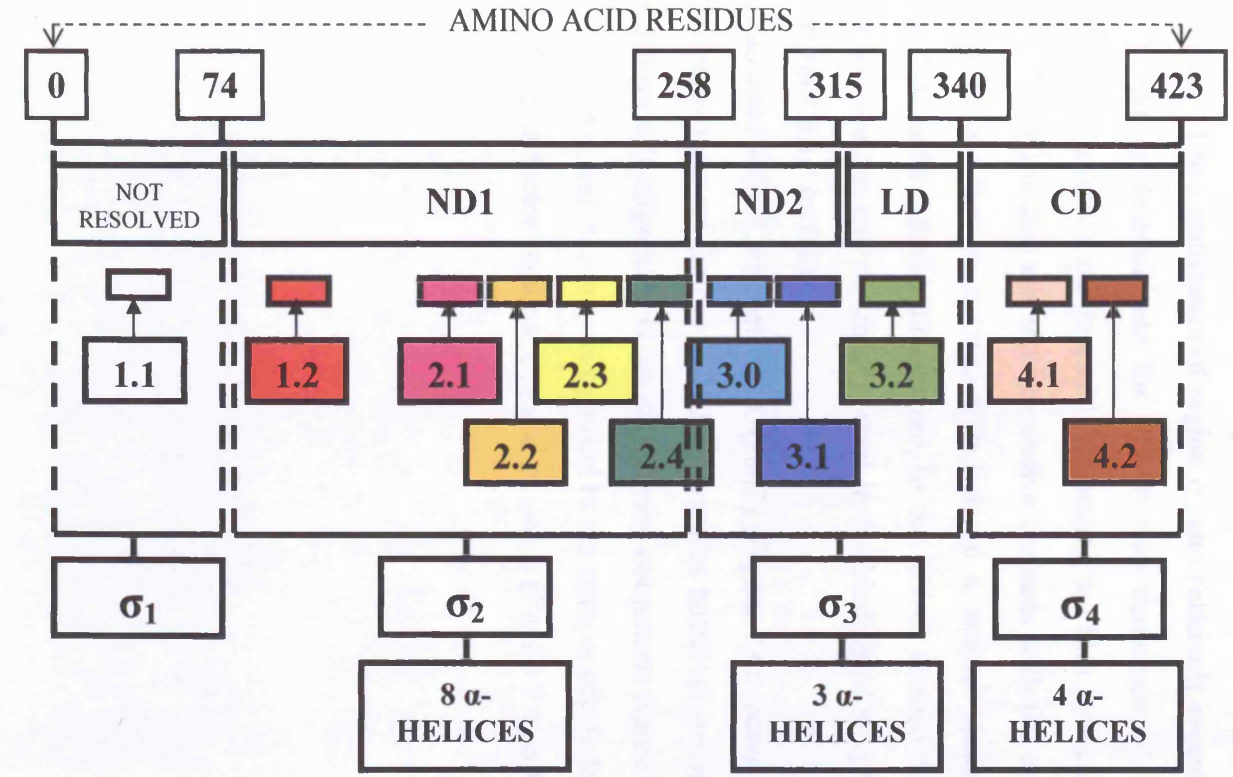
7.7. The σ^{70} protein model.

Borukhov *et al*, (2002) demonstrated a model, from crystallographic data, for the σ^{70} family of proteins that consists entirely of α-helices, connected by either turns or loops. This model can be divided into four structural domains (equivalent to the four σ regions); N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), 'linker' domain (LD) and C-terminal domain (CD) (see Figure 7.3A).

ND1 domain consists of eight α-helices (σ74–254) comprising four helix turn--helix (HtH) motifs; HtH1 (σ75–121), HtH2 (σ94–147), HtH3 (σ123–147) and HtH4 (σ152–203). This domain encompasses sequence region $\sigma_{1.2}$ up to the N-terminal half of region $\sigma_{2.4}$, including the non-conserved segment between regions σ_1 and σ_2. ND1 has a U–shaped structure and is connected to the ND2 domain (σ260–312) by a short linker loop (σ253–260), which lies at the C-terminus of region $\sigma_{2.4}$.

ND2 domain, corresponding to regions $\sigma_{3.0}$, $\sigma_{3.1}$ and $\sigma_{3.2}$, consists of three α-helices that fold into an α-helical bundle. The C-terminal helix of the ND1 domain and the N-terminal helix of the ND2 domain (σ234–281) form a V-shaped structure near the opening of the upstream DNA-binding channel. The inner surface of the V is lined by the residues from the C- and N-terminal α-helices of domains ND1 and ND2, respectively. These α-helices correspond to regions $\sigma_{2.3}$–$\sigma_{2.4}$ which are known to be involved in the specific binding of σ to the -10 and extended -10 elements of the promoter (Figure 7.3B).

The 30 residue LD domain (σ313–339; region $\sigma_{3.2}$) intervenes between the globular N- and C-terminal portions of σ, and has mostly an extended, unfolded conformation. Roughly at its midpoint, the LD domain forms a hairpin loop (σ318–329) that protrudes into the active site cleft, between the 'rudder' loop (β'584–601) and the 'lid' loop (β' 526–538) of the RNAP, which are believed to interact directly with nucleic acids. The rest of the LD domain is stretched along the RNA exit channel of the RNAP holoenzyme complex towards the σ factor CD domain ($\sigma_{4.1}$ and $\sigma_{4.2}$).



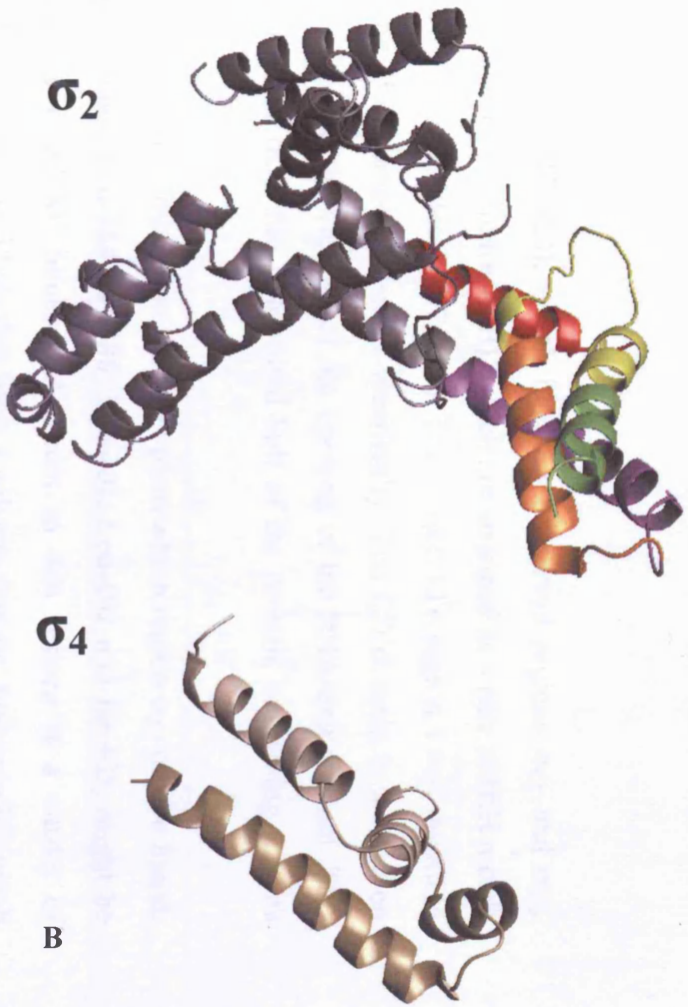


Figure 7.3: A: A schematic representation of the homologous regions typical of σ^{70} family proteins (adapted from Borukhov *et al*, 2002 and Dove *et al*, 2003). **B**: Crystal structure of *E. coli* σ factor, σ_2 (PDB: 1SIG) and σ_4 (PDB: 2H27) color coded as shown in **A** (drawn using PyMOL version 0.98).

The CD domain (σ340–423), which includes conserved regions $\sigma_{4.1}$ and $\sigma_{4.2}$, contains four α-helices (Figure 7.3B), which are arranged as a pair of HtH motifs (HtH5, σ341–376 and HtH6, σ387–418). The conserved σ region 4 was shown to recognize the -35 promoter element specifically. The CD domain is located on the surface of σ^{70} in the vicinity of the opening of the RNA-exit channel. This site is about 57 Å from the N-terminal half of the protein, containing regions $\sigma_{2.3}$–$\sigma_{2.5}$.

Darst *et al,* (1996) also found that the helix upon which region $\sigma_{2.1}$ can be found, containing residues; Leu-384, Ile-388, Phe-401, Leu-402 and Ile-405, might be important for core-RNAP binding. Adjacent to this feature is a cluster of conserved, hydrophobic residues that form a solvent-exposed hydrophobic patch that is suggestive of a protein-binding surface and may also be involved in sigma-core RNAP binding. The significance of region σ_2 and relatively recently also σ_4, as promoter-binding determinants for RNAP was demonstrated by Borukhov *et al,* (2002). The research summarized that among the four structural domains within σ^{70}, ND1 (σ_2) makes the most extensive contacts with the core RNAP along with region σ_4. Region σ_4 was found to be a second binding determinant, of a less extensive tight-binding nature, to the RNAP surface than σ_2. Therefore the two regions, σ_2 and σ_4 are important in the binding of a sigma factor to its cognate core-RNAP surface.

A sequence analysis adapted from Campbell *et al*, (2002) (Figure 7.4) shows an aligned series of 53 Group 1 sigma factors (σ^{70}) from various bacterial species. What can be deduced from this alignment is the extensively ubiquitous nature of sigma factor regions; $\sigma_{2.1}$–$\sigma_{3.0}$ and $\sigma_{4.1}$–$\sigma_{4.2}$ as opposed to the comparatively less ubiquitous presence of sigma factor regions; $\sigma_{1.2}$ and $\sigma_{3.0}$–$\sigma_{3.2}$ (Figure 7.4, color coded boxed areas).

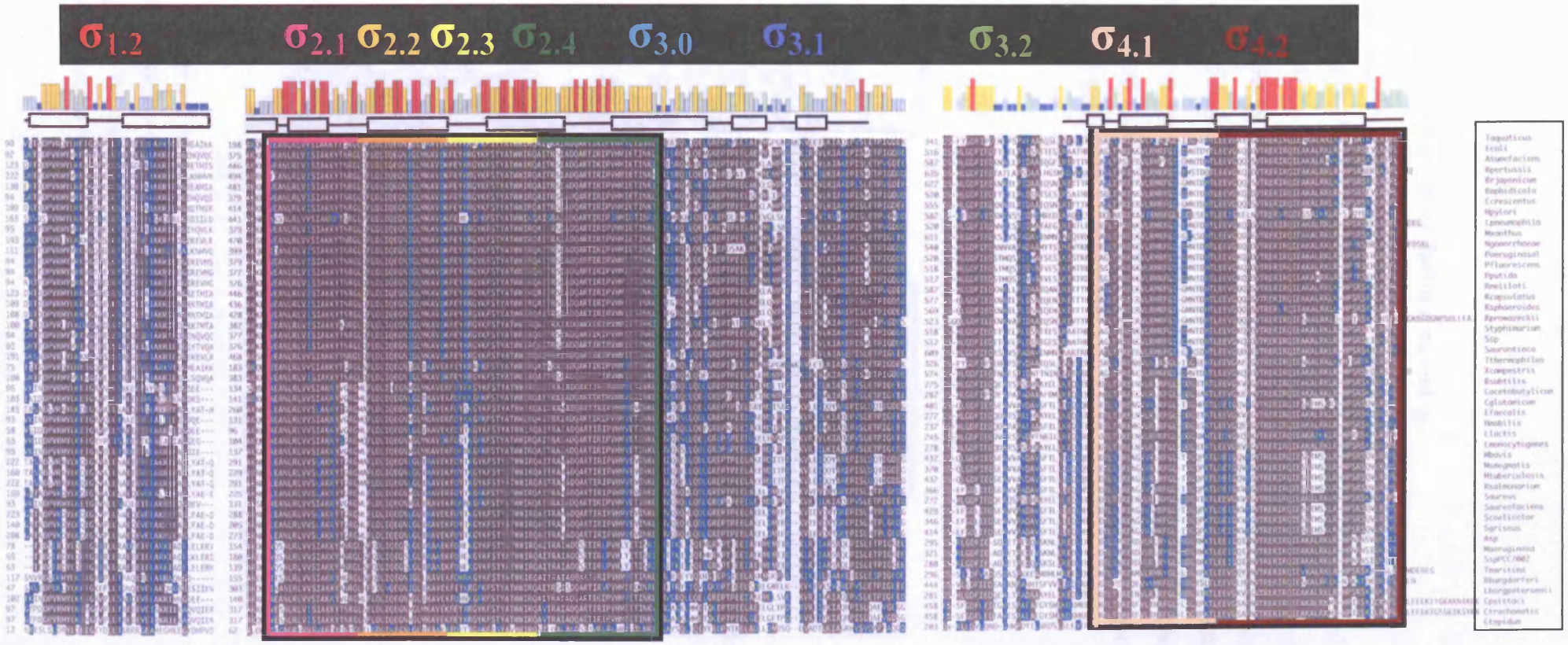


Figure 7.4: Shown is a sequence alignment within conserved regions 1.2–4.2 of 53 primary or group 1 sigma factors (Lonetto *et al.*, 1992; Gruber and Bryant, 1997). The sequences are presented in one-letter amino acid code and are identified by the species in the rightmost column. Numbers at the beginning of each line indicate amino acid positions relative to the start of each mature protein sequence. Amino acid identity in >50 % of the sequences is indicated by a black background, amino acid similarity by a blue background. Groups of residues considered similar are: ST, RK, DE, NQ, FYW, and ILVM. Gaps are indicated by dashes. Helices are indicated above the sequences as rectangles, loops as a solid line. The histogram at the top represents the level of sequence identity at each position. Sequence identity of 100 % is represented by a tall red bar, less than 20 % is represented by a small blue bar, and intermediate levels are represented by orange, light green, and light blue bars. As adapted from Campbell *et al*, 2002 using BioEdit™.

7.8. DNA binding domains in sigma factors: σ_2 and σ_4.

The promoter-binding determinants of σ factors, σ_2 (-10 hexamer element) and σ_4 (-35 hexamer element) are solvent exposed and spaced according to their cognate DNA elements (Figure 5). The σ_3 and σ_4 domains are separated by 45 Å in the holoenzyme. This distance is spanned by an extended 33-residue linker, comprising primarily σ region 3.2 (the $\sigma_{3.2}$ loop), which loops into the RNAP active-site channel and then winds its way out through the RNA exit channel (Murakami and Darst, 2003).

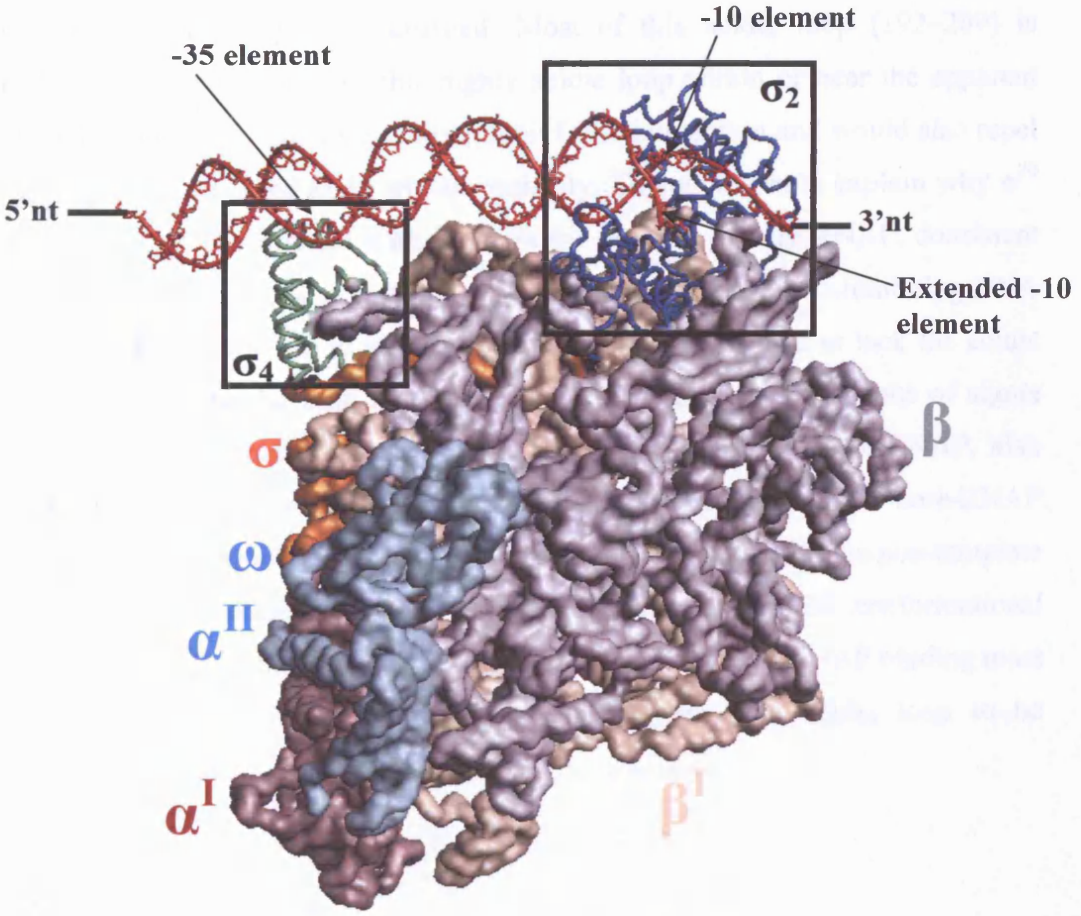


Figure 7.5: Crystal structure of *Thermus aquaticus* RNA polymerase/DNA complex at 6.5 Å resolution (1L9Z). The interaction of sigma factor regions; σ_2 and σ_4, with their cognate DNA elements is shown (boxed). All component RNAP holoenzyme subunits are as indicated and color coded. The rest of the sigma factor adjoined to σ_2 and σ_4 is colored orange.

Binding to core-RNAP induces large relative movements of the σ domains (Callaci *et al*, 1999), converting σ into an active conformation in which the DNA binding determinants in σ_2 and σ_4 are exposed and appropriately spaced to recognize the -10 and -35 elements, respectively, as seen in Figure 7.5 (Murakami and Darst, 2003). The holoenzyme form of RNAP forms transcription-competent open complexes on double-stranded DNA at promoters, suggesting that σ factors are involved in DNA melting at the -10 consensus element perhaps by sequence-specific binding to one of the DNA strands (Helmann and Chamberlin, 1988 and Juang and Helmann, 1994).

The residues implicated in -10 recognition and melting all face into a cleft-like feature. However, potentially occupying this cleft is a highly acidic stretch of residues from 188–209 (numbering based on *E. coli* σ^{70}). In this stretch of 22 residues, 18 are negatively charged. Most of this acidic loop (192–209) is disordered. The presence of this highly acidic loop within or near the apparent DNA-binding cleft might sterically inhibit DNA interaction and would also repel the negatively charged DNA electrostatically. This would help explain why σ^{70} and also σ_2 do not interact with DNA in the absence of core-RNAP, consistent with studies implicating regions of σ^{70} N-terminal of region 2 in inhibiting DNA interactions (Dombroski *et al*, 1993). Since many sigma factors lack the acidic loop, it cannot be the only mechanism by which specific interactions of sigma factors with promoter DNA are inhibited in the absence of core-RNAP, also consistent with the studies of Dombroski *et al*, (1993). Since the σ^{70} core-RNAP complex binds specifically to a single-stranded DNA oligo with the non-template -10 consensus sequence, a mechanism involving substantial conformational changes σ and/or neutralization of the acidic loop upon core-RNAP binding must exist for the auto-inhibition of DNA binding from the acidic loop to be overcome.

7.9. Anti-sigma factors.

Anti-sigma factors modulate the expression of numerous regulons controlled by alternative sigma factors. Anti-sigma factors are themselves regulated by either; secretion from the cell (e.g. FlgM export through the hook-basal body),

sequestration by an anti-anti-sigma (e.g. phosphorylation regulated partner-switching modules), or interaction with extracytoplasmic proteins or small molecule effectors (e.g. transmembrane regulators of extracytoplasmic function sigma factors) (Helmann, 1999 and Campbell *et al,* 2002). Anti-sigma factors have often been used as co-crystallizing binding partners (Table 7.2) for sigma factors and their subsequent structural elucidation.

7.10. Sigma factors in *Mtb.*

In *Mtb* there are at least 13 different sigma factors (Cole *et al*, 1998) and the role of most of them in dormancy is unknown (Hu and Coates, 1999). The abundance of different sigma factors at various stages of *Mtb* growth and replication is influenced by many factors, such as stationary phase and various stress related conditions. The complete genome sequence of *Mtb* reveals two sigma 70 homologues, sigA and sigB, and a subfamily of the sigma 70 class, called the extracytoplasmic function (ECF) family which contains 11 alternative sigma factors (sigC, sigD, sigE, sigF, sigG, sigH, sigL, sigJ, sigK, sigL and sigM) with very little known concerning their physiological functions (Coates and Hu, 2001a).

7.11. Bioinformatic analysis of sigJ.

PFAM; a database which is essentially a large collection of multiple sequence alignments based on hidden Markov models covering many protein domains and families, identifies with confidence the presence of σ_2 within sigJ (Figure 7.6A - shown in bold) and a probable σ_4 region, both located at amino acid residues 5–70 and 110–160 respectively. The same result was obtained when the NCBI conserved domain search program was used to determine the presence of any domains within sigJ that correlated with the NCBI protein database (Figure 7.6B).

The PFAM and NCBI-CD analysis correlates with the σ^{70} alignment analysis carried out by Borukhov *et al,* (2002) in showing that sigJ contains a σ_2 region at the N-terminus, which was found to be conserved amongst other sigma factors from different bacterial species within the same alignment data. Conversely the σ_4 region is not present at the extreme C-terminus, as is common with other bacterial σ^{70} proteins, but is rather comparable to sigG and sigI in that σ_4 is 'supposedly', through low amino acid similarity/identity correlation, found to be close to σ_2, i.e. near the N-terminus. An independent alignment using sigJ was performed; utilizing the NCBI-BLAST program (Altschul, *et al*, 1997) to align the sigJ amino acid sequence against the entire NCBI protein database (Figure 7.7). What is observed is a highly conserved sequence identity comparison between the sigJ amino acid sequence and other ECF group sigma factors with other σ^{70} proteins, from a variety of bacterial sources, exhibiting either complete sequence or σ_2 and/or σ_4 region sequence correlations. When the search is refined by creating a sequence alignment of the sigJ amino acid sequence against the contents of the protein database (PDB), the result is an alignment of poorly correlated protein sequences. This indicates that sigJ bears little structural resemblance, based upon poor *E*-threshold values which are a statistical measure of the number of expected matches in a random database. The lower the *E*-value, the more likely the match is to be significant. *E*-values between 0.1 and 10 are generally dubious, and over 10 are unlikely to have biological significance. The initial NCBI-BLAST provides evidence to support the unique C-terminal characteristic of sigJ and the apparent common region found at the extreme N-terminus by producing an alignment for 19 proteins of *E*-values between 0.21 and 9.00.

The differences in the location of the σ_4 region found in sigJ, relative to the majority of other sigma factors, shows the nature of a C-terminus to be relatively unique in structure and by inference possibly in function as well.

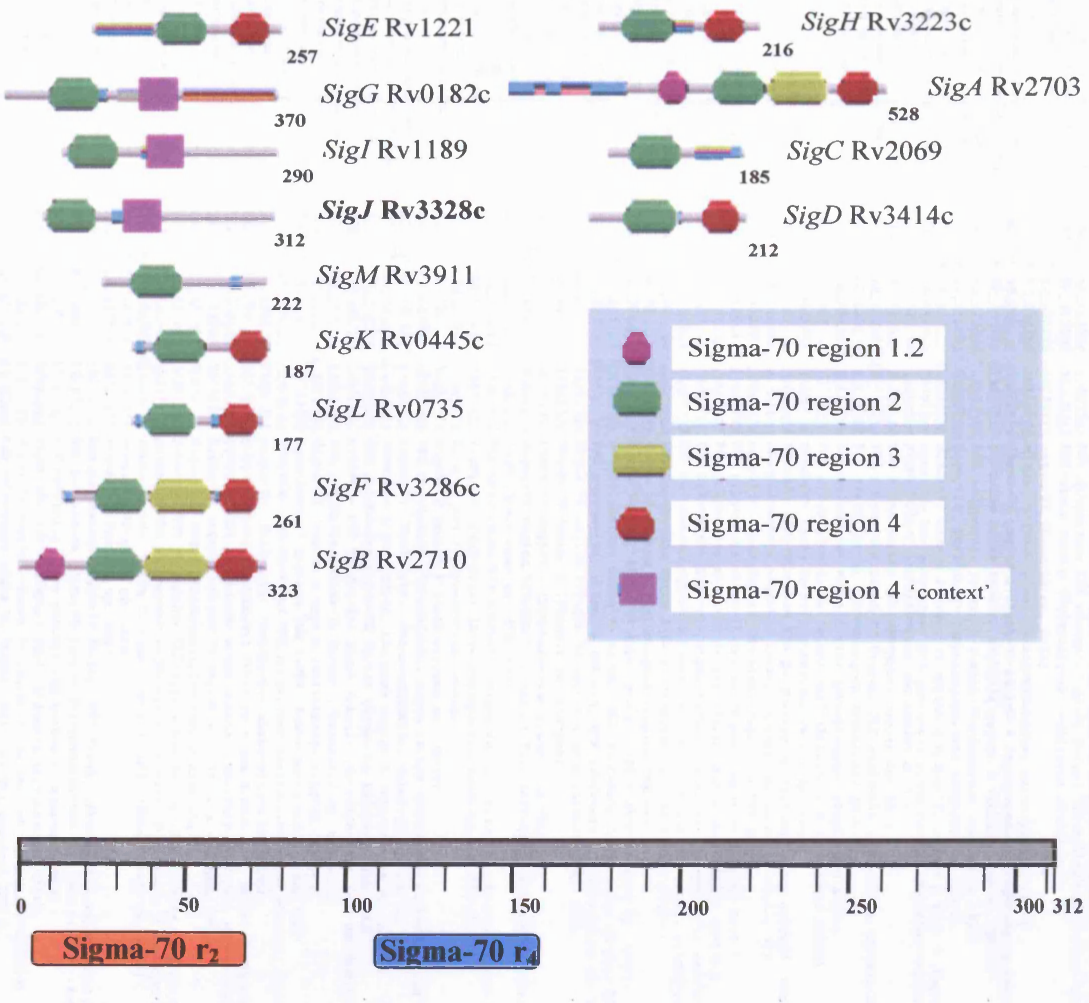
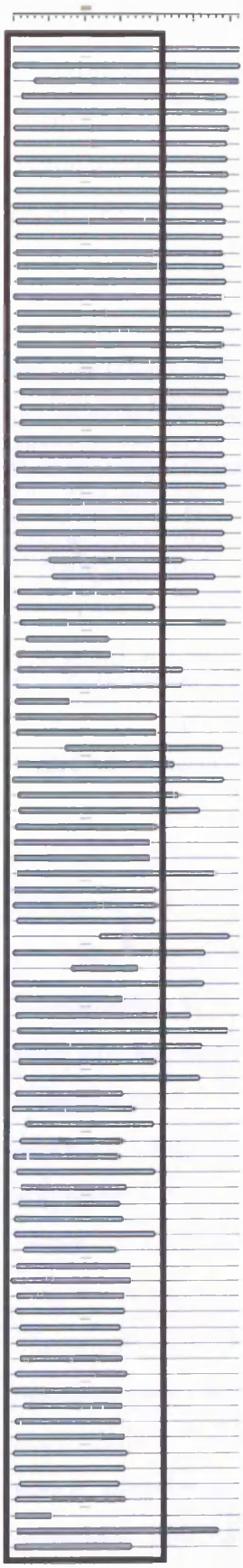


Figure 7.6: **A**: PFAM analyses for the *Mtb* sigma factors (*sigJ* highlighted in bold). **B**: NCBI conserved domain search analysis of sigJ revealing the presence of sigma factor regions; σ_2 and σ_4.



SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
Conserved Domain Database hits				
1586	11	NP_857006	31794513	RNA polymerase sigma-70 factor [Mycobacterium bovis AF2122/97]
1586	12	ZP_008...	81251960	COG1595: DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
1437	12	NP_337960	15842923	RNA polymerase sigma-70 factor [Mycobacterium tuberculosis CDC1551]
1148	10	NP_962380	41409544	RNA polymerase sigma-70 factor [Mycobacterium avium subsp. paratuberculosis K-1(
1076	10	ZP_012...	92911277	Sigma-70 region 2:Sigma-70 region 4 [Mycobacterium sp. JLS]
1072	10	YP_638391	108798194	RNA polymerase, sigma-24 subunit, ECF subfamily [Mycobacterium sp. MCS]
1031	10	ZP_012...	90205133	Sigma-70 region 2 [Mycobacterium vanbaalenii PYR-1]
1006	10	ZP_011...	89338404	Sigma-70 region 2 [Mycobacterium flavescens PYR-GCK]
835	8	YP_704880	111021908	sigma factor, sigma 70 type, group 4 (ECF) [Rhodococcus sp. RHA1]
743	8	AAP69762	37954430	putative ECF sigma factor [Rhodococcus sp. DS7]
510	7	YP_479288	86738888	sigma-24 [Frankia sp. CcI3]
508	7	YP_712306	111221512	Putative RNA polymerase ECF-subfamily sigma factor [Frankia alni ACN14a]
507	7	AAN85484	26541497	RNA polymerase ECF-type sigma factor [Streptomyces atroolivaceus]
500	7	ZP_005...	68233850	Sigma-70 region 2 [Frankia sp. EAN1pec]
462	9	YP_612146	99079992	RNA polymerase, sigma-24 subunit, ECF subfamily [Silicibacter sp. TM1040]
461	3	ZP_006...	69934814	Sigma-70 region 2 [Paracoccus denitrificans PD1222]
461	9	ZP_009...	83858784	putative sigma factor [Oceanicaulis alexandrii HTCC2633]
459	5	YP_645105	108805168	RNA polymerase, sigma-24 subunit, ECF subfamily [Rubrobacter xylanophilus DSM 9
455	7	ZP_009...	84497434	putative sigma factor [Janibacter sp. HTCC2649]
454	7	NP_828600	29833966	RNA polymerase ECF-subfamily sigma factor [Streptomyces avermitilis MA-4680]
453	3	YP_485427	86748931	sigma-24 (FecI-like) [Rhodopseudomonas palustris HaA2]
453	9	YP_630407	108760002	RNA polymerase sigma-70 factor, ECF subfamily [Myxococcus xanthus DK 1622]
452	7	ZP_014...	113943722	Sigma-70, region 4 type 2 [Salinispora tropica CNB-440]
452	7	NP_821838	29827204	RNA polymerase ECF-subfamily sigma factor [Streptomyces avermitilis MA-4680]
451	7	NP_629091	21223312	ECF-sigma factor [Streptomyces coelicolor A3(2)]
451	7	YP_481622	86741222	sigma-24 [Frankia sp. CcI3]
449	3	YP_463954	86157169	sigma-24 (FecI-like) [Anaeromyxobacter dehalogenans 2CP-C]
447	10	ZP_012...	90201420	Sigma-70 region 2:Sigma-70 region 4 [Mycobacterium vanbaalenii PYR-1]
447	8	YP_700910	111017938	probable sigma factor, includes region 2 [Rhodococcus sp. RHA1]
446	9	NP_422060	16127496	RNA polymerase sigma-70 factor [Caulobacter crescentus CB15]
445	3	YP_466335	86159550	sigma-24 (FecI-like) [Anaeromyxobacter dehalogenans 2CP-C]
440	3	ZP_014...	113934163	Sigma-70 region 2:Sigma-70 region 4:Sigma-70, region 4 type 2 [Caulobacter sp.
434	7	ABB90843	82791822	putative RNA polymerase ECF-subfamily sigma factor [Streptomyces davawensis]
265	7	NP_624375	21218815	sigma factor [Streptomyces coelicolor A3(2)]
261	7	NP_627907	21222128	ECF sigma factor [Streptomyces coelicolor A3(2)]
251	9	YP_628485	108759343	RNA polymerase sigma-70 factor, ECF subfamily [Myxococcus xanthus DK 1622]
247	7	NP_631203	21225424	ECF sigma factor [Streptomyces coelicolor A3(2)]
243	7	NP_631160	21225381	RNA polymerase sigma factor [Streptomyces coelicolor A3(2)]
239	9	ZP_009...	83745979	RNA polymerase sigma factor SigJ [Ralstonia solanacearum UW551]
235	7	BAD89288	58530706	putative sigma factor [Streptomyces antibioticus]
224	12	ZP_008...	81252434	COG1595: DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
224	11	NP_857006	31794513	RNA polymerase sigma-70 factor [Mycobacterium bovis AF2122/97]
214	8	YP_119021	54024779	putative sigma factor [Nocardia farcinica IFM 10152]
208	7	NP_629048	21223269	RNA polymerase sigma-70 factor [Streptomyces coelicolor A3(2)]
208	7	NP_824539	29829905	RNA polymerase sigma-70 factor [Streptomyces avermitilis MA-4680]
207	3	ZP_007...	75764707	RNA polymerase ECF-type sigma factor [Bacillus thuringiensis serovar israelensis
203	10	NP_962555	41409719	RNA polymerase sigma-70 factor [Mycobacterium avium subsp. paratuberculosis K-10
202	7	NP_631166	21225386	RNA polymerase sigma factor [Streptomyces coelicolor A3(2)]
202	7	ZP_014...	113946698	sigma-70 region 2 [Salinispora tropica CNB-440]
200	8	YP_705299	111022327	sigma factor, sigma 70 type, group 4 (ECF) [Rhodococcus sp. RHA1]
199	3	YP_634193	108760641	RNA polymerase sigma-70 factor, ECF subfamily [Myxococcus xanthus DK 1622]
198	10	YP_637316	108797119	RNA polymerase, sigma-24 subunit, ECF subfamily [Mycobacterium sp. MCS]
198	10	ZP_012...	92909939	Sigma-70 region 2:Sigma-70 region 4 [Mycobacterium sp. JLS]
198	7	ZP_005...	68233992	Sigma-70 region 2 [Frankia sp. EAN1pec]
196	10	ZP_011...	89342961	Sigma-70 region 2 [Mycobacterium flavescens PYR-GCK]
191	10	ZP_012...	90202073	Sigma-70 region 2:Sigma-70 region 4 [Mycobacterium vanbaalenii PYR-1]
191	7	ZP_009...	84497512	SigG [Janibacter sp. HTCC2649]
190	7	NP_821984	29827350	RNA polymerase ECF-subfamily sigma factor [Streptomyces avermitilis MA-4680]
188	3	YP_466330	86159545	sigma-24 (FecI-like) [Anaeromyxobacter dehalogenans 2CP-C]
185	10	AAA17329	467171	L308_C2_194 [Mycobacterium leprae]
180	7	ZP_006...	71367474	Sigma-70 region 2 [Nocardioides sp. JS614]
176	7	NP_824494	29829860	RNA polymerase ECF-subfamily sigma factor [Streptomyces avermitilis MA-4680]
175	3	YP_465457	86158672	sigma-24 (FecI-like) [Anaeromyxobacter dehalogenans 2CP-C]
174	8	YP_700191	111017219	probable sigma factor, includes region 2 [Rhodococcus sp. RHA1]
172	8	YP_116735	54022493	RNA polymerase sigma-70 factor [Nocardia farcinica IFM 10152]
161	3	BAB49376	14022768	probable RNA polymerase sigma subunit [Mesorhizobium loti MAFF303099]
160	7	ZP_009...	84497124	RNA polymerase sigma-70 factor [Janibacter sp. HTCC2649]
156	7	ZP_014...	113943123	Sigma-70, region 4 type 2 [Salinispora tropica CNB-440]
149	9	ZP_009...	83949629	hypothetical protein ISM_01005 [Roseovarius nubinhibens ISM]
147	7	ZP_003...	62426260	COG1595: DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
145	3	YP_358529	77920714	sigma-24 (FecI-like) [Pelobacter carbinolicus DSM 2380]
143	3	ZP_009...	83375097	putative extracytoplasmic function sigma factor (iron-regulated) transcript
142	3	BAB49441	14022833	probable RNA polymerase sigma subunit [Mesorhizobium loti MAFF303099]
140	3	YP_294923	73540403	Sigma-70 region 2:Sigma-70 region 4 [Ralstonia eutropha JMP134]
139	3	YP_427484	83593732	Sigma-24 (FecI) [Rhodospirillum rubrum ATCC 11170]
139	9	YP_210847	60680703	putative RNA polymerase ECF-type sigma factor [Bacteroides fragilis NCTC 9343]
138	3	YP_468652	86356760	probable RNA polymerase sigma factor protein, ECF family [Rhizobium etli CFN 42]
137	8	YP_701249	111018271	sigma factor, sigma 70 type, group 4 (ECF) [Rhodococcus sp. RHA1]
137	7	YP_482912	86742512	sigma-24 [Frankia sp. CcI3]
135	7	YP_479271	86738871	sigma-24 [Frankia sp. CcI3]
135	9	ZP_009...	83370526	RNA polymerase sigma-70 factor, ECF family [Rhodobacter sphaeroides ATCC 17025]
135	8	NP_939356	38233589	RNA polymerase sigma-70 factor [Corynebacterium diphtheriae NCTC 13129]
135	3	ZP_008...	77954601	ECF subfamily sigma subunit [Marinobacter aquaeolei VT8]
134	5	YP_590900	94968862	sigma-24, ECF subfamily [Acidobacteria bacterium Ellin345]
134	3	ZP_005...	67932095	Sigma-70 region 2:Sigma-70 region 4 [Solibacter usitatus Ellin6076]
134	9	YP_167963	56697594	RNA polymerase sigma-70 factor [Silicibacter pomeroyi DSS-3]
134	7	ZP_003...	62426027	COG1595: DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
133	7	YP_711241	111220447	putative RNA polymerase ECF-subfamily sigma factor [Frankia alni ACN14a]
133	5	YP_644482	108804545	RNA polymerase, sigma-24 subunit, ECF subfamily [Rubrobacter xylanophilus DSM 99
133	8	NP_737787	25027733	RNA polymerase sigma-70 factor [Corynebacterium efficiens YS-314]
133	9	ZP_002...	46201571	COG1595: DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
133	3	YP_258119	70728370	RNA polymerase sigma-70 factor, ECF subfamily [Pseudomonas fluorescens Pf-5]
133	3	NP_840639	30248569	Sigma factor, ECF subfamily [Nitrosomonas europaea ATCC 19718]
132	7	ZP_005...	68231950	Sigma-70 region 2:Sigma-70 region 4 [Frankia sp. EAN1pec]
132	8	YP_700986	111018014	probable sigma factor, includes region 2 [Rhodococcus sp. RHA1]
132	3	YP_548063	91787111	sigma-24 (FecI-like) [Polaromonas sp. JS666]
132	3	YP_463444	86156659	sigma-24 (FecI-like) [Anaeromyxobacter dehalogenans 2CP-C]

Figure 7.7: BLAST Link (BLink) to Protein Alignments and Structures. Pre-computed sequence alignments, generated from routine all-against-all BLAST comparisons performed at NCBI. 160 proteins aligned against the sigJ sequence (top) are displayed with the ubiquitous N-terminus identity found across the alignment highlighted (boxed).

Theoretical secondary structure prediction program, JPred, showed sigJ to be composed of 11 α-helices and 5 β-sheets with prediction confidence being low within amino acid residues 200–310 (Figure 7.8). Other such theoretical structure prediction programs such as PsiPred and Jnet produced similar results.

```
MEVSEFEALRQHLMSVAYRLTGTVADAEDIVQEAWLRWDSPDTVIADPRAWLTTVVSRLGLDKLRSAA
--HHHHHHHHHHHHHHHHHH------HHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHH
HRRETYTGTWLPEPVVTGLDATDPLAAVVAAEDARFAAMVVLERLRPDQRVAFVLHDGFAVPFAEVAE
HH----------------------HHHHHHHHHHHHHHHHHH---HHHHHHHHHHHH----HHHHHH
VLGTSEAAARQLASRARKAVTAQPALISGDPDPAHNEVVGRLMAAMAAGDLDTVVSLLHPDVTFTGDS
HH---HHHHHHHHHHHHHHHH----------HHHHHHHHHHHHHH---HHHHHHHH---EEEEE--
NGKAPTAVRAVRGSDKVVRFILGLVQRYGPGLFGANQLALVNGELGAYTAGLPGVDGYRAMAPRITAI
-------------HHHHHHHHHHHHH--------EEEEEEE----EEEEE------------EEEEEE
TVRDGKVCALWDIANPDKFTGSPLKERRAQPTGRGRHHRN
EE----EEEEEEE----HHHH-------------------

H = ALPHA HELIX
E = BETA SHEET
```

Figure 7.8: Theoretical secondary structure predication of sigJ using the program JPred.

CPHmodels version: 2.0 Server is a structural modelling server which utilises a computer program (XM3) to extract 3D models based on fold recognition/homology modelling, in which a large sequence database of proteins with known structure (PDB-BLAST) is iteratively searched to construct a sequence profile of suitable templates. The server can be accessed via this web address: http://www.cbs.dtu.dk/services/CPHmodels/. A sequence-based alignment is then produced based upon the query sequence applied, using a score reliant on profile-profile comparisons. This produces more accurate alignments than methods based on sequence-profile or sequence-sequence alignment. The alignment score is modified so as to ensure unreliable parts of the alignment is discarded. By adding the amino acid sequence of sigJ to this modelling server, a predictably low threshold scoring alignment is produced. However, by removing poorly correlated sections of the initial query sequence, 20 % identity to PDB deposited proteins is achieved with amino acid residues 10–158, from the full initial 312 amino acid sigJ sequence.

A

PDB entry: **1OR7** chain: A. Score: 54. E: 2e-07

PDB entry: **1OR7** chain: B. Score: 45. E: 1e-04

PDB entry: **2H27** chain: A. Score: 37. E: 0.018

PDB entry: **2H27** chain: D. Score: 37. E: 0.020

PDB entry: **1H3L** chain: A. Score: 37. E: 0.027

PDB entry: **1H3L** chain: B. Score: 36. E: 0.033

B

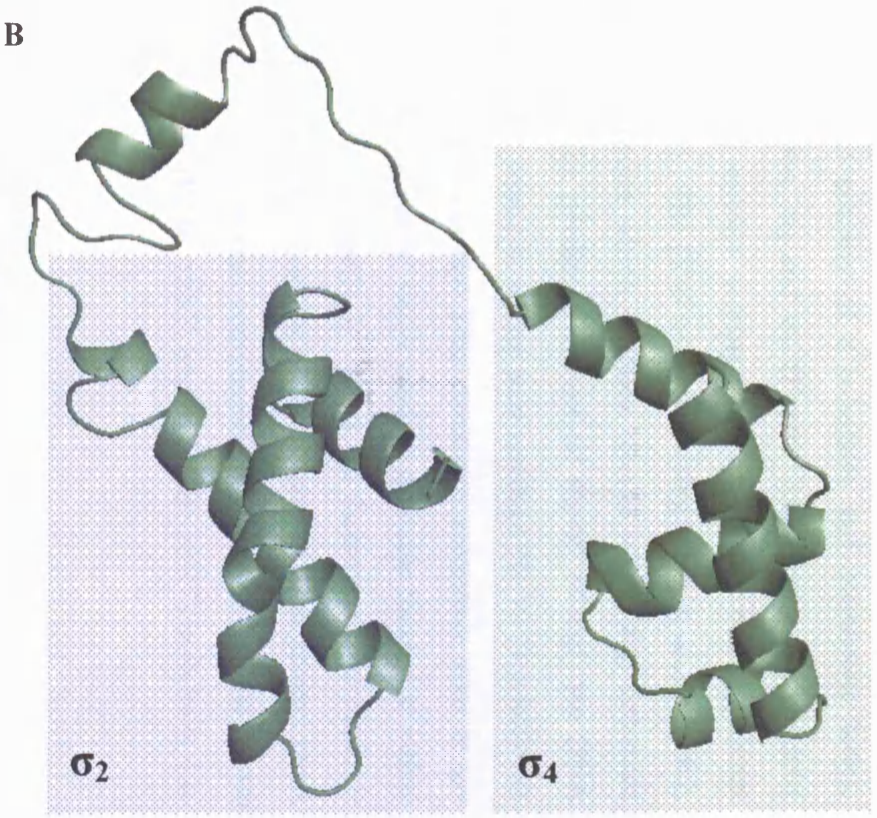


Figure 7.9: **A**: The six best aligned PDB entries to the sigJ sequence query using the CPHmodels version: 2.0 Server. All entries are below the 0.05 E-threshold, making them biologically significant. Entry 1OR7 is used as the template around which residues; 10 – 158 of the amino acid sequence of sigJ is modelled around. **B**: The sigJ theoretical model has DNA binding regions σ_2 and σ_4 highlighted and annotated.

The model created, composed entirely of 8 α-helices and connective loops, identifies DNA binding regions, σ_2 and σ_4 (Figure 7.9A) as a homologous structural motif with 6 PDB entries scoring better than the non-significant threshold of 0.05 (Figure 7.9B). The PDB entry achieving the best score, 1OR7, is then used as a template around which residues 10–158 of the sigJ sequence align around. By examining the predicted theoretical model of sigJ, the DNA binding regions, σ_2 and σ_4, are found at amino acid residues 10–70 (3 α-helices), and 100–158 (4 α-helices), respectively. This correlates extremely well to the previous identification of the σ_2 and σ_4 DNA binding domains located at amino acid residues 5–70 and 110–160 respectively. When the theoretical sigJ model is compared to the secondary structure content of the *E. coli* PDB deposited DNA binding domains, σ_2 and σ_4, the α-helical composition is very similar. The crystal structure of *E. coli* σ factor, σ_2 (PDB: 1SIG) and σ_4 (PDB: 2H27) both have 4 α-helices each.

7.12. Discussion and concluding remarks.

Many of the experimental observations made into sigJ so far suggest a uniquely structured monomeric sigma factor (Coates and Hu, 2001b). Bioinformatic evidence and the *in vitro* and *in vivo* studies carried out by Coates and Hu, (2001b) suggest the existence of a σ_2 and probable σ_4 region, indicating that sigJ deserves to belong as an alternative sigma factor group within the ECF σ^{70} subfamily. The significance of sigJ for *Mtb* survival during a prolonged stationary phase is to allow the bacilli to withstand stress conditions and to tolerate antibiotics by virtue of increased mRNA levels during stationary phase growth, as well as by being strongly up-regulated. This suggests that sigJ is important in controlling the expression of stationary phase-associated genes.

SigJ could invariably be an important drug target for future development and therefore merits an effort into its structural determination. As a result, a process towards structural determination will evolve during the various analytical procedures applied to sigJ described in the following chapter.

Chapter 8

Results: *Mtb* sigma factor J (sigJ)

8.1. Chapter summary.

PolyHis-tagged *Mtb* sigJ protein was expressed according to the protocol suggested by collaborators. The protein was purified both from inclusion bodies and solubilised/refolded or directly from the lysed cell supernatant. Together with NMR spectroscopy, analytical size exclusion chromatography (SEC), freeze/thaw studies and circular dichroism (CD), the biochemical characteristics of sigJ were investigated. The characteristics of sigJ purified from inclusion bodies differed significantly from those of sigJ expressed in the cytoplasm. Whilst the 1D ^{1}H NMR spectrum suggested the presence of globular protein features, the 2D [^{1}H,^{15}N]-HSQC NMR measurements obtained of sigJ yielded predominately disordered spectra with inhomogeneous cross peak line shapes due to extensive line broadening. Attempts to resolve the poor spectral features identified observed in HSQC spectra of sigJ included the use of limited proteolysis to isolate globularly folded domains, known as the σ_2 and σ_4 (-10 and -35 nucleotides upstream of the start site) DNA binding elements. The investigation diversified into using electrophoretic mobility shift assays (EMSA) to examine the putative DNA-binding nature of sigJ, the results of which suggest

that the future application of an appropriate DNA ligand might potentiate further examination by 2D heteronuclear NMR spectroscopy by quenching dynamic structure fluctuations that presently complicate the spectrum.

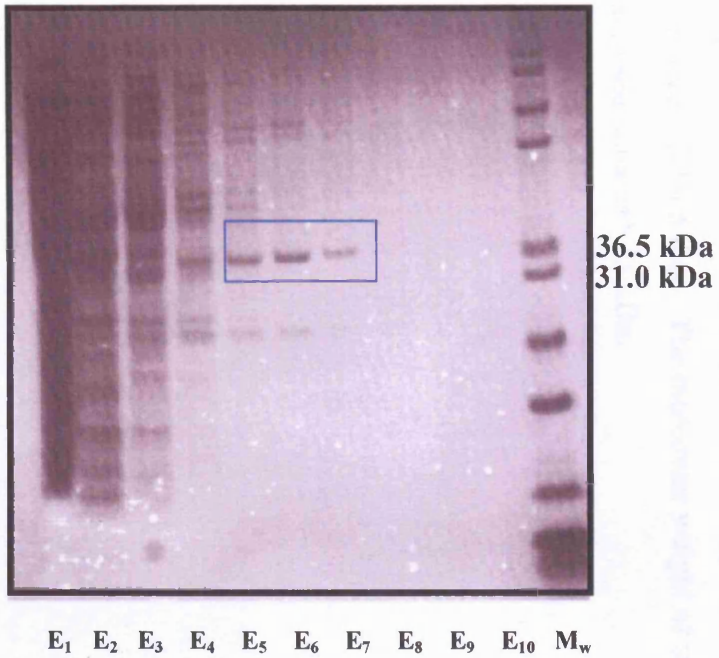
8.2. SigJ – purification and expression.

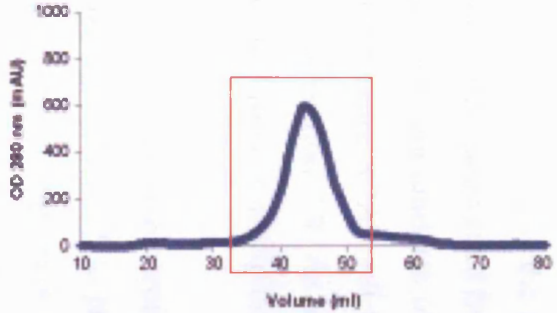
Studying proteins using NMR techniques typically requires multi-milligram quantities in aqueous buffers at millimolar concentrations. In many cases, technical and/or economic limitations can become apparent in the production of proteins and polypeptides to achieve such quantities. In order to carry out a biologically relevant study of sigJ the buffer conditions must be kept as close to physiological pH (pH 7.4) and salt concentration as possible, optimally below 200 mM for NMR spectroscopic study.

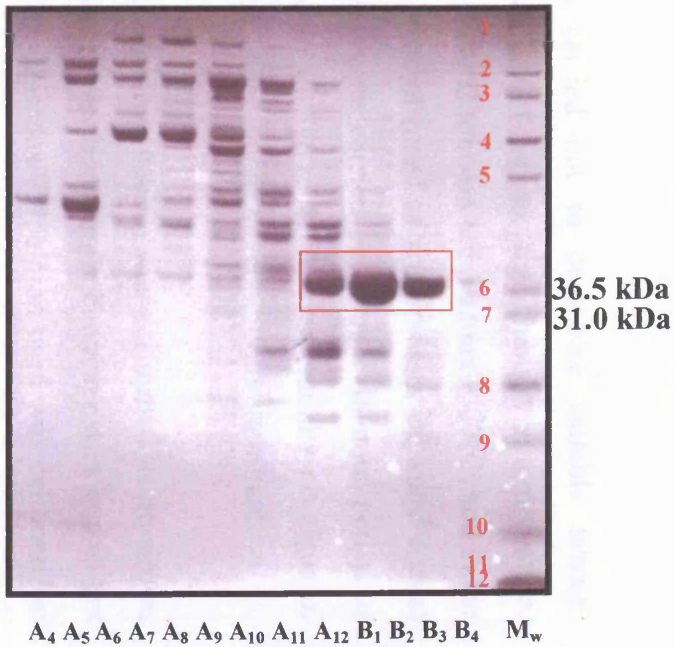
Initially sigJ was expressed and purified from the soluble fraction of the cell lysate using a 1 L scale cell culture as described in Chapter 2.3.2. Suitably consistent, high concentrations (~0.5-0.6 mM) of soluble expressed sigJ were obtained, at a relatively pure level – with few contaminants and at the predicted monomeric molecular weight, i.e. 34 kDa. An immediate scale up of the initial 1 L preparation to a 3 L preparation was then carried out as described in Section 2.3.2. Figure 8.1A shows the protein gel obtained after Ni-NTA chromatography purification (the process of which is described in Section 2.3.3.) for sigJ. The corresponding fractions containing sigJ (Figure 8.1A, blue box) were then concentrated (as described in Section 2.2.) and passed through a Superdex-75 SEC column (as described in Section 2.3.4.). The fractions corresponding to the peak in the chromatogram (Figures 7.1 B^1 and B^2 – red boxes) were then pooled and subsequently concentrated. The final sigJ concentration, displayed as an average range for a number of identical 3 L scale sigJ preparations made during the course of this project, ranged between 1.2–1.4 mM.

The subsequent concentrated sigJ sample was then processed (as described in Section 2.3.1-2.3.1.4), in order to have the N-terminal amino acid sequence confirmed by a commercial sequencing service (MWG™). The successful amino acid sequencing data then allowed the continued use of natively folded sigJ for structural probing using NMR spectroscopy.

Initially our collaborators claimed that higher protein yields of sigJ could be obtained from the insoluble fraction following cell lysis, as opposed to the protein obtained in the soluble part of the lysate. To test this, sigJ was also purified from inclusion bodies and refolded as detailed in Section 2.10. Yields of purified sigJ comparable to those obtained from the purification of the natively folded ('soluble') sigJ were obtained.







A: Key: E_x = Elution fraction, M_w = Molecular weight marker.

B²: Key: A_1 (20 mL) – A_{12} (60 mL)/B_1 (65 mL) – B_4 (80 mL) = Elution's and total collected volumes, M_w = Molecular weight marker.

Red numbering indicates the M_w protein marker sizes for A and B gels, 1: 200 kDa, 2: 116.3 kDa, 3: 97.4 kDa, 4: 66.3 kDa, 5: 55.4 kDa, 6: 36.5, 7: 31 kDa, 8: 21.5 kDa, 9: 14.4 kDa, 10: 6 kDa, 11: 3.5 kDa, and 12: 2.5 kDa.

Figure 8.1: **A**: NI-NTA protein gels of natively folded sigJ, showing protein in the soluble fraction across a range of wash steps of increasing imidazole concentration; E_1 (5 mM imidazole) – E_{10} (500 mM imidazole). Boxed in blue; fractions concentrated. **B¹** and **B²**: SEC Superdex-75 chromatogram and SEC fractional analysis by SDS-PAGE of natively folded sigJ, respectively, showing quantities of monomeric protein (B^2 red box) at the correct molecular range, that was then concentrated.

8.3. Freeze/thaw and analytical size exclusion chromatography (ASEC) tests.

A freeze/thaw experiment was carried out to determine suitable storage conditions and to examine the stability of sigJ to large changes in temperature. The temperatures used varied from -80°C to room temperature (~27°C) and in the presence of varying amounts of glycerol (used as a suppressor of water crystal formation) (Figure 8.2A). All tests were carried out over a 48 hour period as described in Section 2.9. As shown in Figure 8.2B, sigJ exhibits no signs of protein degradation or precipitation under any of the conditions examined.

Analytical size exclusion chromatography (ASEC) was performed using sigJ as described in Section 2.3.5 to probe the hydrodynamic characteristics and to test for the presence of oligomers. The superdex-75 column used was calibrated as described in Section 4.3. ASEC of purified sigJ samples indicates a concentration-independent elution time, consistent with a mono-disperse molecular species (Figure 8.2B). Compared to protein standards the elution time predicts an apparent molecular weight of 45 kDa for sigJ refolded from inclusion bodies (Figure 8.2C), which is larger than the apparent mass of 34 kDa obtained with the natively folded form. The molecular weight of sigJ is predicted from the amino acid sequence is ~33 kDa.

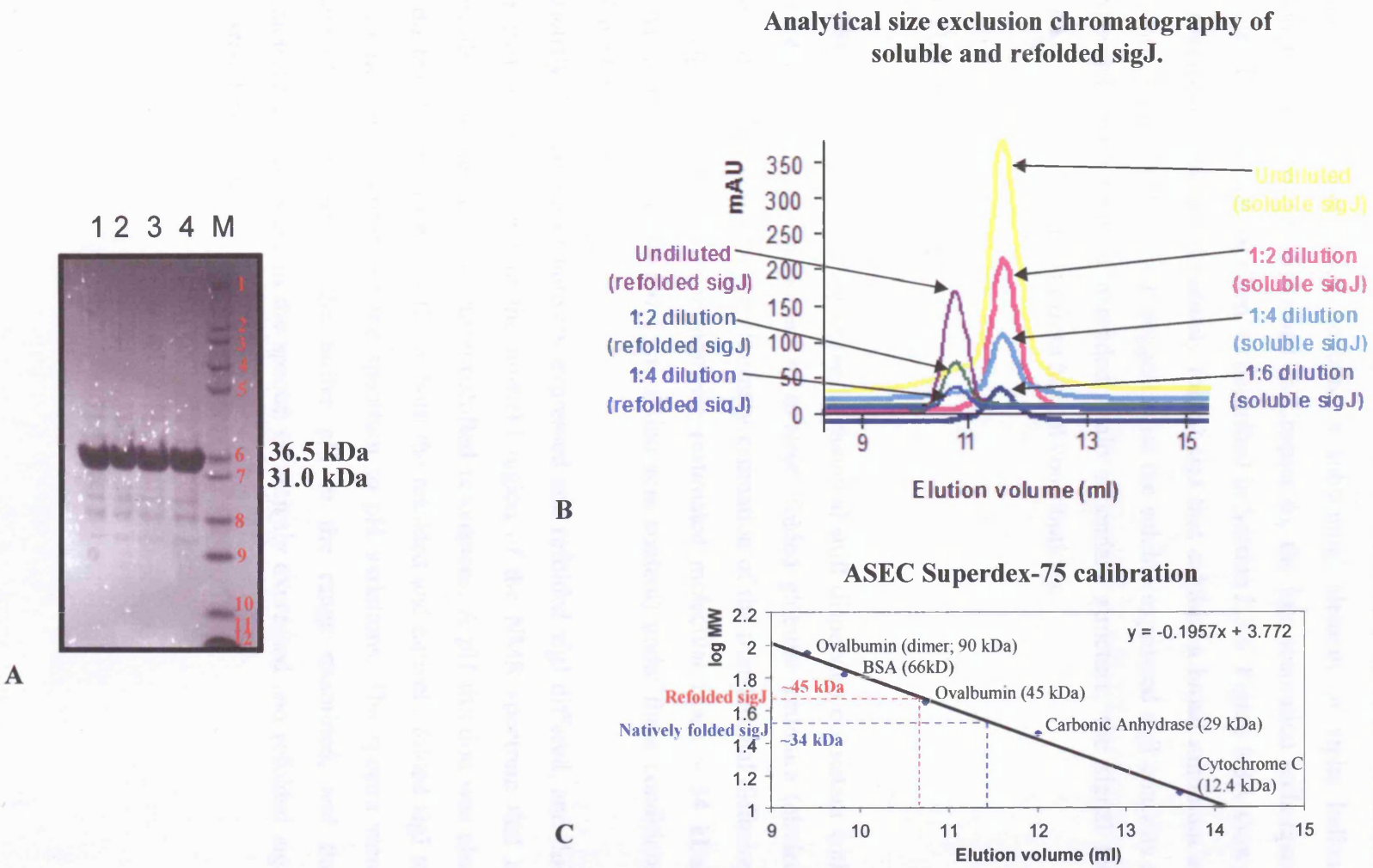


Figure 8.2: **A**: Freeze/thaw test upon sigJ under various temperature and glycerol conditions: Lane 1: 10% (v/v) glycerol, Lane 2: 20% (v/v) glycerol, Lane 3: No glycerol, Lane 4: no freeze thaw but containing glycerol, Lane 5: markers. SigJ concentration used for each lane sample; 1 mM. **B**: ASEC gel filtration cumulative curve data of sigJ: natively folded and refolded forms at various dilution levels from a starting, undiluted concentration, of; 1 mM. **C**: Standard curve generated for the ASEC gel filtration column used in Figure 1B showing the differing apparent MW for both natively folded and refolded sigJ. **A**: Red numbering indicates the M_w protein marker sizes, 1: 200 kDa, 2: 116.3 kDa, 3: 97.4 kDa, 4: 66.3 kDa, 5: 55.4 kDa, 6: 36.5, 7: 31 kDa, 8: 21.5 kDa, 9: 14.4 kDa, 10: 6 kDa, 11: 3.5 kDa, and 12: 2.5 kDa.

8.4. Circular dichroism using sigJ.

To test the results obtained using the various secondary structure prediction programs that suggest sigJ contains a substantial element of alpha helical secondary structure (as described in Chapter 6), the low resolution technique; circular dichroism was applied, as described in Section 2.3.6. Figure 8.3A shows the spectrum obtained of natively folded sigJ that exhibits a broad minimum at 222 nm. Arguably this result suggests that the soluble expressed sigJ contains a substantial component of extended chain secondary structure, the signal for which dominates any present from helical contributions.

8.5. 1D [^{1}H]-NMR of sigJ.

1D [^{1}H]-NMR analysis of sigJ reveals chemical shift dispersion consistent with at least a significant component of ordered, folded globular structure (circled blue and red in Figure 8.4). Preliminary estimation of the translational diffusion time suggests that sigJ is monomeric (estimated molecular mass ~ 34 kDa, similar to the mass predicted from amino acid content) under these conditions (1.0 mM, pH 7.0).

Curiously the spectra of natively expressed and refolded sigJ differed, and this was particularly evident in the up-field region of the NMR spectrum that is typically populated by ring current-shifted resonances. A pH titration was also conducted (Figure 8.4A and B) on both the refolded and natively folded sigJ to determine the sensitivity of the spectrum to pH variations. The spectra were essentially insensitive to the buffer pH in the range examined, and the characteristic differences in the spectra of natively expressed and refolded sigJ persisted throughout.

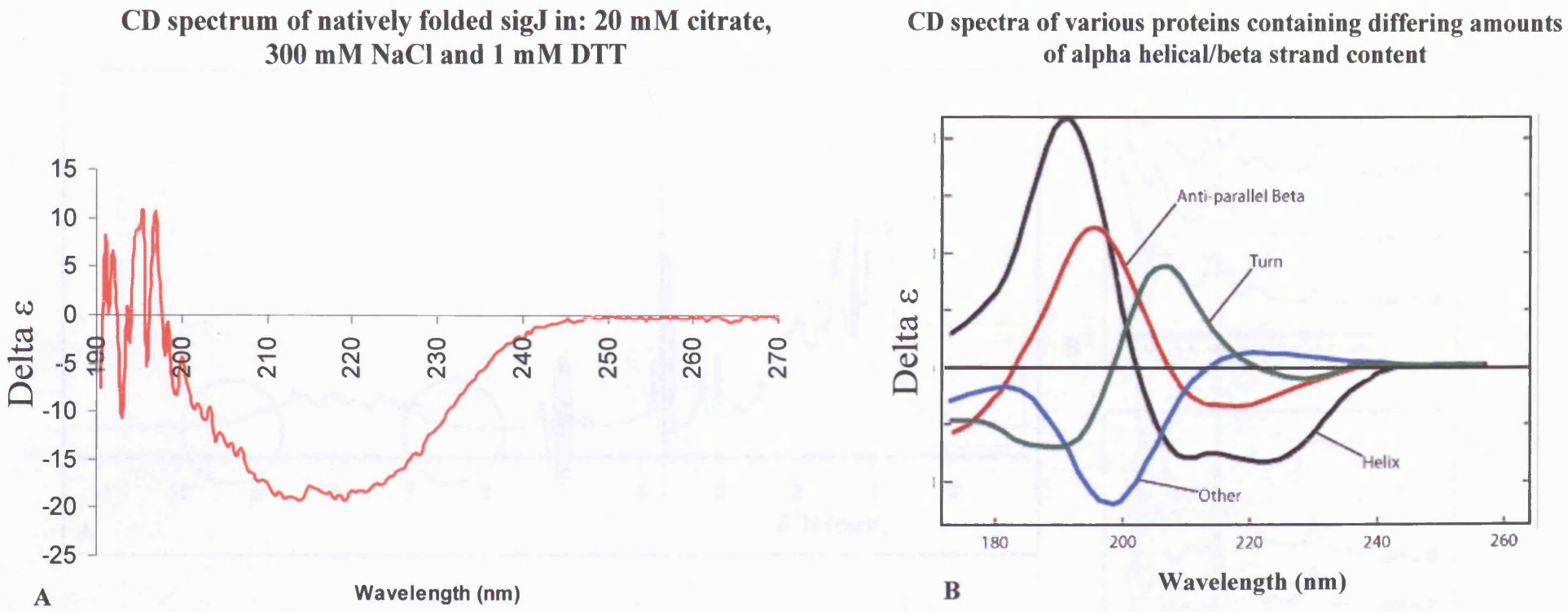


Figure 8.3: **A**: CD spectrum of natively folded (soluble) sigJ (pH 7.4, 25°C). **B**: Typical far-UV representation of different proteins exhibiting different secondary structures (adapted from; Kalodimos, 2004)

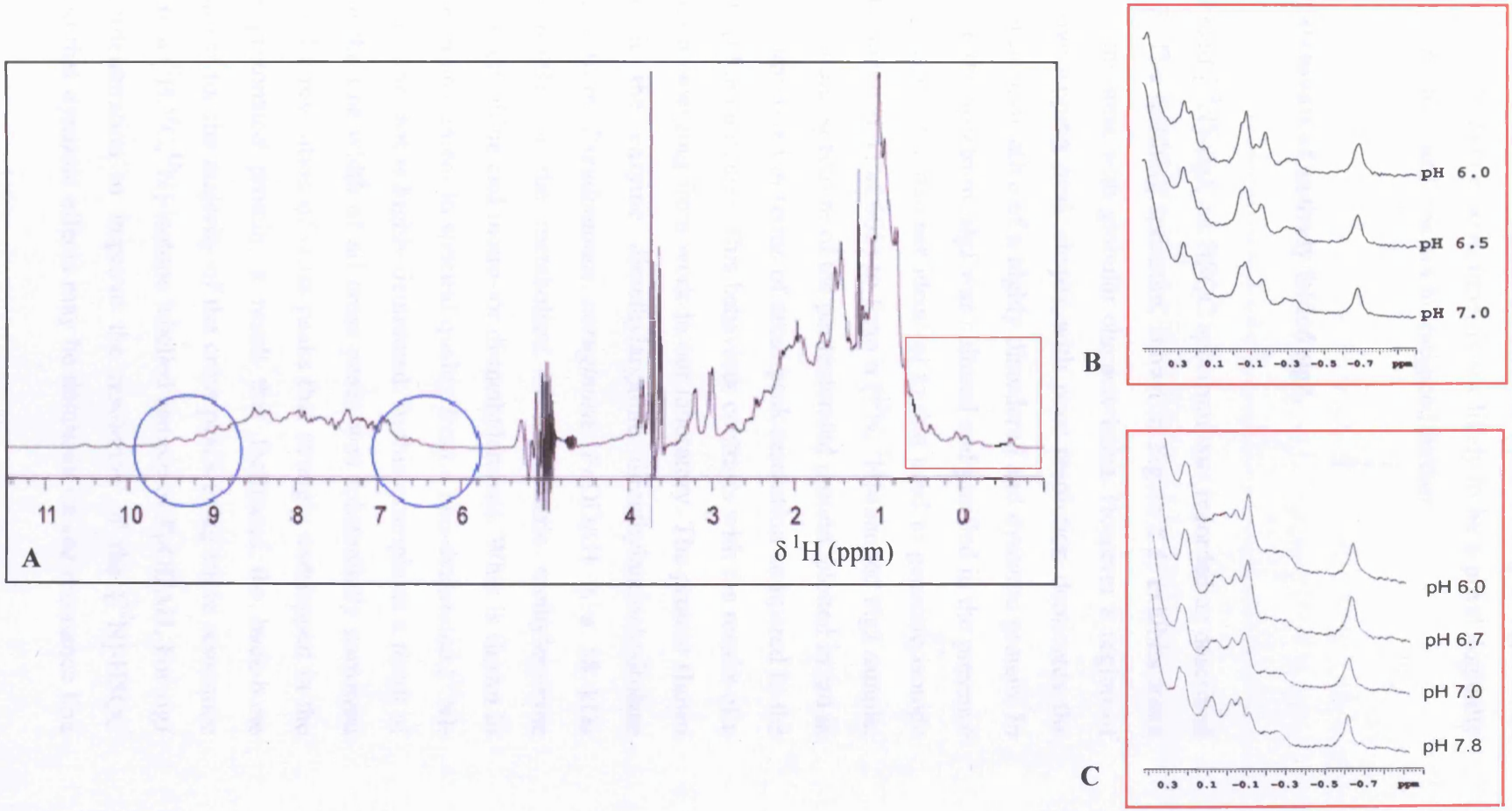


Figure 8.4: **A**: 1D [^{1}H]-NMR spectrum of natively folded sigJ pH 7.4, 1.0 mM, blue circled region; amide part of sigJ, red region; methyl part of sigJ. **B** and **C**: Natively folded and refolded sigJ, respectively, at various pH conditions. All spectra obtained at 500 MHz, at 1 mM, pH 7.4 and 25 °C.

As a result of the differences in behaviour between the natively expressed and refolded sigJ it was concluded that the latter material, particularly in light of its apparent aberrant gel filtration elution time, is not likely to be a physiologically relevant form of the protein, and was not investigated further.

8.6. HSQC measurements of natively folded sigJ.

Using natively folded [^{15}N]-sigJ, an HSQC spectrum was recorded as described in Section 2.3.7.3. The resulting spectrum, shown in Figure 8.5, exhibits some resolved features consistent with globular characteristics. However a region of inhomogeneous, overlapping peak shapes with poor resolution dominates the centre of the spectrum indicative of a highly disordered and dynamic protein. In an effort to improve the spectrum, sigJ was cultured and purified in the presence of deuterium rich media (in a manner identical to that used to generate isotope enriched [^{15}N, ^{2}H]-labelled TPxC60S) to form a [^{15}N, ^{2}H]-labelled sigJ sample. Surprisingly, the resulting spectrum of the perdeuterated protein, plotted in red in Figure 8.5, was not improved in terms of cross peak resolution compared to the protonated sample (blue contours). This behaviour contrasts with the results of a comparable situation emerging from work in our laboratory. The protein shown in Figure 8.5B, is the enzyme dimethylarginine dimethylaminohydrolase (DDAH), obtained from *Pseudomonas aeruginosa.* *Pa*DDAH is a 58 kDa homodimer responsible for the metabolism of asymmetric methylarginine (AMAs) residues to citrulline and mono- or di-methylamines. What is shown in Figure 8.5B is the improvement in spectral quality from a non-deuterated [^{15}N]-*Pa*DDAH protein to one that is highly deuterated. In this example as a result of the perdeuteration the line-width of all cross peaks was substantially narrowed clearly facilitating the resolution of cross peaks that strongly overlapped in the spectrum of the protonated protein, a result that facilitated the back-bone resonance assignment for the majority of the cross peaks using triple resonance methods applied to a [^{2}H,^{13}C,^{15}N]-isotope labelled sample of *Pa*DDAH. For sigJ the failure of perdeuteration to improve the resolution of the [^{15}N]-HSQC spectrum indicated that dynamic effects may be dominant on the resonance line

width. As a result of this extensive line broadening the application of triple resonance measurements sigJ sample was not pursued.

8.7. Limited Proteolysis – Introduction.

As described in Chapter 6, sigma factors appear to be comprised of loosely tethered globular domains, and there is scope that stable sub-domains could be isolated by controlled proteolysis of the full length protein within the linker regions. For example, the ECF sigma factor σ^R is a global regulator of redox homeostasis in the antibiotic-producing bacterium *S. coelicolor* and binds to the anti-sigma factor RsrA. Burton *et al,* (2002) utilized a combination of limited proteolysis, surface-enhanced laser desorption ionization mass spectroscopy, and pull down assays to identify a ~10 kDa N-terminal domain (σ^{RN}). The σ^{RN} domain encompasses the σ_2 region which is predicted to correspond to a folded globular domain. The crystal structure of the complex between σ^{RN} and RsrA was subsequently solved to 2.4 Å resolution. In a similar manner Campbell *et al,* (2002) utilized limited proteolysis to isolate a domain of sigE from *E. coli* encompassing regions $\sigma_{1.2}$, - $\sigma_{4.2}$, which was subsequently sub-cloned and then crystallized and its high resolution structure solved. It appears reasonable that the σ_4 region of sigma factors, which is found to bind the -35 DNA element (reviewed in Gross *et al,* 1998; Craig *et al,* 1998; Colland *et al,* 1998; Burton *et al,* 2002; Campbell *et al,* 2002; Vassylyev *et al,* 2002 and Jain *et al,* 2004) is loosely tethered to the rest of the protein able to freely bind DNA. The σ_4 region is usually linked to region σ_2 via region σ_3 (Murakami *et al,* 2002) which in some sigma factors, including sigJ, is notably absent. There is a reasonable prospect then that the sigJ σ_4 (and possibly σ_2) domain might be accessible by limited proteolysis.

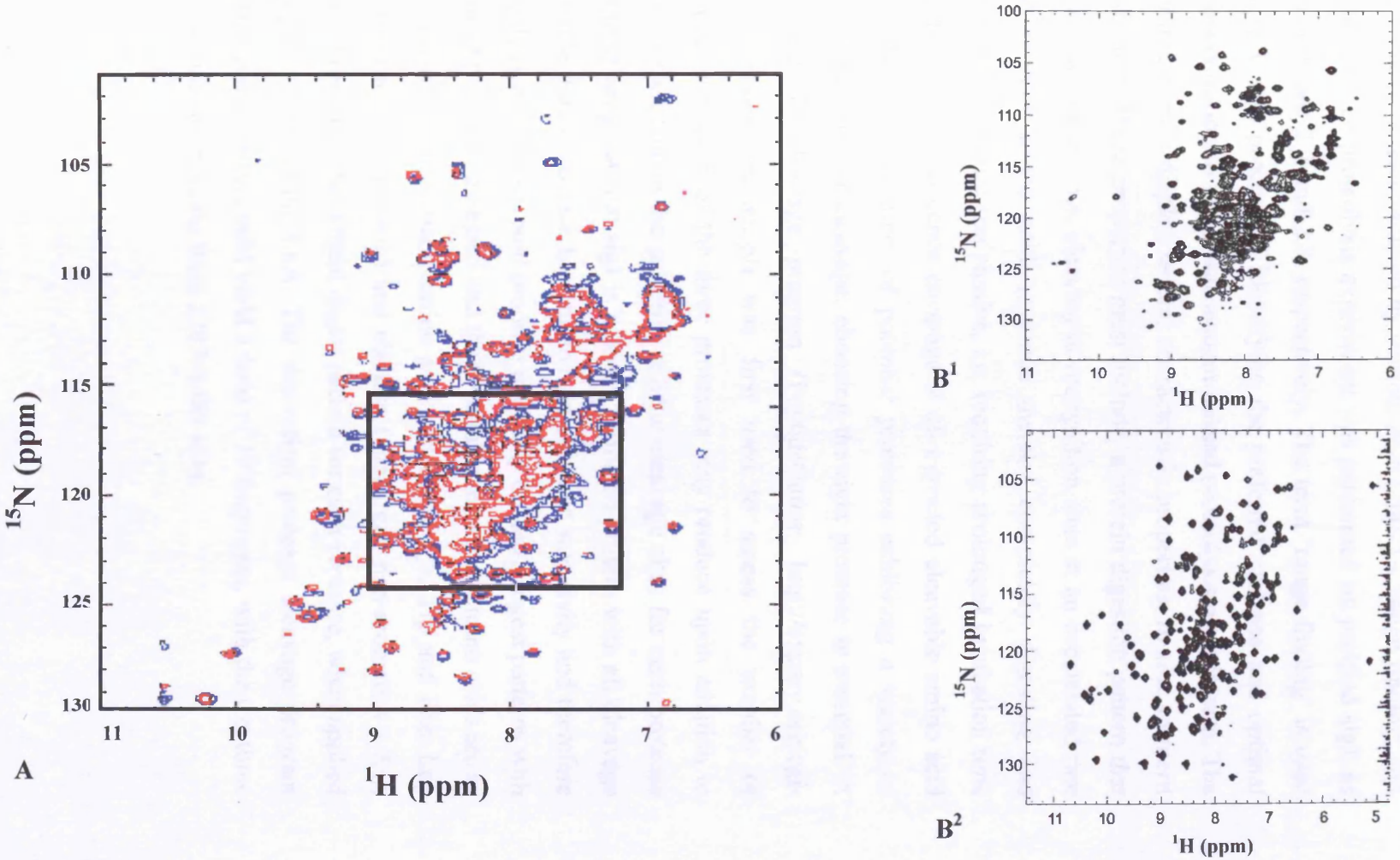
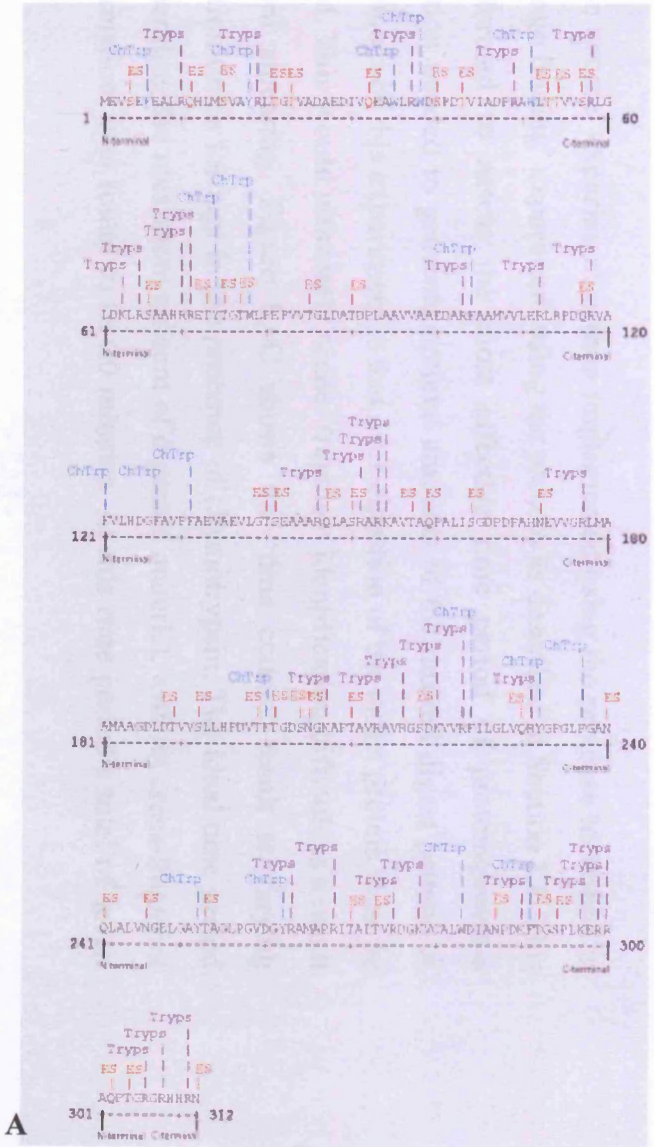
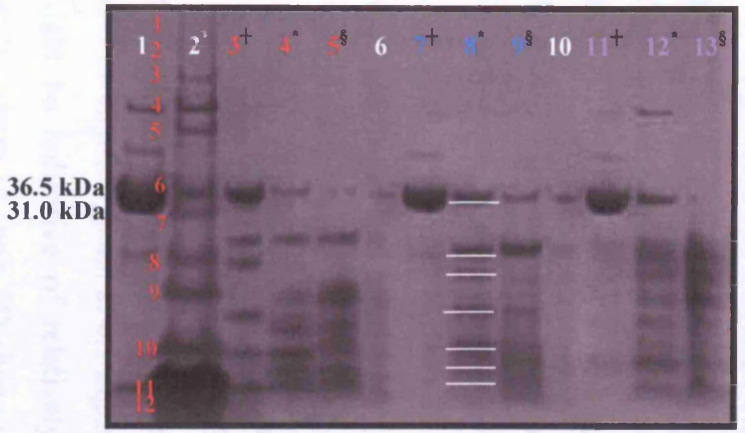


Figure 8.5: **A**: 600 MHz HSQC spectral overlay of; [^{15}N, ^{1}H]-sigJ – blue spectrum and [^{15}N, ^{2}H]-sigJ – red spectrum, both 1 mM and pH 7.4. **B^1** and **B^2**: HSQC spectra of [^{15}N, ^{1}H]-*Pa*DDAH and [^{15}N, ^{2}H]-*Pa*DDAH, respectively, at 1 mM and pH 6.4.

8.7.1. Limited Proteolysis of sigJ.

By applying the principles outlined by Burton, *et al* (2002) a 'range finding' and 'time course' limited proteolysis experiment was performed on purified sigJ, as detailed in Sections 2.7 and 2.8, respectively. The term 'range finding' is used here to refer to the process of identifying the preferred protease and optimal protease concentration to use in subsequent limited proteolysis experiments. The selected protease must display several characteristic properties, *in situ*, to merit its continued use. These properties must include; a protein digestion pattern that is neither too aggressive, i.e. cleaving at recognition sites at an accelerated rate across the protein amino acid sequence thereby completely digesting and unwinding the protein, nor too passive, i.e. requiring prolonged incubation time to achieve full protein sequence coverage at all expected cleavable amino acid residues. With a large number of potential proteases exhibiting a variety of amino acid selectivity for cleavage, choosing the right protease is essential. A theoretical peptide cleavage program (PeptideCutter: http://expasy.org/cgi-bin/peptidecutter/peptidecutter.pl) was first used to assess the number of potential fragments each of the three proteases may produce upon addition to sigJ. This was based upon the selective residue cleavage sites for each protease and with the assumption that sigJ is a linear, unfolded protein with all cleavage sites accessible. Each protease has amino acid cleavage selectivity and therefore each of the three proteases would produce markedly different digest patterns with sigJ. The three proteases assessed and the C-terminal cleavage amino acids are as follows; chymotrypsin (aromatic amino acid residues – Tyr, Trp and Phe, Leu and Glu), trypsin (Lys and Arg) and elastase (uncharged non-aromatics – Ser, Thr, Asn and Gln). The theoretical digest pattern for each protease, when applied to sigJ, is shown in Figure 8.6A. The theoretical protease cleavage program predicts that chymotrypsin could yield a total of 19 fragments, with the top three largest peptide masses ranging from 2387–1489 kDa.





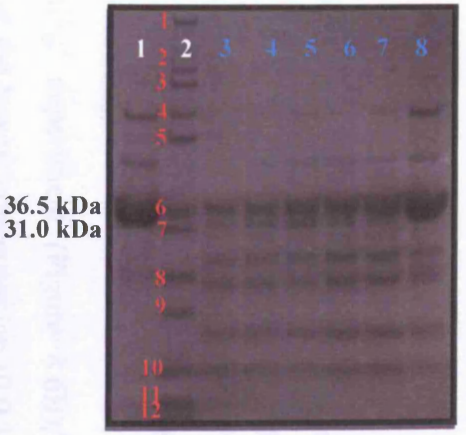


Figure 8.6: **A**: Theoretical protease cleavage of linear, unfolded sigJ using elastase (red), chymotrypsin (blue) and trypsin (magenta). **B**: SDS-PAGE gel of; 'range finding' experiment, using the same three colour-coded proteases from **A** upon sigJ. Protease concentration: ┼: 0.01 mg/ml, *: 0 .1 mg/ml and, §: 1 mg/ml. Lanes; 6 and 10 are empty. White lines indicate seven distinct fragments reproduced in **C**. **C**: SDS-PAGE gel of; 'time course' experiment using chymotrypsin upon sigJ, lanes 3 – 4 correspond to the following incubation times: 5 mins, 10 mins, 20 mins, 30 mins, 1 hour and 2 hours. For **B** and **C** SDS-PAGE gels, lane 1: undigested sigJ and lane 2: protein mass ladder. Red (diagonal) numbering indicates the M_w protein marker sizes for gels in B and C, 1: 200 kDa, 2: 116.3 kDa, 3: 97.4 kDa, 4: 66.3 kDa, 5: 55.4 kDa, 6: 36.5, 7: 31 kDa, 8: 21.5 kDa, 9: 14.4 kDa, 10: 6 kDa, 11: 3.5 kDa, and 12: 2.5 kDa.

The use of trypsin predicts a potential total of 34 fragments with a peptide mass range of 2490–1473 kDa, for the top three largest fragments. Elastase is predicted by far the most proficient with a potential 47 fragments generated and achieving the largest peptide mass range of 2500–1880 kDa for the top three largest possible peptide masses.

From the 'range finding' experiment (Figure 8.6B), elastase creates a well defined digest pattern at the lowest concentration (0.0.1 mg/mL) but proceeds to digest sigJ completely as the protease concentration increases, leaving coalesced small peptide fragments that are indefinable from one another. Trypsin has virtually no effect in digesting sigJ at the lowest concentration used but rapidly cleaves the protein at both 0.1 mg/mL and at 1 mg/mL leaving poorly resolved, indefinable bands on the SDS-PAGE gel. However, when chymotrypsin is used, little reaction is initially seen to occur at the 0.01 mg/mL concentration (Figure 8.6B, lane 7) but at 0.1 mg/mL (Figure 8.6B, lane 8) the protease cleaves sigJ yielding a series of well resolved, clearly defined bands. At the maximum protease concentration (Figure 8.6B, lane 9) 1 mg/mL, the reverse is true for the outcome – at 1 mg/mL, chymotrypsin rapidly cleaves sigJ leaving behind a poorly resolved smear of indefinable bands on the SDS-PAGE gel. It was clear from this 'range finding' experiment that the protease best used for the 'time course' experiment would be chymotrypsin, at 0.1 mg/mL. The presence of such well defined bands might be indicative of relatively stable folded modular components of sigJ that by virtue of their 3D shape would have the effect of shielding the other potential proteolytic cleavage sites.

A 'time course' experiment is then implemented using the protease identified in the 'range finding' experiment using the process as described in Section 2.8. This is performed to assess the most effective time period of protein/protease incubation needed to generate optimal amounts of the desired digest pattern. A requirement of this experiment is that over digestion of the target protein must be avoided. This would ultimately render fragment identification difficult as a result fragment smearing. Figure 8.6C shows the 'time course' result at varying incubation times for sigJ in the presence of chymotrypsin. The ideal time period for generating the maximum amount of digested material, without cross-fragment contamination, was found to be 20 minutes. At this time period a total of seven

fragments can be visually identified from the SDS-PAGE gel (including the presumably intact/undigested sigJ fragment) that persists across the digest pattern in Figure 8.6B, lane 8 (highlighted by white lines), and reproduced in Figure 8.6C. The molecular weights of each of the six digest fragments (not including intact sigJ) are estimated as ranging from between ~10 kDa and ~33 kDa. These seven fragments merit isolation and further detailed analysis by mass spectrometry-based methods. Limitations of time and competing experimental imperatives mean that this extended analysis has not so far been completed to a satisfactory endpoint (preliminary attempts to characterise the fragments have yielded conflicting results, perhaps as a result of contamination of the fragment samples with undigested full length sigJ. Successful separation of full length sigJ would provide additional information that can be used to design a series of isolated fragment-based expression constructs that could subsequently be expressed, purified and assessed using NMR-based experiments.

8.8. DNA binding.

As mentioned previously, the amino acid residues implicated in binding to the DNA promoter -10 element are located in a cleft-like feature of the sigma factor. In this region many sigma factors contain a highly acidic stretch of residues, often mostly disordered in nature. It is this acidic loop region that sterically inhibits DNA interaction and also repels the negatively charged DNA electrostatically making, primarily, DNA binding to σ_2 in the absence of core-RNAP impossible. However, not all sigma factors harbour this feature and based upon information from a collaborator at the St. Georges Hospital Biochemistry Research Department (SGH-BRD), another arm to the EU consortium on *Mtb* research, DNA binding to recombinant sigJ has been investigated.

The data obtained from SGH-BRD was from a discontinued avenue of research, and as such demanded validation of the claims that a sequence of DNA (300 bp in length), upstream of the sigF coding region in *Mtb*, was found to bind sigJ in the absence of core-RNAP. Data to support this claim was scarce however; Figure 8.7A displays the result of a saturation binding curve of the percentage of change in DNA fluorescence as a function of protein concentration. From the

data curve it can be seen that sigJ binds the 300 bp sigF DNA at the first instance at a concentration of ~1 μM and exponentially until a plateau at a sigJ concentration of ~45 μM, where ~80% of the available DNA has bound to the protein. The starting concentration of DNA used during this experiment was not provided.

With sigJ exhibiting very poor quality HSQC spectra as a result of what is believed to be intrinsic structural disorder, DNA binding may help to quench this fluctuating motion by binding to the -10 and -35 DNA binding domains of the protein and thereby 'lock' sigJ into a stable conformation. This expectation fuelled our interest to recapitulate the claim that sigJ could bind *Mtb* chromosomal DNA sequences in the absence of core RNAP.

The 300 bp DNA template which was claimed to bind to sigJ was amplified from genomic *Mtb* DNA using PCR and specifically designed primers (described in Section 2.4.1). By isolating and purifying the desired 300 bp fragment, a modified reproduction of the original binding experiment between sigJ and the DNA (Figure 8.7B) was performed using electrophoretic mobility shift assay (EMSA) as detailed in Section 2.5. The DNA concentrations used during these EMSA experiments ranged from 100 ng/reaction (19.2 μM) mixture to a diluted final concentration of 2.5 ng/reaction (0.48 μM) mixture as used for the last EMSA binding experiment. NOTE: A final concentration of 2.5 ng/reaction mixture is the minimum DNA concentration capable of visualization under UV detection using the Novex® DNA pre-cast retardation gels and staining using ethidium bromide. Several rounds of optimisation were required before the correct concentration ratio of sigJ-DNA concentrations was obtained in producing the results shown. The initial result of the EMSA binding experiment between sigJ and the 300 bp DNA fragment is shown in Figure 8.7B. The migration of the 300 bp fragment is shown to be gradually retarded as sigJ concentration increases, with the first indication of DNA migration inhibition evident at a sigJ concentration of ~0.5 mM. By decreasing the concentration of DNA used in the EMSA binding experiments, while maintaining the same sigJ concentration throughout, an estimate into the binding specificity of sigJ could be made. Figures 8.9C–E show that DNA at the lowest amounts measured (2.5 ng) in the presence of sigJ (45 μM) provides a DNA retardation pattern that is

consistently reproduced suggesting that indeed sigJ has the potential to interact and bind DNA.

However, with DNA binding specificity there are typically two modes of affinity that can be ultimately proposed. These two possible modes include: the presence of a sequence specific DNA feature, much like the DNA binding elements, -10 and -35, that sigma factors are expected to locate to (with amino acid/nucleotide interaction based upon non-polar contacts to hydrogen bonds both with and without solvent molecules acting as bridges), or the entropically favorable structural features of DNA molecules that allows them to orientate and associate with proteins through purely electrostatic interactions in a sequence independent ('non-specific') manner.

With sigJ exhibiting apparently concentration independent DNA binding, it must be questioned whether sigJ is binding to a specific DNA region located somewhere within the 300 bp DNA stretch. In an attempt to further test the DNA-binding characteristics of sigJ, several DNA smaller constructs were made from the proposed 300 bp DNA fragment, as described in Chapter 2.4.1. The four DNA constructs isolate various sub-sections of the initial 300 bp fragment while providing overlapping sections that would ultimately (if displaying differential affinity) single out a specific region of DNA that could be implicated in binding exclusively to sigJ. The EMSA binding experiments of sigJ with the four DNA constructs can be seen in Figure 8.8A-D.

The EMSA results indicate that DNA retardation is experienced with all four DNA constructs, with the initial indication of DNA migration being slowed within a sigJ concentration of ~0.4–0.5 mM. The conclusion to this observation puts into doubt the premise that the binding of DNA to sigJ is exclusively reliant upon a specific DNA sequence-specific recognition element.
To further examine the specific binding nature of sigJ, a randomly chosen 300 bp DNA template obtained from the coding region for the protein HNS (*Salmonella Typhimurium*), was used in a control EMSA binding experiments with sigJ (Figure 8.9A–D).

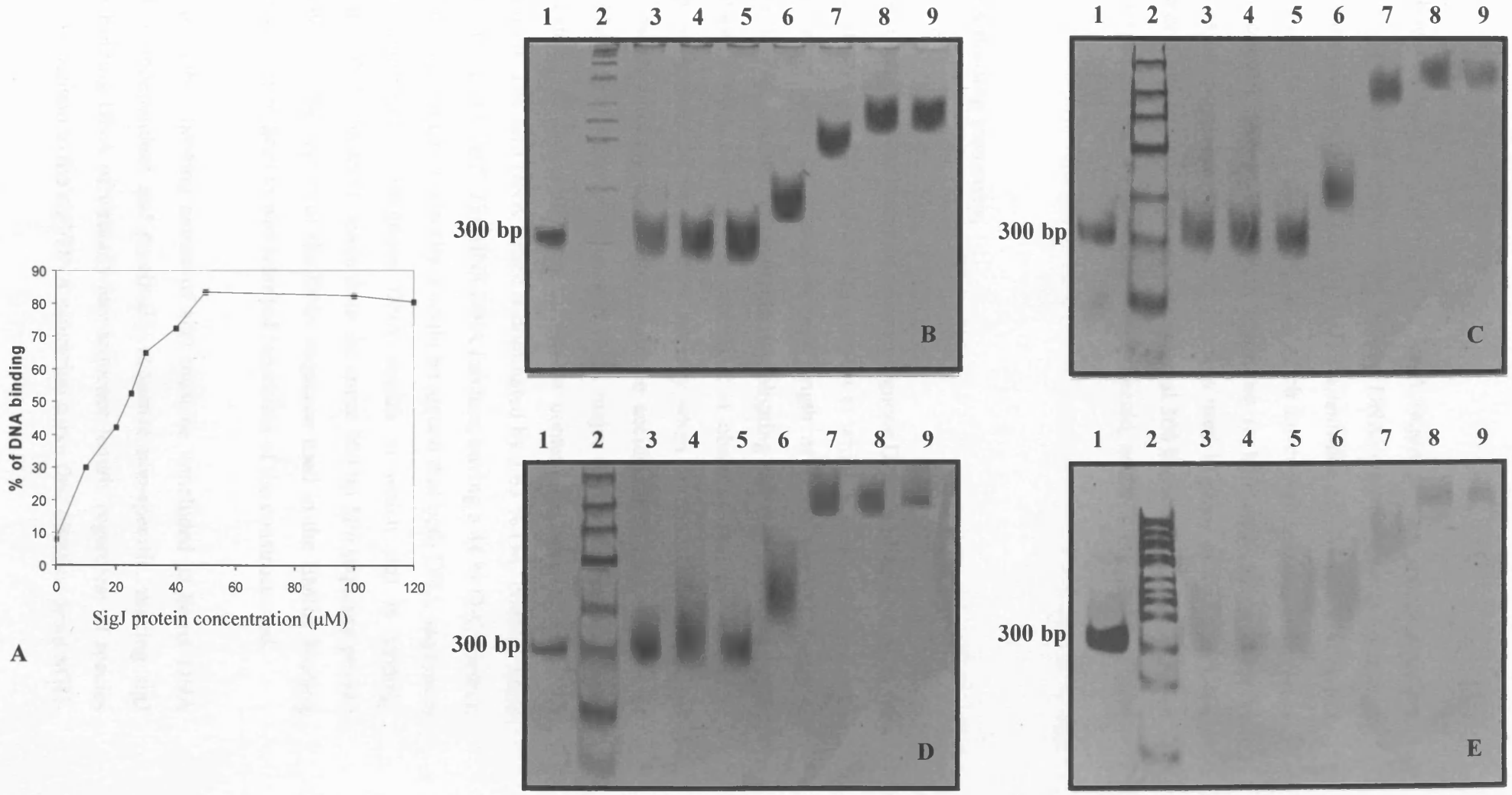


Figure 8.7: **A**: Saturation curve of sigJ in the presence of 300 bp *Mtb* sigF DNA. **B – E**: EMSA Novex® DNA precast retardation gels (6% agarose) of sigJ with 300 bp *Mtb* sigF DNA (varying concentration). DNA concentration for EMSA gel, **B**: 100 ng/lane, **C**: 60 ng/lane, **D**: 5 ng/lane, and **E**: 2.5 ng/lane. Lanes; 3 – 9, contain sigJ at the following concentrations; 0.1 mM, 0.3 mM, 0.4 mM, 0.5 mM, 0.6 mM, 0.8 mM and 1 mM, respectively. Lane 1 is 300 bp *Mtb* DNA with no sigJ present. Lane 2 is LMW ladder.

SigJ is shown to bind the 300 bp HNS DNA sequence with similar apparent affinity as that observed with the *Mtb* 300 bp DNA, with the first indication of DNA migration inhibition evident at a sigJ concentration of ~0.4–0.5 mM and in some instances as with Figure 8.9C, at a much lower sigJ concentration. As a secondary control, the small protein lysozyme (a hydrolase whose function clearly does not involve binding to DNA) was used in place of sigJ in a DNA binding-specificity examination with the original 300 bp *Mtb* DNA. The result of using lysozyme instead of sigJ indicates as expected, no DNA retardation (Figure 8.9E).

8.8.1. DNA binding summary.

With these results, sigJ still remains an unambiguous DNA binding protein in the EMSA paradigm. The DNA structural element to which it binds could be very small, and as a result be present in any length of DNA sequence used in subsequent EMSA binding experiments. By aligning the two DNA sequences used in these EMSA binding experiments it is observed that the sequences exhibit approximately 50 % sequence identity when optimally aligned at full length (it seems unlikely that there would be accidental shorter stretches of sequence identity) as shown in Figure 8.10. A major difference however is the differing extents of the constituent nucleotide content for each of the DNA sequences used. The *Mtb* DNA strand is dominated by a 63 % G-C content while in comparison the 300 bp *S. Typ* HNS DNA has less, having a 44 % G-C content. With 50% overall sequence identity it could be argued that both DNA sequences contain a homologous contiguous DNA stretch to which sigJ is binding. However, the DNA constructs made from the entire 300 bp *Mtb* sequence proved that regardless of the region of the DNA sequence used in the EMSA binding experiments, DNA migration was retarded regardless of the construct used.

Therefore the DNA binding nature of sigJ must be concluded as being DNA concentration independent and practically sequence non-specific, making sigJ capable of binding DNA of virtually any sequence length, regardless of species source. In comparison to the sigJ/DNA saturation curve first obtained from SGH-

BRD, where sigJ (within the range of 1-45 μM) was first observed to bind an indeterminate amount of DNA, the EMSA experiments performed here at least verify that sigJ does indeed bind DNA in a similar manner.

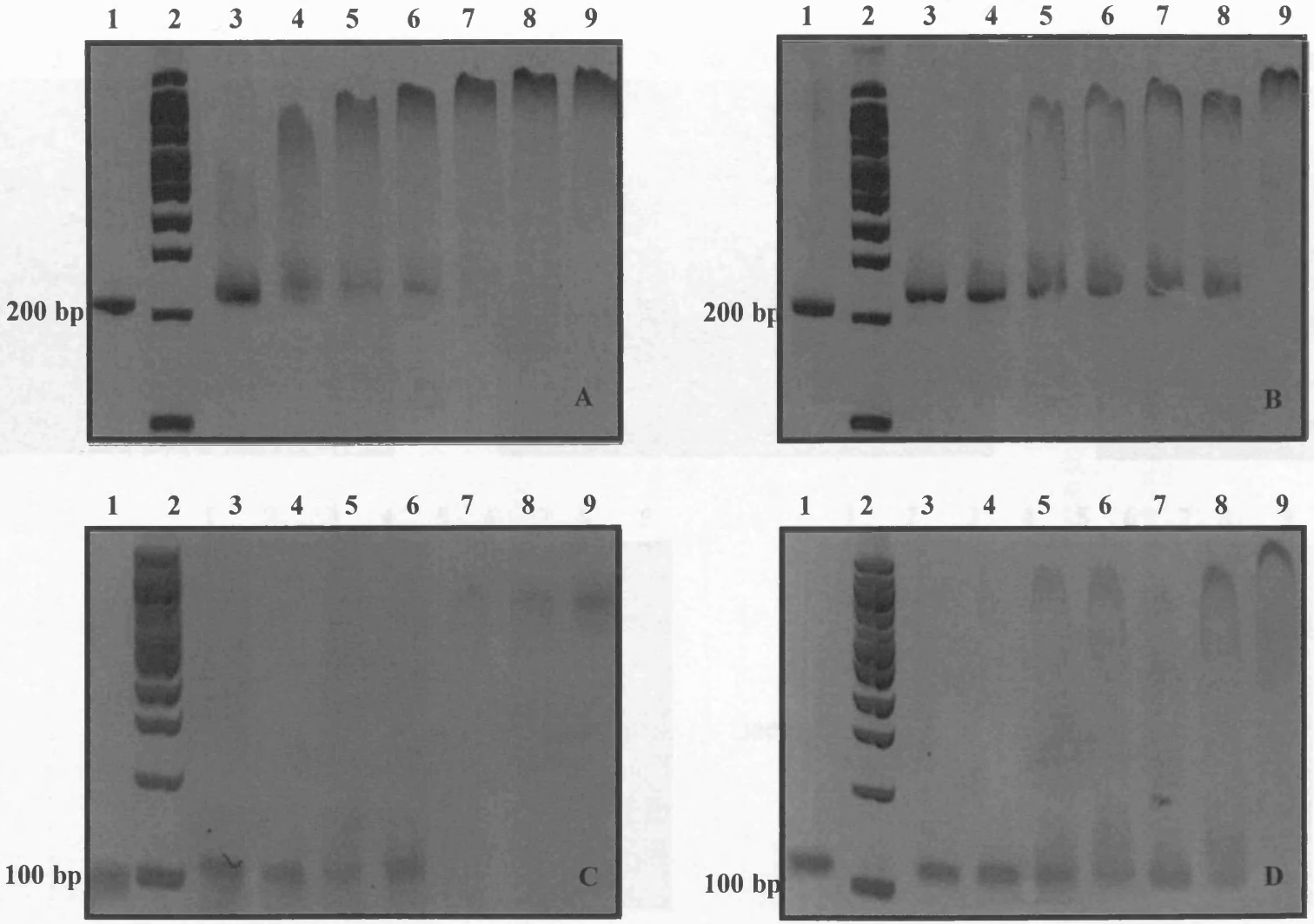


Figure 8.8: **A – D**: EMSA Novex® DNA precast retardation gels (6% agarose) of sigJ with various length *Mtb* sigF DNA constructs. EMSA gel, **A**: PepA (200 bp), **B**: PepB (200bp), **C**: PepC (120 bp), and **D**: PepD (120 bp). Lanes; 3 – 9, contain sigJ at the following concentrations; 0.1 mM, 0.3 mM, 0.4 mM, 0.5 mM, 0.6 mM, 0.8 mM and 1 mM, respectively. Lane 1 is 300 bp *Mtb* DNA with no sigJ present. Lane 2 is LMW ladder.

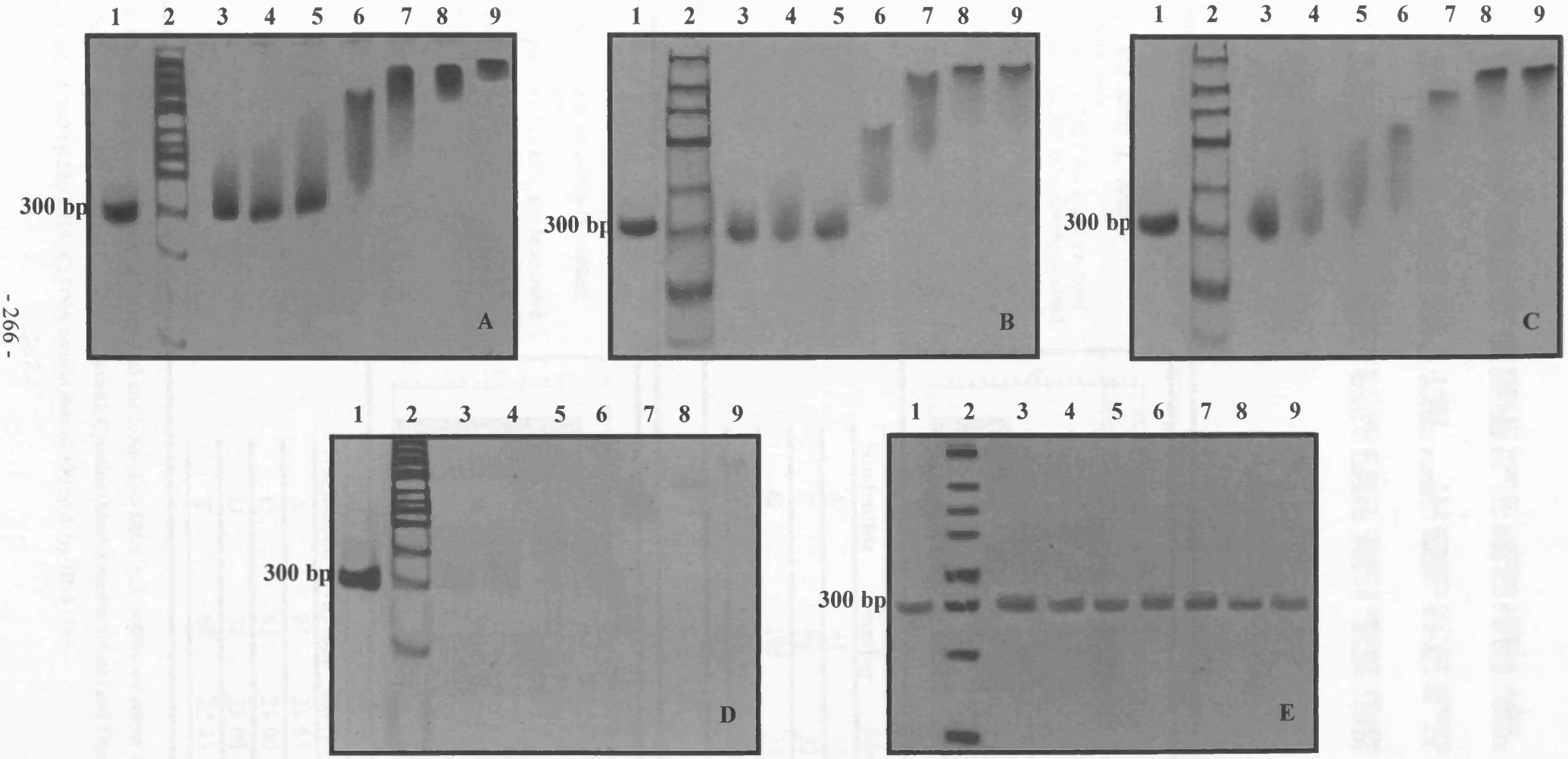
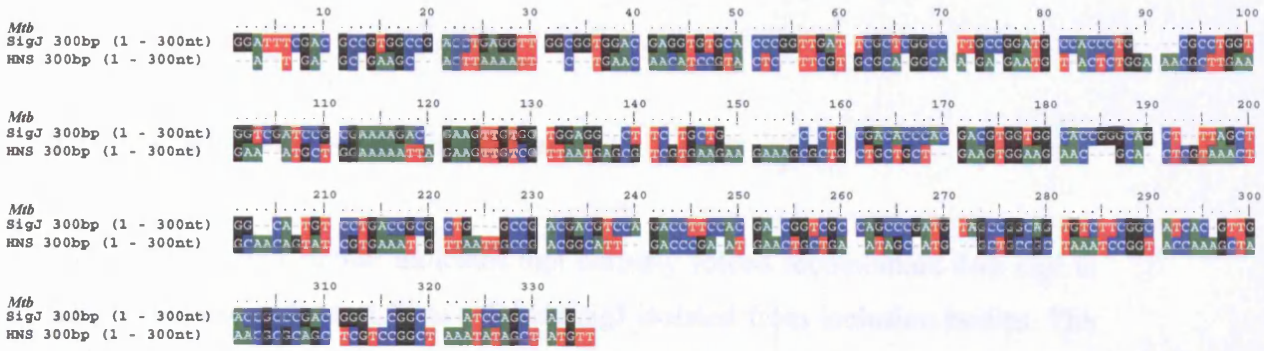
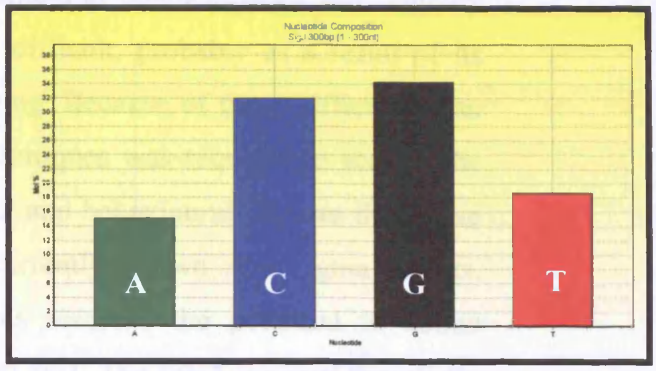


Figure 8.9: **A – D**: EMSA Novex® DNA precast retardation gels (6% agarose) of sigJ with 300 bp HNS (*S. Typhimurium*) (varying concentration). DNA concentration for EMSA gel, **A**: 100 ng/lane, **B**: 60 ng/lane, **C**: 5 ng/lane, and **D**: 2.5 ng/lane. EMSA gel **E**: lysozyme plus 300 bp *Mtb* sigF DNA. Lanes; 3 – 9, contain sigJ/lysozyme at the following concentrations; 0.1 mM, 0.3 mM, 0.4 mM, 0.5 mM, 0.6 mM, 0.8 mM and 1 mM, respectively. Lane 1 is 300 bp *Mtb* DNA with no sigJ present. Lane 2 is LMW ladder.



A

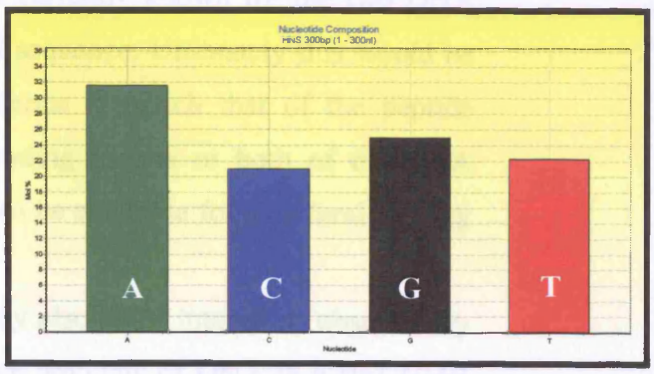
DNA molecule: ***Mtb* 300bp (1 - 300nt)**
Length = 300 base pairs
Molecular Weight = 91,342 kDa, single stranded
Molecular Weight = 18, 323 kDa, double stranded
G+C content = 66.33%
A+T content = 33.67%



Nucleotide	Number	Mol%
A	45	15.00
C	96	32.00
G	103	34.33
T	56	18.67

B

DNA molecule: ***S. typ* HNS 300bp (1 - 300nt)**
Length = 300 base pairs
Molecular Weight = 91,211 kDa, single stranded
Molecular Weight = 18,220 kDa, double stranded
G+C content = 46.00%
A+T content = 54.00%



Nucleotide	Number	Mol%
A	95	31.67
C	63	21.00
G	75	25.00
T	67	22.33

C

Figure 8.10: **A**: Global sequence alignment of 300 bp *Mtb* and *S. typ* HNS DNA. Nucleotides are colour coded to correspond to the block diagrams in **B** and **C**; Adenine (green), Cytosine (blue), Guanine (black) and Thymine (red). **B**: DNA content data of 300 bp *Mtb* DNA. **C**: DNA content data of 300 bp *S. typ* HNS DNA.

8.9. Discussion and concluding remarks.

The work in this Chapter indicates that natively folded recombinant *Mtb* sigJ is spectroscopically distinct from refolded sigJ isolated from inclusion bodies. The natively expressed sigJ is apparently monomeric and contains some globular folded component as witnessed by the appearance of the 1D and 2D NMR spectra. Nevertheless *Mtb* sigJ exhibits structural characteristics that initially made further NMR-based experimentation difficult, probably as a result of its dynamic properties leading to line broadening. Because of these difficulties, an investigation using various biochemical techniques was required to scratch the surface in trying to understand the structural and behavioural features that make sigJ essentially unique amongst the 13 currently known *Mtb* sigma factors. Preliminary limited proteolysis experiments revealed the potential to define relatively stable sub-fragments of full length sigJ. The DNA binding specificity of sigJ was also confirmed as being essentially DNA concentration independent and potentially sequence non-specific, as identified by EMSA-based protein/DNA binding experiments.

Ideally the next step would be isolate some of the peptide fragments obtained from the limited proteolysis of sigJ, as a means to ascertain the amino acid sequence and to compare it to the location currently known for the two DNA binding domains along the sigJ amino acid sequence. Ultimately this would be the prelude to constructing a peptide sequence to match that of the peptide isolated using limited proteolysis corresponding to one or both of the DNA binding domains. Such a construct would then be available for structural probing using NMR and CD spectroscopy.

As for the DNA binding aspect of sigJ it may also be of interest to observe any effects that may occur upon addition of large amounts of DNA to a sigJ NMR sample and the subsequent HSQC measurements taken. Additionally it would be desirable to reconstitute sigJ-dependent action of *Mtb* RNAP to screen for potential sequence-specific sigJ DNA recognition elements.

Appendix:

A.1: The standard genetic code.

Table A1.1:

First Position	Second Position								Third Position
	T		C		A		G		
T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T
	TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys	C
	TTA	Leu	TCA	Ser	TAA	Stop	TGA	Stop	A
	TTG	Leu	TCG	Ser	TAG	Stop	TGG	Trp	G
C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T
	CTC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T
	ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	ATG	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T
	GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

A.2: Recombinant protein expression plasmid.

Shown below are diagrammatical representations of the plasmid vectors used in this project. Each includes the following features; open reading frames, promoters and unique restriction enzyme sites used for cloning.

For full-length TPxWt, TPxC60S and sigJ were each cloned separately into the same plasmid vector-type; [pET22b(+)] as shown in Figure A2.1.

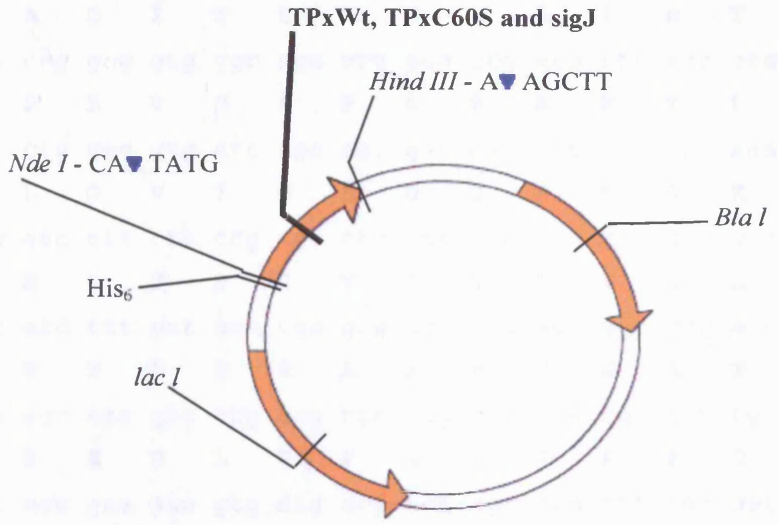


Figure A2.1: Plasmid vector [pET22b(+)] (Novagen®).
Refer to the **Table**, below, for nomenclature description.

Table A2.1: The nomenclature for plasmid vector [pET22b(+)] in Figure A2.1.

ORFs	Promoters
Bla I: ampicillin resistance	T7: T7 promoter
Lac I: Lac repressor	N/A
His_6: His-tag coding sequence	N/A

A.3: TPxWt nucleotide and amino acid sequence.

Mtb TPxWt: Met1-Ala165. Vector designation and size (base pairs): [pLEXTIJ46-1] and 5898 bp. Vector insert positioning (nucleotides): *Nde I* (5204) – *Hind III* (5722).

Shown below is the nucleotide and amino acid sequence (displayed in bold underneath the triplet codon) for TPxWt as confirmed by MWG-Biotech™ sequencing services as described in Chapter 2.2.1.

```
atg gcg cag att acc ctg cgc ggc aac gcg att aac acc gtg ggc gaa
M   A   Q   I   T   L   R   G   N   A   I   N   T   V   G   E
ctg ccg gcg gtg ggc agc ccg gcg ccg gcg ttt acc ctg acc ggc ggc
L   P   A   V   G   S   P   A   P   A   F   T   L   T   G   G
gat ctg ggc gtg att agc agc gat cag ttt cgc ggc aaa agc gtg ctg
D   L   G   V   I   S   S   D   Q   F   R   G   K   S   V   L
ctg aac att ttt ccg agc gtg gat acc ccg gtg tgc gcg acc agc gtg
L   N   I   F   P   S   V   D   T   P   V   C   A   T   S   V
cgc acc ttt gat gaa cgc gcg gcg gcg agc ggc gcg acc gtg ctg tgc
R   T   F   D   E   R   A   A   A   S   G   A   T   V   L   C
gtg agc aaa gat ctg ccg ttt gcg cag aaa cgc ttt tgc ggc gcg gaa
V   S   K   D   L   P   F   A   Q   K   R   F   C   G   A   E
ggc acc gaa aac gtg atg ccg gcg agc gcg ttt cgc gat agc ttt ggc
G   T   E   N   V   M   P   A   S   A   F   R   D   S   F   G
gaa gat tat ggc gtg acc att gcg gat ggc ccg atg gcg ggc ctg ctg
E   D   Y   G   V   T   I   A   D   G   P   M   A   G   L   L
gcg cgc gcg att gtg gtg att ggc gcg gat ggc aac gtg gcg tat acc
A   R   A   I   V   V   I   G   A   D   G   N   V   A   Y   T
gaa ctg gtg ccg gaa att gcg cag gaa ccg aac tat gaa gcg gcg ctg
E   L   V   P   E   I   A   Q   E   P   N   Y   E   A   A   L
gcg gcg ctg ggc gcg
A   A   L   G   A
```

A.4: TPxC50S nucleotide and amino acid sequence.

Mtb TPxC60S: Met1-Ala165. Vector designation and size (base pairs): [pLEXTIJ46-2] and 5898 bp. Vector insert positioning (nucleotides): *Nde I* (5204) – *Hind III* (5722).

Shown below is the nucleotide and amino acid sequence (displayed in bold underneath the triplet codon) for TPxC60S as confirmed by MWG-Biotech™ sequencing services as described in Chapter 2.2.1. The mutation C60S is highlighted in red.

```
atg gcg cag att acc ctg cgc ggc aac gcg att aac acc gtg ggc gaa
M   A   Q   I   T   L   R   G   N   A   I   N   T   V   G   E
ctg ccg gcg gtg ggc agc ccg gcg ccg gcg ttt acc ctg acc ggc ggc
L   P   A   V   G   S   P   A   P   A   F   T   L   T   G   G
gat ctg ggc gtg att agc agc gat cag ttt cgc ggc aaa agc gtg ctg
D   L   G   V   I   S   S   D   Q   F   R   G   K   S   V   L
ctg aac att ttt ccg agc gtg gat acc ccg gtg agc gcg acc agc gtg
L   N   I   F   P   S   V   D   T   P   V   S   A   T   S   V
cgc acc ttt gat gaa cgc gcg gcg gcg agc ggc gcg acc gtg ctg tgc
R   T   F   D   E   R   A   A   A   S   G   A   T   V   L   C
gtg agc aaa gat ctg ccg ttt gcg cag aaa cgc ttt tgc ggc gcg gaa
V   S   K   D   L   P   F   A   Q   K   R   F   C   G   A   E
ggc acc gaa aac gtg atg ccg gcg agc gcg ttt cgc gat agc ttt ggc
G   T   E   N   V   M   P   A   S   A   F   R   D   S   F   G
gaa gat tat ggc gtg acc att gcg gat ggc ccg atg gcg ggc ctg ctg
E   D   Y   G   V   T   I   A   D   G   P   M   A   G   L   L
gcg cgc gcg att gtg gtg att ggc gcg gat ggc aac gtg gcg tat acc
A   R   A   I   V   V   I   G   A   D   G   N   V   A   Y   T
gaa ctg gtg ccg gaa att gcg cag gaa ccg aac tat gaa gcg gcg ctg
E   L   V   P   E   I   A   Q   E   P   N   Y   E   A   A   L
gcg gcg ctg ggc gcg
A   A   L   G   A
```

A.5: Amino acid one and three letter codes.

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

Amino acid nomenclature used in this thesis:

X60: X = amino acid residue number 60.

C60S: Cysteine 60 mutated to Serine.

A.6: Triple resonance chemical shift data for [^{2}H, ^{15}N, ^{13}C]-labelled TPxC60S at pH 7.4.

Table A6.1:

Residue	N	HN	C_α	CO	C_β	Residue	N	HN	C_α	CO	C_β
1 Met	-	-	56.875	175.515	42.658	**81 Val**	123.203	8.836	60.317	175.888	34.592
2 Ala	124.424	8.221	53.071	177.731	19.622	**82 Ser**	117.315	7.832	57.383	64.349	173.408
3 Gln	118.956	8.348	56.408	176.252	29.802	**94 Gly**	106.657	8.204	45.721	174.253	-
4 Ile	121.487	8.109	61.823	176.535	39.337	**95 Ala**	122.147	8.008	53.218	177.766	20.240
5 Thr	118.127	8.177	62.371	174.491	70.307	**96 Glu**	119.544	8.580	58.120	177.682	29.665
6 Leu	124.514	8.203	55.446	177.039	42.774	**97 Gly**	110.296	8.523	45.560	174.518	-
7 Arg	121.398	8.258	56.568	176.844	31.141	**98 Thr**	114.895	7.712	61.596	174.718	70.344
8 Gly	109.636	8.409	45.41	173.887	-	**99 Glu**	124.353	8.800	57.834	176.939	30.786
9 Asn	118.655	8.259	52.264	174.514	39.692	**100 Asn**	114.517	8.682	54.423	172.561	38.698
10 Ala	123.655	8.162	52.791	177.564	20.098	**101 Val**	117.675	7.059	61.401	175.398	34.528
11 Ile	118.732	8.039	60.305	174.735	40.089	**102 Met**	124.029	9.000	53.234	172.511	35.527
12 Asn	123.564	8.645	52.294	175.284	40.362	**103 Pro**	-	-	61.791	175.957	34.041
13 Thr	113.255	8.465	60.123	175.77	71.634	**104 Ala**	122.066	8.932	51.567	175.852	24.656
14 Val	116.643	8.388	62.719	175.672	32.937	**105 Ser**	-	9.462	56.527	-	64.897
15 Gly	107.912	7.626	45.034	172.495	-	**112 Gly**	103.905	9.306	47.711	176.314	-
16 Glu	119.11	8.062	54.863	175.64	32.896	**113 Glu**	119.485	7.395	60.474	179.418	29.474
17 Leu	124.618	8.274	53.908	175.358	42.337	**114 Asp**	121.578	8.512	57.202	178.000	40.064
18 Pro	-	-	62.667	175.891	32.272	**115 Tyr**	116.416	8.201	59.823	173.34	38.08
19 Ala	124.061	8.075	52.239	179.174	19.499	**116 Gly**	105.147	6.92	46.725	174.312	-
20 Val	123.498	8.488	65.282	178.164	31.798	**117 Val**	104.723	7.877	60.552	175.573	33.492
21 Gly	115.789	9.356	44.686	173.241	-	**118 Thr**	120.02	7.232	64.944	174.127	68.899
22 Ser	117.082	7.836	57.107	-	-	**119 Ile**	129.165	9.334	64.018	178.255	37.733
23 Pro	-	-	63.470	177.636	32.149	**120 Ala**	130.572	9.068	54.388	176.631	21.184
24 Ala	128.43	8.501	50.906	175.672	18.671	**121 Asp**	113.084	7.374	53.302	175.308	43.592
25 Pro	-	-	62.972	176.921	31.556	**122 Gly**	105.368	8.415	45.443	173.691	-
26 Ala	127.159	8.566	52.625	177.979	19.168	**123 Pro**	-	-	65.747	178.269	32.899
27 Phe	111.748	8.091	56.625	174.251	43.507	**124 Met**	112.607	8.861	55.337	173.958	32.986
28 Thr	115.52	9.193	61.101	173.09	71.839	**125 Ala**	120.636	7.360	54.059	178.052	18.366
29 Leu	121.855	9.006	53.385	176.343	48.421	**126 Gly**	108.475	8.667	45.527	175.735	-
30 Thr	116.136	8.69	63.368	175.710	70.477	**127 Leu**	118.664	8.147	54.495	175.329	42.337
31 Gly	114.907	8.908	44.173	175.917	-	**128 Leu**	116.358	7.632	54.388	-	41.010
32 Gly	105.591	8.821	47.721	176.973	-	**129 Ala**	120.427	8.268	51.496	175.516	19.822
33 Asp	117.17	7.991	52.635	178.588	40.293	**130 Arg**	114.927	7.457	54.885	176.067	29.369
34 Leu	112.295	8.105	57.603	175.944	38.033	**131 Ala**	129.162	8.186	51.685	174.657	23.358
35 Gly	106.696	8.209	44.190	173.534	-	**132 Ile**	117.115	8.688	60.165	175.354	42.822
36 Val	120.399	8.430	63.111	176.966	34.172	**133 Val**	124.382	8.593	61.633	174.489	36.741
37 Ile	129.707	8.729	61.014	174.698	40.634	**134 Val**	126.22	9.287	61.609	175.041	34.658
38 Ser	120.269	8.442	55.507	175.751	66.001	**135 Ile**	131.241	9.788	61.478	175.885	40.412
39 Ser	119.562	8.215	62.021	174.211	60.876	**136 Gly**	111.033	8.731	44.566	174.803	-
40 Asp	121.727	8.057	56.91	178.101	40.58	**137 Ala**	123.598	8.645	54.951	177.763	18.587

41 Gln	117.553	7.629	57.268	176.584	28.637
42 Phe	117.092	7.204	56.402	174.318	39.585
43 Arg	120.154	6.914	58.293	178.366	29.962
44 Gly	114.257	8.977	45.805	173.581	-
45 Lys	120.48	8.302	54.858	174.519	35.596
46 Ser	115.802	8.423	58.577	174.281	64.81
47 Val	121.775	9.012	60.568	173.421	36.796
48 Leu	128.563	9.072	54.49	173.872	44.594
49 Leu	125.855	9.440	54.097	172.965	42.248
50 Asn	127.504	9.177	51.327	173.397	39.008
51 Ile	125.327	8.711	63.392	175.606	39.937
52 Phe	124.701	9.297	56.262	175.083	44.08
53 Ser	-	-	60.563	175.597	63.640
64 Val	121.133	7.731	65.165	176.839	32.421
65 Arg	120.457	7.856	58.885	178.221	30.440
66 Thr	112.969	7.901	64.347	178.260	58.199
67 Phe	122.076	7.773	60.756	176.054	38.983
68 Asp	118.866	8.346	57.630	178.525	40.473
69 Glu	117.72	8.070	59.559	180.034	30.319
70 Arg	120.059	8.044	59.073	179.671	31.123
71 Ala	122.43	8.446	55.425	180.339	17.976
72 Ala	122.50	8.201	54.893	181.476	18.406
73 Ala	121.08	7.902	54.584	179.082	18.609
74 Ser	111.793	7.383	59.713	174.691	65.026
75 Gly	109.46	7.803	45.211	174.223	-
76 Ala	122.929	7.470	51.726	176.953	20.914
77 Thr	117.218	7.895	63.269	172.899	70.979
78 Val	127.74	8.834	61.585	174.041	34.083
79 Leu	125.667	8.722	52.590	175.877	44.027

138 Asp	114.423	7.944	53.452	176.573	40.188
139 Gly	108.337	8.349	45.180	173.894	-
140 Asn	118.35	8.057	51.622	176.135	39.586
141 Val	122.874	8.983	65.254	176.309	31.749
142 Ala	132.312	9.641	52.815	177.096	27.907
143 Tyr	117.84	7.552	58.99	173.328	41.707
144 Thr	118.308	7.699	60.054	172.486	72.777
145 Glu	118.861	8.460	57.217	173.511	33.611
146 Leu	129.806	8.579	53.859	176.242	44.128
147 Val	128.136	8.145	60.887	175.337	32.036
148 Pro	-	-	64.918	176.27	33.002
149 Glu	116.144	7.00	55.289	176.841	32.542
150 Ile	120.719	8.313	63.081	176.085	38.888
151 Ala	122.050	8.339	52.978	176.968	18.780
152 Gln	118.350	7.373	55.456	175.175	30.666
153 Glu	122.110	8.421	54.348	174.83	31.043
154 Pro	-	-	61.881	177.176	32.716
155 Asn	120.319	10.256	51.868	174.99	36.920
156 Tyr	124.352	7.394	61.166	177.513	38.176
157 Glu	116.424	8.331	59.926	-	29.282
158 Ala	122.462	7.475	54.619	-	121.821
159 Ala	120.821	7.189	54.823	-	18.955
160 Leu	114.77	8.047	57.573	180.371	41.406
161 Ala	122.081	7.965	54.883	180.97	17.918
162 Ala	119.969	7.232	54.145	178.605	18.838
163 Leu	116.177	7.336	55.347	177.307	43.514
164 Gly	107.435	7.704	45.661	173.47	-
165 Ala	128.391	7.670	-	172.127	21.634

A.7: Spin systems identified from unassigned cross peaks found in the TPxC60S HSQC spectrum.

$C_{\alpha,1}$ and $C_{\alpha,i-1}$ chemical shifts are presented for all the triple resonance experiments performed.

Table A7.1:

2D NMR (^{1}H, ^{15}N) and Triple Resonance Experiments (^{1}H, ^{13}C)

(Spectral co-ordinates of nuclei presented in ppm. All first co-ordinates set = $C_{\alpha,i}$, second co-ordinate set = $C_{\alpha,i-1}$)

Spin System	HSQC	HNCA	HN(CO)CA	HNCB	HN(CO)CB	HNCO	HN(CA)CO
1	8.616/131.982	8.604/50.901 8.614/55.824	-	8.630/18.899	-	8.630/177.376 (previous)	8.615/177.185 (weak)
2	7.823/123.415	7.817/61.448	-	7.810/41.938	-	7.834/174.181	7.824/175.830
3	7.669/122.723	7.655/55.858	-	-	-	7.684/176.264	7.694/177.630 7.684/176.245
4	8.198/122.104	8.186/52.083 8.191/53.787	8.178/52.004	8.190/42.703	-	8.198/176.046	8.195/176.185

5	8.109/120.010	8.097/63.438 8.098/61.216	8.102/63.436	8.110/31.216 8.109/33.284	-	8.125/176.802	8.119/176.790 8.118/176.180
6	8.560/119.692	8.527/50.942 8.536/60.453	8.526/50.845	-	-	8.557/177.026	8.558/176.890 8.560/174.209
7	7.332/119.366	7.307/54.184	-	-	-	7.327/173.704	-
8	9.155/118.380	9.137/57.562	-	-	-	9.160/174.493	-
9	7.174/117.740	7.166/52.200 7.164/56.924	-	7.161/18.256	-	7.196/176.571	7.198/178.543
10	8.227/117.253	8.235/52.788	-	-	-	-	8.239/176.636
11	8.073/116.204	8.061/55.516	-	-	-	8.099/173.336	-

12	8.281/115.748	8.280/64.778	-	8.275/51.812	-	8.291/176.654	-
13	7.328/114.376	7.323/55.251 7.326/52.184	7.331/52.229	-	-	7.342/178.600	7.349/175.748
14	5.724/112.085	-	-	-	-	5.745/178.342	-
15	7.033/107.438	7.023/55.759	-	-	-	7.048/178.296	7.052/174.537
16	8.964/114.025	8.964/54.204 8.963/59.386	8.963/54.127	8.969/63.580	-	8.982/176.282	8.976/174.110
17	7.239/117.523	7.229/58.992	-	-	-	7.250/175.736	-
18	7.406/118.640	7.404/61.284	-	-	-	-	-

A.8: Triple resonance chemical shift data for [^{2}H, ^{15}N, ^{13}C]-labelled TPxC60S at pH 6.0.

Table A8.1:

Residue	N	HN	C_α	CO	Residue	N	HN	C_α	CO
1 Met	-	-	-	-	**81 Val**	123.281	8.830	-	-
2 Ala	125.325	8.354	55.302	174.295	**82 Ser**	113.420	7.780	60.384	175.773
3 Gln	120.010	8.400	52.110	177.408	**94 Gly**	106.051	7.856	-	174.447
4 Ile	122.190	8.270	55.471	175.861	**95 Ala**	121.850	7.990	-	-
5 Thr	118.352	8.343	60.703	176.224	**96 Glu**	120.256	8.645	52.437	178.005
6 Leu	124.689	8.351	61.395	174.234	**97 Gly**	110.502	8.555	57.501	177.627
7 Arg	121.350	8.280	54.641	176.916	**98 Thr**	115.824	7.718	45.174	174.538
8 Gly	109.018	8.403	55.873	176.783	**99 Glu**	125.289	8.752	61.326	174.521
9 Asn	118.608	8.256	45.044	173.844	**100 Asn**	114.321	8.599	57.234	176.808
10 Ala	123.650	8.187	52.764	174.604	**101 Val**	117.269	7.062	53.928	172.491
11 Ile	118.746	8.110	52.175	177.463	**102 Met**	123.899	9.021	60.723	175.337
12 Asn	123.444	8.705	59.527	174.671	**103 Pro**	-	-	-	-
13 Thr	112.992	8.484	51.775	175.340	**104 Ala**	122.439	8.961	61.203	176.099
14 Val	116.497	8.416	59.509	175.906	**105 Ser**	114.243	9.496	50.916	175.712
15 Gly	107.932	7.611	61.960	175.692	**112 Gly**	103.954	9.342	61.397	175.888
16 Glu	118.845	8.110	44.662	177.457	**113 Glu**	119.457	7.392	47.383	176.278
17 Leu	124.972	8.243	54.148	175.546	**114 Asp**	121.525	8.491	59.715	179.345
18 Pro	-	-	-	-	**115 Tyr**	116.452	8.219	56.557	177.988
19 Ala	124.081	8.083	62.013	175.854	**116 Gly**	105.169	6.916	59.289	173.318
20 Val	123.429	8.489	51.565	179.143	**117 Val**	104.895	7.851	46.341	174.222
21 Gly	115.823	9.368	64.497	178.131	**118 Thr**	119.960	7.233	59.737	175.574
22 Ser	117.112	7.855	44.325	173.221	**119 Ile**	129.187	9.347	64.390	174.079
23 Pro	-	-	-	-	**120 Ala**	130.588	9.082	63.259	178.190
24 Ala	128.408	8.497	62.796	177.614	**121 Asp**	113.167	7.380	53.687	176.595
25 Pro	-	-	-	-	**122 Gly**	105.296	8.414	52.756	175.261
26 Ala	127.130	8.563	62.322	176.891	**123 Pro**	-	-	-	-
27 Phe	111.625	8.088	51.957	177.919	**124 Met**	112.634	8.861	65.110	178.240
28 Thr	115.366	9.166	56.098	174.238	**125 Ala**	120.661	7.364	54.641	173.935
29 Leu	122.320	9.039	-	173.134	**126 Gly**	108.448	8.663	53.376	178.042
30 Thr	116.165	8.679	52.820	176.289	**127 Leu**	118.711	8.162	45.152	175.666
31 Gly	114.934	8.917	62.735	175.712	**128 Leu**	116.433	7.675	53.869	175.389
32 Gly	105.555	8.829	-	175.871	**129 Ala**	120.288	8.266	53.611	174.723
33 Asp	117.213	7.996	47.356	173.784	**130 Arg**	114.973	7.432	50.810	174.579
34 Leu	112.486	8.105	52.156	176.971	**131 Ala**	128.952	8.167	54.272	176.097
35 Gly	106.663	8.203	56.974	175.891	**132 Ile**	117.372	8.699	51.022	174.626
36 Val	120.416	8.432	43.817	173.520	**133 Val**	124.282	8.590	59.176	175.360
37 Ile	129.627	8.744	62.249	176.951	**134 Val**	126.321	9.300	60.825	174.473
38 Ser	120.422	8.464	60.265	174.675	**135 Ile**	131.332	9.835	60.896	174.990
39 Ser	119.622	8.230	55.009	175.687	**136 Gly**	111.048	8.744	60.693	175.938

40 Asp	121.711	8.051	61.483	177.104	**137 Ala**	123.515	8.671	44.215	175.482
41 Gln	117.490	7.614	56.289	178.028	**138 Asp**	114.409	7.940	54.289	177.768
42 Phe	117.056	7.182	56.644	176.592	**139 Gly**	108.348	8.359	52.948	177.215
43 Arg	120.079	6.905	55.743	174.321	**140 Asn**	118.264	8.059	44.821	173.867
44 Gly	114.171	8.977	57.606	178.333	**141 Val**	122.905	9.011	51.128	176.116
45 Lys	120.224	8.277	45.431	173.568	**142 Ala**	132.335	9.651	64.428	176.290
46 Ser	115.982	8.470	54.184	174.614	**143 Tyr**	117.846	7.560	52.187	177.122
47 Val	122.060	9.032	58.112	174.279	**144 Thr**	118.241	7.699	58.464	173.306
48 Leu	128.468	9.130	59.788	173.468	**145 Glu**	118.573	8.476	59.376	172.477
49 Leu	125.775	9.463	53.829	173.855	**146 Leu**	129.928	8.572	56.624	173.524
50 Asn	127.348	9.216	53.519	173.048	**147 Val**	128.319	8.194	53.260	176.114
51 Ile	125.307	8.716	50.825	174.071	**148 Pro**	116.279	-	-	-
52 Phe	124.719	9.324	62.622	-	**149 Glu**	121.005	6.984	64.201	176.140
64 Val	121.692	7.660	60.601	-	**150 Ile**	121.934	8.347	54.501	176.803
65 Arg	119.349	7.731	65.396	-	**151 Ala**	118.561	8.364	62.308	176.037
66 Thr	113.884	8.045	59.001	-	**152 Gln**	121.971	7.395	52.288	176.869
67 Phe	122.474	7.632	65.450	176.319	**153 Glu**	120.102	8.401	54.705	175.115
68 Asp	118.185	8.421	60.395	-	**154 Pro**	-	-	-	-
69 Glu	118.319	8.287	56.137	-	**155 Asn**	125.296	10.252	61.186	173.536
70 Arg	121.077	7.998	59.040	179.837	**156 Tyr**	122.417	7.454	51.448	176.229
71 Ala	122.340	8.337	58.634	179.581	**157 Glu**	116.560	8.383	59.921	177.851
72 Ala	121.869	8.235	54.670	180.247	**158 Ala**	120.703	7.513	59.270	179.474
73 Ala	121.651	8.003	53.949	171.821	**159 Ala**	114.783	7.148	53.964	180.045
74 Ser	112.180	7.356	53.975	179.105	**160 Leu**	121.931	8.044	54.179	178.803
75 Gly	108.770	7.813	59.152	174.424	**161 Ala**	120.063	7.964	57.073	180.403
76 Ala	122.953	7.269	44.544	174.214	**162 Ala**	116.064	7.231	54.218	180.941
77 Thr	117.597	8.004	51.076	177.068	**163 Leu**	107.534	7.310	53.523	178.642
78 Val	127.868	8.865	56.675	172.986	**164 Gly**	128.353	7.702	54.710	177.319
79 Leu	125.477	8.743	56.700	174.540	**165 Ala**	123.281	7.673	45.318	173.478

A.9: SigJ nucleotide and amino acid sequence.

Mtb SigJ: Met1-Asn312. Vector designation and size (base pairs): [pLEXT-01] and 6045 bp. Vector insert positioning (nucleotides): *Nde I* (5204) – *Hind III* (5978).

Shown below is the nucleotide and amino acid sequence (displayed in bold underneath the triplet codon) for sigJ as confirmed by MWG-Biotech™ sequencing services as described in Chapter 2.2.1.

```
atg gaa gtg agc gaa ttt gaa gcg ctg cgc cag cat ctg atg agc gtg
M   E   V   S   E   F   E   A   L   R   Q   H   L   M   S   V
gcg tat cgc ctg acc ggc acc gtg gcg gat gcg gaa gat att gtg cag
A   Y   R   L   T   G   T   V   A   D   A   E   D   I   V   Q
gaa gcg tgg ctg cgc tgg gat agc ccg gat acc gtg att gcg gat ccg
E   A   W   L   R   W   D   S   P   D   T   V   I   A   D   P
cgc gcg tgg ctg acc acc gtg gtg agc cgc ctg ggc ctg gat aaa ctg
R   A   W   L   T   T   V   V   S   R   L   G   L   D   K   L
cgc agc gcg gcg cat cgc cgc gaa acc tat acc ggc acc tgg ctg ccg
R   S   A   A   H   R   R   E   T   Y   T   G   T   W   L   P
gaa ccg gtg gtg acc ggc ctg gat gcg acc gat ccg ctg gcg gcg gtg
E   P   V   V   T   G   L   D   A   T   D   P   L   A   A   V
gtg gcg gcg gaa gat gcg cgc ttt gcg gcg atg gtg gtg ctg gaa cgc
V   A   A   E   D   A   R   F   A   A   M   V   V   L   E   R
ctg cgc ccg gat cag cgc gtg gcg ttt gtg ctg cat gat ggc ttt gcg
E   R   L   R   P   D   Q   R   V   A   F   V   L   H   D   G
gtg ccg ttt gcg gaa gtg gcg gaa gtg ctg ggc acc agc gaa gcg gcg
F   A   V   P   F   A   E   V   A   E   V   L   G   T   S   E
gcg cgc cag ctg gcg agc cgc gcg cgc aaa gcg gtg acc gcg cag ccg
A   A   A   R   Q   L   A   S   R   A   R   K   A   V   T   A
gcg ctg att agc ggc gat ccg gat ccg gcg cat aac gaa gtg gtg ggc
Q   P   A   L   I   S   G   D   P   D   P   A   H   N   E   V
cgc ctg atg gcg gcg atg gcg gcg ggc gat ctg gat acc gtg gtg agc
V   G   R   L   M   A   A   M   A   A   G   D   L   D   T   V
ctg ctg cat ccg gat gtg acc ttt acc ggc gat agc aac ggc aaa gcg
V   S   L   L   H   P   D   V   T   F   T   G   D   S   N   G
ccg acc gcg gtg cgc gcg gtg cgc ggc agc gat aaa gtg gtg cgc ttt
K   A   P   T   A   V   R   A   V   R   G   S   D   K   V   V
att ctg ggc ctg gtg cag cgc tat ggc ccg ggc ctg ttt ggc gcg aac
R   F   I   L   G   L   V   Q   R   Y   G   P   G   L   F   G
```

cag ctg gcg ctg gtg aac ggc gaa ctg ggc gcg tat acc gcg ggc ctg

A N Q L A L V N G E L G A Y T A

ccg ggc gtg gat ggc tat cgc gcg atg gcg ccg cgc att acc gcg att

G L P G V D G Y R A M A P R I T

acc gtg cgc gat ggc aaa gtg tgc gcg ctg tgg gat att gcg aac ccg

T V R D G K V C A L W D I A N P

gat aaa ttt acc ggc agc ccg ctg aaa gaa cgc cgc gcg cag ccg acc

D K F T G S P L K E R R A Q P T

ggc cgc ggc cgc cat cat cgc aac

G R G R H H R N

References:

Abragam, A. 1961. 'The Principles of Nuclear Magnetism.' Oxford University Press, Oxford.

Ahearn, J. M., Bartolomei, M. S., West, M. L., Cisek, L. J., and Corden, J. L. 1987. 'Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II'. *Journal of Biology*, Vol: 262, 10695–10705.

Alphey. M. S. Bond, C. S. Tetaud, E. Fairlamb, A. H. and Hunter, W. N. 2000. 'The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2-Cys peroxirexins.' *Journal of Molecular Biology*, Vol: 300, 903–916.

Altschul, S. F. Madden, T. L. Schäffer, A. A. Zhang, J. Zhang, Z. Miller, W. and Lipman D. J. 1997. 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.' *Nucleic Acids Research.* Vol: 25, 3389-3402.

van Amsterdam, I. M. C. Ubbink, M. Jeuken, L. J. C. *et al.* 2001. 'Effects of dimerization on protein electron transfer.' *Chemistry – A European journal*, Vol: 7, 2398-2406.

Baker, L. M. and Poole, L. B. 2003. 'Catalytic mechanism of thiol peroxidase from Escherichia coli. Sulfenic acid formation and over-oxidation of essential CYS61.' *Journal of Biological Chemistry*, Vol: 278, 9203–9211.

Barne, K.A., Bown, I.A. Busby, S. J. W. and Minchin, S. D. 1997. 'Region 2.5 of the *Escherichia coli* RNA polymerase σ70 subunit is responsible for the recognition of the 'extended -10' motif at promoters'. *EMBO J.* Vol: 16, 4034–4040.

Barry, C. E. Crane, D. D. and Yuan, Y. 1996. 'Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: Function of the mycobacterial α-crystallin homolog'. *Journal of Bacteriology,* Vol: 178, 4484–4492.

Benison, G. Nyarko, A. and Barbar, E. 2006. 'Heteronuclear NMR identifies a nascent helix in intrinsically disordered dynein intermediate chain: implications for folding and dimerization.' *Journal of Molecular Biology*, Vol: 362, 1082-1093.

Beer, M. H. Fletcher, A. J. Jones, T. V. *et al.* 2003. 'The Merck Manual of Medical Information — Second Home Edition.' Merck Publishing, Pages: 1125–1130.

Berjanskii, M. V. Neal, S. Wishart, D. S. 2006. 'PREDITOR: a web server for predicting protein torsion angle restraints.' *Nucleic Acids Research*, Published online. Web Server Issue. W63-W69. DOI: 10.1093/nar/gkl341.

BioEdit v4.0.6 Biological Sequence Alignment Editor for Windows. 1997. Accessed via: http://www.mbio.ncsu.edu/BioEdit/bioedit.html

Bishai, W. R. Dick, J. D. Parish, N. M. 1998. 'Mechanisms of latency in *Mycobacterium tuberculos*is'. *Trends in Microbiology*, Vol: 6, 107–112.

Bishai, W. R. and Manabe, Y. C. 2000. 'Latent *Mycobacterium tuberculosis* – persistence patience, and winning by waiting.' *Nature Medicine*, Vol: 6, 1327–1329.

Borukhov, S., Laptenko, O., Lee, Jookjung., Sekine, S., Vassylyev, D. G., and Yokoyama, S. 2002. 'Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6Å resolution'. *Nature*, Vol: 417, 712–719.

Boucher, W. 2002. AZARA version 2.7.

Brockhurst, K. and Little, G. 1973. 'Reactions of papain and of low-molecular-weight thiols with some aromatic disulphides. 2,2′-Dipyridyl disulphide as a convenient active-site titrant for papain even in the presence of other thiols.' *Biochemistry*, Vol: 133, 67–80.

Bryk, R. Griffin, P. and Nathan, C. 2000. 'Peroxynitrite reductase activity of bacterial peroxiredoxins.' *Nature*, Vol: 278, 10816–10823.

Bryk, R. Lima, C. D. Erdjument-Bromage, H. *et al.* 2002. 'Metabolic enzymes of mycobacteria linked to antioxidant defence by a thioredoxin-like protein.' *Science.* Vol: 295, 1073–1077.

Burton, N., Buttner, M. J., Jakimowicz, P. Kleanthous., Lawson, D.M., C. Li, W., Paget, M.S.B., and Stevenson, C.E.M. 2002. 'Identifying the structure of the anti-sigma factor binding domain of the disulphide-stress regulated sigma factor σ^R from *Streptomyces coelicolor'. Journal of Molecular Biology,* Vol: 323, 225–236.

Callaci, S., Heyduk, E., and Heyduk, T. 1999. 'Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of the sigma 70 subunit.' *Molecular Cell*, Vol: 3, 229–238.

Campbell, E. A., Muzzin, O., Chelnov, M., Sun, J.L., Olsen, A., Weinman, O., Trester-Zedlitz, M., and Darst, S.A. 2002. 'Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit'. *Molecular Cell*, Vol: 3, 527– 39.

Campbell, E.A. Tupy, J.L. Gruber, T.M. Wang, S. Sharp, M.M. Gross, C.A. Darst, S.A. 2003. 'Crystal structure of *Escherichia coli* σE with the Cytoplasmic Domain of its Anti-σ RseA'. *Molecular Cell*, Vol: 11, 1067-1078.

Cavanagh, J. Fairbrother, W. J. Palmer, A. G. I. and Skelton, N. J. 1996. 'Protein NMR spectroscopy: Principles and practice.' Published by Academic Press.

Cha, M. K. Kim, H. K. Kim, I. H. 1996. 'Mutation and Mutagenesis of thiol peroxidase of *Escherichia coli* and a new type of thiol peroxidase family.' *Journal of Bacteriology*. Vol: 178, 5610–5624.

Chae, H. Z. Yim, M. B. Rhee, S. G. Chock, P. B. and Stadtman, E. R. 1994. 'On the protective mechanism of the thiol-specific antioxidant enzyme against the oxidative damage of biomacromolecules.' *Journal of Biological Chemistry*, Vol: 269, 1621–1626.

Chan, W. T. McKinney. J. D. Swenson, D. *et al.* 2000. 'Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase.' *Nature*, Vol: 406, 735–738.

Chen, H. A. Pfhul, M. McAlister, M. S. B. and Driscoll, P. C. 2000. 'Determination of p*K*a Values of Carboxyl Groups in the N-Terminal Domain of Rat CD2: Anomalous p*K*a of a Glutamate on the Ligand-Binding Surface.' *Biochemistry*, Vol: 39, 6814-6828.

Chi, Y. Kumar, T. K. Wang, H. M. Ho, M. C. Chiu, I. M. and Yu, C. 2001. 'Thermodynamic characterization of the human acidic fibroblast growth factor: evidence for cold denaturation.' *Biochemistry*, Vol: 40, 7746–7756.

Choi, H. J. Kang, S. W. Yang, C. H. Rhee, S. G. and Ryu, S. E. 1998. 'Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution.' *Nature Structural Biology,* Vol: 5, 400–406.

Choi, J. Choi, S. Cha, M. K. Kim, I. H. and Shin, W. 2003. 'Crystal structure of *E. coli* thiol peroxidase in the oxidised state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins.' *Journal of Biochemical Chemistry*, Vol: 278, 49478–49486.

Coates, A. R. M. and Y, Hu. 2001a. 'Transcription of two sigma 70 homologue genes, *sigA* and *sigB*, in stationary-phase *Mycobacterium tuberculosis'*. *Journal of Bacteriology,* Vol: 181, 469–476.

Coates, A. R. M. and Hu, Y. 2001b. 'Increased levels of sigJ mRNA in late stationary phase cultures of *Mycobacterium tuberculosis* detected by DNA array hybridization'. *FEMS Microbiology Letters,* Vol: 202, 59–65.

Cole. S. T. Brosch, R. Parkhill, J. Garnier, *et al.* 1998. 'Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence'. *Nature,* Vol: 6685, 537–544.

Colland, F. Orsini, G. Brody, E.N. Buc, H. Kolb, A. 1998. 'The bacteriophage T4 AsiA protein: a molecular switch for sigma 70-dependent promoters'. *Molecular Microbiology*, Vol: 27, 819-829.

Craig, M.L. Tsodikov, O.V. McQuade, K.L. Schlax, P.E.J. Capp, M.W. Saecker, R.M. Record, M.T.J. 1998. 'DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA.' *Journal of Molecular Biology*, Vol: 283, 741-756.

Creighton T, E. 2000. 'Protein folding coupled to disulphide-bond formation.' In *Mechanisms of Protein Folding* 2nd ed. Editor RH Pain. Oxford University Press.

Cummings, M. R. and Klug, W. S. 1994. *Concepts of Genetics, 5th Ed.* Prentice Hall, New York.

Darst, S. E., Marlhotra, A., and Severinova, E. 1996. 'Crystal structure of a σ^{70} subunit fragment from *E.coli* PNA polymerase'. *Cell*, Vol: 1, 127–136.

Declercq, J. P. Evrard, C. Clippe, A. Stricht, D. V. Bernard, A. and Knoops, B. 2001. 'Crystal structure of human peroxiredoxin 5, a novel type of mammalian

peroxiredoxin at 1.5 Å resolution.' *Journal of Molecular Biology*, Vol: 311, 751 – 759.

Delaglio, F. Grzesiek, S. Vuister, G. W. Zhu, G. Pfeifer, J. and Bax, A. 1995. 'NMRPipe: a multidimensional spectral processing system based on UNIX pipes.' *Journal of Biomolecular NMR*, Vol: 6, 277–293.

DeLano, W. L. 2006. 'The PyMOL Molecular Graphics System.' DeLano Scientific, San Carlos, CA, USA. Accessed via: http://www.pymol.org

DeMaio, J. Zhang, Y. Ko, C. Young, D. B. and Bishai, W. R. 1996. 'A stationary-phase stress-response sigma factor from Mycobacterium tuberculosis'. *Proc. Natl. Acad. Sci. USA.* Vol: 93, 2790–2794.

DeWit, D. M. Wootton, J. Dhillon, and A. Mitchison. 1995. 'The bacterial DNA content of mouse organs in the Cornell model of dormant tuberculosis.' Tuberculosis and. Lung Disease, Vol: 76, 555-562.

Dietz, K. J. Horling, F. Konig, J. Baier, M. 2002. 'The function of the chloroplast 2-cysteine peroxiredoxin in peroxide detoxification and its regulation.' *Journal of Experimental Botany*, Vol: 53, 1321–1329.

Dombroski, A. J., Walter, W. A. and Gross, C. A. 1993. 'Amino-terminal amino acids modulate σ-factor DNA-binding activity'. *Genes and Development*, Vol: 7, 2446–2455.

Dove, S. L., Darst, S., and Hochschilds, A. 2003. 'Region 4 of σ as a target for transcription regulation.' *Molecular Microbiology*, Vol: 48, 863–874.

Ebright, R. H. 2000. 'RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II.' *Journal of Molecular Biology,* Vol: 304, 687–698.

Egwim, G. L. and Gruber, H. J. 2001. 'Spectrophotometric Measurement of Mercaptans with 4,4'-Dithiodipyridine.' *Analytical Biochemistry*, Vol: 288, 188–194.

Electronic Source 1. BUPA's Health Information Team. 2003. 'BUPA fact sheet: TUBERCULOSIS'. http://hcd2.bupa.co.uk/fact_sheets/html/Tuberculosis.html Accessed: June 2004.

Electronic Source 2. *DHS (Department of Human Services).* 'Management, Control and Prevention of Tuberculosis – Guidelines for Health Care Providers (2002- 2005).' 2002. Published by: Rural and Regional Health and Aged Care Services Division. Australia.
http://www.health.vic.gov.au/ideas/diseases/tb_mgmt_guide.htm Accessed: April 2004.

Electronic source 3. W.H.O Media Center Fact Sheet. 'Tuberculosis'. 2002. http://www.who.int/mediacentre/factsheets/who104/en/print.html Accessed: June 2004.

Eletsky, A. Kienhofer, A. and Pervushin, K. 2001. 'TROSY NMR with partially deuterated proteins'. *Journal of Biomolecular NMR.* Vol: 20, 177-180.

Evrard, C. Capron, A. Marchand, C. Clippe, A. Wattiez, R. Soumillion, P. *et al.* 2004. 'Crystal structure of a dimeric oxidised form of human peroxiredoxin 5.' *Journal of Molecular Biology*, Vol: 337, 1079–1090.

Fenton, M. S. and Vermeulen, M. W. 1996. 'Immunopathology of Tuberculosis: Role of Macrophages and Monocytes.' *Infection and Immunity*, Vol: 64, 683–690.

Fenton, M. S, Lee, S. J. and Gralla, J. D. 2000. '*Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ^{70}'. *EMBO J*, Vol: 19, 1130–1137.

Fernandez, C and Wider, G. 2003. 'TROSY in NMR studies of the structure and function of large biological macromolecules'. *Current Opinion in Structural Biology*, Vol: 13, 570-580.

Fiaux, J. Bertelsen, E. B. Horwich, A. L. Wuthrich, K. Nature. 2002. 'NMR analysis of a 900K GroEL-GroES complex.' *Nature*, Vol: 11, 207-211.

Flohe, L. Brigelius-Flohe, R. and Hatfield, D. L. 2001. 'Selenium. Its molecular biology and role in human health'. Kluwer Academic Publishers, London. Pages; 157–178.

Garrett, D. S. Seok, Y-J. Liao, D-I. Peterkofsky, A. Gronenborn, A. M. and Clore, G. M. 1997. 'Solution structure of the 30 kDa N-terminal domain of Enzyme I of the *Escherichia coli* phosphoenolpyruvate system by multidimensional NMR.' *Biochemistry*, Vol: 36, 2517–2530.

Gill, S. C. and von Hippel, P. H. 1989. 'Calculation of protein extinction coefficients from amino acid sequence data.' *Analytical Biochemistry*, Vol: 182, 319–326.

Grasberger, B. Gronenborn, A. M. and Clore, G. M. 1992. 'Analysis of the backbone dynamics of interleukin-8 by ^{15}N relaxation measurements.' *Journal of molecular biology*, Vol: 230, 364-372.

Grimshaw, C. E. Whistler, R. L. Cleland, W. W. 1979. 'Ring opening and closing rates for thiosugars.' *Journal of the American Chemistry Society*, Vol: 101, 1521–1532.

Gronenborn, A. M. and Clore, G. M. 1995 'Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy.' *Critical Review of Biochemistry and Molecular Biology*, Vol: 30, 351–385.

Gross, C., Lonetto, M. and Losick, R. 1992. 'Transciptional Regulation'. (eds. McKnight, S. R. and Yamamoto, K.R. Cold Harbor Laboratory Press. Pages; 129–176.

Goto, N. K. and Kay, L. E. 2000. 'New developments in isotope labelling strategies for protein solution NMR spectroscopy.' *Current Opinion on Structural Biology*, Vol: 10, 585–592.

Grzesiek, S. and Bax, A. 1993. 'Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins.' *Journal of Biomolecular NMR*, Vol: 3, 185–204.

Grzesiek, S. Anglister, J. Ren, H. and Bax, A. 1993. 'C-13 Line Narrowing By H-2 Decoupling in H-2/C-13/N-15-Enriched Proteins - Application to Triple-Resonance 4d J-Connectivity of Sequential Amides.' *Journal of the American Chemical Society*, Vol: 115, 4369–4370.

Guimaraes, B. G. Souchon, H. Honore, N. Saint-Joanis, B. Brosch, R. Shepard, W. Cole, S. T. Alzari, P. M. 2005. 'Structure and Mechanism of the Alkyl Hydroperoxidase Ahpc, a Key Element of the *Mycobacterium Tuberculosis* Defence System Against Oxidative Stress.' *Journal of Biological Chemistry*, Vol: 280, 25735.

Handwerger, S. and Tomasz, A. 1985. 'Antibiotic tolerance among clinical isolates of bacteria'. *Annual Review of Pharmacology and Toxicology*, Vol: 25, 349–380.

Harris, W. 2001. 'WHO Facilitator Guide: The natural history of pulmonary tuberculosis. Social mobilization and training department of control prevention and eradication communicable diseases cluster'.
http://www.who.int/tb/publications/who_cds_tb_2003_314/en/ Accessed: July 2004.

Helmann, J. D., and Chamberlin, M. J. 1988. 'Structure and function of bacterial sigma factors.' *Annual Review of Biochemistry*, Vol: 57, 839–872.

Helmann J. D., and Juang, Y. 1994. 'A promoter melting region in the primary σ factor of *Bacillus subtilis*.' *Journal of Molecular Biology*, Vol: 235, 1470-1488.

Helmann, J. D. 2002. 'The Extracytoplasmic Function (ECF) Sigma Factors.' *Advances in Microbial Physiology*, Vol: 6, 47–110.

Higgins, D. Thompson, J. Gibson, T. Thompson, J. D. Higgins, D.G. Gibson, T. J. 1994. 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, Vol: 22, 4673–4680.

Hirotsu, S. Abe, Y. Okada, K. Nagahara, N. Hori, H. Nishino, T. and Hakoshima, T. 1999. 'Crystal structure of a multifunctional 2-Cys peroxiredoxin heme-binding protein 23 kDa/proliferation-associated gene product.' *National Academy of Science USA*, Vol: 96, 12333–12338.

Hofmann, B. Hecht, H. J. Flohe, L. 2002. 'Peroxiredoxins.' *Biological Chemistry*, Vol: 383(3-4), 347–364.

Holm, L. and Sander, C. 1995. 'Dali: a network tool for protein structure comparison.' *Trends in Biochemical Science*, Vol: 20(11), 478–480.

Hu. Y. and Coates, A. R. M . 1999. 'Transcription of two sigma 70 homologue genes, sigA and sigB, in stationary-phase *Mycobacterium tuberculosis*.' *Journal of Bacteriology*, Vol: 181, 469–476.

Hu, K. Eletsky, A. and Pervushin, K. 2003. 'Backbone resonance assignment in large protonated proteins using a combination of new 3D TROSY-HN(CA)HA, 4D TROSY-HACANH and ^{13}C-detected HACACO experiments.' *Journal of Biomolecular NMR*, Vol: 26, 69–77.

Invitrogen™ NuPage Technical Guide, supplied with pre-cast gels or PDF file available via:
http://catalog.invitrogen.com/index.cfm?fuseaction=viewCatalog.viewProductDetails&productDescription=5186

Jaeger, T. Budde, H. Flohe, L. Menge, U. Singh, M. Trujillo, M. and Radi, R. 2003. 'Multiple thioredoxin-mediated routes to detoxify hydroperoxides in *M. tuberculosis*.' *Achieves of Biochemistry and Biophysics*, Vol: 423, 182–191.

Jain, D. Nickels, B.E. Sun, L. Hochschild, A. Darst, S.E. 2004. 'Structure of a ternary transcription activation complex'. *Molecular Cell*, Vol: 13, 45-53.

Jones, S. and Thornton, J. M. 1996. 'Principles of protein-protein interaction.' *Proceedings of the National Academy of Science USA*, Vol: 93, 13-20.

Juang, Y. L., and Helmann, J. D. 1994. 'A promoter melting region in the primary sigma factor of *Bacillus subtilis*: identification of functionally important aromatic amino acids.' *Journal of Molecular Biology*, Vol: 235, 1470–1488.

Kay, L. E. Keifer, P. Saarinen, T. 1992. 'Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity.' *Journal of the American Chemical Society*, Vol: 114, 10663–10665.

Kalodimos, C. 2004. 'Circular dichroism spectroscopy of biomolecules'. http://www.newark.rutgers.edu/chemistry/grad/chem585/lecture1.html Accessed: July 2004.

Kirley, T. L. 1989. 'Determination of three disulfide bonds and one free sulfhydryl in the beta subunit of (Na,K)-ATPase.' *Journal of Biological Chemistry*. Vol: 264, 7185–7192.

Kitano, K. Kita, A. Hakoshima, T. Niimura, Y. and Miki, K. 2005. 'Crystal structure of decameric peroxiredoxin (AhpC) from Amphibacillus xylans.' *Proteins: Structure, Function and Genetics*, Vol: 59, 644–647.

Kim, K, Rhee, S. G. Stadtman, E. R. 1985. 'Nonenzymatic cleavage of proteins by reactive oxygen species generated by dithiothreitol and iron.' *Journal of Biological Chemistry*, Vol: 260, 15394-15397.

Kraulis, P. 1991. 'ANSIG: A program for the assignment of protein ^{1}H 2D NMR spectra by interactive graphics.' *Journal of Magnetic Resonance*, Vol: 24, 627–633.

Lee, S. P. Hwang, Y. S. Kim, Y. J. Kwon, K. S. Kim, H. J. Kim, K. Chae, H. Z. 2001. 'Cyclophilin a binds to peroxiredoxins and activates its peroxidase activity.' *Journal of Biological Chemistry*, Vol: 276, 29826–29832.

Levitt, M. H. 2001. 'Spin dynamics: Basics of Nuclear Magnetic Resonance'. Publisher: John Wiley and Sons, Pages; 513-517.

Li, S. Peterson, N. A. Kim, M. Y. Kim, C. Y. Hung, L. W. Yu, M. *et al.* 2005. 'Crystal structure of AhpE from M. tuberculosis, a 1-Cys peroxiredoxin.' *Journal of Molecular Biology*, Vol: 346, 1035–1046.

Magalhães, B. S., Harris, R., Plevin, M. J. and Driscoll, P. C. 2004 'Letter to the Editor: Backbone ^{1}H, ^{13}C, and ^{15}N Resonance Assignments for a 29 kDa Monomeric Variant of *Pseudomonas Aeruginosa* Dimethylarginine Dimethylaminohydrolase.' *Journal of Biomolecular NMR*, Vol:. 29, 463–464.

Malhotra, A. Severinova, E. Darst, S.A. (1996). 'Crystal structure of a σ^{70} subunit fragment from *E.coli* RNA polymerase'. *Cell*, Vol: 87, 127-136.

Markus, M. A. Dayie, K. T. Matsudairat, P. and Wagner, G. 1994. 'Effect of deuteration on the amide proton relaxation rate in proteins. Heteronuclear NMR experiments on villin 14T'. *Journal of Magnetic Resonance*, Vol: 105, 192-195.

Marr, M. T. and Roberts, J. W. 1997. 'Promoter Recognition As Measured by Binding of Polymerase to Nontemplate Strand Oligonucleotide.' *Science*, Vol: 276, 1258–1260.

Master, S. S. Springer, B. Sander, P. Boettger, E. C. Deretic, V. Timmins, G. S. 2002. 'Oxidative stress response genes in *Mycobacterium tuberculosis*: role of ahpC in resistance to peroxynitrite and stage-specific survival in macrophages.' *Journal of Microbiology*, Vol: 148, 3139–3144.

Mohan, K. P. M. Barve, M. Chatterjee, A. and Hosur, R. 2006. 'pH driven conformational dynamics and dimer-to-monomer transition in DLC8.' *Protein Science*, Vol: 15, 335-342.

Mrksich, M and Whitesides, G.M. 1996. 'Using self-assembled monolayers to understand the interactions of man-made surfaces with proteins and cells'. *Annual Review of Biophysics and Biomolecular Structure*, Vol: 25, 55–78.

Murakami, K., and Darst, S. A. 2003. 'Bacterial RNA polymerases: the wHole story'. *Current Opinion in Structural Biology*, Vol: 13, 31-39.

Nardell, E. 'Pathogenesis of tuberculosis.' 1993. In Reichman, L.B. and Hirschfield, E. (ed), *Lung biology in health and disease*. Marcel Dekker, Inc. New York.

Niimura, Y. Poole, L. B. Massey, V. 1995. '*Amphibacillus xylanus* NADH oxidase and *Salmonella typhimurium* alkyl-hydroperoxide reductase flavoprotein components show extremely high scavenging activity for both alkyl hydroperoxide and hydrogen peroxide in the presence of *S. typhimurium* alkyl-

hydroperoxide reductase 22-kDa protein component.' *Journal of Biological Chemistry*, Vol: 270, 25645–25650.

Nishiyama, J. and Kuninori, T. 1992. 'Assay of thiols and disulfides based on the reversibility of *N*-ethylmaleimide alkylation of thiols combined with electrolysis.' *Analytical Biochemistry*, Vol: 200, 230-234

Nogoceke E, Gommel D. U, Kiess M, Kalisz H. M, and Flohe L.1997. 'A unique cascade of oxidoreductases catalyses trypanothione-mediated peroxide metabolism in *Crithidia fasciculata.*' *Journal of Biological Chemistry*, Vol: 378(8), 827-836.

Orme, I. M. 'Immunity to Mycobacteria.' 1993. *Current Opinion in Immunology,* Vol: 5, 497–502.

Packer, L. 1995. Methods in enzymology, Biothiols, Parts A and B, Academic Press, New York. Vol: 251 and 252.

Pervushin, K. Riek, R. Wider, G. Wuthrich, K. 1997. 'Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.' *Proceedings of the National Academy of Science USA*, Vol: 94, 12366-12371.

Plevin, M. J. Magalhaes, B. S. Harris, R. *et al.* 2004. 'Characterisation and manipulation of the *Pseudomonas aeruginosa* dimethylarginine dimethylaminohydrolase monomer-dimer equilibrium.' *Journal of Molecular Biology*, Vol: 341, 171-184.

Poole, L.B. Reynolds, C. M. Wood, Z. A. *et al.* 2000. 'AhpF and other NADH:peroxiredoxin oxidoreductases, homologues of low Mr thioredoxin reductase.' *European Journal of Biochemistry*, Vol: 267, 6126–6133.

Poole, L. B. 2005. 'Bacterial defences against oxidants: mechanistic features of cysteine-based peroxidases and their flavoprotein reductases.' *Archives of Biochemistry and Biophysics*, Vol: 433, 240–254.

Reiner, C. K. Kada, G. Gruber, H. J. 2002. 'Quick measurement of protein sulfhydryls with Ellman's reagent and with 4, 4'-dithiodipyridine.' *Bioanalytical Chemistry*, Vol: 373, 266–276.

Rho, B-S. Hung, L-W. Holton, J. M. Vigil, D. Kim, S. Park, M. S. Terwilliger, T. C. Pedelacq, J. D. 2006. 'Functional and structural characterisation of a thiol peroxidase from *M. tuberculosis*.' *Journal of Molecular Biology*, Vol: 361, 850–863.

Riek, R. Pervushin, K. and Wüthrich, K. 2000. 'TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution.' *Trends in Biochemical Science*. Vol: 25, 462–468.

Rouhier, N. Gelhaye, E. Jacquot, J. P. 2002. 'Glutaredoxin-dependent peroxiredoxin from poplar: protein-protein interaction and catalytic mechanism.' *Journal of Biological Chemistry*, Vol: 277, 13609–13614.

Rhee, S. G. Kang, S. W. Chang, T. S. Jeong, W. Kim, K. 2001. 'Peroxiredoxin, a novel family of peroxidases.' *IUBMB Life*, Vol: 52, 35–41.

Riener, C. Kada, G. and Gruber, H. J. 2002. 'Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine.' *Analytical Bioanalytical Chemistry*, Vol: 373, 266–276.

Rosenkrands, I. King, A. Weldingh, K. Moniatte, M. Moertz, E. and Andersen, P. 2000a. 'Towards the proteome of *M. tuberculosis*.' *Journal of Electrophoresis*, Vol: 21, 3740–3756.

Rosenkrands, I. Weldingh, K. Jacobsen, S. Hansen, C. V. Florio, W. Gianetri, I. and Andersen, P. 2000b. 'Mapping and identification of *M. tuberculosis* proteins

by two-dimensional gel electrophoresis, microsequencing and immunodetection.' *Journal of Electrophoresis*, Vol: 21, 935–948.

Salzmann, M. Pervushin, K. Wider, G. *et al.* 2000. 'NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments.' *Journal of American Chemical Society*, Vol: 122, 7543-7548.

Sarma, G. N. Nickel, C. Rahlfs, S. Fischer, M. Becker, K. and Karplus, P. A. 2005. 'Crystal structure of a novel *Plasmodium falciparum* 1-Cys peroxiredoxin.' *Journal of Molecular Biology*, Vol: 346, 1021–1034.

Sattler, M. and Fesik, S. W. 1996. 'Use of deuterium labelling in NMR: overcoming a sizeable problem.' *Structure*, Vol: 4, 1245–1249.

Schroder, E. Littlechild, J. A. Lebedev, A. A. Errington, N. Vagin, A. A. and Isupov, M. N. 2000. 'Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution.' *Structure with folding and design*, Vol: 8, 605–615.

Sherman, D. R. Mdluli, K. Hickey, M. J. *et al.* 1996. 'Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis.*' *Science*, Vol: 272, 1641–1643.

Sorenson, M. K. Ray, S. S. Darst, S. A. 2004. 'Crystal structure of the flagellar σ/Anti-σ complex σ^{28}/FlgM reveals an intact σ factor in an inactive conformation'. *Molecular Cell*, Vol: 14, 127-138.

Spera, S. and Bax, A. 1991. 'Empirical correlation between protein back-bone conformation and C_α and C_β ^{13}C nuclear magnetic resonance chemical shifts.' *Journal of American Chemical Society*, Vol: 113, 5490-5492.

Stehr, M. Hecht, H. J. Jager, T. Flohe, L. and Singh, M. 2006. 'Structure of the inactive variant C60S of *M. tuberculosis* thiol peroxidase.' *Biological Crystallography*, Vol: 62, 563–567.

Storz, G. Christman, M. F. Sies, H. Ames, B. N. 1989. 'Spontaneous mutagenesis and oxidative damage to DNA in *Salmonella typhimurium*.' *Proceedings of the National Academy of Sciences of the United States of America*, Vol: 84 (24), 8917–8921.

Syzperski, T. Mills, J. L. Perl, D, Baldach, J. 2006. 'Combined NMR-observation of cold denaturation in supercooled water and heat denaturation enables accurate measurement of DeltaC (p) of protein unfolding.' *European Biophysics Journal*, Vol: 35, 363–393.

Thacker, S. B. 2003. 'Treatment of Tuberculosis.' *CDC – MMWR*. Vol: 52: N^{o}: RR-11.

Thomas, G. 2007. 'Tuberculosis.' W.H.O. fact sheet, N^{o}: 104.

Trujillo, M. Mauri, P. Benazzi, L. Comini, M. *et al.* 2006. 'The mycobacterial thioredoxin peroxidase can act as a one-cysteine peroxiredoxin.' *Journal of Biological Chemistry*, Vol: 281, 20555–20566.

Tugarinov, V. Choy, W. Y. Orekhov, V. Y. and Kay, L. E. 2005. 'Solution NMR-derived global fold of a monomeric 82-kDa enzyme.' *Proceedings of the National Academy of Science, USA*. Vol: 102, 622–627.

Vassylyev, D.G. Sekine, S. Laptenko, O. Lee, J. Vassylyeva, M.N. Borukhov, S. Yokoyama, S. (2002). 'Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6A resolution'. *Nature*, Vol: 417, 712-719.

Venters, R. A. Farmer, B. T. 2nd, Fierke, C. A. and Spicer, L. D. 1996. 'Characterising the use of perdeuteration in NMR studies of large proteins: ^{13}C,

^{15}N and ^{1}H assignments of human carbonic Anhydrase II.' *Journal of Molecular Biology*, Vol: 264, 1101-1116.

von Hippel, P. 1998. 'An integrated model of the transcription complex in elongation, termination and editing'. *Science*, Vol: 281, 660–665.

Wan, X. Y, Zhou, Y, Yan, Z. Y, Wang, H. L, Hou, Y. D, and Jin, D. Y. 1997. 'Scavengase p20: a novel family of bacterial antioxidant enzymes.' *FEBS Lett,* Vol: 407, 32–36.

Wayne, L. G. and Sramek, H. A. 1994. 'Metronidazole is Bactericidal to Dormant Cells of *Mycobacterium tuberculosis*.' *Antimicrobial Agents and Chemotherapy*. Vol: 38, 2054–2058.

Weischel, A. Gasdaska, J. R. Powis, G. Montfort, W. R. 1996. 'Crystal structures of reduced, oxidised, and mutated thioredoxins: evidence for a regulatory homodimer.' *Structure*, Vol: 4, 735–751.

Weldingh, K. Rosenkrands, I. Jacobsen, S. Rasmussen, P. B. Elhay, M. J. and Andersen, P. 1998. 'Two-dimensional electrophoresis for analysis of M. tuberculosis culture filtrate and purification and characterisation of six novel proteins.' *Infection and Immunology*, Vol: 66, 3492–3500.

Whitmore, L. and Wallace, B. A. 2004. 'DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.' *Nucleic Acids Research*, Vol: 32, web server issue: W668-73.

Wishart D. S. and Sykes, B. D. 1994. 'Chemical shifts as a tool for structure determination.' Methods in Enzymology, Vol: 239, 363-392.

Wishart, D. S. Bigam, C. G. Yao, J. *et al.* 1995a. '^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR.' *Journal of Biomolecular NMR*, Vol: 6, 135-140.

Wishart, D. S. Bigam, C. G. Holm, A. *et al.* 1995b. '^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbour effects.' *Journal of Biomolecular NMR*, Vol: 5, 67-81.

Wood, Z. A. Poole, L. B. Hantgan, R. R. and Karplus, P. A. 2002. 'Dimers to doughnuts: redox-sensitive oligomerisation of 2-Cys peroxiredoxins.' *Biochemistry*, Vol: 41, 5493–5504.

Wood, Z. A. Schroder, E. Robin, H. J. and Poole, L. B. 2003a. 'Structure, mechanism and regulation of peroxiredoxins.' *Trends in Biochemical Science*, Vol: 28, 32–40.

Wood, Z. A. Poole, L. B. and Karplus, P. A.2003b. 'Peroxiredoxin evolution and the regulation of hydrogen peroxide signalling.' *Science*, Vol: 300, 650–653.

Yamazaki, T. Lee, W. Arrowsmith, C. H. Muhandiram, D. R. and Kay, L. E. 1994A. 'A suite of triple resonance NMR experiments for the back-bone assignment of ^{15}N, ^{13}C, ^{2}H labelled proteins with high sensitivity.' *Journal of the American Chemical Society*, Vol: 116, 11655–11666.

Yamazaki, T. Lee, W. Revington, M. Mattellio, D. L. Dahlquist, F. W. Arrowsmith, C. H. and Kay, L. E. 1994B. 'An HNCA pulse scheme for the back-bone assignment of ^{15}N, ^{13}C, ^{2}H labelled proteins: application to a 37-kDa *Trp* repressor-DNA complex.' *Journal of the American Chemical Society*, Vol: 116, 6464–6465.

Zhang, O. Kay, L. E. Olivier, J. P, Forman-Kay, J. D. 1994. 'Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques.' *Journal of Biomolecular NMR*, Vol: 4, 845-858

Zhang, Y. Heym, B. Allen, B. Young, D. Cole, S. 1992. 'The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*.' *Nature*, Vol: 358, 591–593.

Zhou, Y. Wan, X. Y. Wang, H. L. Yan, Z. Y. Hou, Y.D. Jin, D. Y. 1997. 'Bacterial scavengase p20 is structurally and functionally related to peroxiredoxins.' *Biochemical and Biophysical Research Communications*, Vol: 233, 848–852.