# Biophysical analysis of MarR family transcription factors from *Streptomyces coelicolor*

## Tracey A. Holley

Thesis submitted to the
University of East Anglia
for the degree of Doctor of Philosophy

Department of Biological Chemistry
John Innes Centre
Norwich, UK

**September 2012**

**Date**

I certify that the work contained in the thesis submitted by me for the degree of PhD is my own original work, except where due reference is made to the author, and has not been submitted by me for a degree at this or any other university.

**Signed**

**Tracey A Holley**

## Abstract

Streptomycetes represent a rich source of potentially useful therapeutics. However, in a laboratory environment, a large proportion of their secondary metabolite gene clusters remain silent, and thus their full metabolic potential is not realised. One way to unlock these so-called "cryptic" clusters may be to manipulate specific regulatory proteins. But, the vast majority of these regulators are uncharacterized, making it difficult to do this in a rational way. The MarR family of regulatory proteins (MFRs) represents an important group of transcriptional regulators. Named after the first to be characterized in *Escherichia coli*, the <u>m</u>ultiple <u>a</u>ntibiotic <u>r</u>esistance <u>r</u>epressor (MarR), the members of this family are involved in a number of biological processes which include antibiotic resistance, virulence, oxidative stress responses and pathogenesis and are therefore of significant clinical interest. There are 42 of these MFRs present in the genome of *Streptomyces coelicolor*, a prolific producer of secondary metabolites, the majority of which are uncharacterized.

In this work, the crystal structures of three of these MFRs from *Streptomyces coelicolor*, Sco3914, Sco4122 and Sco5413, have been determined using anomalous dispersion methods. The structure of Sco5413 was determined to 1.25 Å resolution, the highest resolution for an MFR at this point. Genetic and biophysical analyses of these transcription factors have also been performed and the DNA footprints of Sco3914 and Sco4122 have been defined. Disruption mutants for each of the genes have been created and phenotypic analysis revealed a distinct phenotype for Sco5413. The structures do not reveal a ligand for any of the structures, but comparisons with other MFR structures provide some evidence towards a conserved ligand binding site amongst MFRs.

# Contents

**Table of Contents**

# Contents

# Contents

# Contents

## List of Figures

## List of Tables

# List of Tables

## Abbreviations

| | |
|---|---|
| Å | Angstrom ($10^{-10}$ m) |
| λ | X-ray wavelength |
| a,b,c, α,β,γ | unit cell dimensions |
| Act | Actinorhodin |
| ASU | Asymmetric unit |
| βME | β-mercaptoethanol |
| bp | base pairs |
| CCD | Charge coupled device |
| Da | Dalton |
| Diamond | Diamond Light Source Synchrotron facility, Oxfordshire |
| DLS | Dynamic Light Scattering |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribose Nucleic Acid |
| DTNB | 5,5'-dithiobis-(2-nitrobenzoic acid) |
| DTT | 1,4 dithiothreitol |
| *E. coli* | *Escherichia coli* |
| EDTA | Ethylenediamine tetraacetic acid |
| EtOH | Ethanol |
| HEPES | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| F | Structure factor amplitude |
| $F_{calc}$ | Calculated structure factor amplitude |
| $F_{obs}$ | Observed structure factor amplitude |
| GlcNAc | N-Acetyl-D-Glucosamine |
| H | α-helix |
| HCl | Hydrochloric acid |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| kDa | kilo Dalton |
| LB | Luria Bertani media |
| LMW | Low molecular weight |
| M9 | Minimal media |

| | |
|---|---|
| MAD | Multi wavelength anomalous dispersion |
| MALDI-TOF | Matrix assisted laser desorption ionisation - time of flight |
| MFR | MarR family of regulators |
| MS | Mass spectrometry |
| NCS | Non-crystallographic symmetry |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| PDB | Protein data bank |
| Red | Undecylprodigiosin |
| ReDCaT | Reusable DNA Capture Technology |
| RMSD | Root mean squared deviation |
| rpm | Revolutions per minute |
| *S. coelicolor* | *Streptomyces coelicolor* |
| SAL | Salicylate |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SeMet | Selenomethionine |
| SFM | Soya flour mannitol |
| SIRAS | Single isomorphous replacement with anomalous scattering |
| SPR | Surface plasmon resonance |
| TCEP | Tris(2-carboxyethyl)phosphine |
| Tris | Tris(hydroxymethyl)aminomethane |
| W | loop (wing) |

## Acknowledgements

There have been many people who have helped me to get to this point and I would like to offer my most sincere thanks to many, perhaps not all of them mentioned here.

First and foremost, I am truly indebted to my supervisor, David Lawson, who has provided excellent guidance and support throughout. Your extreme patience and depth of knowledge are admirable and I thank you for introducing me to the wonderful world of crystallography. I also express my sincere gratitude to my co-supervisor, Mervyn Bibb, and members of my supervisory committee, Richard Morris and Clare Stevenson. Extra special thanks go to Clare, who welcomed me to the lab as an undergraduate, and has supported and encouraged me throughout.

During my four years of study, I have had the privilege of working with many excellent scientists, too many to mention everyone by name, but I would like to state that it has been a true pleasure to work with many members, past and present, of the Biological Chemistry and Molecular Microbiology departments of the John Innes Centre.  I would especially like to thank Sandra Greive, Karl Syson, Tung Le, Ngat Tran, Stuart King, and Dan Tromans for sound advice and a good dose of humour.

Of course, I have not completely existed in a scientific bubble of bacteria for the past few years and, on a more personal level, I must thank my mother, Nuala, and my sisters, Sam and Emma. Your love, support and regular updates from the fatherland have reminded me of how important you all are to everything I do. And to my father, Roger Holley, to whose memory this thesis is dedicated: you always encouraged me to follow science and I sorely wish that you could have known that I did take your advice, eventually.

Finally, I would never have started any of this without my husband, the wonderful Finlay Scott. I can't ever thank you enough for nurturing my interest in science, encouraging me to return to education and for enduring me over the past few months (years?). And you make the best corn tortillas this side of Mexico. To you, I owe so very much.

1

# Chapter 1

# Introduction

## 1.1    Introduction

The widespread emergence of new forms of bacteria which are resistant to antibacterial compounds is a cause for concern and investigation. Bacteria have evolved strategies to enable them to evade destruction from previously very effective antibiotics such as the fluoroquinolones and vancomycin. Some of the proteins involved in this resistance have also been implicated in regulating virulence genes (MgrA, for example, which is a global regulator in *Staphylococcus aureus*), and so are of increasing clinical importance in the efforts to control pathogenic bacteria. Understanding the roles and mechanisms of these proteins will assist the development of new methods of fighting and controlling the diseases caused by these organisms. The existence of structural homologues in saprophitic bacteria which are also prolific producers of antibiotics, such as streptomycetes, offers a valuable platform for the study of the regulatory systems and signalling networks in operation. The fact that streptomycetes produce a vast number of natural products is also of interest as the biosynthetic gene clusters offer a huge resource which could be exploited. However, the regulation of the pathways involved in these gene clusters is complicated and the vast majority of the numerous regulatory proteins involved remain uncharacterized. This work details the genetic and biophysical analysis of selected transcription factors belonging to the MarR family from *Streptomyces coelicolor*.

## 1.2    *Streptomyces coelicolor*

*Streptomyces coelicolor* is the most studied of the actinomycetes and, as such, it is the ideal model system for the study of streptomycetes in general. Its complete genome was published in 2002 and its chromosome is over 8 Mb long with a G+C content of 72.1% and is predicted to contain 7825 protein encoding genes (Bentley *et al.*, 2002). The streptomycetes produce a wide range of antibiotics and antifungal compounds as well as immunosuppressants and anti-tumour agents which are useful for cancer treatments. They produce over two thirds of the natural antibiotic agents currently in use and therefore present a wealthy natural resource. *Streptomyces coelicolor* itself produces four

distinct antibiotic compounds: the blue-pigmented actinorhodin; red-pigmented undecylprodigiosin; methylenomycin (Mmy) and calcium-dependent antibiotic (CDA). The pigmented antibiotics offer a useful tool for the study of phenotypical responses to environmental changes and mutations.

The regulation of secondary metabolite biosynthesis in *Streptomyces* is the result of a complex interplay between a variety of signals and sensors. It is difficult to fully comprehend the method of regulation and indeed the precise roles of the secondary metabolites in a laboratory setting. Streptomycetes inhabit the soil and therefore exist in an environment where they frequently encounter and must respond to threats and stresses which are quite different to those encountered during a laboratory experiment. However, the interplay between the signals and response elements presents an interesting challenge to those who wish to understand and exploit the regulatory networks and cascades which make the streptomycetes such prolific producers of antibiotics.
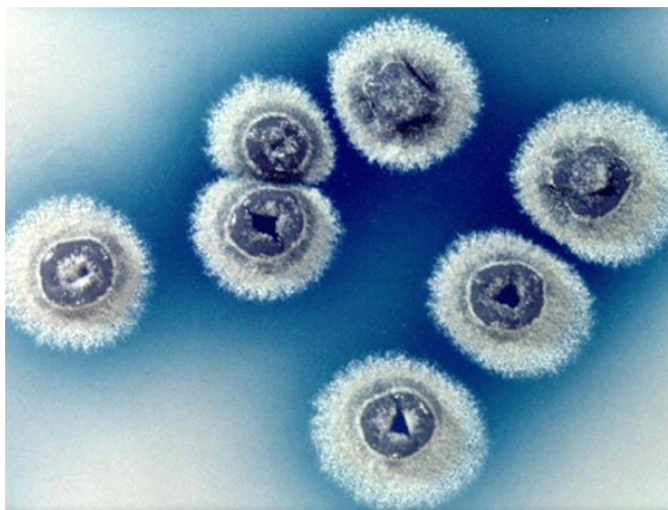


**Figure 1.1** Colonies of *Streptomyces coelicolor* producing blue-pigmented actinorhodin

In liquid medium, antibiotic production generally occurs during the stationary phase (Bibb, 2005) whereas on agar, antibiotic production can overlap with morphological differentiation (Takano *et al.*, 1992). *Streptomyces coelicolor* contains several pleiotropic

genes which have a role in the regulation of antibiotic biosynthesis. Amongst these are *afsR* and *absB*. AfsR is a protein which exhibits pleiotropic regulatory activity upon antibiotic biosynthetic genes, but only when grown on certain media. *absB* mutants were found to be deficient in antibiotic synthesis but were able to sporulate (Adamidis & Champness, 1992), indicating that this gene is involved in the specific regulation of antibiotic synthesis.

Antibiotic production can be triggered by the presence of small, diffusible molecules such as γ-butyrolactone, a hormone involved in quorum sensing. These small molecules can also be produced in response to physiological and environmental stress factors. The phosphorylated guanine nucleotide, ppGpp, has been studied in relation to its influence upon antibiotic production and its role in regulatory networks (Bibb, 1996).

## 1.3   Control of antibiotic biosynthesis

Antibiotic biosynthesis occurs usually after the growth phase, particularly when there is a nutritional limitation in the immediate environment (Martin & Demain, 1980). Soil bacteria such as streptomycetes are particularly adept producers of antibiotics. Their environment would necessitate this as they often encounter other soil dwelling microbes and need to compete for resources. Prior to this stationary/linear growth phase, bacteria can be susceptible to their own antibiotic. *Streptomyces antibioticus*, for example, is inhibited 50% by 4 mg ml$^{-1}$ of streptomycin during the growth phase, whereas it produces about 120 mg ml$^{-1}$ during idiophase (Martin & Demain, 1980). Resistance occurs generally as the antibiotic is produced. This is further supported by the findings of a study upon the transcription of *mmy* and *mmr* during the biosynthesis of methylenomycin (Hobbs *et al.*, 1992). This study also observed the inhibition of actinorhodin synthesis by phosphate and gave methylenomycin as the sole antibiotic produced. The inhibition of actinorhodin production was purported to be at the level of transcription as no *actIII* transcripts were found in the culture containing 11 mM phosphate whereas both *actIII* and *mmy* transcripts were found in media containing 1 mM phosphate.

The Abs⁻ phenotype revealed that mutations in the *absA* and *abaB* loci blocked antibiotic synthesis but not sporulation (Adamidis *et al.*, 1990). This work determined that there are genes which are specifically involved in synthesis of antibiotics and, unlike the *bld* mutants, do not also affect morphological differentiation. This phenotype is not conditional upon the media used and as such, it provides a stepping stone to understanding how the antibiotic synthesis genes are regulated independent of genes associated with sporulation. Interestingly, it was also found that the *absA* mutants were as resistant to methylenomycin as the wild type strain, despite not producing antibiotic itself. Antibiotic resistance usually occurs following antibiotic synthesis and it is generally induced by the antibiotic or precursor molecules.

Another locus, *absB*, also gives rise to the Abs⁻ phenotype upon mutation (Adamidis & Champness, 1992). Production of all four antibiotics was also blocked in this mutant. A mutation in this locus gave a mutant which was able to sporulate like the WT but was deficient in antibiotic synthesis. By cloning multiple copies of the *actII-ORF4* gene into these mutants, production of actinorhodin was restored, suggesting that this gene is somehow regulated by *absB*. *absA* encodes a two-component signal transduction system (Brian *et al.*, 1996) which is linked to the transcription of *actII-ORF4* and *redD* which code for *act*II-ORF4 and RedD: pathway specific transcriptional activators in the actinorhodin and undecylprodigiosin biosynthetic pathways respectively (Gramajo *et al.*, 1993, Takano *et al.*, 1992). They are OmpR-like DNA-binding proteins whose expression is regulated in a growth-phase dependent mechanism. AbsA1 is a sensor kinase whilst AbsA2 is a response regulator. *absB* encodes RNase III, a homologue of RNaseII from *E. coli* (Price *et al.*, 1999), which is a double-stranded RNA specific endoribonuclease. This enzyme is required for the regulation of antibiotic biosynthesis and for the formation of sporulation septa (Sello & Buttner, 2008). Its specific mode of regulation has not been determined, but it has been suggested that it could degrade mRNA which encodes a repressor or activate other transcripts of genes involved in regulation (Aceti & Champness, 1998). It has recently

been shown that it regulates its own expression by cleaving its own transcript (Xu *et al.*, 2008) and participates in a post-transcriptional autoregulatory feedback loop with AdpA (Xu *et al.*, 2010).

Mutations of other loci involved in antibiotic synthesis, such as *afsR* and *abaA*, also revealed regulation systems which act independently of sporulation. *afsR* encodes a transcriptional activator with ATPase activity which is phosphorylated by a kinase, such as AfsK. AfsR is a pleiotropic regulator of genes involved in Act, Red and CDA production and, as it is phosphorylated by several kinases, it is capable of integrating a variety of signals (Bibb, 2005, Horinouchi, 2003). The *abaA* mutation abolishes actinorhodin production and reduces the production of CDA and undecylprodigiosin, but synthesis of methylenomycin is unaffected (Fernandez-Moreno *et al.*, 1992).

The phosphorylated guanine nucleotide, ppGpp, is a secondary metabolite which is also known as a stringent factor. Synthesis of ppGpp occurs under conditions of amino acid starvation by the ribosome associated ppGpp synthetase, RelA. The synthesis of ppGpp was induced under conditions of sufficient nutrition by modification of the N-terminal of the ppGpp synthetase gene, relA (controlled by a thiostrepton-inducible promoter) (Hesketh *et al.*, 2001). These induced concentrations of ppGpp were linked to the induction of transcription of *actII-ORF4*. There was no link found to the production of undecylprodigioisin. A *ΔrelA* mutant strain of *S. coelicolor*, M570, was found to be unable to produce the pigmented actinorhodin and undecylprodigiosin. It was concluded that this RelA protein is involved in activating regulatory proteins in the antibiotic synthesis pathways. A later study showed that Sco4336 was induced by the production of ppGpp (Hesketh *et al.*, 2007). It is possible that this is related to the increase in transcription of *actII-ORF4* and *cdaR* regulatory genes. There is some evidence to suggest that *sco4336* transposon mutants have a reduced ability to synthesize Act but only on certain media. This implies a global role for ppGpp synthesis in morphological differentiation and antibiotic production in *S. coelicolor* (2).

## 1.4 MarR Family of transcriptional Regulators (MFRs)

The MarR family of proteins (MFRs) constitute a group of transcriptional regulators which regulate a diverse range of genes and consequent cellular processes. To date, there are over 13000 MFR sequences which have been annotated in genome databases (http://pfam.sanger.ac.uk/). The structures of several members of the MarR family (MFR) have been solved by X-ray diffraction and they all contain the characteristic winged helix-turn-helix (wHTH) DNA binding motif as well as being triangular in shape. They exist as homodimers in solution and typically bind anionic lipophilic effector molecules, such as salicylate, which modify the dimer's affinity for DNA (Wilkinson & Grove, 2006). Salicylate has been found to induce the expression of the *mar* operon of *Escherichia coli* (Cohen *et al.*, 1993b) and MarR was the first structure to have been solved with this ligand bound. Many subsequent studies of MFRs have focused on this ligand and several other structures have since been determined in complex with salicylate. The members of this family are transcriptional regulators which are involved in a variety of processes including antibiotic resistance, pathogenesis and aromatic catabolic pathways. Mutations in the *mar* locus were first implicated in the amplifiable resistance to multiple, unrelated antibiotics in *E. coli* (George & Levy, 1983) and clinical isolates of these drug-resistant mutants revealed multiple mutations in the *marR* gene (Maneewannakul & Levy, 1996). Thus, the role of MarR in regulating this operon is clearly of clinical interest. Over half of the MFRs which have been characterized are known to be autoregulatory (Wilkinson & Grove, 2006). Typically, they bind to pseudopalindromic sequences at sites which can overlap the -35 and/or -10 promoter regions of their target genes (Martin & Rosner, 1995, Martin & Rosner, 2004, Wilkinson & Grove, 2006). This mode of binding would make it a steric inhibitor of RNA polymerase binding and thus a repressor of transcription. The majority of MFRs are repressors, but some have been shown to activate genes, for example, OhrR in *Streptomcyes coelicolor* (see section 1.4.2).

The *mar* locus has been well characterized, genetically and functionally (Cohen *et al.*, 1993a). Introduction of a mutant *marR* gave a slight increase in resistance to tetracycline, chloramphenicol and norfloxacin, strongly suggesting that MarR is a repressor of the *marRAB* operon (Cohen *et al.*, 1993a). Further evidence that MarR is the repressor of *marO* was provided by incorporating multiple copies of the operator sequence in the genome which led to the reduced repression of the *marRAB* operon. MarA was found to have significant sequence identity to transcriptional activators (or positive regulators) found in *E. coli* such as AraC and RhaR (involved in carbohydrate metabolism) and SoxS, which is involved in oxidative stress response. The similarities between SoxS and MarA were more marked as they were both relatively short proteins whilst the carboxy-terminal ends of these proteins have a higher homology to the other proteins of this family. The C-terminal domains are thought to be the DNA-binding domains and they have the characteristic helix-turn-helix motifs. This work follows on from an earlier study which concluded that Mar mutants had a reduced expression of the outer membrane porin, OmpF (Cohen *et al.*, 1989). The study went on to investigate the role of *marA* upon this phenotype: by inserting a Tn5 transposon into *marA*, it was found that OmpF expression was restored. *marA* affects the expression of *micF*, an antisense RNA which in turn decreases the amount of *ompF* mRNA available for translation (Mizuno *et al.*, 1984). By fusing *lacZ* to *ompF*, it was determined that *marA* affects *ompF* post-transcriptionally. However, insertion of Tn5 into *marA* of Mar mutants only partially restored OmpF expression whilst it restored antibiotic susceptibility to that of the wild type strain. Clearly there are other factors involved in antibiotic resistance in Mar strains.
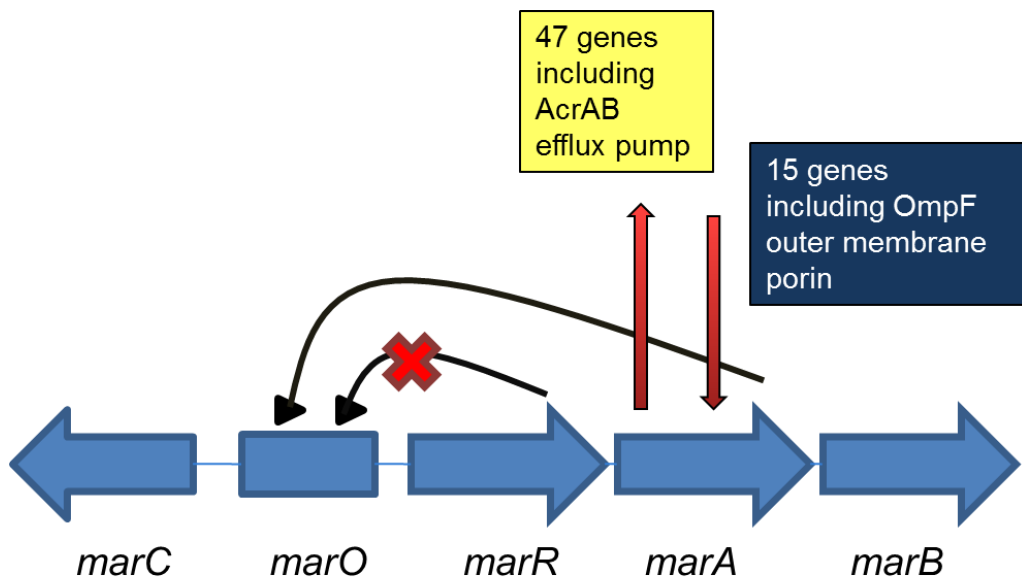
**Figure 1.2** Diagramatic representation of the mar operon in *E. coli*.

Several efflux pumps have been shown to contribute to the Mar phenotype: a study of the AcrAB efflux pump showed that the *marR1* mutants did not exhibit the increased resistance of the Mar phenotype in the absence of this system, thereby asserting that it is necessary for antibiotic resistance in the Mar mutants (Okusu *et al.*, 1996). In the *marR1* mutant, binding of this repressor to the *marRAB* operon was prevented resulting in the transcription of the operon. In the Δ*acrAB* strain, this MarR mutation was not seen to increase antibiotic resistance. In fact, it had the same level of resistance to five different antibiotics (with the exception of tetracycline) as did the Δ*acrAB* strain with wild type MarR.

## 1.4.1 Structure of MarR

The structure of MarR, shown in Figure 1.4, has been solved to 2.3 Å resolution (Alekshun *et al.*, 2001) and it was the first reported structure of a MarR protein. MarR was shown to crystallize as a dimer and each subunit is an αβ-protein which can be divided into two domains. An α-helical interface composed of the N- and C- terminal helices which

intertwine and connect the two subunits. The DNA-binding domain was recognized by the characteristic arrangement of secondary elements: H1-S1-H2-H3-S2-W1-S3 (Figure 1.3) (Gajiwala & Burley, 2000) and the electropositive surface residues of this region.
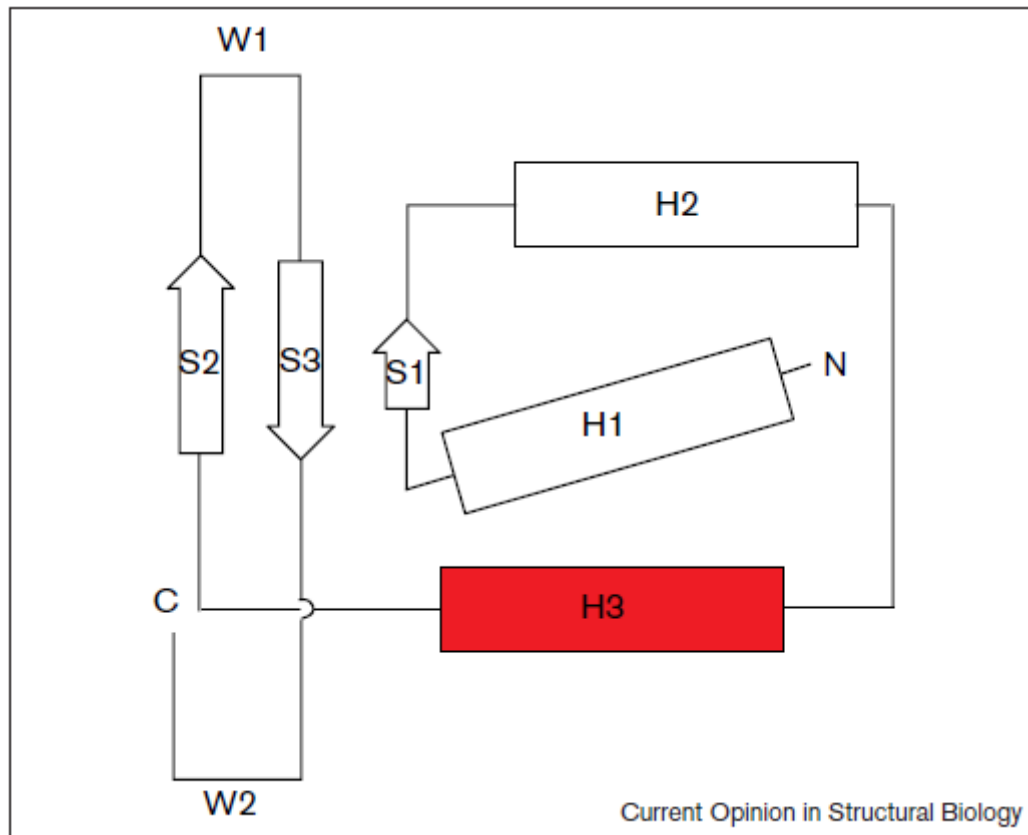


**Figure 1.3** Topology of the winged helix motif as presented by Gajiwala and Burley, 2000.

In *E. coli* MarR, α4 was proposed to be the DNA recognition helix (equivalent of H3 in the model depicted in Figure 1.3) and W1 is also implicated in DNA-interactions, based on mutations of residues in these regions which resulted in the abolition of DNA binding (Alekshun *et al.*, 2000). R94C was found to abolish DNA-binding. This residue, along with D92, has been found to be highly conserved amongst the MarR family and the OhrR structure bound to DNA reveals the equivalent R94 interacts with the phosphate backbone of the minor groove. However, without a crystal structure of the protein bound to DNA, the precise interactions between residues and nucleotides in this protein is

unknown. MarR also has a pair of salt-bridges formed between R73 and D67' (and the reciprocal pair) which stabilize the interaction between these two subunits.
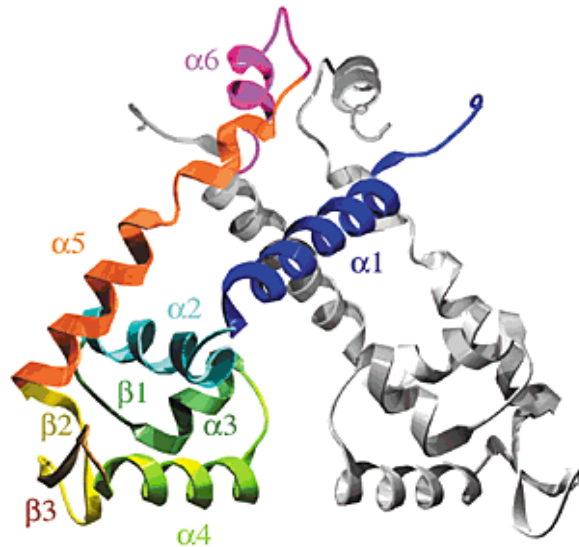


**Figure 1.4**   A ribbon diagram of MarR as presented by Alekshun and Levy, 2001. The secondary structures of one monomer are labelled and coloured.

Salicylate bound to MarR was determined clearly in this structure: two molecules can be observed per monomer. This contrasts with the study of MTH313 (Saridakis *et al.*, 2008) which found only one salicylate bound per monomer. This binding site was found to be between the recognition helix and the dimer helical interface on each subunit of MTH313. In other words, they are not equivalent and the study reports that the first binding site induces a large conformational change, but not the second. Both salicylate binding sites of MarR are positioned either side of the DNA-recognition helix so that they could affect DNA binding. Indeed, salicylate has been shown to activate the *mar* operon (Cohen *et al.*, 1993b) and, specifically, it has been shown to inhibit MarR (Alekshun & Levy, 1999). This salicylate-bound MarR structure cannot be modelled onto B-DNA (Wilkinson & Grove, 2006) or superimposed onto the OhrR-*ohrR* complex, strongly suggesting that it is in a conformation which does not allow DNA binding. In the MTH313 models, the large, asymmetrical structural change which occurred on binding salicylate causes the DNA-binding lobes to move apart by about 7 Å. The authors suggest that this open structure is

the inactive form which is incapable of DNA-binding. Comparing this structure to salicylate-bound MarR, they go on to argue that perhaps the MarR structure is actually in the active conformation as the separation of its lobes is similar to the apo (active) form of MTH313. This has since been disputed by Wilke *et al*. (2008): superimposing salicylate-bound MTH313 onto DNA-bound *Bacillus subtilis* OhrR reveals that is is in fact well configured for DNA binding. Whichever conclusion is correct, it is evidently clear that the MarR proteins have significant structural flexibility which affects their DNA-binding abilities.

## 1.4.2 MarR proteins have structural flexibility – HucR

HucR is a MarR homologue from *Deinococcus radiodurans*, an extremely resilient bacterium whose genome encodes orthologues of nearly every known stress response protein. HucR binds to a pseudopalindromic sequence: 8 bp half-sites separated by 2 bp. This binding site is located in a shared promoter region between *hucR* and a putative uricase which could mean that there is a mechanism of simultaneous corepression between these two genes (as the substrate for uricase is uric acid which is an efficient antagonist of DNA binding by HucR) (Wilkinson & Grove, 2004). Uric acid is also a scavenger of reactive oxygen species, suggesting that HucR plays a role in resistance of *D. radiodurans* to oxidative stress. The structure of this MarR family member reveals that it has an extra N-terminal α-helix which participates in dimerization (Bordelon *et al.*, 2006). The presence of two stacked histidines (His51 and His51') at the dimer interface suggests that these may participate in the pH-dependent DNA-binding mechanism. Circular dichroism spectra show that there is modification of the secondary structure of DNA upon binding of HucR (Wilkinson & Grove, 2005). Although there is no structure of HucR with ligand bound, the authors of this study concluded that there must be significant structural change upon binding of uric acid as the crystals cracked when transferred to mother liquor solution containing uric acid. Using site-directed mutagenesis and molecular-modeling, the urate-binding site of HucR has been defined and Trp20 (located in the extra helix) and Arg80 (in the DNA-recognition helix) shown to be the critical residues for ligand interaction (Perera *et al.*, 2009). The hypothesis is that HucR has the potential for

conformational flexibility observed in other MarR proteins and that this structural rearrangement is initiated by the binding of the ligand, urate, which propagates to the DNA recognition helices. Following the determination of the required residues for urate-binding and communication of the presence of this ligand to the DNA-recognition helices, a new subfamily of MarR homologues has been proposed. This is based on the predicted conservation of function denoted by a urate-binding signature: the urate responsive transcriptional regulator (UrtR), and the family includes HucR, PecS from *Agrobacterium tumefaciens* (Perera & Grove, 2010b) and MftR from *Burkholderia thailandensis* (Grove, 2010). Bioinformatic analysis of this subfamily found that the vast majority of the genes divergently transcribed to UrtRs annotated in sequence databases are membrane transporters. This proves interesting relating to the MFRs in *S. coelicolor*: Sco3133 is predicted to be a member of the UrtR subfamily and it is transcribed divergently to a putative *trans*-aconitate methyltransferase. The designation of such subfamilies indicates the growing wealth of information about MFRs and this should prove useful in exploiting their regulatory functions for clinical purposes.

In this work, a structural homology model based on HucR was used in the structural solution of two MFRs. Particular attention is paid to the presence of the extra helix of HucR and how this relates to two of the models presented here.

### 1.4.3  Organic hydroperoxide resistance in bacteria – thiol switches

Some of the MFRs which have been characterized are known to participate in the regulation of multi-drug efflux pumps. MgrA is a MarR homologue which is a key global regulator of virulence and resistance genes in the human pathogen, *Staphylococcus aureus*. It is a pleiotropic regulator which regulates a vast number of genes (~350) in this organism (Luong *et al.*, 2006). The transcription of multiple multidrug efflux pumps, including NorA and NorB (Truong-Bolduc *et al.*, 2005), is regulated by MgrA and it is therefore responsible for mediating resistance to fluoroquinolones and vancomycin. It is located 7 kb from *norA* and 700 kb from *norB* (Truong-Bolduc *et al.*, 2003). Microarray

13

analysis has shown that MgrA can both negatively and positively regulate the same genes at different time points (Luong *et al.*, 2006). The suggestion is that there are other regulators involved in other pathways during different stages of the organism's growth and that some genes are indirectly regulated by this protein. Indeed, it has also been observed that MgrA can act as a repressor of genes where other proteins act as activators and vice versa. MgrA represses *norB* directly and activates *abcA* directly, whereas NorG (a GntR-like transcriptional regulator) behaves as a direct activator for *norB* and a direct repressor for *abcA*. Thus the complexity of the systems at work increases, but unravelling them offers much potential.

The crystal structure of MgrA gave valuable clues as to its mechanism, suggesting that it operates by sensing reactive oxygen species via the oxidation of a cysteine located in the dimer interface on each monomer. Cys12 is the only cysteine residue found in the MgrA sequence and is located in helix α1. It can hydrogen bond with Ser113 and Tyr38 of the other monomer and therefore bears a similarity to the lone cysteine (Cys15) found in the oxygen sensing repressor, OhrR from *Bacillus subtilis* (Fuangthong & Helmann, 2002). The hydrogen bonds between Cys12 and the other two residues could be broken by oxidation, which could cause a conformational change resulting in DNA dissociation. This was found to be the case in this study which revealed that the MgrA-DNA complex *in vitro* did dissociate upon addition of a variety of oxidants, whilst the complex reformed upon addition of a dithiothreitol (DTT) (Chen *et al.*, 2006). An MgrA C12S mutant remained in complex with the DNA under oxidizing conditions, confirming the oxidation of this specific residue in the wild type protein. This represents a strategy of regulating multiple pathways involved in defence and resistance with one protein which can sense the threat. It also offers a potential target for future drug development.

The OhrR repressor in *Bacillus subtilis* participates in sensing organic hydroperoxides through conserved cysteine residues (Fuangthong & Helmann, 2002). It binds to the *ohrA* operator and represses this gene which encodes a protein involved in organic

hydroperoxide resistance (Fuangthong *et al.*, 2001). Reactive oxygen species (ROS) such as organic hydroperoxides can cause irreversible damage to cells and DNA and so a mechanism of sensing the presence of these molecules and rendering them harmless has emerged in bacteria. Its mechanism varies between species and has been found to be dependent upon the number of cysteine residues. These become oxidized which give rise to changes in conformation and thus affect binding to DNA. Such a mechanism can involve the formation of a reversible disulphide bridge, as is the case with OxyR in *E. coli* and MexR in *Pseudomonas aeruginosa*. In the case of BsOhrR (*Bacillus subtilis*), however, the mechanism involves a single cysteine residue (Cys15) which becomes oxidized to form cysteine-sulfenic acid. Conversely, OhrR from *Xanthomonas campestris* (XcOhrR) has three conserved cysteines, two of which can form intersubunit disulphide bonds upon oxidation: Cys22 and Cys127' (Panmanee *et al.*, 2006), which gives rise to a substantial conformational change and significant reorientation of the α4 helices rendering the protein incapable of binding to DNA.

OhrR also exists in *S. coelicolor* and falls into the same category as BsOhrR in that it has a single oxidizable cysteine, Cys28 (Oh *et al.*, 2007). In this organism, OhrR works as a repressor of its own transcription and that of the divergently transcribed gene, *ohrA*, and *ohrR* when reduced by binding cooperatively to the operator region between the two genes and sterically blocking transcription by RNA polymerase (RNAP). Upon oxidation, it induces *ohrA* by derepression whilst *ohrR* is activated by the oxidized OhrR. In the oxidized state, it has a reduced affinity for DNA but a dimer may still remain bound and acts as an activator by promoting the recruitment of RNAP. This dual role is different to the role of OhrR in *Bacillus subtilis* which, as far is known, only represses the *ohrA* gene under reducing conditions. The ability of Cys28 to sense organic peroxides and the regulatory response of OhrR were established through the creation of a C28S mutant. The affinity of this mutant for the DNA probe was unaffected following treatment with an oxidant, indicating the role of this residue in the redox regulation of the protein.

The actual mechanisms involved in some of these reactive cysteine proteins have been described in several studies which have found that there is a sulfenic acid intermediate involved in the reaction, and this applies to both single and multiple cysteine sensors. The difference is the subsequent reaction whereby the disulphide bond is formed between the single cysteine and a low molecular weight (LMW) thiol, or the two cysteines present in the protein molecule react with each other to form a disulphide bond.

SarZ, another single cysteine MFR which is redox active, undergoes a cysteine oxidation reaction when exposed to organic hydroperoxides (OHPs) which disrupts DNA binding (Poor *et al.*, 2009). Cys13 forms a sulfenic acid intermediate which was shown to be insufficient for derepression. It undergoes a further reaction with LMW thiols present in the cell to form a mixed disulphide (S-cysteinylation). This inactivates the protein and thus derepresses the gene. The crystal structures of the reduced (SarZ-SH), sulfenic acid (SarZ-SOH) and disulphide-modified (SarZ-BT) forms of SarZ demonstrate how the disruption of a hydrogen-bonding network surrounding the Cys13 caused by disulphide formation and the subsequent movement of Phe117' culminates in the conformational change which results in the dissociation of the protein from DNA. This allosteric mechanism defines the complete picture for the single cysteine regulatory proteins.



**α4 recognition helices**
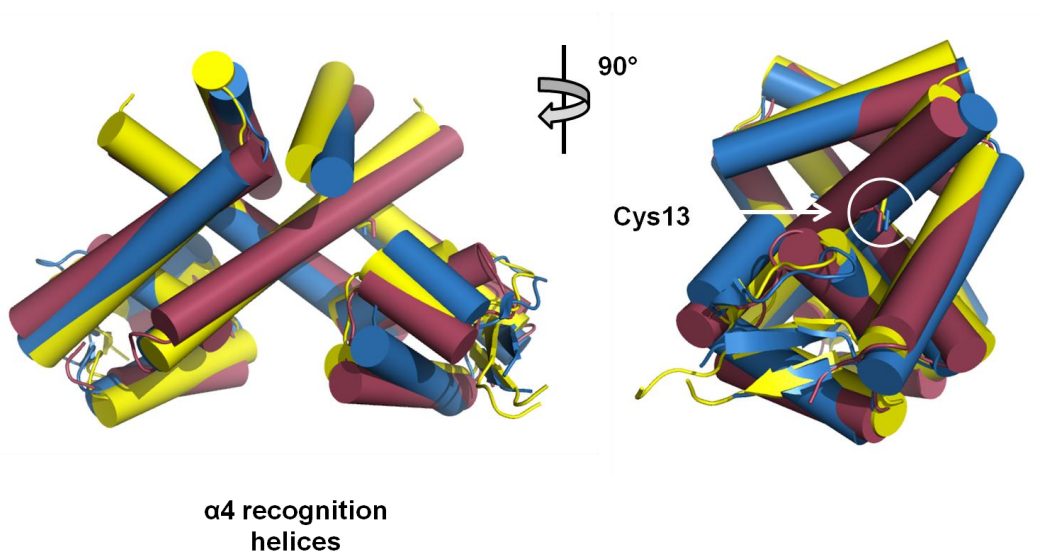
**Figure 1.5**  Superimposed crystal structures of SarZ in the reduced (dark red), mixed disulphide (blue) and sulfenic acid (yellow) forms.

Most recently, a study on the MarR type regulator, HypR, has provided further insights into the mechanisms of the dual cysteine class of repressors. The crystal structures of both the reduced and oxidized forms of this protein have been determined and revealed structural changes in the protein which result in the movement of the α4 helices towards each other by 4 Å (Palm *et al.*, 2012). HypR is activated by sodium hypochlorite (NaOCl) stress which induces the formation a Cys14-Cys49' intersubunit disulphide bond. This results in the transcription of *hypO* which encodes a putative nitroreductase. Cys14 is well conserved in other N-terminal cysteine containing MarR homologues. This work also determined the $pK_a$ values of the cysteines and found Cys14 has a lower value of 6.36 which would mean that it is probably deprotonated at physiological pH, making it particularly reactive and therefore useful to an oxidative stress response.

### 1.4.4  Protein-DNA interactions

The MFRs characterized to date control gene expression by interacting with pseudopalindromic DNA sequences (Wilkinson & Grove, 2006). Genetic analysis of the *ohrA* operon reveals an inverted repeat which is reported to be the binding site of OhrR (Fuangthong *et al*., 2001). It was also found that OhrR will tolerate imperfect inverted repeats in the binding region (up to three mismatches). This is another feature of the MarR homologues: they generally bind to pseudopalindromic sequences and also, many of them, to their own promoters (Perera & Grove, 2010a).
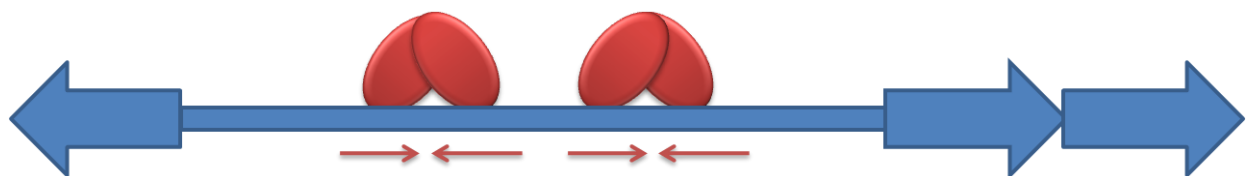


**Figure 1.6**  A model for the binding of two MarR dimers on DNA between two divergently transcribed genes.  The red arrows represent the (pseudo)palindromic sequences which the proteins recognize.

MepR is a MarR homologue in *Staphylococcus aureus* which has been shown to bind to its own promoter. It represses its own transcription as well as that of the multidrug efflux pump, MepA (Kaatz *et al.*, 2006). Gel mobility shift assays revealed that MepR binds upstream of its gene. There is also evidence that it binds the *mepA* gene as a dimer of dimers whilst it binds its own operator sequence as a single dimer (Kumaraswami *et al.*, 2009). It is induced by binding to most, but not all, MepA substrates and this interaction alleviates repression of *mepA* (Kaatz *et al.*, 2006). The interaction of protein with substrate at the *mepR* operator is different and remains to be clarified. What is interesting is that many of the substrates are cationic and it has a region which corresponds to the ArmR hydrophobic binding pocket of MexR. However, there is as yet no structure of this protein with a ligand bound. The structural data for this protein reveal that it is in a particularly open conformation which the authors deem unsuitable for DNA-binding. This was further exemplified by superimposing the structure onto DNA-bound BsOhrR.

MexR, however, has been shown to undergo a significant conformational change upon the binding of its repressor, the peptide, ArmR. Binding of this molecule results in the rearrangement of the dimerization region which consequently affects the DNA-binding regions, displacing the DNA recognition helices by 13 Å (from 29 Å to 16 Å in the apo structure with an RMSD value for the dimers of 2.8 Å) (Chen *et al.*, 2010). This apo-MexR conformation is reportedly capable of binding DNA. Strikingly, the superimposition of the oxidized form of MexR with the reduced MexR gives a relatively small RMSD value of 1.6 $Å^2$ with only small movement of the DNA recognition helices, instead inducing a localized shift in the HTH region which prevents DNA binding due to steric clashes. The main conclusion from this study is that there are two distinct cellular regulatory pathways present in *P. aeruginosa* to regulate multidrug resistance.

SlyA, an important transcription and virulence factor in *Salmonella typhimurium*, has been shown to recognize a partially palindromic site within its own promoter region (Stapleton *et al.*, 2002) and at least five SlyA dimers bind in this region. The structure of SlyA in complex with a 22 base pair (bp) DNA duplex (Dolan *et al.*, 2011) derived from its promoter region sequence which contains the 12 bp pseudopalindrome, TTAGCAAGCTAA, its high affinity binding site (Stapleton *et al.*, 2002). The structure was solved to 2.96 Å resolution and is visually very different to the structure of the apo SlyA (refined to 2.4 Å resolution). The separation between the α4 recognition helices in the apo structure is 15 Å and these are held in place by intermolecular salt bridges between E59 and R65'. The binding of the α4 recognition helices results in a widening of the major groove at the centre of the duplex by up to 2.3 Å. In this structure, R65 has been shown to provide the sole direct contact between the DNA and the recognition helix (via bidentate H-bonds with guanine 14), whilst the α2 helix also participates in an electrostatic interaction with the N-terminal positive dipole and the DNA backbone (Figure 1.7). Van der Waals contacts are also made by P61 and S62. Further contacts can be seen between the β-wing and α3 helix with the minor groove and these display some differences between each subunit, relecting the asymmetry of the sequences flanking the 12 bp pseudopalindrome. Specifically, R86 (equivalent to OhrR R94) which interacts with a pair of sugar rings of two specifically positioned nucleotides. The proposal is that this residue has an important role in the indirect readout of the DNA by "measuring" the width of the minor groove via its interactions with these sugar moieties, allowing SlyA to recognize DNA shape rather than sequence. The number of contacts with the DNA backbone outnumber the contacts with bases of the operator, again pointing towards a mechanism of indirect readout of the high-affinity site.
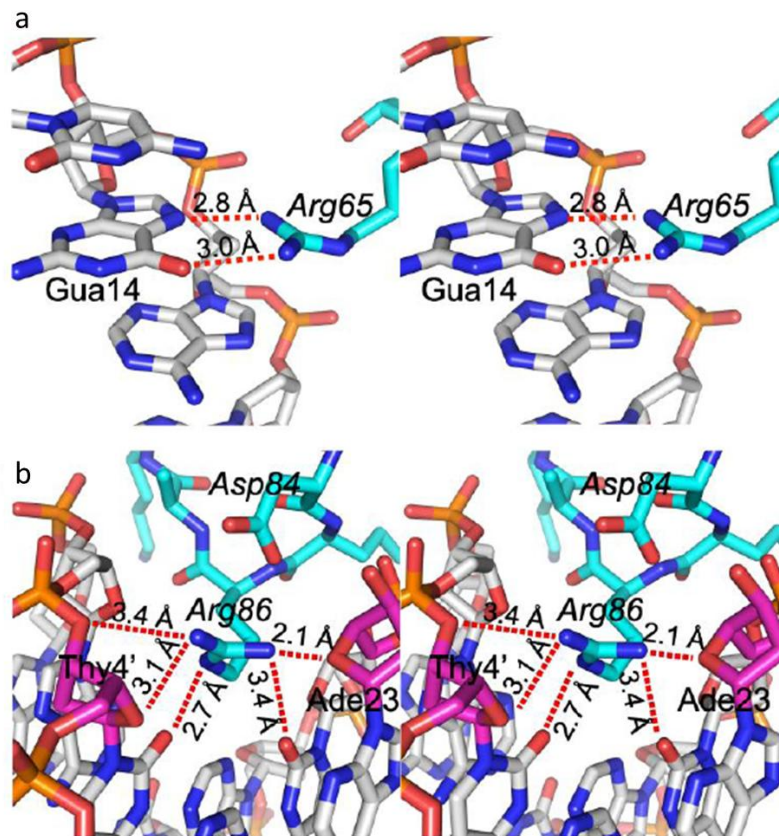
**Figure 1.7** SlyA-DNA contacts from crystal structure (PDB code 3Q5F). (a) Stereo view of hydrogen bond contacts made by Arg65 to guanine 14; (b) stereo view of contacts made by Arg86 with minor groove. (Dolan *et al.*, 2011).

This study also addressed the binding of small molecule ligands and their effects on DNA-binding, following up the work on the unpublished salicylate bound SlyA structure, PDB code 3DEU. The crystal structures of both the DNA and salicylate bound SlyA dimers were superimposed at the monomer level with an RMSD value of 1.6 Å. This does not suggest a marked conformational change, but one of the salicylates from the 3DEU model was seen to occupy a position otherwise occupied by guanine 14 in the 3Q5F DNA complex structure. The superimposition also revealed that each α4 DNA recognition helix undergoes a rotation of ~35° around its axis which would disrupt protein-DNA contacts in this region via steric clashes. This supports the EMSA data which found that salicylate binding occludes DNA-binding (Stapleton *et al.*, 2002), but at high concentrations (25 mM). One of the main conclusions from this 2011 study is that the recognition of DNA shape as opposed to direct sequence reading will be a significant feature of MFR-DNA interactions. This is how MFRs can act as global regulators: their structural flexibility

contributes to their ability to recognize and interact with specific DNA sequences with some degree of degeneracy.

SarA is an MFR found in *Staphylococcus aureus* (Cheung *et al.*, 2008). It possesses the characteristic HTH and β-hairpin winged motifs which are involved in DNA interactions based on the model of its structural homologue, SarR (Liu *et al.*, 2001). Mutations of conserved basic residues in the winged region revealed that R84 and R90 are required for interactions with DNA (Liu *et al.*, 2006) whereas conserved basic residues (K63 and K69) in the HTH region were found not to be essential. This is consistent with the findings for several other MarR-like proteins. The acidic residues, D8 and E11, have a role in the activation of SarA. It has been suggested that they coordinate divalent cations such that two dimers are connected via a divalent cation bridge. Gel retardation assays confirmed the existence of these multimeric protein-DNA complexes. The significance of these multimers in the regulation of transcription was probed by complementation of a D8A-E11A mutant with the *sarA* mutant phenotype. It failed to complement this mutant and so it has been proposed these residues are necessary for activation of SarA. In the absence of divalent cations, the acidic residues could repel each other and so prevent the complexation of several homodimers on the DNA. This could affect the topology of the DNA for correct binding of RNA polymerase on the promoter.

AbsC, a pleiotropic regulator of antibiotic production in *Streptomyces coelicolor*, binds to its own promoter (work carried out in Mervyn Bibb's group, JIC) which is also the case with MarR. R91 was found to be vital to DNA-binding as the R91W mutant does not retard the 198 bp fragment containing the *absC* promoter in a gel-shift assay. This residue is highly conserved in MarR homologues and in the OhrR-DNA complex (Figure 1.8), R94 can be seen to interact with the O2 and O4' of thymine 7 located in the minor groove (Hong *et al.*, 2005b). This residue mirrors that of R86 in SlyA, which also enters the minor groove of DNA.  In OhrR, D92 is also located in the wing domain of the protein and interacts with C5' of adenine 25 and the phosphate backbone of thymine 26, as well as

making an electrostatic interaction with R94 (again, the equivalent D84 in SlyA also interacts with R86). It has been suggested that D92 is also conserved amongst the MarR family and this can be seen in some of the homologues selected for this research project and this provides a useful model for the DNA-binding mechanism of the MarR family. Oxidation of this complex by organic hydroperoxides results in the induction of OhrR which in turn leads to *ohrA* transcription.
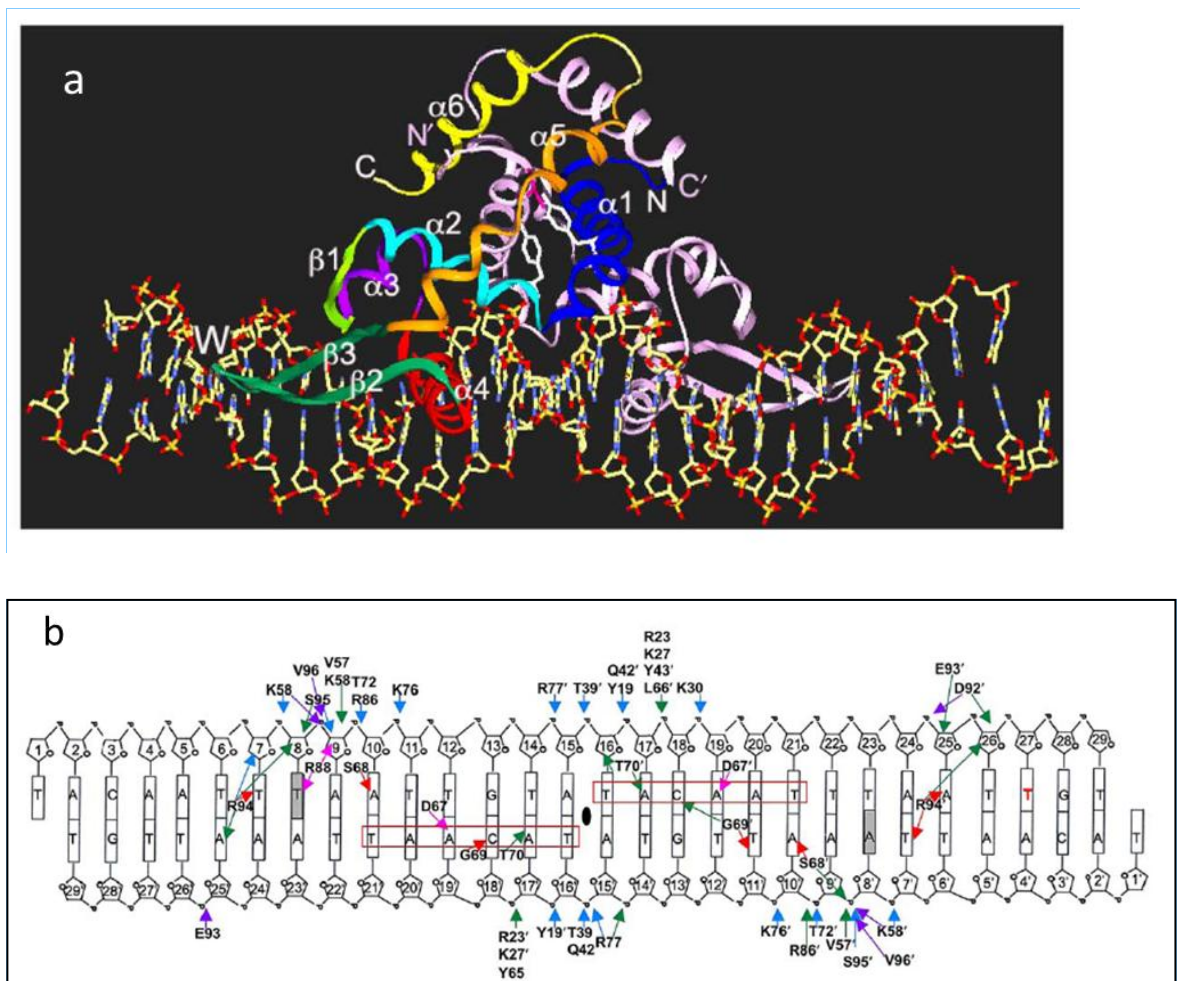


**Figure 1.8** (a) OhrR-*ohrA* operator complex (taken from Hong *et al*., 2005b). (b) Schematic of the protein:DNA interactions (Hong *et al.*, 2005b). Base specific contacts are indicated by the red arrows; blue and purple arrows indicate DNA backbone contacts; green arrows, Van der Waals contacts.

The current data for MFRs in complex with a ligand and/or with DNA point towards a mechanism which links ligand binding with attenuation of DNA-binding. The binding of a ligand is communicated to the DNA-binding domain resulting in a movement of the α4

recognition helices such that DNA-binding is no longer favourable. Thus, MFRs must have considerable conformational flexibility in order to bind DNA sequence with specificity. Such conformational flexibility may also be extended to the DNA helix, as the circular dichroism spectroscopy data for the alteration of DNA secondary structure by HucR has demonstrated (Wilkinson & Grove, 2005).

### 1.4.5 MarR family proteins involved in the regulation of virulence factors

RovA is a member of the MarR/SlyA family which is involved in the regulation of the virulence factors in the human pathogens, *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* (Cathelyn *et al.*, 2006). It has been shown to regulate the invasion factor invasion (*inv*) and appears to have a more global role in the virulence of *Y. Pestis*, causative agent of the bubonic plague, which lacks the *inv* gene. This pathogen requires a very complex regulatory system to coordinate integration of environmental signals with regulation of gene expression, given that it infects very different hosts: fleas and humans. Its success depends on tight regulation and this study concluded that RovA is an important global regulator of genes associated with virulence in *Y. Pestis*.

MfbR is involved in regulating virulence factors in the necrosis-inducing plant pathogen, *Dickeya dadantii* (Reverchon *et al.,* 2010).  This protein represses its own transcription and acts as an activator of genes involved in virulence, specifically plant-cell wall degrading enzymes. MfbR is pH responsive and is unable to activate these virulence genes under acidic conditions during the early stages of infection. At an advanced stage of infection, the apoplast is more alkaline and this activates MfbR and in turn, upregulates the transcription of the virulence genes.

### 1.4.6 Effector molecules of MarR proteins

There appears to be some diversity in the types of effectors which members of the MarR family bind. Identifying effector molecules is perhaps the most challenging aspect of

studying MFRs; to date, very few of these ligands have been conclusively identified. MarR itself was crystallized in complex with salicylate and two binding sites were determined for this interaction. Whilst the Mar phenotype is induced by the binding of salicylate (Alekshun & Levy, 1999, Martin & Rosner, 1995), the protein was crystallized with a high concentration of salicylate (250 mM) and thus the physiological relevance of these sites has been questioned. Nevertheless, subsequent studies of MFRs have tended to focus on salicylate as a potential ligand and a few have been found to complex with this molecule.

The structure of MTH313 from *Methanobacterium thermoautotrophicum* has been solved both in its apo form and complexed with salicylate (Saridakis *et al*., 2008). This structure revealed two asymmetrical binding sites for salicylate (Figure 1.9), but only one of these, site 1 (SAL1), was considered to be biologically relevant as it affects the position of the DNA recognition helix in that monomer, resulting in an asymmetrical conformational change. The position of this site is at the dimer interface between helices α1',α3 and α7 and is composed of predominantly hydrophobic residues with a small number of charged/polar residues. Salicylate has been shown to disrupt DNA-binding of MTH313 and the suggestion in this work was that this is due to a large, asymmetrical conformational change.
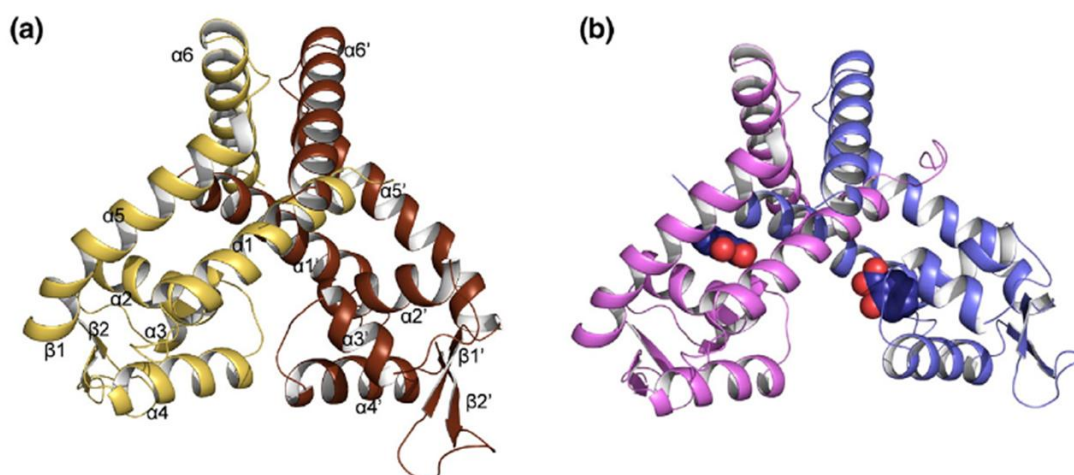


**Figure 1.9** The crystal structure of MTH313 (a) apo form and (b) in complex with salicylate; image shows the two, asymmetrical binding sites. Image taken from Saridakis *et al.*, 2008.

ST1710, from *Sulfolubus tokodaii* (Kumarevel *et al.*, 2009), and SlyA, from *Salmonella enterica* (PDB ref. 3DEU; Le Trong *et al.*, 2008, unpublished), are two other MFRs which have both been crystallized in complex with salicylate. Three salicylate molecules per monomer were determined for SlyA and interestingly, one of these sites corresponded to the SAL1 site in MTH313 (Perera & Grove, 2010a) whilst ST1710 has one salicylate molecule bound per monomer in symmetrically opposed sites, again in a comparable location.

The evidence is accumulating for a conserved ligand binding site amongst MFRs. TcaR is a multifunctional repressor from *Staphylococcus epidermis* and has been implicated in resistance to antibiotics such as methicillin and teicoplanin (Brandenberger *et al.*, 2000). It appears to have a dual regulatory function; activating another global regulator, SarS and repressing the transcription of a probable adhesin, SasF (McCallum *et al.*, 2004) and has a direct effect on the production of the protective biofilm which is necessary for the survival of this pathogen. A comprehensive structural study of TcaR in complex with several antibiotic molecules provides further information regarding a conserved ligand binding site and the allosteric mechanisms of MFR regulation (Chang *et al.*, 2010). TcaR was crystallized with salicylate, penicillin G (PnG), ampicillin, methicillin and kanamycin whilst the inhibition of TcaR by these compounds, in addition to gentamycin, streptomycin and chloramphenicol, was tested using electrophoretic mobility shift assay (EMSA). The study presented eight distinct (non-equivalent) binding sites on the dimer for salicylate and these were reported to drive an asymmetrical conformational change which would likely disfavour DNA-binding. Two sites were determined for PnG, but only one of these was considered to be biologically relevant and this one is located between the helical dimerization and DNA-binding domains and induces an asymmetrical conformational change similar to that observed in the salicylate structure. Each of the other antibiotics in the crystal complexes were found to have two binding sites, the first

of which was suggested to be the most biologically important as it could regulate TcaR activity. These findings are consistent with a conserved binding pocket for MFRs.
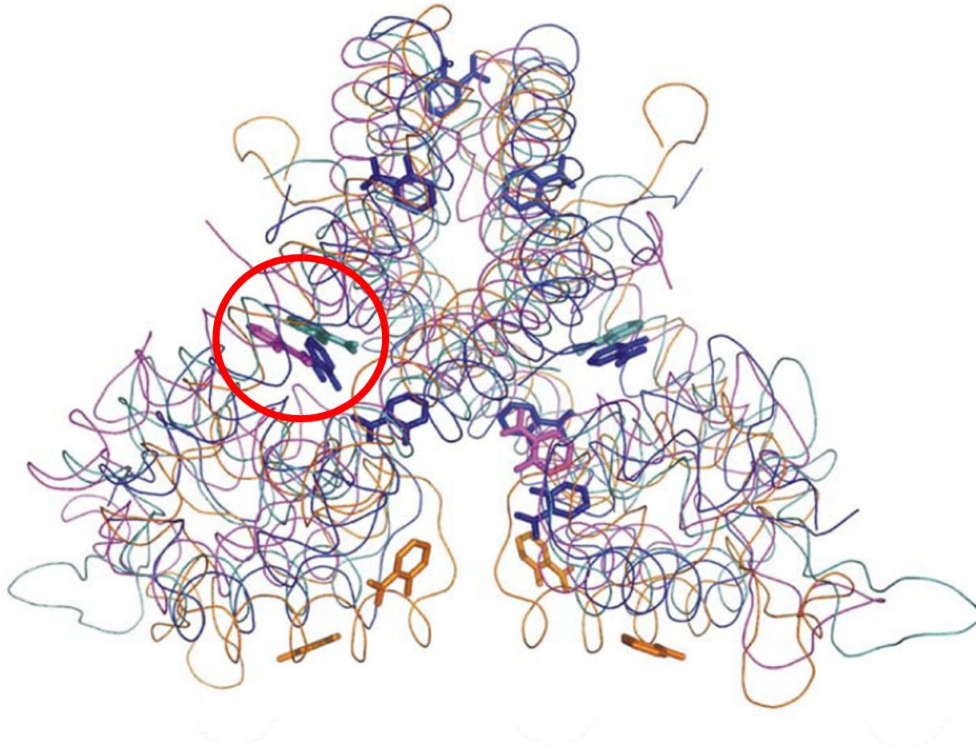


**Figure 1.10** Comparison of the salicylate binding sites on four MFR structures: TcaR – blue; *E. coli* MarR – orange; MTH313 – magenta and ST1710 – teal. The putative conserved SAL1 binding site is circled.

Whereas the DNA binding affinity of MarR and TcaR is significantly reduced by salicylate, MexR suffers no such inhibition in its DNA binding upon the addition of salicylate (Lim *et al.*, 2002). MexR is a member of the MarR family of proteins and it negatively regulates the expression of a multidrug efflux system, such as MexAB-OprM. *Pseudomonas aeruginosa* is a gram-negative human pathogen which can prove fatal to patients with compromised immune systems. It evades antibiotic treatment through its outer membrane which has a low permeability, and through the expression of efflux pumps (which are hyperexpressed in the *mexR* mutants). The presence of two redox-active cysteines in this protein is the subject of a study which concludes that it senses oxidative stress through these cysteines (Cys30 and Cys62) (Chen *et al.*, 2008, Chen *et al.*, 2006). These form intermonomer disulphide bonds upon oxidation which result in a

conformational change, causing the protein to dissociate from the promoter DNA. The *mexAB-oprM* operon is thus activated and the drug efflux system switched on. MexR is also inactivated by an antirepressor peptide, ArmR, which associates with the protein allosterically (Wilke *et al.*, 2008). The crystal structure of the MexR dimer in complex with a fragment of ArmR reveals that this peptide binds in a hydrophobic cavity in the centre of the protein and induces a conformational change which renders MexR incapable of binding DNA. Superimposition of the MexRll-ArmRc complex with the DNA-binding domain of apo-MexR dimer CD shows that the α4 recognition helix is displaced by 13 Å whilst the wing is displaced by 18 Å from its expected position in the minor groove. Further evidence that the binding of ArmR prevents DNA-binding is provided by the presence of E50 of ArmR which projects between the two MexR DNA-binding domains. This negatively charged residue would result in electrostatic repulsion and so weaken the interaction between MexR and the DNA. The pseudo-symmetrical binding site overlaps with the small effector molecule binding sites found in other MarR homologues (Figure 1.10), but the interactions of the peptide with the protein surface are multiple and therefore stabilize this conformation. Interesting comparisons with other MarR homologues were made in this study (Wilke *et al*., 2008). Whilst there are obvious similarities such as the overall structure and the importance of wing residues R83 and R91 to DNA-binding, there are significant differences which bring into question the current theories on effector binding and mechanisms. Not least is the criticism of the salicylate-bound MTH313 structure which is proposed by Saridakis *et al.* (2008) to be in the inactive form. Wilke *et al.* (2008) state that this structure is better configured for DNA-binding when superimposed onto the OhrR-*ohra* operator complex than is the apo-form of the protein. The structure of MTH313 has been presented in both its apo form and also complexed with salicylate (Saridakis *et al.*, 2008). The two structures were compared and it was found that the DNA binding lobe undergoes a large, asymmetrical conformational change which, according to Wilke *et al.* (2008), prevents it binding its promoter region. However, electrophoretic mobility shift assays would indicate that salicylate does act as an inactivator of MTH313 (Saridakis *et al.*, 2008).

The variety of ligands which can associate with different MFRs is becoming apparent. MexR, mentioned previously, can bind a peptide effector and LVIS553 has been shown to specifically recognize novobiocin as an effector (Pagliai *et al.*, 2010). The *Streptococcus pneumoniae* adhesion competence regulator, AdcR, is the first metal-dependent member of the MarR family to be structurally characterized. Its structure was solved to 2 Å resolution and reveals two zinc ions bound per monomer (Guerra *et al.*, 2011). This structure presents an interesting new insight into allosteric mechanisms in MarR-DNA interactions. The superimposition of the helix-turn-helix motifs of DNA-bound SlyA with Zn(II)-bound AdcR has a root mean square deviation value of only 0.5 Å, suggesting that this protein is in a conformation for DNA binding. Thus, binding of an effector in this case induces high affinity DNA binding.

The MarR proteins bind DNA as dimers and display the same characteristic fold despite typically having less than 20% amino acid sequence identity (Nichols *et al.*, 2009). This consists of a winged helix-turn-helix DNA binding motif with a dimerization domain which is mostly helical. It is this helical dimerization domain which is affected by ligand binding (or oxidation of key residues), resulting in movement of the two DNA binding regions such that the protein dissociates from the DNA. The arms of the interface and the wHTH domains were shown to reorientate upon ligand binding in MTH313 (Saridakis *et al.*, 2008). As more structures of protein-ligand complexes become available, so the knowledge database required for interpretation of the mechanisms at work grows, increasing the opportunities to develop novel therapeutics to target these repressors.

## 1.5    MarR homologues in *Streptomyces coelicolor*

There are 42 MarR homologues in *Streptomyces coelicolor* (Figure 1.11) and taking its very complex life cycle into consideration, this is perhaps not overly surprising. However, this presents an interesting challenge to characterize these proteins to uncover their physiological roles as well as attempting to use this information to make some more global assumptions about MarR mechanisms and gene regulation. Much of the work that

has been done so far has focused on OhrR (discussed above) and AbsC, which has been linked to the gene cluster controlling the expression of coelibactin (Hesketh *et al.*, 2009) and specifically the zinc-responsive Zur repressor (Kallifidas *et al.*, 2010). A mutation in AbsC results in the zinc-dependent pleiotropic defect in the production of both actinorhodin and undecylprodigiosin and AbsC and Zur have been shown to tightly regulate coelibactin production by direct interactions with the coelibactin cluster. What is still not known is the identity of the ligand for AbsC. Work is ongoing to determine the identity of this ligand through structural analyses (Stevenson *et al.*, 2007). Sco6704 is the most recent of the 42 MFRs in *S. coelicolor* to be characterized and it has been linked to the β-ketoadipate pathway and is responsive to protocatechuate and *p*-hydroxybenzoate (Davis & Sello, 2010).
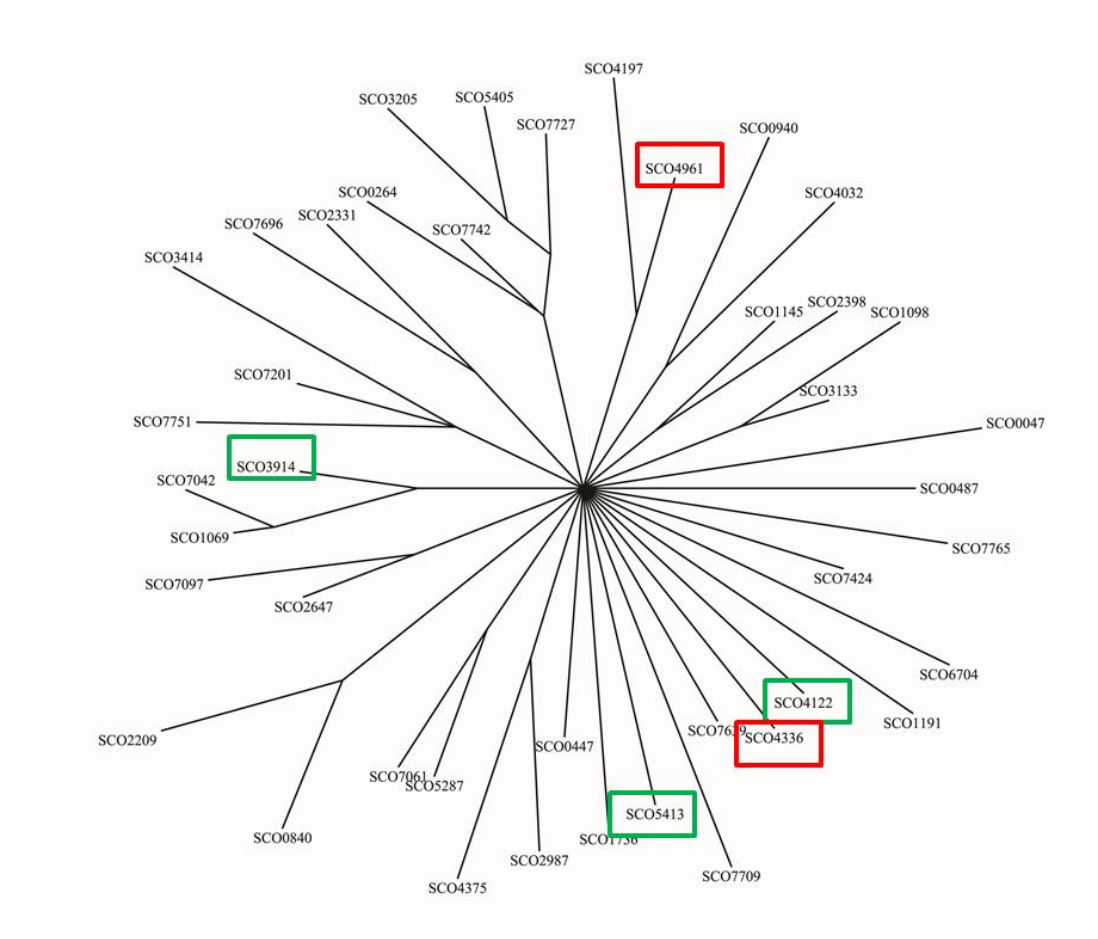


**Figure 1.11** Phylogenetic tree showing the 42 MarR homologues in *Streptomyces coelicolor*. The five which were selected for this study are boxed. The three which feature in this thesis are in green.

It is, of course, possible that each of the MarR proteins could potentially bind a variety of ligands, each of which could elicit a very different response across homologues from a multitude of species. Evidently, there is much scope for research to uncover the common and also the specific mechanisms in operation of each protein. The clinical importance of MarR proteins due to their functions in multidrug resistance is obvious and the continued discovery of the diversity of their functions and mechanisms is increasingly relevant in developing an effective armoury against pathogenic bacteria.

## 1.6    Aims of the project

Despite the growing body of biological information for the MarR family of transcriptional repressors, their mode of action and specific inhibitors remains elusive.  This project aims to perform biophysical analyses of several of the MFRs present in *Streptomyces coelicolor* in an effort to understand their biological function as well as gaining further structural information about MarR family proteins in general. The specific aims of the project included: 1) the generation of deletion mutants with phenotypic analyses of these in an effort to determine biological functions; 2) cloning of the MFR genes into an expression vector and the overexpression of the protein for downstream analyses; 3) crystallization and structural determination of the MFRs and subsequent analyses leading to 4) generation of site-directed mutants for further biophysical and structural analyses; 5) identification of the regulons of each of the MFRs using a variety of methods, including SPR. The information obtained from these objectives will contribute to the understanding of the MarR family of proteins as well as transcription factors in general.

Five of the 42 MFR homologues in *S. coelicolor* were initially selected for the purpose of this study and three of these have featured in this thesis.

# Chapter 2

# Materials and Methods

## 2.1 Molecular Biology Methods

### 2.1.1 Bacterial Strains and Plasmids

**Table 2.1** *E. coli* strains used in this work

| Strain | Genotype | Antibiotic resistance | Reference |
|---|---|---|---|
| DH5α | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* | None | (Sambrook *et al.*, 2001) |
| OmniMAX™ 2 | F′ *pro*AB *lac*I$^q$ *lac*ZΔM15 *T*n10(Tet$^R$ ) Δ*(ccdA*B) *mcr*A Δ*(mrr hs*dRMS-*mc*rBC) Φ 80(*lac*Z)ΔM15 Δ*(la*cZYA-*ar*gF)U169 *en*dA1 *rec*A1 *su*pE44 *th*i-1 *gyr*A96 *re*lA1 *to*nA *pa*nD | None | |
| ET12567/pUZ8002 | *dam13::*Tn*9 dcm6 hsdM hsdR recF143 zjj201::*Tn*10 galK galT22 ara14 lacY1 xyl5 leuB6*  *thi*l *tonA*31 *rpL*136 *his*G4 *tsx*78 *mtl*i *glnV*44 F$^-$ | Cm, Tet | (MacNeil *et al.*, 1992)  (Datsenko and Wanner, 2000) |
| BL21(DE3)pLysS | F–*, ompT, hsdS$_B$ (r$_B$–, m$_B$–), dcm, gal*, λ(DE3) [*lac*I *lacUV5*-T7 gene 1 *ind*1 *sam*7 *nin*5]), pLysS [Cm$^r$] | Cm | (Studier & Moffatt, 1986) |

**Table 2.2** *Streptomyces coelicolor* A3 (2) strains used in this work

| Strain | Genotype | Source/ Reference |
|---|---|---|
| M145 | Wild type | (Kieser *et al.,* 2000) |

| | | | |
|---|---|---|---|
| M145 Δ*sco3914::apr* | M145 Δ*sco3914::apr* | | This work |
| M145 Δ*sco4122::apr* | M145 Δ*sco4122::apr* | | This work |
| M145 Δ*sco4961::apr* | M145 Δ*sco4961::apr* | | This work |
| M145 Δ*sco5413::apr* | M145 Δ*sco54134::apr* | | This work |

**Table 2.3** Plasmids and cosmids used and constructed in this study

| Plasmid | Vector | Insert | Genotype/Use | Antibiotic resistance | Reference/Source |
|---|---|---|---|---|---|
| Supercos 1 | Supercos 1 | None | *Neo, bla* | Carb, Kan | Stratagene |
| pUZ8002 | RK2 | None | RK2 derivative with defective *oriT* | Kan | (Paget *et al.*, 1999) |
| pOPINF | pTriEx2 | | Overexpression vector for N-terminally His-tagged protein (3CProtease cleavable tag) | Carb | (Berrow *et al.*, 2007) |
| pIJ773 | pBluescript | *aac(3)IV*(Apra[R]) + *oriT* | Apramycin resistance gene flanked by FLP recognition sites | Apr, Carb | (Gust *et al.*, 2003) |
| pGem | pGem | None | Cloning vector | Carb | Promega |
| pLysS | pLysS | None | Expression of T7 lysozyme | Cm | (Moffatt & Studier, 1987) |

| pIJ10257 | pMS81 | 330-bp *ermEp\** (*Kpn*I-*Pst*I) with ribosome binding site and multicloning site<br><br>from pIJ8723 | Integrative *Streptomyces* expression vector | Hyg | (Hong *et al.*, 2005b) |

### 2.1.2   Buffer and solutions

| Solution/buffer | Compostion | |
| --- | --- | --- |
| 1 x SDS-PAGE running buffer | Tris base | 25 mM |
| | Glycine | 200 mM |
| | SDS | 0.1 % |
| Cosmid isolation solution I | Glucose | 50 mM |
| | Tris-HCl (pH 8) | 25 mM |
| | EDTA | 10 mM |
| Cosmid isolation solution II | NaOH | 0.2 M |
| | SDS | 1 % (w/v) |
| Cosmid isolation solution III | Sodium acetate (pH 4.8) | 3 M |
| | Acetic acid | |
| *E. coli* cell lysis buffer | Tris-HCl (pH 8) | 25 mM |
| | NaCl | 0.5 M |
| | Imidazole | 40 mM |
| | 0.2 % Tween-20 | 2 ml |
| | DNAse | 0.1 mg ml$^{-1}$ |
| | Complete protease inhibitor tablets (Roche) | According to the manufacturer's recommendations |
| *S. coelicolor* cell lysis buffer | Tris-HCl (pH 8) | 10 mM |
| | NaCl | 50 mM |
| | Lysozyme | 10 mg ml$^{-1}$ |
| | Protease inhibitor | 1 tablet |

| Immunoprecipitation (IP) buffer | Tris-HCl (pH 8) | 100 mM |
|---|---|---|
| | NaCl | 250 mM |
| | Triton X-100 | 0.5 % (w/v) |
| | SDS | 0.1 % (w/v) |
| | Protease inhibitor | 1 tablet |
| IP elution buffer | Tris-HCl pH 7.6 | 50 mM |
| | EDTA | 10 mM |
| | SDS | 1 % (w/v) |
| Buffer A | Tris-HCl (pH 8) | 25 mM |
| | NaCl | 0.5 M |
| | Imidazole | 40 mM |
| Buffer B | Tris-HCl (pH 8) | 25 mM |
| | NaCl | 0.5 M |
| | Imidazole | 500 mM |
| Gel filtration buffer | HEPES pH 7.5 | 50 mM |
| | NaCl | 250 mM |
| Tag cleavage buffer | Tris-HCl (pH 8) | 25 mM |
| | NaCl | 0.5 M |
| | B-mercaptoethanol | 2 mM |
| Lysozyme solution (genomic DNA extractions) | Sucrose | 0.3 M |
| | Tris-HCl (pH 8) | 25 mM |
| | EDTA (pH 8) | 25 mM |
| | Lysozyme (Sigma) | 2 mg ml$^{-1}$ |
| SDS stock solution | 10% (w/v) in $H_2O$, filtered through a 0.45$\mu$m membrane. | |
| TAE buffer | Tris | 40 mM |
| | Acetic acid | 1.142 % |
| | EDTA | 1 mM |
| TBS/Tween (TBST) | Tris-HCl (pH7.6) | 20 mM |
| | NaCl | 137 mM |
| | Tween | 0.1 % (w/v) |
| TES buffer | Tris-HCl (pH 8) | 10 mM |
| | EDTA | 1 mM |
| | NaCl | 1 M |
| Western blot blocking solution | TBS/Tween with the addition of 10 % skimmed milk powder | |
| Western blot transfer buffer (pH 9.9) | $NaHCO_3$ | 10 mM |
| | $Na_2CO_3$ | 3 mM |
| | Methanol | 20 % (w/v) |
| Running buffer for wet transfer | Tris | 25 mM |
| | Glycine | 192 mM |
| | Distilled water | Up to 1000 ml |

### 2.1.3 Solid media

| Medium | Composition | | Instructions for preparation |
|---|---|---|---|
| L-Agar | Agar | 10 g | The ingredients, except agar, were dissolved, in the distilled water and 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks containing 2 g agar. The flasks were closed and autoclaved. |
| | Difco Bacto tryptone | 10 g | |
| | NaCl | 5 g | |
| | Glucose | 1 g | |
| | Distilled water | Up to 1000 ml | |
| Difco nutrient agar (DNA) | Difco Nutrient Agar | 4.6 g | Difco Nutrient Agar was placed in each 250 ml Erlenmeyer flask and distilled water was added. The flasks were closed and autoclaved. |
| | Distilled water | 200 ml | |
| Tryptone Soya Agar (TSA) | Oxoid Tryptone Soya Broth powder | 30 g | The Tryptone Soya Broth powder was dissolved in the distilled water and 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks each containing 2.0 g Lab M agar. The flasks were closed and autoclaved. |
| | Distilled water | Up to 1000 ml | |
| | Lab M  agar | 10 g | |
| Mannitol soya flour medium (SFM) | Agar | 20 g | The mannitol was dissolved in the water and 200 ml aliquots poured into 250 ml Erlenmeyer flasks each containing 2 g agar and 2 g soya flour. The flasks were closed and autoclaved twice (115 °C/15min), with gentle shaking between the two runs. |
| | Mannitol | 20 g | |
| | Soya Flour | 20 g | |
| | Tap water | Up to 1000 ml | |
| Minimal medium for *Streptomyces* (MM) | L-asparagine | 0.5 g | The ingredients except agar were dissolved and the pH adjusted to 7.0 -7.2. 200 ml aliquots were dispensed into 250 ml Erlenmeyer flasks containing 2 g Lab M agar. The flasks were closed and autoclaved. Before use the medium was re-melted and 20 ml 10 % mannitol (or another carbon source) added. |
| | $K_2HPO_4$ | 0.5 g | |
| | $MgSO_4.7H_2O$ | 0.2 g | |
| | $FeSO_4.7H_2O$ | 0.01 g | |
| | Lab M Agar | 10 g | |
| | De-ionised water | Up to 1000 ml | |
| | 10 % Mannitol | 20 ml | |

| Medium | Composition | | Instructions for preparation |
|--------|-------------|---|------------------------------|
| R2 | Sucrose | 103 g | The ingredients were dissolved, in the distilled water and 80 ml aliquots were dispensed into 250 ml Erlenmeyer flasks containing 2.2 g Difco Bacto agar. The flasks were closed and autoclaved. At the time of use, the medium was re-melted and the remaining autoclaved solutions were added to each flask in order listed: |
| | $K_2SO_4$ | 0.25 g | |
| | $MgCl_2.6H_2O$ | 10.12 g | |
| | Glucose | 10 g | |
| | DifcoCasamino acids | 0.1 g | |
| | Difco Bacto agar | 22 g | |
| | Distilled water | Up to 800 ml | |
| | At the time of Use, add:- | | |
| | $KH_2PO_4$ (0.5 %) | 1 ml | |
| | $CaCl_2.2H_2O$ (3.68 %) | 8 ml | |
| | L-proline (20 %) | 1.5 ml | |
| | TES buffer (5.73 %, adjusted to pH 7.2) | 10 ml | |
| | Trace elements solution I | 0.2 ml | |
| | NaOH (1 N) | 0.5 ml | |
| R2YE | | | R2YE was made in the same way as R2 with 5 ml 10 % Difco yeast extract solution added once the medium had been re-melted. |
| R5 | Sucrose | 103 g | The ingredients were dissolved, in the distilled water and 100 ml aliquots were dispensed into 250 ml Erlenmeyer flasks each containing 2.2 g Difco Bacto agar. The flasks were closed and autoclaved. At the time of use, the medium was re-melted and the remaining autoclaved solutions were added to each flask in the order listed: |
| | $K_2SO_4$ | 0.25 g | |
| | $MgCl_2.6H_2O$ | 10.12 g | |
| | Glucose | 10 g | |
| | Difco Casamino acids | 0.1 g | |
| | Trace elements solution I | 2 ml | |
| | Difco yeast extract | 5 g | |
| | TES buffer | 5.73 g | |
| | Difco Bacto agar | 22 g | |
| | Distilled water | Up to 1000 ml | |
| | At time of use, add:- | | |
| | $KH_2PO_4$ (0.5 %) | 1 ml | |
| | $CaCl_2.2H_2O$ (5 M) | 0.4 ml | |
| | L-proline (20 %) | 1.5 ml | |
| | NaOH (1 N) | 0.7 ml | |

| Medium | Composition | | Instructions for preparation |
|---|---|---|---|
| Trace elements solution I (for R2/R2YE/R5) | $ZnCl_2$ | 40 mg | Trace elements solution was filter-sterilised and stored at 277 K. |
| | $FeCl_3.6H_2O$ | 200 mg | |
| | $CuCl_2.2H_2O$ | 10 mg | |
| | $MnCl_2.4H_2O$ | 10 mg | |
| | $Na_2B_4O_7.10H_2O$ | 10 mg | |
| | $(NH_4)_6Mo_7O_{24}.4H_2O.$ | 10 mg | |
| | Distilled water | Up to 1000 ml | |

### 2.1.4 Liquid media

| Medium | Composition | | Instructions for preparation |
|---|---|---|---|
| L (Lennox) broth (LB) | Difco Bacto tryptone | 10 g | The ingredients were dissolved, in the distilled water and autoclaved. |
| | Difco yeast extract | 5 g | |
| | NaCl | 5 g | |
| | Glucose | 1 g | |
| | Distilled water | Up to 1000 ml | |
| 2 X YT medium | Difco Bacto tryptone | 16 g | The ingredients were dissolved, in the distilled water and 10 ml aliquots were dispensed into universals and autoclaved. |
| | Difco yeast extract | 10 g | |
| | NaCl | 5 g | |
| | Distilled water | Up to 1000 ml | |
| SOB medium | Tryptone | 20 g | After dissolving the solutes in water, 10 ml 250 mM KCl was added and the pH was adjusted to pH 7 with 5 N NaOH. The volume was then made up to 1000 ml with deionised water and autoclaved. |
| | Yeast extract | 5 g | |
| | NaCl | 0.5 g | |
| | Distilled water | Up to 1000 ml | |
| SOC medium | | | SOC medium is identical to SOB medium except that after autoclaving, 20 ml of sterile 1 M solution of glucose and 5 ml of sterile 2 M $MgCl_2$ were added. |
| M9 Minimal media | $Na_2HPO_4,$ | 6 g | The ingredients were dissolved, in the distilled water and autoclaved. |
| | $KH_2PO_4,$ | 3 g | |
| | NaCl | 0.5 g | |
| | $NH_4Cl$ | 1 g | |
| TSB (Tryptone soya broth) | Oxoid Tryptone Soya Broth powder | 30 g | The ingredients were dissolved, in the distilled water and |

| | | | autoclaved. |
|---|---|---|---|
| | Distilled water | Up to 1000 ml | |
| NMMP (Minimal liquid medium) | $(NH_4)2SO_4$ | 2 g | 80 ml aliquots were dispensed and autoclaved. |
| | Difco Casaminoacids | 5 g | |
| | $MgSO_4.7H_2O$ | 0.6 g | |
| | PEG 6000 | 50 g | |
| | Minor elements solution | 1 ml | |
| | Distilled water | 800 ml | |
| | At time of use, add:- $NaH_2PO_4/K_2HPO_4$ buffer (0.1M, pH6.8) | 15 ml | |
| | Carbon source (20%) | 2.5 ml | |
| YEME (Yeast extract-malt extract medium) | Difco Bacto-peptone | 3 g | The ingredients were dissolved, in the distilled water and autoclaved. |
| | Difco yeast extract | 5 g | |
| | Oxoid malt extract | 5 g | |
| | Glucose | 10 g | |
| | Sucrose | 340 g | |
| | Distilled water | To 1000 ml | |
| | At time of use, add:- $MgCl_2.6H_2O$ (2.5 M) | 2 ml | |

## 2.1.5   Antibiotics

| Antibiotic | Solvent | Concentration in media (µg/ml) |
|---|---|---|
| Carbenicillin (Carb) | Water | 100 |
| Kanamycin (Kan) | Water | 50 |
| Chloramphenicol (Cm) | Ethanol (100%) | 25 |
| Apramycin (Apr) | Water | 50 |
| Nalidixic acid (Nal) | Water | 20 |

### 2.1.6    Preparation of electro-competent cells and transformation of DNA

A single colony was used to inoculate 10 ml LB and grown overnight with shaking at 310 K.  100 µl of this preculture was inoculated into 10 ml SOB and grown at 310 K (303 K for BW25113/pIJ790) for 3-4 h with shaking at 200 rpm until an $OD_{600}$ of ~0.6 was attained. The cells were recovered by centrifugation at 3000 x g for 5 min at 277 K in a Sorvall GS3 rotor. After decanting the medium the pellet was resuspended by gentle mixing in 10 ml ice-cold 10% glycerol.  The cells were centrifuged as before and washed in 10% glycerol a further two times. After decanting the supernatant, cells were resuspended in the remaining ~ 100 µl of 10% (v/v) glycerol.

For each transformation, 50 µl electro-competent cell suspension was mixed with ~ 100 ng DNA. Electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a GenePulser II (Bio-Rad) set to: 200 Ω, 25 µF and 2.5 kV. The expected time constant is 4.5 – 4.9 ms. After electroporation, 1 ml ice cold LB was added immediately to the shocked cells and incubated with shaking (200 rpm) for 1 h at 310 K (303 K for BW25113/pIJ790 transformed with cosmid). Transformants were selected by spreading onto LB agar containing the appropriate antibiotic.

### 2.1.7    Preparation of chemically competent cells and transformation of DNA

Strains of *E. coli* were made competent for the uptake of plasmid DNA using a modified version of the calcium chloride method (Mandel & Higa, 1970).  A single bacterial colony was grown in 10 ml of LB at 310 K to early log phase (OD 600 nm approximately 0.6) and then incubated on ice for 20 min.  The cells were centrifuged for 10 min at 3,000 x g and the supernatant discarded.  The cells were then re-suspended in 5 ml of ice cold 100 mM $CaCl_2$ and left on ice for a further 15 min.  The cells were centrifuged again and re-suspended in 1 ml of 100 mM $CaCl_2$, 15% glycerol and left on ice for 15 min. Cells were stored in 200 µl aliquots at 193 K.

For transformation, 1 µl of the appropriate DNA miniprep plasmid solution, typically containing 10-20 µg of DNA, was added to a 200 µl aliquot of the competent cells and left on ice for 30 min. The cells were heat shocked for 90 seconds at 315 K and placed back on ice for a few minutes prior to incubating with 1 ml of LB at 310 K for approximately 1 h before plating approximately 200 µl, onto solid LB agar petri plates containing the appropriate antibiotic selection. (For the pOPINF transformants following each InFusion reaction, cells were pelleted by centrifugation at 13000 rpm and the entire pellet resuspended in 150 µl LB as there was a very low rate of successful transformants). The plates were then incubated at 310 K overnight.

### 2.1.8    Plasmid isolation from *E. coli*

Qiagen miniprep kits were used according to the manufacturer's instructions (and the buffer components are also described in the instructions). A 10 ml overnight LB culture grown from a single colony harbouring the plasmid of interest was harvested by centrifugation at 13000 rpm for 10 min. The cell pellet was resuspended in buffer P1 and underwent alkaline lysis (buffer P2). The lysate was then neutralised and adjusted to a high salt condition by the addition of buffer N3. The tube was centrifuged in a microcentrifuge at 13000 rpm to remove cell debris and precipitated protein. The supernatant was then applied to a silica membrane mounted in a microcentrifuge tube where it was washed under high salt and ethanolic buffer conditions. The DNA was eluted from the column in ultrapure water or elution buffer. Plasmid DNA was stored at 253 K.

### 2.1.9    DNA electrophoresis on agarose

DNA samples were separated using 1% agarose gels which were prepared by dissolving 1 g agarose in 100 ml TAE (40 mM Tris-acetate, 2 mM EDTA) buffer by heating for 2 minutes in a microwave. The agarose was allowed to cool until hand warm and then 1 µg ml$^{-1}$ of ethidium bromide was added. The gel was poured into a horizontal gel running apparatus, containing a comb for well formation and allowed to set. The DNA sample was

mixed with loading dye prior to loading into a well. A commercially available 1 kb DNA ladder was also loaded into a well as a reference. The gel was run in TAE buffer at 100 V and the DNA was visualised by UV radiation.

### 2.1.10 Restriction digestion of DNA

Restriction enzyme digestion of cosmids, plasmids or genomic DNA was carried out according to the enzyme manufacturer's instructions. In the case of double digests, an appropriate buffer was selected after consulting the manufacturer's literature (Roche or Invitrogen). The reaction volume was usually 20 µl for analytical digests and 50 µl for preparative digests. Unless otherwise instructed, digests were typically carried out for 1 h at 310 K and contained the following reaction components:-

Total volume – 20 µL

| | |
|---|---|
| dH$_2$O | 12 µL |
| Buffer | 2 µL |
| DNA | 5 µL |
| Enzyme | 1 µL |

### 2.1.11 DNA extraction from agarose gels

DNA fragments separated in agarose gels were excised from the gel using a clean scalpel and purified using the Qiaquick™ gel extraction kit, following the manufacturer's instructions. Briefly, the agarose gel slice containing the DNA fragment of interest was dissolved in a neutral pH, high salt buffer provided with the kit and applied to a silica gel membrane mounted in a microcentrifuge tube. The column was washed and the DNA fragment was eluted in water or elution buffer.

### 2.1.12 Ligation of DNA

Fragments to be ligated were purified from solution using a Qiagen PCR purification kit or from an agarose gel using a Qiagen gel extraction kit. The purifications were carried out according to the manufacturer's instructions and the DNA was eluted in water. A ligation

reaction volume of 10 µl was carried out with an insert: vector molar ratio of 3:1. Ligations were carried out overnight at 289 K using high concentration T4 DNA ligase (Promega).

### 2.1.13  DNA Sequencing

DNA was sequenced by The Genome Analysis Centre (TGAC) in Norwich. Plasmid samples were submitted to TGAC in water at a concentration of ~200 ng µl$^{-1}$ and the sequencing was performed using a T7 forward promoter primer (supplied by TGAC) or gene specific primers (supplied with sample). Results were analysed using Vector NTI® (Invitrogen) or Chromas Lite (Technelysium).

### 2.1.14  Growth and storage of *E. coli*

*E. coli* was grown at 310 K overnight on L-agar or shaking in LB broth (303 K for BW25113/pIJ790).  For selection of plasmid-containing cells, the appropriate antibiotics were added at the appropriate concentrations. For long-term storage, overnight LB cultures of *E. coli* strains were mixed with an equal volume of 40% (v/v) glycerol and stored at 193 K.

### 2.1.15  General PCR protocol

PCR from *Streptomyces coelicolor* M145 genomic DNA was typically carried out using 10 ng DNA as template (genomic DNA prepared according to section 2.2.4). PCR from plasmid DNA was typically carried out using 0.5 µl of a standard plasmid preparation as template.

PCR mix for 50 µl volume:-

| | |
|---|---|
| 5% (v/v) DMSO | 2.5 µl |
| 10x PCR reaction buffer | 5 µl |
| 5 mM dNTP mix | 2 µl |

| | |
|---|---|
| For primer (100 pmol μl$^{-1}$) | 1 μl |
| Rev primer (100 pmol μl$^{-1}$) | 1 μl |
| Template (50 – 200 ng μl$^{-1}$) | 1 μl (use 2 to 3 x this for genomic DNA) |
| Taq polymerase | 1 μl |
| Water | 33 μl |

General purpose PCR programme:-

1.    95°C for 4 min
2.    95 °C for 45 sec
3.    55°C for 45 sec
4.    72°C for 1 min 30 sec
5.    Go to step 2 26 more times
6.    72°C for 7 min
7.    END

Times were adjusted according to product lengths and the annealing temperature in step 3 was optimized for the specific set of primers.

### 2.1.16  Colony PCR

Colony PCR was used for the rapid screening of recombinant plasmids from *E. coli* colonies during cloning. The PCR was carried out as described for general PCR, except that template DNA was substituted with *E. coli* cells and the initial denaturation cycle was extended for an additional 1 minute.

### 2.1.17  Site-directed mutagenesis methods

Initially, the site-directed mutagenesis by PCR methods was attempted. This approach utilized the QuikChange Lightning Kit (Stratagene). The appropriate oligos were designed

and ordered (Invitrogen) and were diluted with water to produce a 125 ng µl$^{-1}$ solution. PCR reactions were set up, one for each mutation, as follows:-

| | |
|---|---|
| 1 µl | Pfu turbo 2.5 U µl$^{-1}$ (Stratagene) |
| 5 µl | 10 x Pfu turbo buffer |
| 1 µl | (2 mM) dNTP mix |
| 1.25 µl | Forward primer (125 ng) |
| 1.25 µl | Reverse primer (125 ng) |
| 1 µl | Template DNA |

The template DNA was pOPINF*sco3914* at 90 ng µl$^{-1}$.   The PCR reactions were run using the PCR program below:-

1.    95°C for 2 min
2.    95°C for 30 sec
3.    60°C for 30 sec
4.    68°C for 10 min
5.    Go to step 2 18 more times
6.    68°C for 10 min
7.    END

Many variations of this PCR program and of the reaction components were used during the multiple attempts to obtain a mutated plasmid product. Unfortunately, this method did not work and so the mutants were synthesized by GenScript and subcloned into the pOPINF expression vector.

### 2.1.18 Electrophoretic mobility shift assay (EMSA)

The EMSA DNA probes spanning the entire *sco4961-4962* and *sco5413-5414* intergenic regions were amplified by PCR using primers and then 5'-end labelled using [γ$^{32}$-P] ATP and T4 polynucleotide kinase (New England Biolabs). Binding of Sco5413 to DNA was

carried out in 20 µl EMSA Buffer [20 mM Tris, pH 8.0, 1 µg poly(dI-dC), 1 mM EDTA, 100 mM NaCl, 0.5 mM DTT, 8% (v/v) glycerol] containing 0.1 nM radiolabelled DNA (approximately 8000 cpm) and varying amounts of Sco5413. After incubation at 303 K for 5 min, the binding reaction mixtures were loaded on 5% (w/v) native polyacrylamide gels and run in TBE buffer at 120 V for 45 min. EMSA data were collected and analysed on a PhosphoImager (FujiFilm) using Multi Gauge image analysis software (FujiFilm).

Non-denaturing 5 % polyacrylamide gel:-

1.5 ml 40% acrylamide (37.5:1)
1.2 ml 10xTBE
9.3 ml dH$_2$O
15 µl TEMED
150 µl 10% (w/v) APS

### 2.1.19 Chromatin Immunoprecipitation on Chip (ChIP on Chip)

Wild type M145 and M145Δ*sco5413* spores were pregerminated as described in section 2.2.1 and used to inoculate 4 x 50 ml (one for each timepoint) 50:50 TSB/YEME. After 16 hours, formaldehyde was added to a final concentration of 1% (v/v) and the culture incubated for a further 30 minutes. The addition of formaldehyde induces cross-linking of any proteins that are attached to the genomic DNA. Glycine was added to a final concentration of 125 mM to stop the reaction and incubation continued for a further 5 minutes at room temperature. The cells were then harvested by centrifugation at 7000 rpm for 5 minutes at 277 K and then washed twice with 25 ml of ice-cold PBS buffer. This was repeated for the samples taken after 90 hours of growth.

For the immunoprecipitation of the protein, the cells were resuspended in 750 µl of lysis buffer and incubated for 25 minutes at 298 K in a water bath. The samples were then placed on ice and 750 µl of IP buffer was added and incubated for a further 2 minutes. The slurry was transferred to a 2 ml tube and sonicated 8 times for 15 seconds at 50%

power. The tubes were incubated on ice between each sonication cycle. This step was performed to shear the genomic DNA into ~ 500 bp fragments. These fragments were treated twice with phenol/chloroform before checking on agarose gel. The samples were then centrifuged at 14000 rpm for 10 minutes at 277 K to remove debris and the supernatant transferred to a new tube and the centrifugation repeated. 725 µl and 25 µl of the lysate were transferred to a new tube for the 'IP' and 'total DNA' samples, respectively.

The extract was pre-cleared using ~1/10th vol (75 µl) of an equilibrated 50% protein A-Sepharose slurry. This was incubated for 1 hr at 277 K on a rotating wheel and then centrifuged for 15 min at full speed at 277 K and 700 µl of supernatant was transferred to a new tube. 1/10th-volume of the antibody stock (50-75 µl) was added and incubated overnight at 277 K on a rotating wheel. 1/10th vol (75µl) of an equilibrated 50% protein A-Sepharose slurry was added to recover antibody-protein-DNA complexes. This was incubated for 4 hr at 277 K on a rotating wheel and then centrifuged for 5 min at 3500 rpm and washed twice for 15 minutes with gentle agitation with 1 ml 0.5 x IP buffer, and then twice with 1 x IP buffer. 2 x 500 µl was transferred to a fresh tube after the first wash to remove DNA/protein absorbed to the side of the tube.

To elute the protein, 150 µl of IP elution buffer was added to the beads and to 10 µl of the "total DNA" sample set aside before IP. This was mixed and incubated at 338 K overnight. The following morning, the tubes were inverted several times. The beads were removed by spinning at maximum speed for 5 minutes. The supernatant was removed and the remainder of the DNA "re-extracted" with ~ 50 µl TE and incubation at 338 K for 5 minutes. 2 µl of proteinase K 10 mg ml$^{-1}$ was added and incubated for 1.5 hrs at 328 K. The DNA was extracted using phenol/chloroform, extracting the organic phase with TE buffer, and purified using the QiaQuick kit (following the manufacturer's instructions). The concentration of the DNA was determined using the using the ND-1000 spectrophotometer (NanoDrop®). The DNA labelling and hybridization to the DNA

microarrays were performed by Oxford Gene Technology (OGT). The data were analyzed using the OGT ChIP Browser.

## 2.2 *Streptomyces coelicolor* cultivation methods and genetic manipulation

### 2.2.1 Growth and storage of *S. coelicolor*

For the generation of spore stocks, *S. coelicolor* spores were streaked out onto mannitol soya flour solid medium (SFM). The plates were incubated at 303 K for up to 7 days and spores were harvested according to the protocol described in (Kieser, 2000). Spore stocks were stored in 20 % (w/v) glycerol at 253 K or 193 K.

For liquid culturing of *S. coelicolor*, 100 µl (~$10^8$) spores were heat shocked at 323 K for 10 min in 5 ml TES buffer. An equal volume of 2 x YT medium was added and this was incubated at 310 K for 5-6 h. The small, emerging germ tubes were microscopically visible at this stage. The germinated spores were harvested by centrifugation at 3000 x g on a benchtop centrifuge and then resuspended in 500 µl of TES buffer and vortexed to disperse the aggregating germlings. TSB liquid medium was inoculated to $OD_{450}$ of 0.03-0.05. Cultures were grown with shaking at 303 K until late stationary phase. Mycelia were harvested by centrifugation at 3000 x g in a benchtop centrifuge.

### 2.2.2 Cosmid isolation from *E. coli*

A 10 ml overnight culture of *E. coli* containing the cosmid of interest was harvested by centrifugation and the pellet resuspended by vortexing in 100 µl solution I (50 mM Tris-HCl, pH 8; 10 mM EDTA). Cosmid isolation from *E. coli* was carried out by alkaline lysis as described by Sambrook & Russell, (2001). The cell pellet from 1.5 ml of culture was resuspended by vortexing in 100 µl solution I (50 mM Tris-HCl, pH 8; 10 mM EDTA). 200 µl solution II (200 mM NaOH; 1 % SDS) was added immediately and the tube inverted ten times. 150 µl solution III (3 M potassium acetate, pH 5.5) was then added and mixed by inverting the tube five times. The tube was then centrifuged at 16000 x g in a

microcentrifuge for 5 minutes at room temperature. The supernatant was mixed with 400 µl phenol/chloroform, vortexed briefly to mix and then centrifuged at 16000 x g in a microcentrifuge for 5 minutes. The upper phase was then transferred to a 1.5 ml tube, 600 µl of ice cold isopropanol was added and DNA precipitation was achieved by placing the tube on ice for 10 minutes followed by centrifuging at 16000 x g in a micro centrifuge for 5 minutes. The pellet was washed with 200 µl 70% ethanol and centrifuged at 16000 x g in a microcentrifuge for 5 minutes. The tube was left open for 5 minutes at room temperature to allow the pellet to dry prior to resuspending in 50 µl 10 mM Tris-HCl (pH 8).

### 2.2.3   Genomic DNA extractions from *S. coelicolor*

2 µl of *S. coelicolor* spore stock was used to inoculate 10 ml SOC and grown overnight with shaking at 303 K. The mycelia were recovered by centrifugation at 3000 x g for 5 min at 277 K in a Sorvall GS3 rotor. The mycelium was resuspended in 500 µl of lysozyme solution and then incubated at 310 K for 60 min. 50 µl of 10 % (w/v) SDS was then added and incubation was continued at 310 K for a further 15 minutes. The sample was then vortexed until the viscosity of the solution had decreased. 300 µl phenol/chloroform/isoamyl alcohol was added and vortexed briefly until completely mixed. The sample was centrifuged in a microcentrifuge on 16000 x g for 8 minutes. Following centrifugation, the aqueous phase was removed to a fresh Eppendorf tube. To decrease viscosity, 200 µl water was added to the supernatant. The supernatant was extracted twice with phenol/chloroform/isoamyl alcohol. Precipitation of the DNA was achieved by the addition of 0.1 volumes of 3 M sodium acetate and 1 volume of ice-cold isopropanol and incubating on ice for 10 minutes. The precipitated DNA was then recovered by centrifugation in a microcentrifuge at 16000 x g for 5 minutes. The DNA pellet was then washed with 70% (v/v) ethanol before dissolving in 500 µl elution buffer at room temperature for 2 hours.  DNAse free RNAse was then added to a final concentration of 40 µg ml$^{-1}$ and the sample was incubated at 310 K for 30 minutes. The DNA sample then underwent two phenol/chloroform/isoamyl alcohol extractions.  The DNA was precipitated and washed as before and was then dissolved in 100 µl elution

buffer. The DNA concentration was determined by spectrophotometry using the Nanodrop.

### 2.2.4 Growth curves

For the generation of growth curves, wild type M145 and M145Δ*sco5413* spores were pregerminated and used to inoculate 3 x 50 ml flasks of minimal media supplemented with 10 mM N-acetyl glucosamine. These were incubated at 303 K with shaking at 200 rpm. 1 ml samples (triplicates) were collected at 6 hour intervals for 72 hours. The samples were stored at 253 K and processed at the same time.

To assess the cell density, the cells were defrosted and the optical density at 450 nm was measured and plotted. To assess the total protein content of the cells, the pellets were resuspended in 200 µl 1M NaOH and boiled at 373 K for 5 minutes. They were vortexed and boiled again for a further 5 minutes before centrifuging at 13000 rpm for 6 minutes to remove cell debris. A standard curve was generated to calculate the protein concentrations in the experimental samples using bovine serum albumin (BSA) standards of 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4 ng/ml which were prepared in cuvettes in 0.5 ml volumes.  The *DC* protein assay (Bio-Rad) was used according to the manufacturer's instructions, with an equivalent amount of protein buffer in each BSA standard, sample and blank.

## 2.3 REDIRECT™ technology – PCR targeting to generate deletion mutants

### 2.3.1 Amplification of disruption cassette by PCR

The pIJ773 template was amplified using gene specific primers for each deletion mutant. These were designed in accordance with the criteria specified by (Gust *et al.*, 2003) and are shown in Table 2.4. The cassette is flanked by the gene-specific 39 nt homology extensions and the inclusion of *oriT* (RK2) allows conjugal transfer of the PCR targeted cosmid DNA into *S. coelicolor* (Figure 2.1).
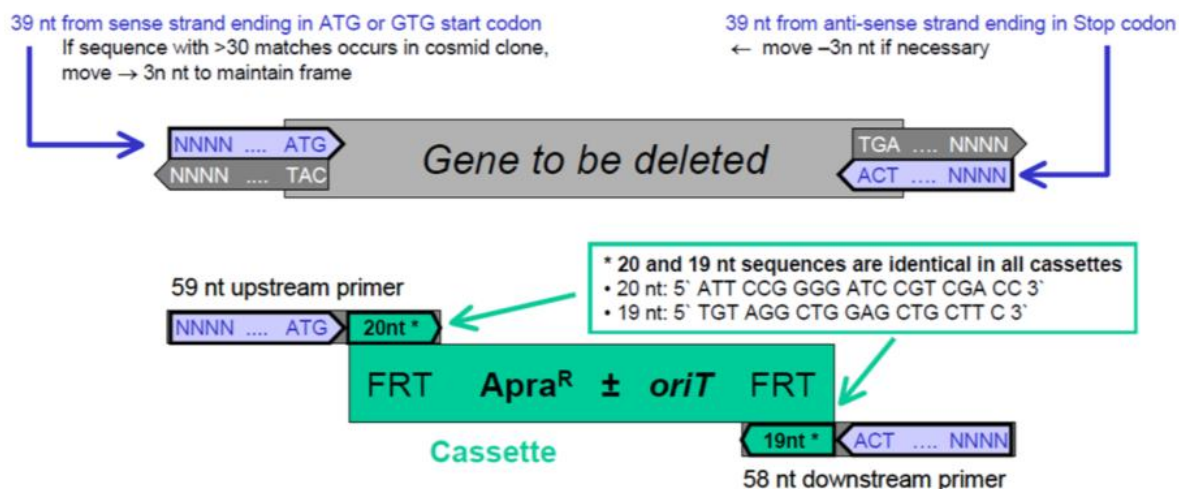
**Figure 2.1** Designing long PCR primers to amplify the pIJ773 disruption cassette to make a gene deletion (Gust *et al.*, 2003).

All PCR amplifications were performed using the Expand high fidelity PCR system according to the manufacturer's instructions (Roche). Reaction components:-

| | | |
|---|---|---|
| Primers (100 pmoles µl⁻¹) | 0.5 µl each | 50 pmoles each |
| Template DNA (100 ng µl⁻¹) | 0.5 µl | 50 ng ~ 0.06 pmoles |
| Buffer (10x) | 5 µl | 1 x |
| dNTPs (10 mM) | 1 µl each | 50 µM each |
| DMSO  (100%) | 2.5 µl | 5% |
| DNA polymerase (2.5 U µl⁻¹) | 1 µl | 2.5 Units |
| Water | 36 µl | |
| Total volume | 50 µl | |

PCR program:-

| | | |
|---|---|---|
| 1. Denaturation: | 94°C, 2 min | |
| 2. Denaturation: | 94°C, 45 s | |
| 3. Primer annealing: | 50°C, 45 s | 10 cycles |
| 4. Extension: | 72°C, 90 s | |

5. Denaturation:      94°C, 45 s
6. Primer annealing:  55°C, 45 s      15 cycles
7. Extension:         72°C, 90 s
8. Final extension:    72°C, 5 min

5 µl of the PCR product was used for analysis by agarose gel electrophoresis. To remove enzymes and unincorporated primers and dNTPs, the remaining 45 µl of the PCR product was purified using the Qiagen PCR purification kit according to the manufacturer's instructions. The PCR product was eluted from the column with 20 µl of water to give a DNA concentration of approximately 100 ng $\mu l^{-1}$.

**Table 2.4** Primers used to amplify the pIJ773 disruption cassette. The sequence which is complementary to the disruption cassette is shown in red.

| Oligo name | Sequence |
|---|---|
| **Sco3914 For** | ACCCTGGTCTCAGCCACTCCTGGACGGAGGATCCACATGATTCCGGGGATCCGTCGACC |
| **Sco3914 Rev** | TAGAACGGGCCGACGCCAGGTTGCCCGGACAGGCTCTATGTAGGCTGGAGCTG CTTC |
| **Sco4122 Forward** | CTTCAAAGCAAAGGAGTTGCGTGCCATACTGCGGCCATGATTCCGGGGATCCGTCGACC |
| **Sco4122 Reverse** | CCCGCAGAGGGCTGTCAAGCGTCGCTGTCAAGATCTTTATGTAGGCTGGAGCTGCTTC |
| **Sco4336 Forward** | GGTATTTCGCCAGACTAATGAGTTACGCTAAGGAACATGATTCCGGGGATCCGTCGACC |
| **Sco4336 Reverse** | CGTACTCAAAAGGGTTCGCCTCCTTGCGAGAACTGCTTATGTAGGCTGGAGCTGCTTC |
| **Sco4961 Forward** | GAGCCCCCGGCCATGCCCACAACACCTGAAATGTCGATGATTCCGGGGATCCGTCGACC |
| **Sco4961 Reverse** | AGCCCGGGAAGGCTCCGGAAATGCGGGAAGGGTCGTTCATGTAGGCTGGAGCTGCTTC |
| **Sco5413 Forward** | TGACCGGCATGTGACCGGCCGGTAAGGTCTCTTGCCGTGATTCCGGGGATCCGTCGACC |
| **Sco5413 Reverse** | GTCGGGCGACTCGCCGGGCCCGCCCGGCCCGCGCGCTCATGTAGGCTGGAGCTGCTTC |

**Table 2.5** Primers used to verify positive transformants using PCR (Checking primers).

| | | |
|---|---|---|
| **3914 Forward** | CGTTCCGATCCCGAAGGCCC | 172 bp upstream of gene |
| **3914 Reverse** | ACCCTGAGGCGTTCGAAGGC | 170 bp downstream of gene |
| **4122 Forward** | CCCACGTGGATCCGGCGCA | 177 bp upstream of gene |
| **4122 Reverse** | GTTACGGTCCCGTGGCAA | 73 bp downstream of gene |
| **4336 Forward** | CGGGTGAACTGCGGGTTT | 93 bp upstream of gene |
| **4336 Reverse** | CGTTCTGCGAGGGGCGGCGG | 129 bp downstream of gene |
| **4961 Forward** | CGCCGTGCGCCGTTGTGAC | 144 bp upstream of gene |
| **4961 Reverse** | TGCGAACCCCGTTCCGGAAA | 105 bp downstream of gene |
| **5413 Forward** | GTGTGACGGACGGTCTCACG | 135 bp upstream of gene |
| **5413 Reverse** | GGTGACGTACGCCATGACCC | 126 bp downstream of gene |

**Table 2.6** Primers used to amplify disruption cassette (for verifying)

| | | |
|---|---|---|
| **pIJ773_P1** | ATTCCGGGGATCCGTCGACC | Disruption cassette For |
| **pIJ773_P2** | TGTAGGCTGGAGCTGCTTC | Disruption cassette Rev |

## 2.3.2  Introduction of *S. coelicolor* cosmids into *E. coli* BW25113/pIJ790 by electroporation

*E. coli* BW25113 is a λ-RED recombination-proficient host strain and contains pIJ790, which carries the antibiotic resistance marker *cat* (chloramphenicol resistance) and a temperature sensitive origin of replication (replicates at 303 K, not at 310 K). BW25113/pIJ790 was grown overnight at 303 K in 10 ml LB containing chloramphenicol (25 μg ml$^{-1}$). 100 μl of the overnight culture were inoculated into 10 ml SOB containing chloramphenicol (25 μg ml$^{-1}$) and grown for 3-4 h at 303 K with shaking at 200 rpm to an OD$_{600}$ of ~0.6. Electrocompetent cells were generated from this culture and transformed with 100 ng cosmid DNA by electroporation. Transformants were selected by spreading onto L-agar containing carbenicillin (100 μg ml$^{-1}$), kanamycin (50 μg ml$^{-1}$) and chloramphenicol (25 μg ml$^{-1}$) and incubating overnight at 303 K.

### 2.3.3 PCR targeting of the *S. coelicolor* cosmid

10 ml SOB (without $MgSO_4$) containing carbenicillin (100 µg $ml^{-1}$), kanamycin (50 µg $ml^{-1}$) and chloramphenicol (25 µg $ml^{-1}$) were inoculated at a concentration of 1% with an overnight culture of *E. coli* BW25113/pIJ790 containing the cosmid of interest. 100 µl 1 M L-arabinose was added to a final concentration of 10 mM to induce the λ RED recombination system. The culture was grown for 3-4 h at 303 K shaking at 200 rpm to an $OD_{600}$ of ~0.6 and electrocompetent cells were prepared. 50 µl cell suspension was then mixed with ~100 ng (1-2 µl) of PCR product and electroporated. Selection for gene disruption in the cosmid of interest was carried out on LB agar containing carbenicillin (100 µg $ml^{-1}$), kanamycin (50 µg $ml^{-1}$) and apramycin (50 µg $ml^{-1}$) overnight at 310 K.

**Table 2.7** Cosmids used for the generation of the deletion mutants by PCR targeting

| Target gene | *S. coelicolor* cosmid |
|---|---|
| *sco3914* | StQ11 |
| *Sco4122* | StD72A |
| *Sco4336* | StD19 |
| *Sco4961* | StK13 |
| *Sco5413* | St8F4 |

### 2.3.4 Confirmation of gene disruption by restriction digest and PCR

The gene disruption was confirmed by restriction analysis. The reaction volume was usually 20 µl for analytical digests. Cosmid DNA of pIJ773 transformants was isolated from 10 ml 310 K overnight LB cultures, containing carbenicillin (100 µg $ml^{-1}$), kanamycin (50 µg $ml^{-1}$) and apramycin (50 µg $ml^{-1}$). Alkaline lysis followed by phenol/chloroform extraction produced DNA suitable for restriction analysis. Restriction digests were analysed by agarose gel electrophoresis.

The positive transformants were also verified by PCR using the checking primers (Table 2.5) which anneal 100-200 bp upstream and downstream of the recombination region. These PCR products were analysed by agarose gel electrophoresis.

### 2.3.5    Transfer of mutant cosmids into *S. coelicolor*

As *S. coelicolor* carries a methyl-specific restriction system it was necessary to passage the targeted cosmids containing an apramycin resistance-*oriT* cassette through a non-methylating *E. coli* host. To achieve this, the cosmids were introduced by transformation into the non-methylating *E. coli* ET12567 containing the RP4 derivative pUZ8002. The cosmid was then transferred to M145 by conjugation. 10 µl of an M145 spore stock was added to 500 µl of 2xYT broth and heat-shocked at 323 K for 10 minutes. These were then mixed with 500 µl of the *E. coli* ET12567 cell suspension and centrifuged briefly. Most of the supernatant was removed and the cells resuspended in the remaining ~50 µl. A dilution series ($10^{-1}$ - $10^{-4}$) was prepared from this in 100 µl water and plated on SFM agar supplemented with 10 mM MgCl$_2$ and no antibiotics. These were incubated at 303 K for 16-20 hours after which time each plate overlaid with 1 ml of water containing 0.5 mg nalidixic acid and 1.25 mg apramycin. Incubation was continued for a further 24 hours. Exconjugants were subsequently selected using the appropriate antibiotic. Replica-plating of each SFM agar plate with single colonies onto DNA plates containing nalidixic acid (25 µg ml$^{-1}$) and apramycin (50 µg ml$^{-1}$) with and without kanamycin (50 µg ml$^{-1}$) allowed double crossover exconjugants (kanamycin$^S$ and apramycin$^R$) to be identified. Kanamycin$^S$ clones were picked from the DNA plates and streaked for single colonies on SFM agar containing nalidixic acid (25 µg ml$^{-1}$) and apramycin (50 µg ml$^{-1}$). The kanamycin sensitivity was confirmed by replica plating onto DNA plates containing nalidixic acid (25 µg ml$^{-1}$) with and without kanamycin (50 µg ml$^{-1}$). Purified kanamycin sensitive strains were then verified by PCR.

### 2.3.6    Phenotypic analysis of deletion mutants

All of the deletion mutants generated in this work were phenotypically characterized on solid media. Particular attention was paid to the production of pigmented antibiotic (actinorhodin and/or undecylprodigiosin), sporulation and the relative sizes of the colonies. Spore stocks were streaked for single colonies on to the following media: SFM,

R5, DNA, minimal media (MM) and L agar, all of which were supplemented with different carbon sources (1% glucose/1% glycerol/1% mannitol/10 mM N-acetyl glucosamine). Plates were incubated for seven days and the growing bacteria were monitored daily (by visual inspection) for phenotypic differences between the deletion mutant strains and the wild type strain M145. The results of this analysis are summarised in the relevant chapters.

## 2.4    Protein preparation methods

### 2.4.1    Protein overexpression

The five selected MarR homologue genes were amplified from *Streptomyces coelicolor* M145 genomic DNA using the primers shown in Table 2.8 and the PCR program detailed in section 2.1.15. These fragments were checked and purified from 1% agarose and then cloned into the pOPINF vector (with an N-terminal 3C Precision protease cleavable hexahistidine tag) which had been digested with HindIII and KpnI using the In-Fusion™ PCR Cloning Kit (Berrow *et al.*, 2007). This method, which uses PCR fragments with 5' ends which are complementary to the vector sequence, is outlined in Figure 2.2. The In-Fusion Dry-Down mix was incubated for 30 minutes at 315 K and then immediately diluted with 20 µl TE buffer (10 mM Tris-HCl pH 8 and 1 mM EDTA pH 8). The reaction can then be frozen or used to transform immediately.

**Table 2.8** Oligos used to amplify sequences from *Streptomyces coelicolor* genomic DNA. The sequences which are homologous to the pOPINF vector are in red; the gene specific sequences are in black.

| Oligo name | Sequence |
|---|---|
| **Sco3914 Forward** | AAGTTCTGTTTCAGGGCCCGACAGCGACGGACCCCGCG |
| **Sco3914 Reverse** | ATGGTCTAGAAAGCTTTAGGCGGGCACCGGTGCC |
| **Sco4122 Forward** | AAGTTCTGTTTCAGGGCCCGGGCGACACCTCCGGCGCC |
| **Sco4122 Reverse** | ATGGTCTAGAAAGCTTTAATCGAGGTCGCTGAGCCGGCC |
| **Sco4336 Forward** | AAGTTCTGTTTCAGGGCCCGCCGGACCTCAAGCATGGCGACG |
| **Sco4336 Reverse** | ATGGTCTAGAAAGCTTTACTTGTGTGCGAGCTTCTCCAGCACGG |
| **Sco4961 Forward** | AAGTTCTGTTTCAGGGCCCGGACATGACGACCGCCGGTGACACC |
| **Sco4961 Reverse** | ATGGTCTAGAAAGCTTTAGGAGCCCGAGGGGGCG |

| Sco5413 Forward | AAGTTCTGTTTCAGGGCCCGCCGAAGCCCCTCAGTCTCTCCTTCG |
|---|---|
| Sco5413 Reverse | ATGGTCTAGAAAGCTTTAGTCCTCGTCGAAGTCGCCCGC |
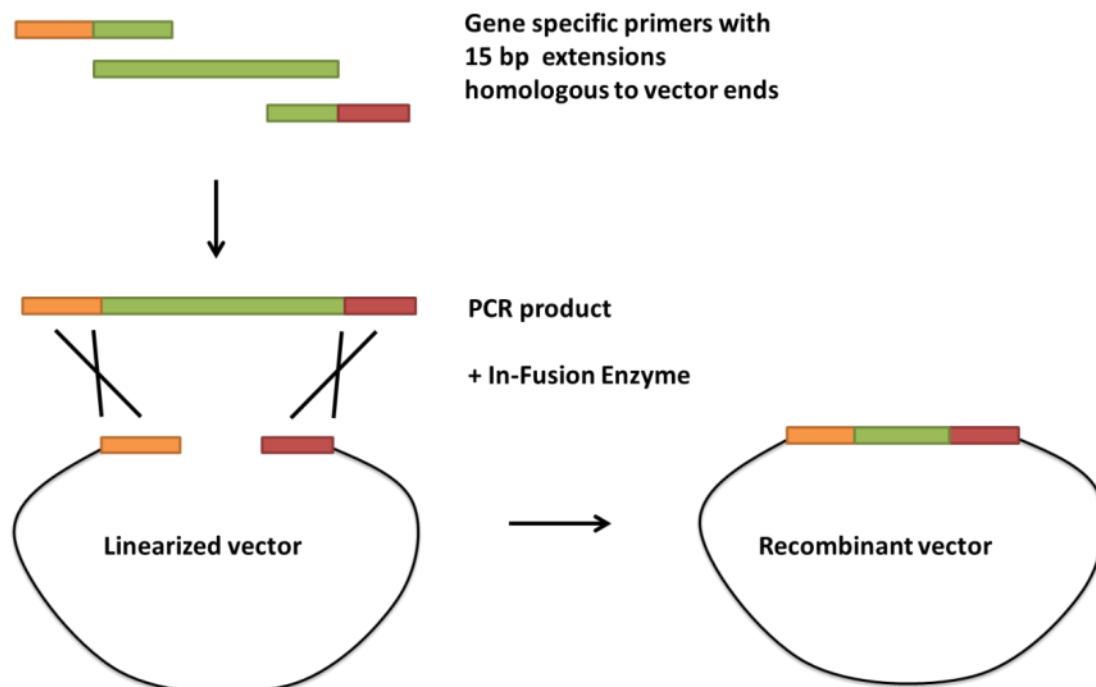


**Figure 2.2** Ligation-independent cloning using the the In-Fusion™ PCR Cloning Kit (Clontech).

DH5α and OmnimaxII (Invitrogen) cells were transformed with this construct and plated on LB agar containing 0.02% X-gal; 1 mM IPTG and 100 μg ml$^{-1}$ carbenicillin. The resultant white colonies were checked by PCR using the gene-specific primers and those containing a band of the correct size were sequenced. These pOPINF constructs were purified from the *E. coli* subcloning strain then introduced into chemically competent *E. coli* expression strain BL21(DE3)(pLysS) cells by transformation (see section 2.1.7). The proteins were overexpressed in Luria Bertani media (10 g l$^{-1}$ NaCl, 10 g l$^{-1}$ tryptone and 5 g l$^{-1}$ yeast extract) (Miller, 1972). Protein expression was then induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the cells were left shaking at 310 K, (or at a reduced temperature to improve protein solubility) for between 6-18 h.  The cells were harvested by centrifugation in a Sorvall Evolution centrifuge (7 min, 5500 rev min$^{-1}$, 277 K, SLC-4000 rotor) and stored at 254 K.

### 2.4.2    Cell lysis

The cell pellet was resuspended in lysis buffer (50 mM Tris pH 8, 0.5 M NaCl, 40 mM imidazole and 1% Tween-20) containing a Complete EDTA-free protease-inhibitor cocktail (Roche) and Deoxyribonuclease I was added to a concentration of approximately 20 μg ml$^{-1}$ resuspended cells. The lysis of each preparation was performed either using a sonicator or a cell disruptor and the cell debris removed by centrifugation in a Sorvall Evolution centrifuge (25 min, 16000 rev min$^{-1}$, 277 K, SS34 rotor).

### 2.4.3    Purification by immobilized metal ion affinity chromatography (IMAC)

The supernatant was subsequently filtered (0.22 μm filter) to remove any remaining cell debris and then applied to a 1 ml Ni$^{2+}$ charged  His-Trap Chelating HP column (GE Healthcare) which had been pre-equilibrated with buffer A (50 mM Tris pH 8, 0.5 M NaCl, 40 mM imidazole). All protein purifications were performed at 277 K. The column was washed with 20 ml of buffer A, to elute loosely bound proteins, before eluting the protein of interest by a linear gradient to 0.5 M imidazole in buffer A. The fractions containing the protein were analyzed using SDS-PAGE and then pooled and concentrated using an Amicon-Ultra 10 kDa cutoff concentrator (Millipore).

### 2.4.4    Size exclusion gel filtration chromatography

Further purification of the proteins was performed by applying the concentrated samples to a Superdex-75 HiLoad HP gel filtration column (GE Healthcare) which had been pre-equilibrated with 25 mM HEPES pH 7.5, 0.2 M NaCl. This method separates the proteins on the basis of size. The protein sample was loaded onto the column at 1 ml min$^{-1}$ and an isocratic elution was then applied for 1 column volume. The fractions containing the protein were again analyzed using SDS-PAGE and concentrated (concentrations of each protein are specified in the relevant sections.) Size exclusion chromatography was also used to estimate the oligomeric state of the protein by comparing the elution volume

against a standard curve of proteins of known molecular weight.  The Superdex 75 HiLoad HP gel filtration column (GE Healthcare) was calibrated using the low molecular weight protein standards kit (GE Healthcare) according to the manufacturer's instructions.  A calibration curve was then used to estimate the size of the protein of interest.

### 2.4.5   Analysis of protein samples by SDS-PAGE

SDS-PAGE was used to analyse protein samples by monitoring the migration of the proteins through a polyacrylamide gel based on their molecular size.  Gels were poured and run using the BioRad Mini Protean apparatus. These gels consisted of a stacking gel, resolving gel and running buffer, which were prepared by the Laemmli method (Laemmli, 1970). The stacking gel had a polyacrylamide concentration of 3.0% and the polyacrylamide concentration of the resolving gel was 15%.  A 15% resolving gel was prepared by mixing 3.2 ml Milli-Q (MQ) water (Millipore), 2.5 ml resolving buffer (1.5 M Tris-HCl pH 8.8, 0.4% (w/v) SDS) and 5.0 ml of a 30% acrylamide / bisacrylamide solution. 50 µl of a 10% ammonium persulphate (APS) solution and 8 µl TEMED were added to start the polymerization and the mixture was then poured between the glass plates separated by two 1 mm spacers.  A layer of water was poured on top of the gel to obtain a flat surface and the gel was allowed to polymerise for approximately 30 minutes.  The stacking gel was then prepared by mixing 3.2 ml MQ water, 1.3 ml stacking gel buffer (0.5 M Tris-HCl pH 6.8, 0.4% SDS) and 0.5 ml of a 30% acrylamide / bisacrylamide solution.  25 µl of a 10% ammonium persulphate (APS) solution and 8 µl TEMED were added to start the polymerization and the mixture was then poured on top of the resolving gel.  A comb was added for well formation and the gel was left to polymerise for approximately 30 min.  Gels were stored for up to one week at 277 K.  The protein samples were prepared by the addition of sample buffer (400 mM Tris-HCl, 6.4 % SDS, 160 mM EDTA, 32% glycerol, 16% β-mercaptoethanol and 4% bromophenol blue) and heating to 373 K for approximately 5 minutes prior to loading in the wells.  A low range molecular mass ladder was also loaded.  The gels were run using BioRad Mini Protean apparatus with a running buffer of 2.5 mM Tris-HCl, 192 mM glycine and 0.1% SDS at a constant voltage of 250 V.

The gels were stained with instant blue stain (Novexin) according to the manufacturer's instructions, prior to destaining in MQ water.

### 2.4.6 Protein concentration

The concentration of protein was estimated using Bradford reagent in a method based on the Bradford assay (Bradford, 1976). Duplicate samples diluted in water were added to 1 ml of the Bradford reagent and the absorbance was then measured at 595 nm using a spectrophotometer. Protein concentrations were then estimated using a standard curve prepared using Bovine Serum Albumin.

### 2.4.7 Cleavage of the histidine tag

The histidine tag was removed by proteolytic cleavage by human rhinovirus His-tagged 3C protease (Novagen), which cleaves after glutamine at its cleavage site (LEVLFQ↓GP). The protease itself contains a 6x histidine tag which allows its own purification by IMAC and also aids separation of the cleaved protein of interest from the reaction solution. To remove the tag, the protein was first diluted to a concentration of 1-2 mg ml$^{-1}$ in tag cleavage buffer (buffer A with 2 mM β-mercaptoethanol – see section 2.1.2) and 12 µg of 3C protease were added per mg of protein. This was incubated at 277 K for ~ 16 hours and the sample applied straight onto a nickel affinity column which had been pre-equilibrated with tag cleavage buffer. The flow-through was collected and then the column was washed with 100% buffer B (0.5 M imidazole) to elute any uncleaved sample and the protease. Elution samples were analysed by SDS-PAGE.

### 2.4.8 Dynamic Light Scattering (DLS)

The quality of protein purified was assessed prior to crystallization using a Dynapro Titan Molecular Sizing instrument with temperature-controlled microsampler (Wyatt). DLS measures the intensity of light scattered by the sample molecules as a function of time and from this, if the temperature and the solvent are known, the average size of the

molecules can be estimated. It can also establish if the sample molecules are aggregating in solution and the polydispersity of the solution. Proteins that exist as monodisperse solutions have been shown to be more likely to crystallise than proteins that aggregate randomly (D'Arcy, 1994). All protein samples that were tested using DLS were filtered through a 0.1 µm centrifugal filter (Millipore) prior to analysis of 13 µl of the protein sample in a quartz cuvette. DLS was carried out according to the manufacturer's instructions.

### 2.4.9   Selenomethionine incorporation using metabolic inhibition

This method was adapted from (Doublie, 1997). A 50 ml culture of *E. coli* strain BL21 (DE3) pLysS, containing the plasmid of interest, was grown overnight at 310 K in LB media with appropriate antibiotic selection. The culture was then centrifuged and the cell pellet re-suspended in MQ water. This was repeated a further two times to remove any trace of LB media. The cells were then re-suspended in water and used to inoculate 2 litres of minimal (M9) media supplemented with 0.2% (w/v) glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.001% (w/v) thiamine and 20 ml of an amino acid stock. The amino acid stock (20 ml) contained 40 mg of valine, phenylalanine, isoleucine, leucine, aspartic acid, glutamic acid, lysine, arginine, serine, threonine, tyrosine, histidine, glutamine and tryptophan. The cells were left to grow with shaking at 310 K until an absorbance at 600 nm of approximately 0.4 was reached. At this point a further 100 mg of threonine, lysine and phenylalanine, 50 mg of leucine, isoleucine and valine and 60 mg of selenomethionine were added. The flasks were then shaken at 310 K for a further 45 min prior to the addition of 1 mM IPTG. The cells were then left shaking overnight at 310 K and subsequently centrifuged, lysed and the protein purified in an identical manner as for the wild type protein.

### 2.4.10  Western blot

50 µl spore suspension were aliquoted into sterile eppendorfs and centrifuged for 1 minute at 13000 rpm. The glycerol was removed and the pellet was resuspended in 1 ml

TES buffer. The spores were then heat shocked at 323 K for 10 minutes and cooled in cold water. The spores were then added to 10 ml 2xYT in universals containing baffles and incubated at 303 K with shaking at 250 rpm for up to 7 hours. The spores were checked using microscopy to observe the emerging germ tubes. The germlings were centrifuged for 8 minutes at 3000 rpm and resuspended in 1 ml TES buffer and vortexed to separate aggregated germlings. The $OD_{450}$ of each sample was checked (using 1:100 and 1:1000 dilutions) in order to determine the volume required to inoculate with and $OD_{450}$ of 0.03-0.05. 50 ml of TSB:YEME was inoculated with the suspension and incubated at 303 K with shaking at 250 rpm. 3 ml of this inoculum were collected every 6 hours for 48 hours. The spores were centrifuged at 3000 rpm for 1 minute and frozen at 253 K. These pellets were resuspended in 0.4 ml of sonication buffer (20 mM Tris-HCl pH 8; 5 mM EDTA and protease inhibitor) and sonicated for 5 cycles of 15 seconds each. The supernatant was collected by centrifuging at 13000 rpm at 277 K for 15 minutes and the protein concentration determined by Bradford assay. 10 µl of protein sample (with 4x sample buffer) was loaded per well (a stock protein solution of 150 µg ml$^{-1}$ was prepared) and the proteins were separated on 15 % SDS-polyacrylamide gel electrophoresis gels at 90 V in the stacking gel and 140 V in the resolving gel.

**Wet transfer**

Once removed, both the gel and the Hybond C membrane (Amersham Pharmacia Biotech) were soaked in 1x running buffer for 20 minutes before blotting. The proteins were then blotted onto the membrane by wet transfer at 100 V for 1 hour at 277 K in 1x running buffer. The membranes were then removed and placed face up in 25 ml of blocking solution and incubated overnight at 277 K.

**Washing and antibody incubation**

The membranes were washed in TBST buffer for 15 minutes. Several dilutions of antibody were prepared in blocking solution: 1:500; 1:400; 1:300 and 1:200. The membranes were transferred to each of these solutions and incubated for 1 hour at room temperature. They were then washed twice in TBST for 15 minutes and incubated in 10 ml fresh blocking solution and secondary antibody (ECL Rabbit IgG – 1:10000 dilution) for 1 hour at

room temperature and then washed three times in TBST (10 minutes each wash). The blot was developed using ECL enhanced chemiluminescence system (GE Healthcare Life Sciences) and were exposed to X-ray film between 15 seconds to 5 minutes.

### 2.4.11  Protein identification using mass spectrometry

This was performed by the proteomics facility at the John Innes Centre. Samples were made up in 50/50 water/methanol + 0.1% (v/v) formic acid, loaded into coated borosilicate glass capillaries (Micromass, Manchester, UK) and sprayed into the nano-electrospray ion-source of a Q-ToF 2 mass spectrometer (Micromass, Manchester, UK) in positive ion mode. Charge-states from the resulting spectra were deconvoluted using the MaxEnt algorithm (Micromass, Manchester, UK) to enable determination of exact molecular weights.

## 2.5  Protein Crystallography Methods

### 2.5.1  Crystallization

All crystallization trials were set up using the vapour diffusion method.  A drop containing the protein solution is mixed with a precipitant solution and then is equilibrated against a reservoir containing the precipitant solution at a higher concentration than in the drop. Equilibration between the drop and the reservoir proceeds so that the precipitant concentration increases in the drop.  In an ideal case, this will lead to a level of saturation that will result in nucleation and the production of protein crystals. Crystallizations were performed at a constant temperature of 291 K. The initial screening was carried out with an OryxNano robot (Douglas Instruments Ltd) in 96-well sitting-drop MRC plates (Molecular Dimensions) using a variety of commercial screens (Molecular Dimensions and Qiagen). The drops consisted of 0.3 µl protein solution mixed with 0.3 µl precipitant solution with a total reservoir volume of 50 µl.

### 2.5.2  Crystallization optimization

Any promising crystallization conditions observed in the screens were optimised.  All optimisations were carried out using hanging drop vapour diffusion using 24 well plates (Molecular Dimensions).  The drops typically contained 1 µl protein and 1 µl precipitant placed on a plastic coverslip which was then suspended and sealed with vacuum grease over a well containing 1 ml of precipitant.  Many crystallization optimisations were set up to obtain the best crystals for data collection. These also included additive screens (Hampton Research) as well as co-crystallization with halide compounds in an attempt to obtain derivative crystals for anomalous diffraction experiments.

### 2.5.3  Crystal soaks

Crystals were soaked with a variety of heavy atom compounds (potassium dicyanoaurate; thimerosal; potassium hexachloroplatinate) and halide solutions (Dauter *et al.*, 2000). Typically the heavy atom was dissolved in an artificial crystallization solution at a concentration of 1-10 mM, but potassium iodide was used at a concentration of 1 M. Soaks were performed for varying lengths of time from 30 min (or 10 seconds for the iodide solution) to several hours at 291 K and the heavy atom solutions were kept in the dark to prevent any photochemical reactions (Ducruix & Giege, 1992).

### 2.5.4  Cryoprotection of crystals for data collection

A cryoprotectant is necessary so that the crystals can be cooled to cryogenic temperatures without the formation of ice.  Typically for cryoprotection, approximately 10-15% of the volume of water in the crystallization solution was replaced with ethylene glycol or glycerol.  This was either done in the crystallization well conditions, or post-crysallization, in which the crystals were transferred to the cryoprotectant using litholoops (Molecular Dimensions) for an optimised length of time, and were then cooled by plunging into liquid nitrogen and stored the mounted litholoops (Molecular

Dimensions) for transport to the synchrotron. In some cases, further optimisation of the cryoprotectant was required.

### 2.5.5 Data collection

X-ray data were either collected 'in house' using a Mar 345-image plate detector (X-ray Research) mounted on a Rigaku RU-H3RHB rotating anode X-ray generator (operated at 50 kV and 100 mA) fitted with osmic confocal optics and a copper rotating anode (Cu $K_{\alpha}$; $\lambda$ = 1.542 Å), or at the Diamond Light Source (DLS) Synchrotron facility in Oxfordshire. Data were collected both onsite and remotely at DLS. Several stations were used: i04-1 (fixed wavelength monochromatic beamline); the microfocus beamline, i24, and the tunable beamlines, i02, i03 and i04. The diffracted X-rays were recorded on a MAR 345 image plate in house and on either charge coupled device (CCD) or photon counting pixel detectors at the synchrotron.

### 2.5.6 Data processing

Data were integrated with iMOSFLM (Leslie *et al.*, 2007) or XDS (Kabsch) and scaled and merged using either SCALA (Evans, 2006) or, again, with XDS/Xia2 (Winter, 2010). The downstream processing was carried out using the CCP4 suite (Winn *et al.*, 2011) or PHENIX (Adams *et al.*, 2010). In all datasets, a subset of the data comprising 5% of the reflections was set aside for calculation of the free R factor (Kleywegt & Brunger, 1996, Brunger, 1993) during model refinement.

### 2.5.7 Single isomorphous replacement with anomalous scattering (SIRAS) and Single wavelength anomalous diffraction (SAD)

For this technique, crystals are derivatised by the introduction of a heavy atom. This may be done by soaking the crystal in a crystallization solution containing the heavy atom. In an ideal case only a few heavy atoms are incorporated into the crystal and the crystal lattice is unperturbed (isomorphous) by the soaking. The heavy atom atom structure can

then be determined using a Patterson analysis. The structures of Sco3914 and Sco5413 were determined using this method as the native and derivative crystals belonged to the same space groups with virtually identical cell dimensions.

The SAD technique relies upon all of the measurements being taken from the same derivative crystal. This has been used to solve the structure of selenomethionine-labelled Sco4122. A fluorescence scan is performed on the crystal at the synchrotron to determine the f' (dispersive component) and f'' (absorptive) component as a function of the energy, in addition to ascertaining the wavelength of the absorption edge.  The X-ray wavelength is then chosen for data collection.  Commonly, a peak wavelength is chosen to maximise the f'' value. It is important to collect high quality, complete data to solve the structure. The programs SOLVE (Terwilliger & Berendzen, 1999a, Terwilliger & Berendzen, 1999b) and PHENIX (Adams *et al.*) can all be used to locate the heavy atoms and obtain the required phase information. All of the phases determined in this work were obtained using PHENIX.

### 2.5.8   Molecular replacement

Molecular replacement is a technique that exploits the use of known similar structures to determine an initial estimate of the phase.   It is commonly used to solve different crystal forms of the same molecule or to solve structures that have high structural similarity.  If the search model and the new protein are isomorphous, as could be the case for soaking a small ligand into the protein crystal, the phases from the search model can be used directly to compute the electron density ($\rho$) for each point (x,y,z) in the unit cell from the native intensities of the new protein. If the phasing model is not isomorphous with the desired structure, as would be the case with a structurally similar protein, the problem is more difficult.  It is necessary to find the position and orientation of the phasing model in the new unit cell that would give phases that were most like those of the new protein. The program PHASER (McCoy, 2007) was used from the CCP4 suite of programs to try to solve the structures using molecular replacement.  PHASER was used to run rotation and

translation functions, combined with maximum likelihood. None of the initial structures in this thesis could be solved by molecular replacement (Sco3914C16S was solved by this method using the Sco3914 model as a template), so experimental phasing methods were required.

### 2.5.9 Phase improvement, model-building and refinement

The protein model is improved and completed by alternating cycles of electron density modification, (re)building and refinement. In some cases, automatic model building was carried out using the program BUCCANEER (Cowtan, 2006) or PHENIX Autobuild (Terwilliger *et al.*, 2008). The electron density maps and protein model were visualised using the program COOT (Emsley & Cowtan, 2004). Manual model building of the protein model was carried out by manually altering it to fit the electron density. Where appropriate, atoms were modelled into positive density and removed from negative density. Rigid body and restrained refinement of the protein structure was performed using REFMAC5 (Murshudov *et al.*, 1997) and fitting and refinement of the water molecules was done using ARP/wARP (Perrakis *et al.*, 1999). Cycles of manual model building and refinement were continued until the model adequately accounted for the observed electron density, had acceptable stereochemical properties and no further improvements were observed in $R_{work}$ and $R_{free}$ values.

### 2.5.10 Structure validation

Automatic validation of the structure was carried out using the online program Molprobity (Chen *et al.*, Davis *et al.*, 2007). One of the outputs of Molprobity is the Ramachandran plot (Ramachandran *et al.*, 1963) and this monitors whether the main chain torsion angles are in a favourable conformation. The presence of amino acids in the disallowed regions often highlights problems in the local area of the structure. However, residues in the disallowed region can be genuine and may be residues of functional or structural significance.

Further analysis of the structures were performed by a visual inspection and also using a variety of online tools such as PDBSum (Laskowski, 2001) and the PISA server (Krissinel, 2007).

## 2.6 Surface Plasmon Resonance (SPR)

SPR is a valuable technique for detecting and measuring intermolecular interactions. In this experiment, it was used to qualify and quantify the interactions between the proteins and their cognate DNA sequences by immobilizing the DNA. The method used here involves a streptavidin coated chip which can bind a biotin-labelled DNA molecule. The protein of interest is then flowed across the immobilized DNA (the association phase) and the response is measured for this and the subsequent dissociation phase.

### 2.6.1 DNA Sequences and Bioinformatics

All of the intergenic sequences used in the SPR experiments were obtained from StrepDB. The *in silico* analysis of these sequences to locate putative binding sites was performed using the MEME Suite (Motif based sequence analysis tools) (Bailey *et al.*, 2009) and these searches were carried out by Govind Chandra. The search results for each of the proteins will be discussed in the relevant chapters.

### 2.6.2 Testing the whole intergenic region

The initial step involved amplifying the whole of the intergenic sequence by PCR in order to establish if there is any binding to this region. This was achieved by covalently attaching a biotinylated single-stranded DNA (ssDNA) linker to the streptavidin chip surface which could be used to capture double-stranded DNA (dsDNA) molecules which contain a single-stranded sequence at the 5' end (which is complementary to the linker). This method (Reusable DNA Capture Technology - ReDCaT) was developed by Clare Stevenson, David Lawson and Sandra Greive and was used to test all of the sequences used in this study.

Biotinylated ssDNA linker sequence (1):-

5' CGCATACGCAAACGCTAA – bio 3'

The intergenic sequences required the generation of a 20 bp 5' overhanging ssDNA complementary in sequence to the 20 bp 3' biotinylated oligo, which is immobilized on the streptavidin chip surface. This single-stranded overhanging sequence was achieved by the inclusion of an abasic site (1',2'-Didoxyribose or dSpacer from Integrated DNA Technologies, Inc) in the forward PCR oligo which blocks transcription by DNA polymerase. The amplification was carried out in 3 steps as detailed in Figure 2.3. Taq polymerase was used as there is no exonuclease activity which would otherwise remove the single-stranded overhanging bases. The sequence specific oligos used are detailed in the relevant chapters.



**PCR Step**
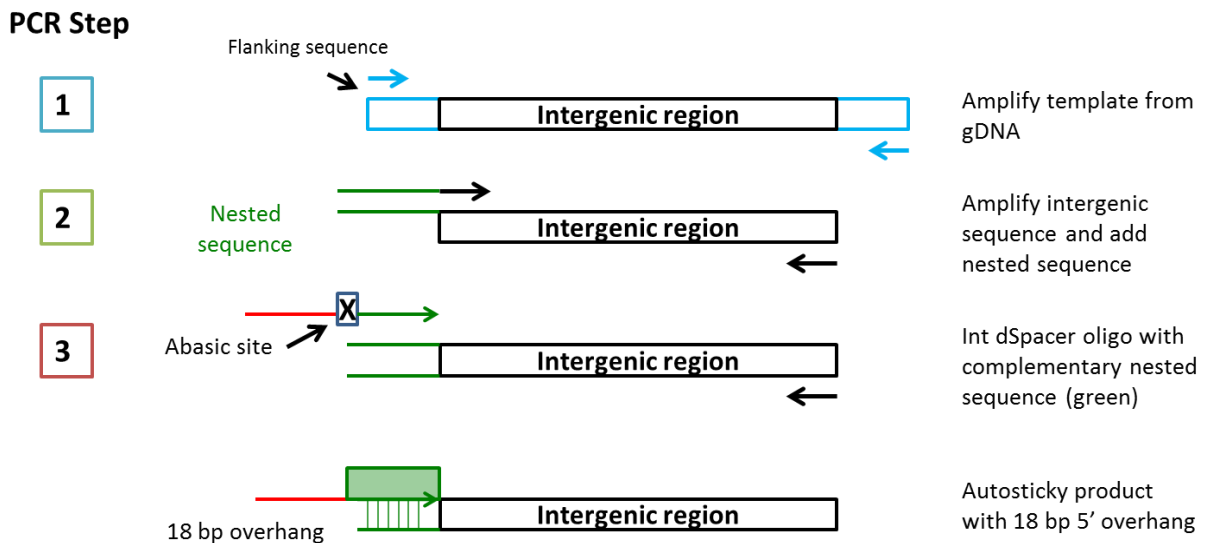
**Figure 2.3** Autosticky PCR – generation of a 5' overhang on the intergenic sequence of interest. The 5' overhang is complementary to the 3' biotinylated oligo which has been immobilized on a streptavidin chip surface.

**PCR Step 1**    The first PCR step involved the amplification of a larger template which incorporates the intergenic region of interest. This was necessary to avoid amplification of non-specific DNA and a BLAST search of the genome was performed to find the optimal primers to minimize the possibility of non-specific PCR product.

Reaction mix

| | |
|---|---|
| DMSO | 2.5 µl |
| 10x PCR reaction buffer | 5 µl |
| 10 mM dNTP mix | 1 µl |
| For primer (100 pmol µl$^{-1}$) | 1 µl |
| Rev primer (100 pmol µl$^{-1}$) | 1 µl |
| Template (50 – 200 ng µl$^{-1}$) | 1 µl (use 2 to 3 x this for genomic DNA) |
| Taq polymerase | 1 µl |
| Water | 37.5 µl |

PCR programme:

1. 95°C for 4 min
2. 95°C for 45 sec
3. 55°C for 45 sec
4. 72°C for 1 min 30 sec
5. Go to step 2 26 more times
6. 72°C for 7 min
7. END

Times can be shortened for short product lengths

Annealing temperature in 3 may need optimising depending on the primer sequence.

**PCR Step 2** The intergenic region was then amplified from this template fragment to include a "nested" sequence which is complementary to the 3' end of the forward oligo used in the final stage.

**PCR Step 3** The nested intergenic region was amplified again using the reverse oligo used in stage 2 and a forward oligo with an abasic site for the generation of the single stranded sequence. For this PCR reaction, the annealing temperature was set to 45°C.

For PCR steps 1 and 2, the PCR product was purified from 1% agarose gels using the Qiagen gel purification kit according to the manufacturer's instructions.

**Loading the DNA**

The biotinylated ssDNA oligo was immobilized on the chip surface by using the wizard for immobilization on the Biacore T100 machine. The chip was washed three times with 50 mM NaOH and 1 M NaCl in order to clear the chip surface prior to loading a 10 nM biotinylated oligo solution. All of the flow cells were subsequently blocked with three 60 second injections of biotin polyethyleneoxide iodoacetamide (Sigma). A 1 µM solution of the autosticky PCR product was captured by injecting across the experimental flow cell 4 (fc4), aiming for a capture level of 50-100 RU. As the generation of ssDNA overhangs is inefficient using the dspacers, it is necessary to use a slow rate of injection (10 µl min$^{-1}$) over a longer period of time (~240 seconds). A 239 bp control sequence, kindly donated by Sandra Greive, was used in order to account for any non-specific binding and this was captured on the control flow cell 2:-

Control sequence

5'  TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCAGCAGGACG CACTCGCTCTTAAAAATTAGCCCTGAAGAAGGGCAGCATTCAAAGCAGAAGGCTTTGGGGTGTGT GATACGAAACGAAGCATTGGCGGAATTCCAATCCCGAAACAGTTCGCAGGTAATAGTTAGAGCC TGCATAACGGTTTCGGGATTTTTTATATCTGCACAACAGTTTTTTAT 3'

The reference cells (fc1 and fc3) contained no linker and were blocked with biotin. The experimental set up is shown in Figure 2.4.
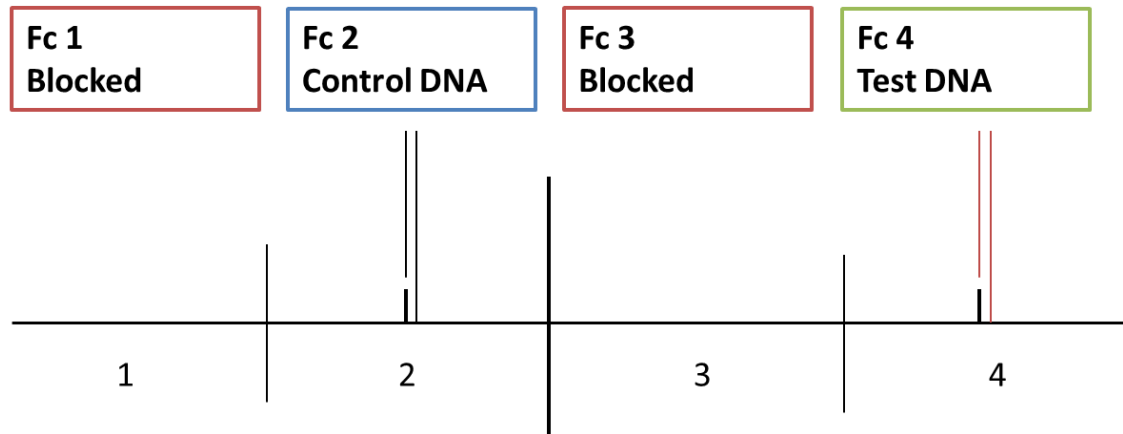
**Figure 2.4** Four flow cell chip experiment to test protein interaction with whole intergenic region

### 2.6.3   Oligos used for the testing of the whole intergenic regions in SPR

Biotinylated oligo (this is immobilized on the streptavadin chip)

5' CGC ATA CGC AAA CGC TAA – biotin 3'

Forward oligo with abasic site (idSp)

5'TTA GCG TTT GCG TAT GCG /idSp/ TTA TCA AAA AGA GTA TTG AC 3'

Forward oligo with nested sequence

5' TTA TCA AAA AGA GTA TTG AC (~ 20 bases of target sequence) 3'

The PCR product from the final reaction was purified by the ethanol extraction method. The sample was pooled and precipitated with 1/10 volume of sodium acetate and 2 volumes of 100% ethanol and incubated overnight at 277 K. The reaction was then centrifuged at 13000 rpm at room temperature for 30 minutes and the supernatant was removed immediately afterwards. The pellet was rinsed with 70% ethanol and allowed to dry for approximately 5 minutes. The DNA was dissolved gently in 30 μl water.

### 2.6.4  DNA footprinting

This technique again utilized the reusable streptavidin chip to screen a number of DNA sequences to identify the cognate sequence. This time, however, the control sequence used above was omitted and, instead, the reference cell was loaded with the biotinylated linker sequence. All of the following experiments were carried out at 293 K using a Biacore T100/T200. The oligos were ordered from Sigma at a concentration of 100 μM.

**Stage 1**

In order to define the cognate binding sequence, it was necessary to fragment the intergenic region into overlapping sequences. The intergenic regions between *sco5413-5414*; *sco3914-3915* and *sco4121-4122*, were fragmented into shorter length sequences which incorporated ~22 base pair overlaps between each sequence (specific lengths are detailed in the relevant chapters). This was achieved using the Perl Overlapping Oligo Program (POOP) which was developed by Tung Le (unpublished work). The maximum length of 30 base pairs was chosen as DNA synthesis becomes expensive and less accurate above 50 bases (accounting for the extra 20 base pairs included on the reverse strand for annealing to the linker).

The streptavidin chip was prepared by loading the biotinylated single-stranded linker onto the chip surface (fc1 and fc2). Typically, a capture level of 500 response units (RU) was set for the footprinting assay. For the affinity assay, this was adjusted to 100 RU (although the actual attained was 166 RU) as a lower response value improves the accuracy of such measurements (Stockley & Persson, 2009). The ssDNA linker was captured on both the control and experimental flow cells (in order to account for any non-specific binding to the linker sequence). The single stranded linker sequence is shown below:-

5' bio- GCAGGAGGACGTAGGGTAGG 3'

This sequence is a "random" sequence, distinct from the linker used for testing the whole intergenic region, which was designed by Clare Stevenson to minimize secondary

structure and self-dimer formation. The reverse complement of this sequence was then attached to the 3' end of each of the reverse oligos (sequence specific oligos are detailed in the relevant chapters) during the oligo synthesis (Sigma).

In this experiment, the complementary ssDNA sequences were annealed by mixing 60 µl of the forward oligo with 50 µl of the reverse oligo to give a 45 µM stock solution. This mixture was heated to 95°C for 10 minutes and then allowed to cool to room temperature. A larger volume of the forward oligo was used in order to maximize the potential for all of the linker strands (reverse) to form dsDNA and to reduce the opportunities for ssDNA to anneal to the chip. This was diluted to the working concentration of 1 µM with HBS-EP$^+$ buffer. Each dsDNA sequence was then captured on the chip by running it across the analytical flow cell (fc2) containing the ssDNA linker at a flow rate of 10 µl min$^{-1}$. The DNA captured was then washed with 0.5 M NaCl in order to remove any unbound DNA from the chip's surface. The protein was then run across both the control and analytical flow cells (60 seconds at 30 µl min$^{-1}$) and the results displayed as a percentage of the theoretical $R_{max}$ for each cycle. The theoretical $R_{max}$ was calculated as follows:-

$$R_{max} = \text{MW protein/ MW DNA x DNA captured (RU) x stoichiometry (1) x 0.78}$$

An adjustment value for DNA of 0.78 was used to calculate the theoretical $R_{max}$ value as nucleic acids give a slightly higher RU than does protein (Buckle *et al.*, 1996, Crouch *et al.*, 1999).

Following the protein injection, the protein was allowed to dissociate for 300 seconds before a second salt wash (0.5 M for 60 seconds at 30 µl min$^{-1}$) and a final wash with 1 M NaCl and 50 mM NaOH (60 seconds at 10 µl min$^{-1}$) to remove both the protein and the DNA and regenerate the chip for the next cycle. Two replicates of each protein

concentration were run against each sequence. The outcome of this stage determined the sequences for stage 2.

**Stage 2**

This stage involved taking the sequence(s) which elicited the highest response in stage 1 and eliminating 2 base pairs from each end in turn to narrow down the binding site (Figure 2.5). In order to account for any effects from the linker sequence, such as stabilization of binding due to the extra bases, the linker was subsequently attached to the reverse complement of the sequence to test the left hand boundary. The right hand and left hand boundaries were tested separately: two base pairs at a time were removed from the right hand side of the entire sequence, followed by two base pairs at a time from the left (the entire sequence). The specific sequences are detailed in the relevant chapters. The oligos were synthesized by Sigma and annealed using the protocol described previously. Again, three concentrations of the protein were run (in duplicate) and the interactions plotted as a percentage of the calculated $R_{max}$. This information was analysed to determine the right and left hand boundaries of the sequence. The resultant sequence, with which the protein has the most significant interaction, was used in stage 3 to investigate the strength of this interaction.

**Entire intergenic sequence**
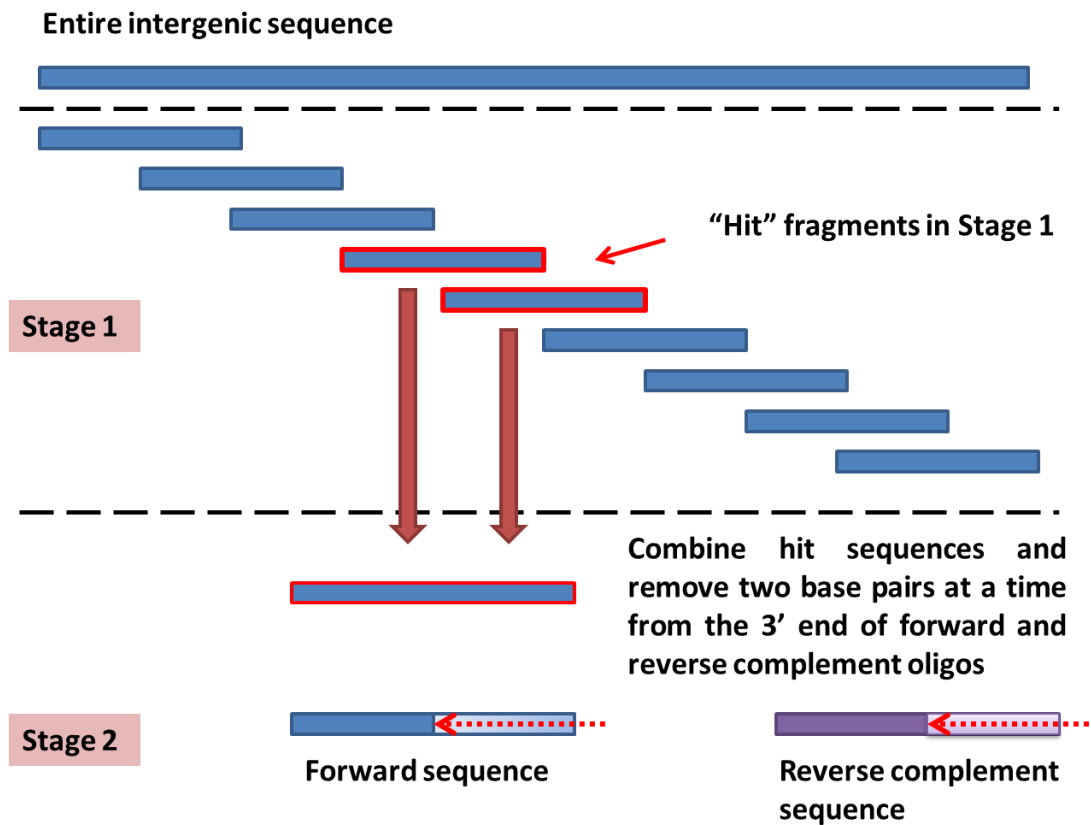


**Figure 2.5** DNA footprinting, stages 1 and 2.

**Stage 3**

The affinity of the protein for the cognate sequence, which was determined in stage 2, was investigated. The "footprint" was determined by eliminating the defined number of base pairs from each end of the sequence and this new sequence was synthesized by Sigma. On flow cell 3 (the reference cell), a ssDNA sequence which was complementary to the ssDNA linker sequence was captured in order account for any non-specific interactions with the linker. The sequence being tested was captured on flow cell 4 and the chip was regenerated after each cycle. This time the protein was injected across the DNA at a flow rate of 100 µl min$^{-1}$ for a period of 210 seconds to allow saturation of the DNA to occur. A dilution series was set up to include protein concentrations from 200 mM to 0.4 mM and each concentration was tested in triplicate. The responses were plotted against concentration and the affinity calculation was carried out using the Biacore T200 data evaluation software.

## 2.7 Determination of pKa values of cysteine residues in Sco3914

Sco3914 has two cysteines: Cys16 and Cys95. Cys16 is located in what would appear to be a putative ligand binding pocket, whereas Cys95 is located on the loop which could potentially make contacts with DNA. This experiment sought to determine the p$K_a$ of each of the cysteines in Sco3914 in an attempt to understand their relative reactivities. This information could also suggest the relevance of these cysteines to ligand detection and DNA-binding.

### 2.7.1 DTNB assay to determine pK$_a$ of C16 and C95

Ellman's reagent, or 5,5'-dithiobis(2-nitrobenzoate) (DTNB), undergoes a thiol-disulfide interchange in the presence of free thiols to generate the mixed disulfide and 2-nitro-5-thiobenzoate (TNB$^{2-}$). The stoichiometry of protein thiol to TNB formed is 1:1 and TNB has an intense absorption at 412 nm so it is possible to calculate the pK$_a$ by measuring the change in absorbance over time over a range of pH values. This was adapted from a similar experiment carried out on the MarR homologue, HypR  (Palm *et al.*, 2012).
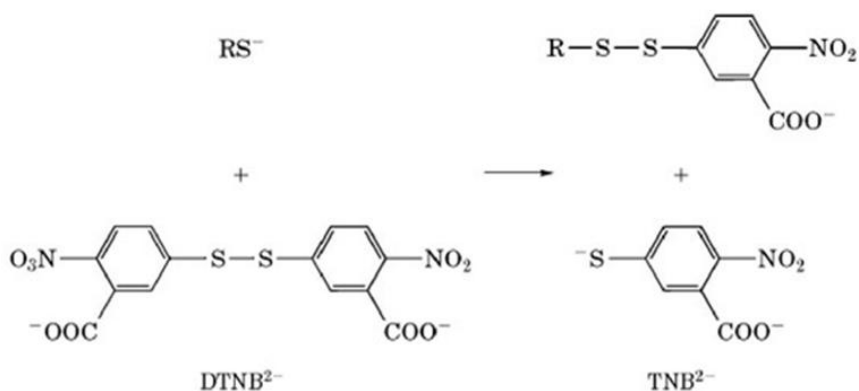


**Figure 2.6**  The reaction between DTNB and the cysteine residue (RS$^-$) in its thiolate form.

In this experiment, the purified Sco3914C16S and Sco3914C95S mutants were used to determine the pK$_a$ of the Cys95 and Cys16 residues respectively by reaction with DTNB. 5 µM of each protein were reacted with 20 µM DTNB in phosphate/citrate buffer (50 mM phosphate, 50 mM citrate, 100 mM NaCl) across the pH range  5.0-11.0 (5.0 ; 6.0; 6.5;

7.0; 7.5; 8.0; 8.5; 9.5; 10.0; 10.5; 11.0). The pH dependent rates of reaction of each cysteine with the DTNB were measured as a change in absorbance at 412 nm at room temperature on a Lambda 21 spectrophotometer (Perkin Elmer) over a period of 200 seconds. The first order kinetic constants were calculated by a first fit of time-dependent absorbance change by non-linear regression (using SigmaPlot). A second fit of the initial rate of each reaction against the pH was used to calculate the $pK_a$ using the following equation in KaleidaGraph (Synergy Software):-

$$K_{obs} = k_{max} * k_a / K_a + [H^+]$$

The data were plotted in KaleidaGraph (Synergy Software) and the p$K_a$s of each cysteine with their associated error values calculated.

## 2.8 Selecting the MarR proteins in *Streptomyces coelicolor*

The initial part of the project involved selecting the proteins to be characterized. This was done using a variety of bioinformatics tools to choose 5 MFRs from the total of 42 in the *S. coelicolor* genome. The first method was using the microarray time course data for the *Streptomyces coelicolor* strains M570 (*relA* null mutant which is deficient in the production of act and red (Sun *et al.,* 2001) and M600 (which is similar to M145 but differs in that it produces antibiotics at different time points) to analyse the expression profiles of each of the 42 genes. The data were viewed using GeneSpring (Agilent Technologies) software, with the assistance of Andy Hesketh and Maureen Bibb, and the main criteria were whether a protein was highly expressed or displayed a fluctuation in its expression level at a particular time point (e.g. 36 hours when sporulation occurs). The object was to select a protein which, when deleted from the genome, would yield a phenotype which differed to the wild type. The expression profiles for all 42 homologues were analysed, looking for any fluctuations over time, and the raw data were also used as this reveals the level of expression (on a log scale). The values of 0-50 were considered to be indicative of low expression. Also analysed were the microarray data for the orthologues in *Streptomyces venezuelae* to check that these also had a similar expression profile.
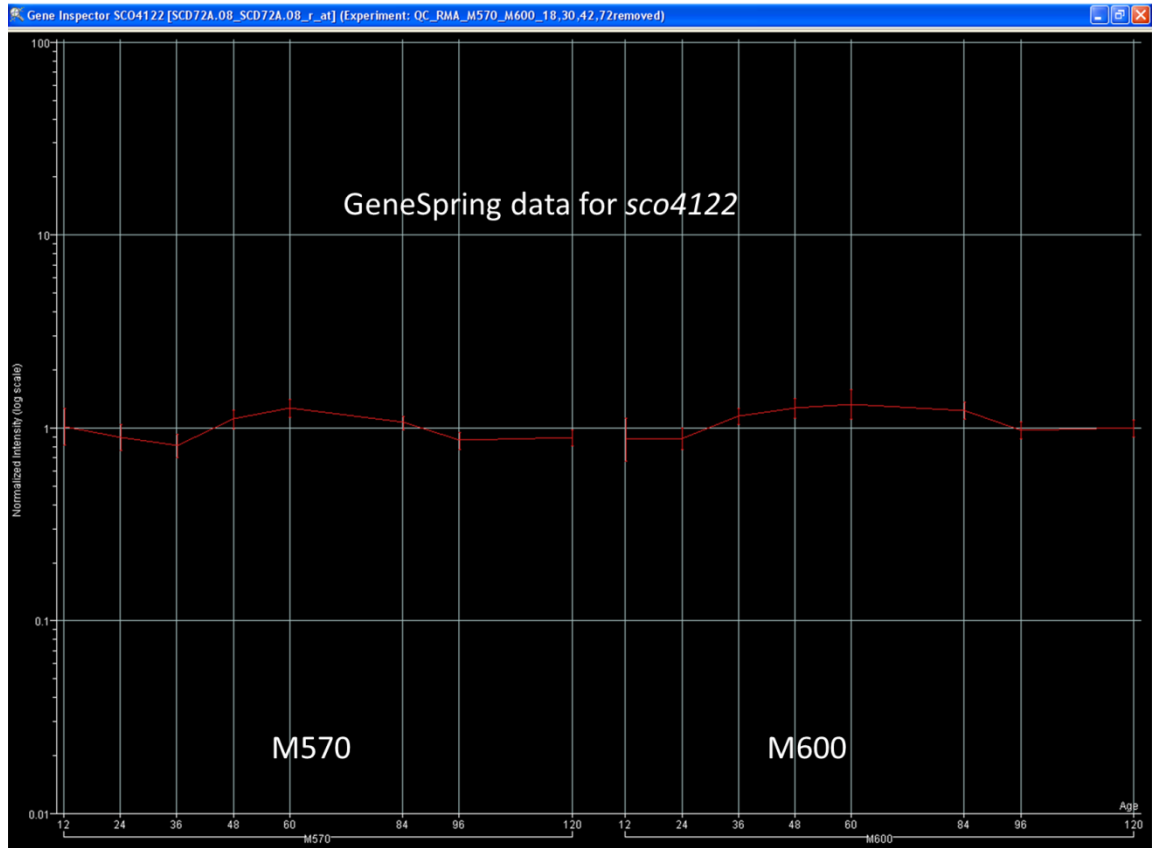
**Figure 2.7** Microarray data from GeneSpring showing the expression profile of *sco4122* over the course of a 5 day experiment.

The proteins were selected for their expression under the conditions measured in the microarray experiments and also their sequence identities to other species of streptomycete. The latter were obtained from the *Streptomyces* orthologues table (compiled by Govind Chandra, John Innes Centre), searching specifically for MFRs which had good conservation of identity across several species of streptomycete (*S. griseus; S.coelicolor; S. scabies; S. avermitilis* and *S. venezuelae*). Finally, the Streptomyces Annotation Server, StrepDB (http://strepdb.streptomyces.org.uk/), was also used to look for regions of synteny across the annotated genomes.

# Chapter 3

# Sco3914

## 3.1 Introduction

Sco3914 is one of the 42 putative transcriptional regulators of the MarR family of regulators (MFRs) in *Streptomyces coelicolor*. Bioinformatic analysis predicts it to be a mainly helical protein of 154 amino acids, with the winged helix-turn-helix motif which is characteristic of some regulatory DNA-binding proteins. Sco3914 has been cloned, expressed, purified and crystallized and the structure solved to 1.5 Å resolution. A deletion mutant was generated in the host organism but no visible phenotype observed under the conditions tested. The DNA footprint of Sco3914 has been determined in this work by using Surface Plasmon Resonance (SPR) and the kinetics of this interaction also quantified. There is potential for this protein to be crystallized with its cognate DNA sequence in order to identify the precise protein-DNA interactions.
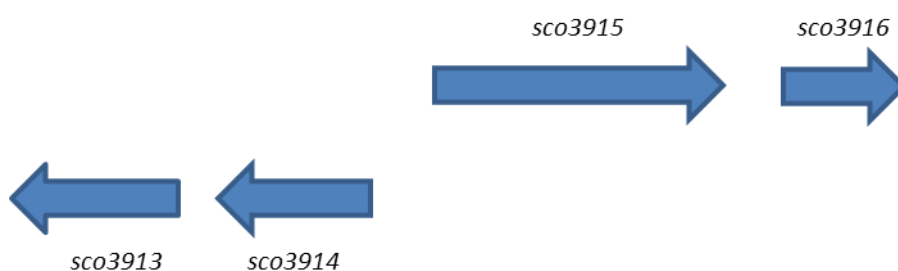


**Figure 3.1** A representation of the genomic context of s*co3914* in the *Streptomyces coelicolor* genome. The divergently transcribed gene, *sco3915*, is a probable transmembrane efflux protein. Sco3914 binds to the intergenic region between these two genes.

The *sco3914* gene is 465 bp in length and has a GC content of 69%. It is an extremely well conserved gene amongst streptomycetes and has 68 orthologues amongst 112 actinobacterial genomes and *E. coli* and *Bacillus subtilis* (bioinformatics analysis done by Govind Chandra, John Innes Centre). Its close homology to orthologues in other published sequences of streptomycetes can be seen in Table below.

**Table 3.1** The sequence identities shared between the orthologues from other Streptomycetes and *Streptomyces coelicolor* Sco3914. Sav – *Streptomyces avermitilis*; Ssc – *Streptomyces scabius*; Sgr – *Streptomyces griseus*: Sven – *Streptomyces venezuelae*.

| | Amino acid sequence %ID | | | | DNA Sequence %ID | | | |
|---|---|---|---|---|---|---|---|---|
| **Sco** | **Sav** | **Ssc** | **Sgr** | **Sven** | **Sav** | **Ssc** | **Sgr** | **Sven** |
| Sco3914 | Sav4281 86% | Scab4619 87% | Sgr3665 84% | Sven3693 84% | Sav4281 87% | Scab46191 85% | Sgr3665 82% | Sven3693 86% |

As MFRs are often autoregulatory (Rosenfeld *et al.*, 2002), the intergenic region upstream of *sco3914* was a logical starting place to look for Sco3914 operator sites. Moreover, since the adjacent gene, *sco3915*, is divergently transcribed from *sco3914,* Sco3914 could potentially regulate this gene as well. Sco3915 is annotated in the StrepDB server as a probable transmembrane efflux protein. Further bioinformatic analysis reveals that it shares a high sequence identity (>50%) with transmembrane efflux proteins or major facilitator superfamily protein members from a variety of streptomycetes and other actinobacteria including *Nocardia* and proteobacteria such as *Burkholderia*. The prevalence of this protein in these species would suggest that it may be important in protecting the cell against toxic molecules and so, if Sco3914 were to regulate it, then this would be of significant interest. This is a hypothesis which would warrant further investigation. In this study, the interactions between Sco3914 and this intergenic sequence were explored using Surface Plasmon Resonance (SPR).

The results obtained for the biophysical, genetic and structural characterization of Sco3914 will be discussed in this chapter.

## 3.2  Protein Expression and Purification

The protein was cloned and expressed according to the method described in Chapter 2. Typically, Sco3914 elutes from the nickel affinity column with a fairly broad peak profile (Figure 3.2). The fractions containing the protein were analysed by SDS-PAGE pooled and

applied to the size exclusion gel filtration column (GE Healthcare Superdex 75 16/60). The purity of the protein following the nickel affinity column purification and the size exclusion gel filtration chromatography column can be seen in Figure 3.2 and Figure 3.3.
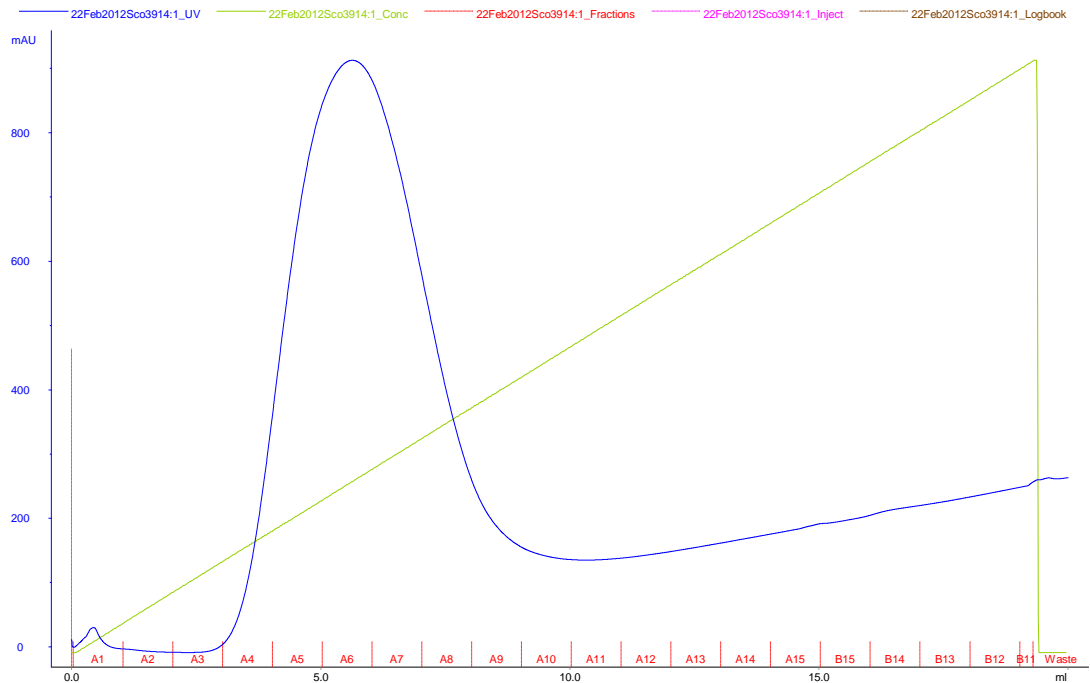


**Figure 3.2** Elution profile of Sco3914 from nickel affinity column chromatography. The protein elutes at a concentration of ~100 mM imidazole.
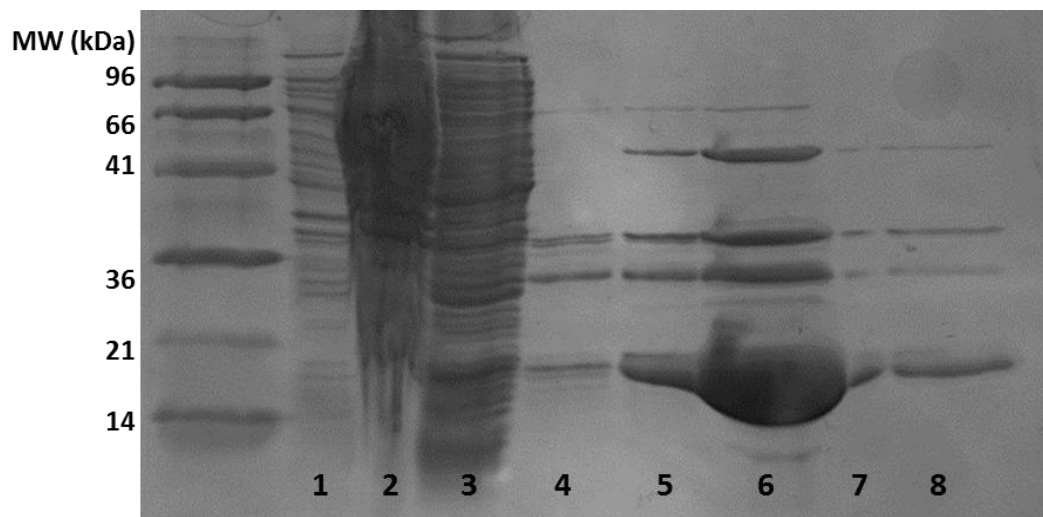


**Figure 3.3** 15% SDS-PAGE after nickel affinity column purification. The columns contain: (1) Lyzed pre-induction pellet; (2) Lyzed post-induction pellet; (3) Supernatant; (4) Fraction A4; (5) A5; (6) A7; (7) A8; (8) A9.

The samples containing Sco3914 were pooled and the concentration of the sample determined to be 1.5 mg ml$^{-1}$ by the Bradford assay. The optimum concentration for the His-tag cleavage with 3CP protease is between 1 and 2 mg ml$^{-1}$, and so the sample was buffer exchanged into buffer A (Chapter 2) containing 2 mM β-mercaptoethanol (βME).

### 3.2.1   His-tag cleavage with 3C PreScission Protease

The His-tag was cleaved from the purified protein using HRV 3C protease (Chapter 2). The cleaved protein sample eluted from the column in the tag cleavage buffer (Figure 3.4) whilst the protease and the non-cleaved sample eluted after the application of 100% buffer B (500 mM imidazole). The cleaved sample was judged to be >95% pure by SDS-PAGE analysis (Figure 3.5).
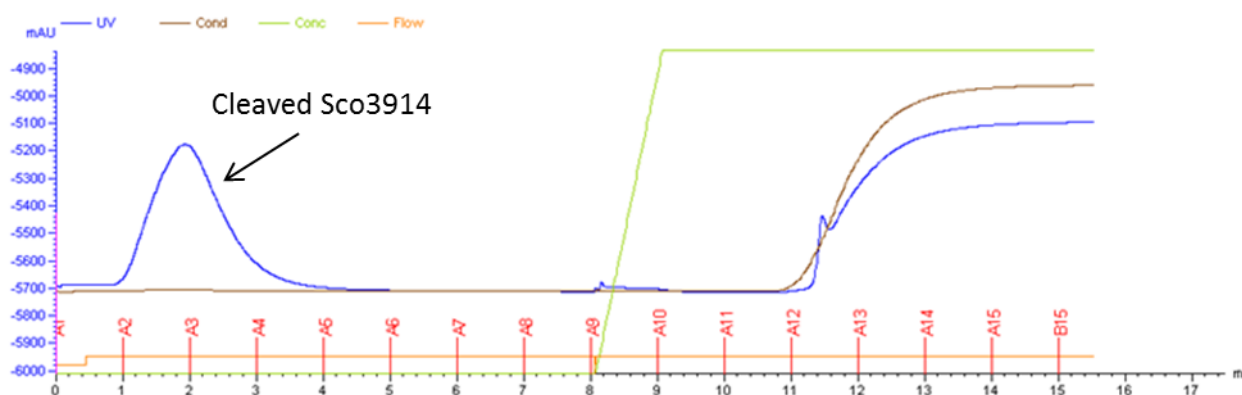


**Figure 3.4** Elution profile of Sco3914 from the nickel affinity column post-tag cleavage. Cleaved Sco3914 elutes in fractions A2-A4 whilst the absorbance that is observed in fractions A12-B15 is mainly due to the imidazole in Buffer B.
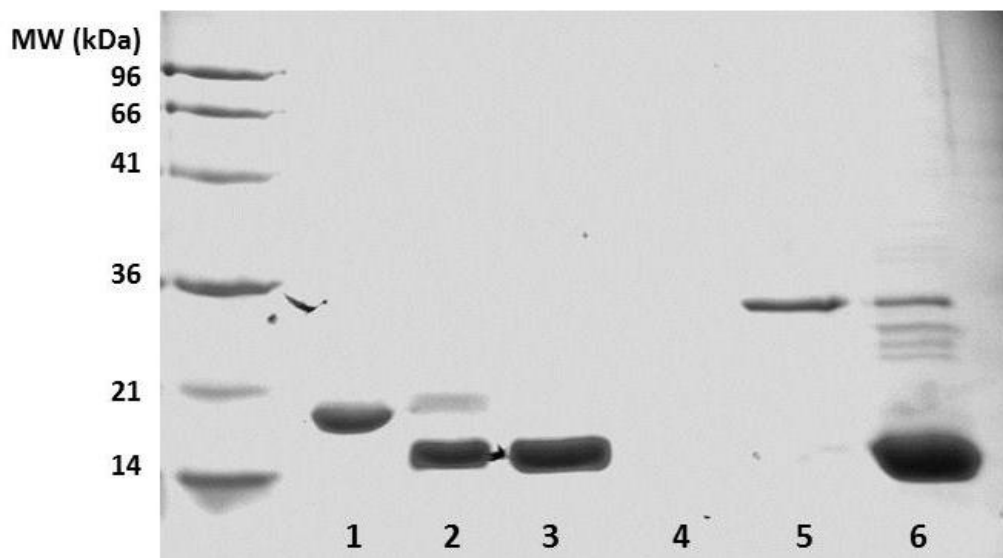
**Figure 3.5** 15% SDS-PAGE after purification of tag-cleaved sample using nickel-affinity column purification. The columns contain: (1) Uncleaved Sco3914; (2) cleavage reaction sample (16 hours after start of reaction); (3) A2; (4) A6; (5) A13; (6) Sco3914 (tag cleaved sample from previous preparation).

The cleaved protein samples were pooled and dialyzed into 20 mM Tris-HCl pH 7.5 and 0.2 M NaCl buffer and applied to the Superdex 75 (16/60) gel filtration column (GE Healthcare). The protein elutes in two oligomeric states, of which the most abundant species (represented by the highest peak in Figure 3.6) corresponds to a protein of 45210 Da (using the standard curve generated by the calibration of this column). Although this is larger than the calculated molecular weight of 33444 Da for the tag-cleaved dimer, it is in good agreement with gel filtration results for the tagged protein which eluted at a volume which corresponded to a molecular weight of ~ 55 kDa (calculated MW of tagged dimer is 37416 Da). Dynamic light scattering (Figure 3.7) was also carried out on the tagged protein (method described in Chapter 2). The data were analyzed using the DYNAMICS V6 software (Wyatt Technology Corp.) and showed a polydispersity of 26% and a molecular size estimate of 52 kDa, comparable to the 55 kDa estimated from the gel filtration column. All three values are larger than expected for the Sco3914 dimer, but this could be due to the shape of the protein.
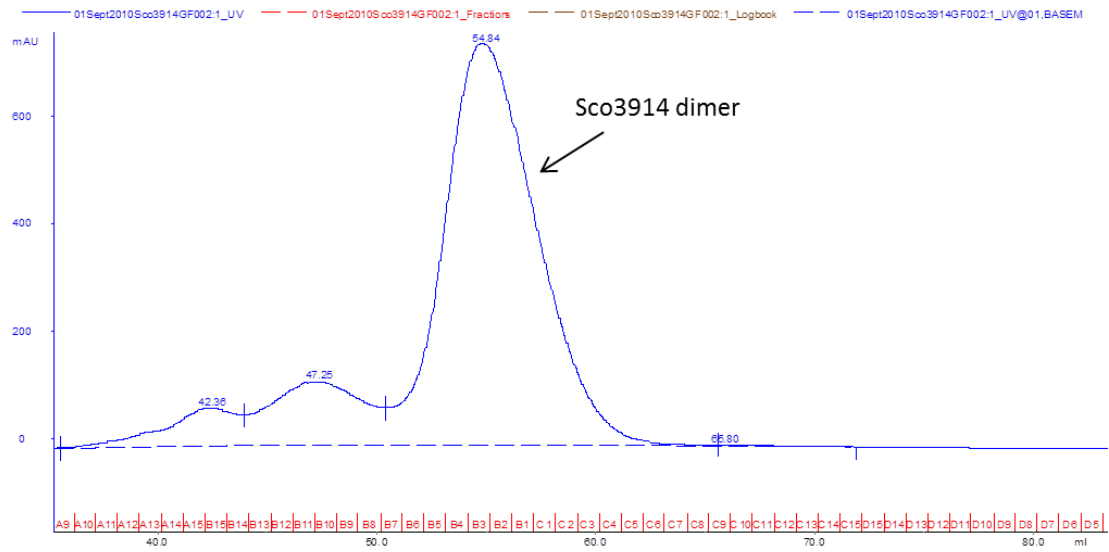
**Figure 3.6** Elution profile of Sco3914 following size exclusion gel purification.
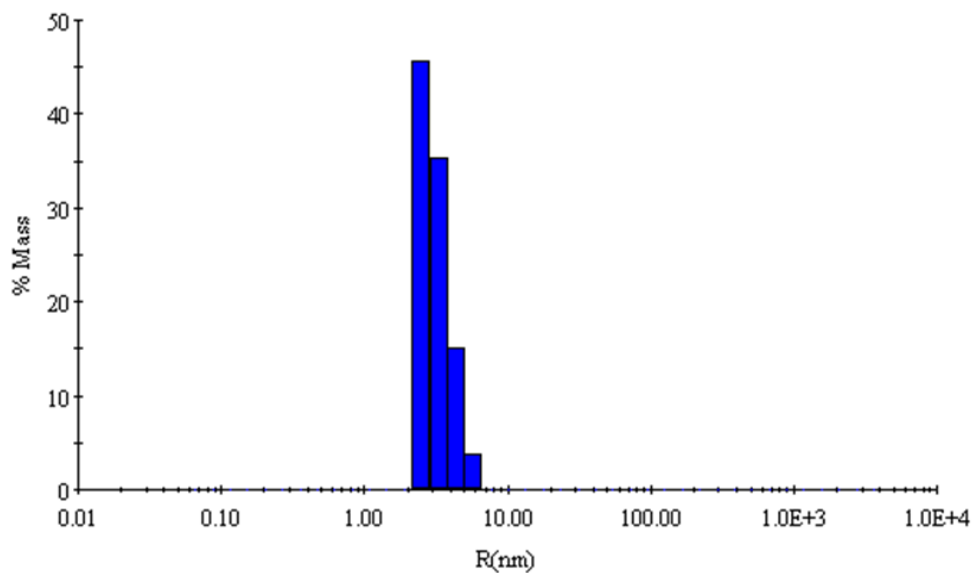


| Item | R (nm) | %Pd | MW-R (kDa) | %Int | %Mass |
|--------|--------|------|----------|-------|-------|
| Peak 1 | 3.2 | 26 | 52 | 62.1 | 99.9 |
| Peak 2 | 28.6 | 19.7 | 8624 | 17.7 | 0 |
| Peak 3 | 133.3 | 18.4 | 315461 | 14.4 | 0 |
| Peak 4 | 3619.4 | 0 | 7.14E+08 | 5.8 | 0 |

**Figure 3.7** Dynamic Light Scattering profile of Sco3914. The peaks represent the species of a particular size (radius) as a percentage of the total mass present in solution. There is a predominant species with an estimated size of 52 kDa present in the solution.

## 3.3 X-ray Crystallography and Data Collection

### 3.3.1 Data collection

Initial crystal screening of the His-tagged protein did not produce crystals in any of the 5 screens tested. An analysis of the tagged sequence was performed in order to determine the possible cause of this reluctance to crystallize. The sequence of Sco3914 plus the 6-His tag is shown below:-

3CP cleavage site

==MAHHHHHHSSGLEVLFQ==GPTATDPALTALSQGWCALSLLHGRIEAHIERALQTGHDLSVREYSLLDVLSRQHD
GDGGHLQMKQVADAVVLSQSATTRLVTRLEDRGLLERYLCPTDRRGIYTNVTPAGLQLLDEARPTNDTALREAL
DDAAANPDLAPLVRAVETLKAPVPA
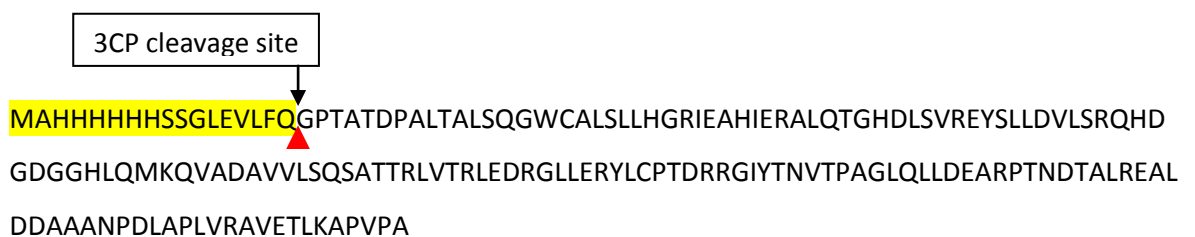
3CP proteolytic cleavage of the His-tag gives the following sequence which differs from the native sequence in that the initial methionine is replaced by two residues: a glycine and a proline:-

==GP==TATDPALTALSQGWCALSLLHGRIEAHIERALQTGHDLSVREYSLLDVLSRQHDGDGGHLQMKQVADAVVL
SQSATTRLVTRLEDRGLLERYLCPTDRRGIYTNVTPAGLQLLDEARPTNDTALREALDDAAANPDLAPLVRAVET
LKAPVPA
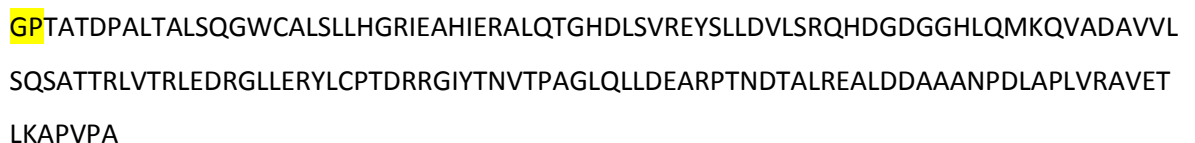
The Disopred server (http://bioinf.cs.ucl.ac.uk/disopred/) (Ward *et al.*, 2004) predicted that there was a high probability of disorder in the first 24 residues at the N-terminus. Disorder prediction analysis of the tag cleaved sequence suggested that only the first five residues would be disordered.

Disordered profile plot



a

Disordered profile plot
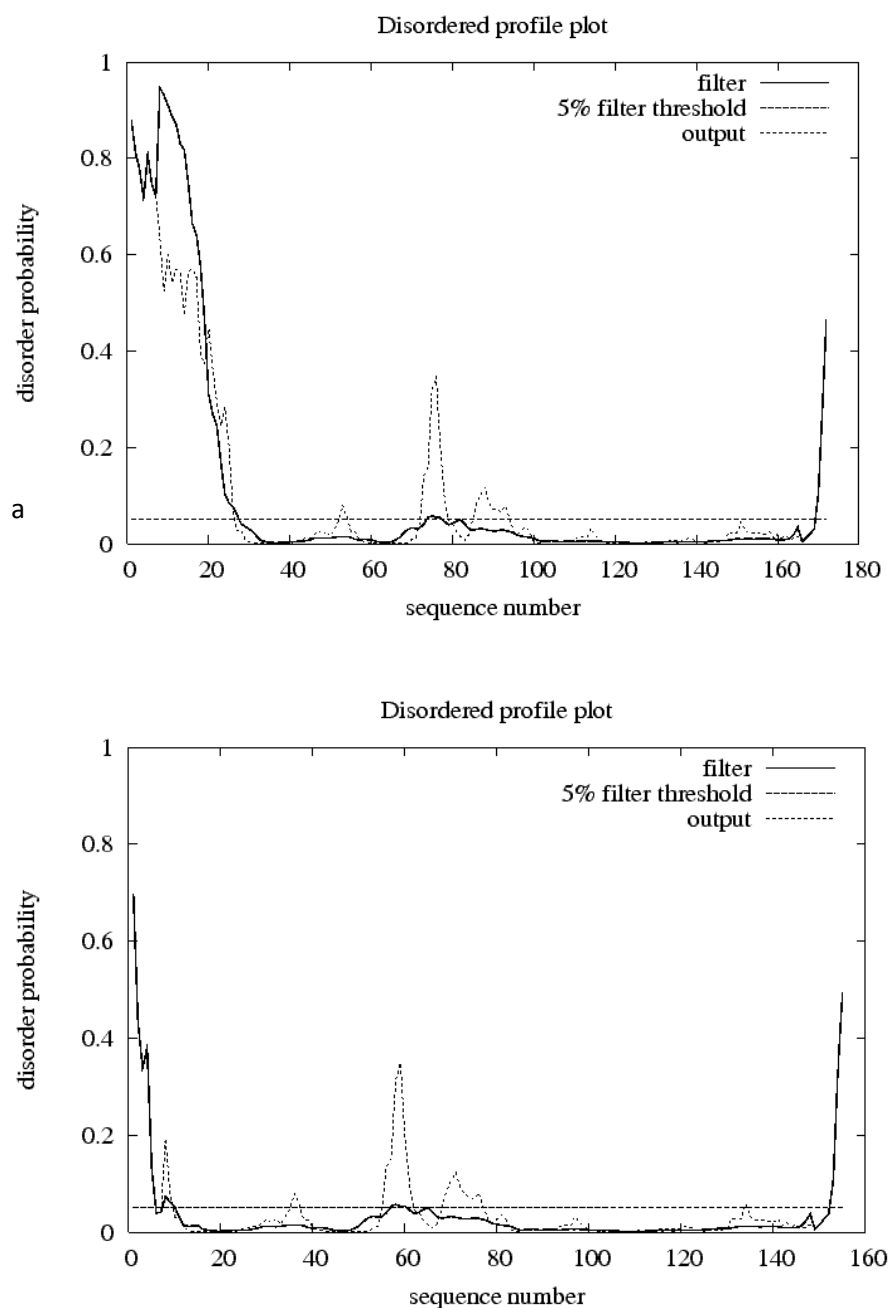


**Figure 3.8** Disopred server predictions of the disorder probability of (a) tagged sequence of Sco3914; (b) tag cleaved sequence of Sco3914. This analysis suggests that removing the His-tag may decrease the disorder of the N-terminus and so increase the potential to crystallize the protein.

Removing the tag yielded crystals (Figure 3.9) in several conditions of the PACT screen (Molecular Dimensions). The protein:well solution ratio was 1:1 (0.6 µl total volume) and

88

the protein was filtered prior to dispensing using a 0.1 µm Ultrafree-MC centrifugal filter device (Millipore).
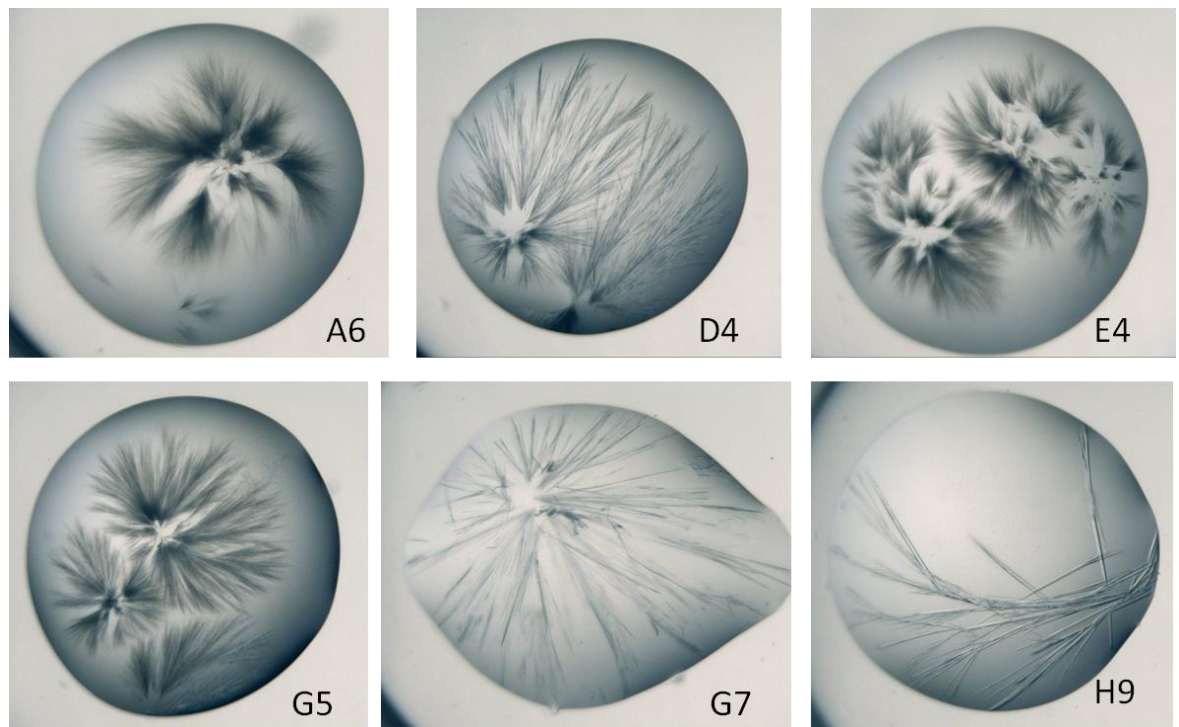


**Figure 3.9** Crystals of Sco3914 grown in a variety of conditions in the PACT crystallization screen.

The H9 condition (containing 25% PEG 3350, 0.1 M Bis-Tris propane pH7 and 0.2 M Na Malonate pH 5.5) was optimized in order to improve crystal quality. Optimizations were set up in 24-well hanging drop plates with 1 ml well solution. 1 µl of protein and 1 µl precipitant solution were combined in the hanging drop. 15% glycerol was added to the well solutions prior to crystallization in order to provide a cryoprotectant for the crystal. Sco3914 crystals tolerated this additive well and the diffraction of the crystals was slightly improved. The morphology of the crystals also seemed to improve as they appeared to be more robust than the ones from the initial screen (Figure 3.10).
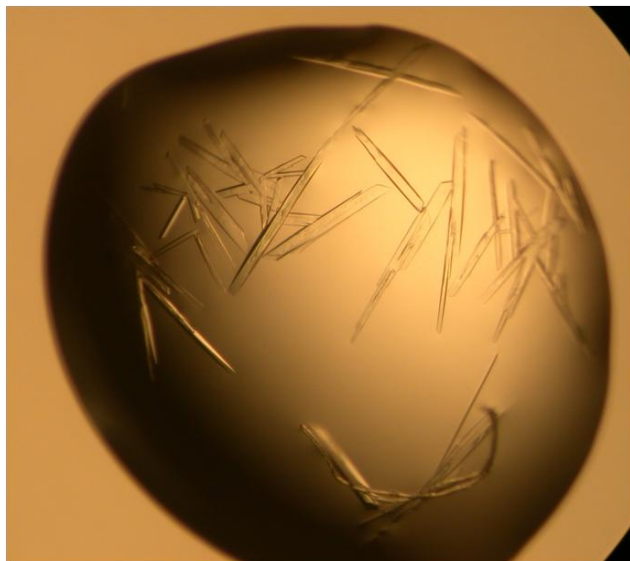
**Figure 3.10** Sco3914 crystals in the optimized PACT H9 condition (25% PEG 3350; 0.1 M Bis-Tris propane pH7; 0.2 M Na Malonate pH 5.5 and 15% glycerol).

Many of these rod shaped crystals were grown from this condition and attempts were made to obtain derivative crystals for structural determination. These attempts included soaks with a variety of heavy atom compound solutions, co-crystallization and soaking with halide solutions and preparation of selenomethionine-substituted protein (described in Chapter 2). A successful derivative dataset was collected from a crystal which had been soaked for ~10 seconds in the well solution containing 1 M potassium iodide (Dauter *et al.*, 2000). The incorporation of halide ions into the crystal occurs rapidly and so this substitution did not damage the crystal. This dataset was collected in-house using a Mar345 image-plate detector (MarResearch) mounted on a Rigaku RU-H3RHB rotating-anode X-ray generator (operated at 50 kV and 100 mA) fitted with Osmic confocal optics and a copper target (Cu K$\alpha$; $\lambda$ = 1.542 Å). Iodide has a significant anomalous signal at this wavelength (Evans & Bricogne, 2002). A highly redundant dataset comprising 2385 x 1° images (~6.6 full rotations) was then recorded to 2.44 Å resolution in-house. Sco3914 crystals belonged to space group P1 with approximate cell parameters of a = 42, b = 59, c = 69 Å and $\alpha$ = 90, $\beta$ = 95, $\gamma$ = 108°. The native data were collected on beamline i04 at the Diamond Light Source (UK) to a maximum resolution of 1.5 Å using an ADSC Quantum 315

CCD detector. The collection was carried out in two passes: a low resolution pass at 2.5 Å resolution and a high resolution pass at 1.5 Å resolution.
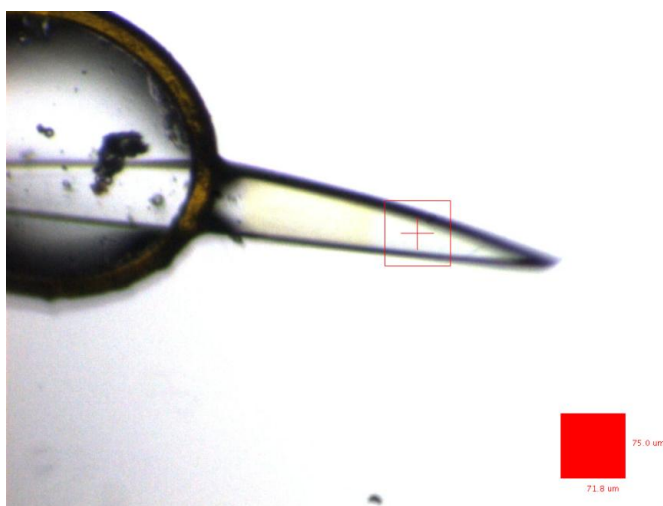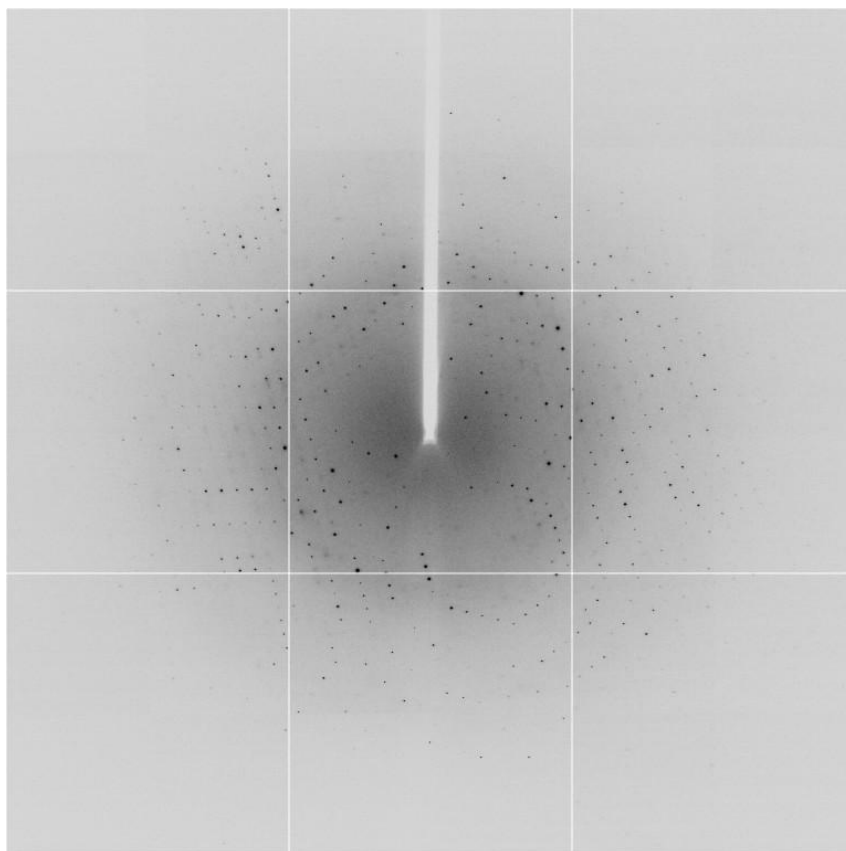


**Figure 3.11**  (a) Diffraction image of Sco3914 native crystal (data collected at Diamond Light Source Synchrotron station i04); (b) native crystal in the loop. The crystal was approximately 700 x 75 x 50 µm in size. The red box indicates the dimensions of the beam.

### 3.3.2    Structural determination

All data were processed using MOSFLM (Leslie *et al.*, 2007) and integrated using SCALA (Evans, 2006). The native and iodide datafiles were fed into PHENIX (Adams *et al.*, 2010) which used the AutoSol wizard (Terwilliger *et al.*, 2009) to solve the structure by the Single Isomorphous Replacement with Anomalous Scattering (SIRAS) method. This wizard uses the datafiles containing the amplitudes of structure factors and identifies the heavy atom sites (using HySS) and calculates the associated phases (using SOLVE). Three good iodide sites were identified, giving a figure of merit of 0.793 to 2.46 Å resolution following density modification (using RESOLVE). There are 4 copies of the monomer in the asymmetric unit (giving a corresponding solvent content of 48%). These phases were then fed into AutoBuild (Terwilliger *et al.*, 2008) which, through many cycles of building and refinement, generated a partial model. This partial model was quite incomplete (401 residues had been fitted) yet the predominantly α-helical structure characteristic of MarR family proteins was apparent. However, the electron density was quite fragmented, even though the existing density was of a reasonable quality, and this would make a manual rebuild of the four chains quite time-consuming.

The density modification performed by RESOLVE had not determined the NCS operators for the expected four-fold NCS and so the next step was to determine these NCS operators in order to improve phases by density modification using four-fold averaging. Several structural templates had already been generated by the PHYRE server (Kelley & Sternberg, 2009) during previous attempts at solving the structure by molecular replacement. One of these models was derived from HucR from *Deinococcus radiodurans* (PDB code 2FBK (Bordelon *et al.*, 2006)). Using a combination of the SSM superpose (Rigali *et al.*, 2006) option and manual superposition in COOT (Emsley *et al.*), four copies of this template were placed onto the partial model from PHENIX, giving two recognisable MarR-like dimers in the asymmetric unit. The experimental  phases from PHENIX were calculated to 2.46 Å resolution (the resolution of the iodide dataset), but it was possible to extend these phases to 1.5 Å resolution by running rigid body refinement with the partial model against the native dataset in REFMAC5 (Murshudov *et al.*, 1997). The

phases were improved by density modification (the NCS operators were obtained from the four docked molecular replacement templates) using four-fold averaging in PARROT (Cowtan), giving a FOM of 0.553 to 1.5 Å resolution. These phases were used to calculate electron density maps which were improved enough to generate a model from scratch using BUCCANEER (Cowtan, 2006) with 578 residues and $R_{work}$ and $R_{free}$ values of 0.332 and 0.358 respectively. This model was used to derive the NCS operators for another cycle of density modification in conjunction with the resultant phases, giving a FOM of 0.800. Again, the electron density map was visibly improved and this time BUCCANEER was able to build a model of 609 residues from scratch with $R_{work}$ and $R_{free}$ values of 0.290 and 0.312, respectively. This model and map were good enough to perform several iterations of model-building in COOT in combination with refinement in REFMAC5. Water molecules were fitted using Arp/wARP (Perrakis *et al.*, 1999).

| Native dataset | Iodide dataset |
|---|---|
| *P*1 | *P*1 |
| 1. 5 Å | 2.4 Å |

**PHENIX** Autosol

8 iodide sites

FOM 0.381 to 2.46 Å

**RESOLVE**

Density modification

FOM 0.793 to 2.29 Å

**PHENIX** Autobuild

401 residues fitted

$R_{work}$ = 0.358  $R_{free}$ = 0.408
to 2.44 Å

**COOT**

Dock 4 copies of PHYRE
HucR template onto
PHENIX model

**REFMAC5**

RB Refine at 1. 5 Å

$R_{work}$ = 0.469  $R_{free}$ = 0.471
to 1.5 Å

Density improvement in
**PARROT** with 4-fold
averaging

FOM 0.553 to 1. 5 Å

x 2

**BUCCANEER**

1.  509 residues fitted
2.  600 residues fitted

1. $R_{work}$ = 0.332  $R_{free}$ = 0.358
to 1.5 Å
2. $R_{work}$ = 0.290  $R_{free}$ = 0.312
to 1.5 Å

Further refinement and
model-building in **COOT**
and **REFMAC5**  and
water fitting with
**ARP/wARP**

$R_{work}$ = 0.149  $R_{free}$ = 0.192
to 1.5 Å

**Figure 3.12**  Structure solution steps for Sco3914

Following further refinement, extra density surrounding the cysteines (Cys16 and Cys95) on each monomer was observed (Figure 3.13). Attempts to model the extra density as oxidized cysteine residues did not account for all the residual density. It was postulated that these were β-mercaptoethanol (βME) adducts, as this was used as a buffer component for the His-tag cleavage with 3C protease. Using the modelling options in COOT, it was possible to edit the PDB file to incorporate these modifications, changing the cysteine (CYS) to *S,S*-(2-hydroxyethyl)thiocysteine (CME). Subsequent refinement reduced the difference density significantly and the modified residue fitted the electron density well. There is a small amount of negative density in the difference density map, but this may be due to radiation damage of some of these modified residues or even disorder that was not modelled.
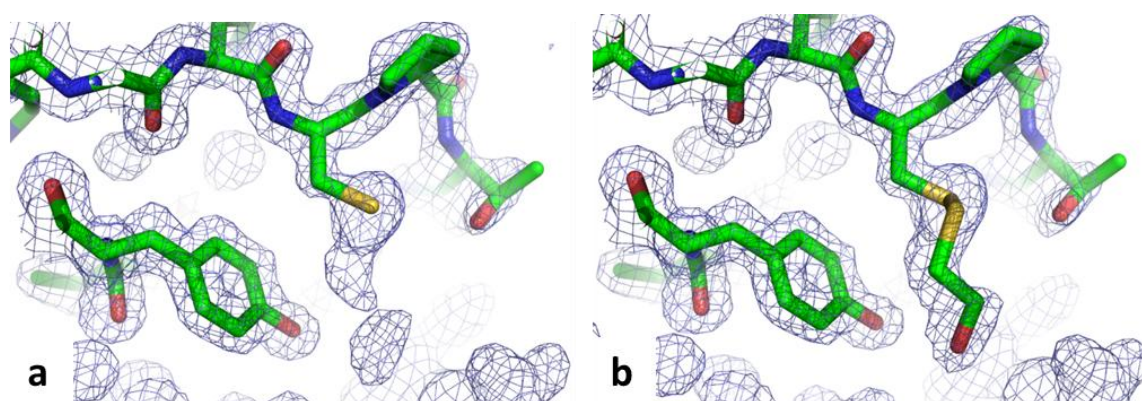


**Figure 3.13** Extra electron density surrounding the cysteine 95 molecule (chain A) in Sco3914. (a) Map at early stages of refinement with Cys95; (b) final map with β-mercaptoethanol modified cysteine (CME95). $2mF_{obs} - dF_{calc}$ maps at 1.5 Å resolution and contoured at 1.0 σ.

**Table 3.1** Summary of X-ray data for Sco3914

| Dataset | Native | Iodide derivative |
|---|---|---|
| **Number of crystals** | 1 | 1 |
| **Beamline** | I04, Diamond Light Source, UK | Rigaku RU-H3RHB rotating-anode X-ray generator |
| **Wavelength (Å)** | 0.98 | 1.54 |
| **Detector** | Quantum 315 CCD | MAR 345 image plate |
| **Rotation range per image (°)** | 1.0/0.5 | 1.0 |
| **Exposure time per image (s)** | 0.25/0.5 | 600 |
| **Total rotation range (°)** | 360/360 | 2385 |
| **Resolution range (Å)** | 38.38-1.50 (1.58-1.50) | 22.89-2.44 (2.57-2.44) |
| **Space Group** | $P1$ | $P1$ |
| **Cell parameters (Å)** | $a$ = 41.67, $b$ = 59.28, $c$ = 68.69 $\alpha$ = 90.15, $\beta$ = 95.17, $\gamma$ = 107.88 | $a$ = 41.77, $b$ = 59.14, $c$ = 68.91 $\alpha$ = 89.66, $\beta$ = 95.65, $\gamma$ = 108.05 |
| **Estimated mosaicity (°)** | 0.24 | 1.00 |
| **Total no. of measured intensities** | 359794(52249) | 542904(55481) |
| **Unique reflections** | 94894(13729) | 21246(2342) |
| **Multiplicity** | 3.8 (3.8) | 12.8 (12.3) |
| **Mean $I/\sigma(I)$** | 10.8 (3.4) | 35.0 (13.8) |
| **Completeness (%)** | 95.1 (94.2) | 90.9 (68.1) |
| **$R_{\text{merge}}$**[†] | 0.071 (0.341) | 0.095 (0.252) |
| **$R_{\text{meas}}$**[‡] | 0.082 (0.396) | 0.099 (0.263) |
| **Wilson $B$ value (Å$^2$)** | 19.1 | 25.5 |

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. [‡] $R_{\text{meas}} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the $i$th observation of reflection $hkl$, $\langle I(hkl) \rangle$ is the weighted average intensity for all observations $i$ of reflection $hkl$ and $N$ is the number of observations of reflection $hkl$.

**Table 3.2** Summary of refined model parameters for Sco3914

| | |
|---|---|
| Resolution range (Å) | 69.38 - 1.50 |
| $R_{work}$/ $R_{free}$[†] | 0.1488/0.1918 |
| Ramachandran plot: favoured/allowed/disallowed[‡] (%) | 98.7/100/0 |
| R.m.s. bond distance deviation (Å) | 0.013 |
| R.m.s. bond angle deviation (º) | 1.5 |
| No. of protein residues: chain A/chain B/chain C/chain D | 154/154/154/154 |
| No. of water molecules | 602 |
| Mean $B$ factors: protein/water/overall (Å$^2$) | 20.5/32.4/21.9 |

[†] The R-factors $R_{work}$ and $R_{free}$ are calculated as follows: $R = \sum(|\ F_{obs} - F_{calc}\ |)/\sum|\ F_{obs}\ | \times 100$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitudes, respectively. [‡] As calculated using MOLPROBITY (Chen *et al.*, 2010, Davis *et al.*, 2007).

Structure validation was carried out using the validation tools in COOT and Molprobity (Chen *et al*., 2010, Davis *et al*., 2007). The Ramachandran plot (Figure 3.14) indicates that there are no outliers in the structure. The Molprobity analysis is good, scoring in the 98th percentile (Table 3.3).

**Table 3.3** Results from Molprobity analysis of Sco3914 structure.

| | | | |
|---|---|---|---|
| All-Atom Contacts | Clashscore, all atoms: | 3.62 | 96th percentile[*] (N=598, 1.50Å ± 0.25Å) |
| | Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms. | | |
| Protein Geometry | Poor rotamers | 0.00% | Goal: <1% |
| | Ramachandran outliers | 0.00% | Goal: <0.2% |
| | Ramachandran favored | 98.64% | Goal: >98% |
| | Cβ deviations >0.25Å | 0 | Goal: 0 |
| | MolProbity score^ | 1.15 | 98th percentile[*] (N=4836, 1.50Å ± 0.25Å) |
| | Residues with bad bonds: | 0.00% | Goal: 0% |
| | Residues with bad angles: | 0.00% | Goal: <0.1% |

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

## MolProbity Ramachandran analysis

sco3914_coot-CME-coot-70_refmac6FH.pdb, model 1
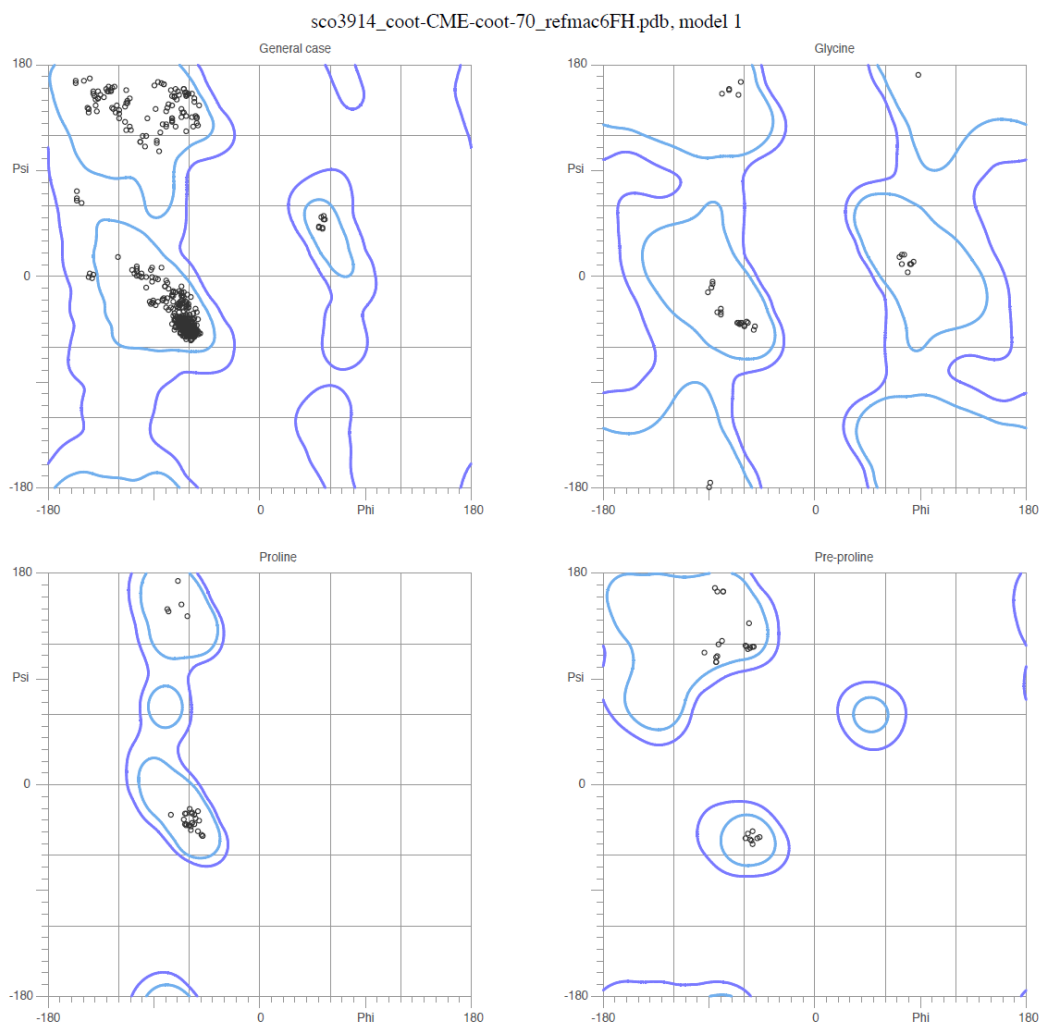


98.7% (593/601) of all residues were in favored (98%) regions.
100.0% (601/601) of all residues were in allowed (>99.8%) regions.

There were no outliers.

http://kinemage.biochem.duke.edu                    Lovell, Davis, et al. Proteins 50:437 (2003)

**Figure 3.14** Ramachandran plot (from Molprobity) for the crystal structure of Sco3914. There are no residues in disallowed regions whilst 98.7% of residues are in favoured regions.

### 3.3.3 Structural analysis

Sco3914 has the typical MarR features: it is a helical-rich dimer with the triangular MarR shape (and there are two dimers present in the asymmetric unit of this structure). It has the topology α1-α2-α3-α4-β1-β2-α5-α6, with α4 constituting the recognition helix in the helix-turn-helix motif, whilst the wing consists of two antiparallel β-strands, β1 and β2 and the intervening loop. These are the DNA-binding domains in most MarR homologues. The dimerization interface comprises the α1, α5 and α6 helices from each monomer which intertwine with each other.
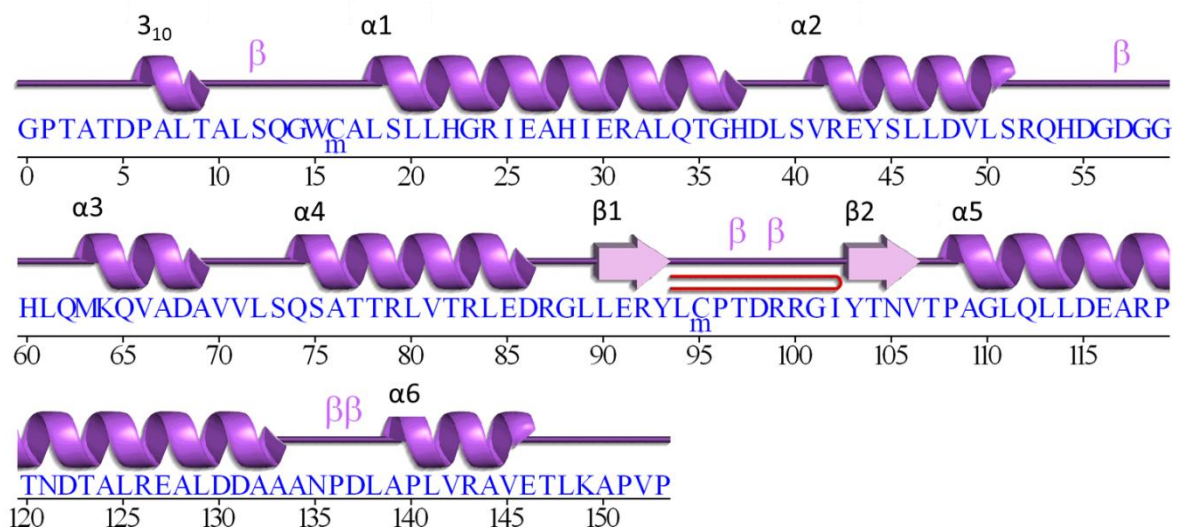


**Figure 3.15** Secondary structure analysis of Sco3914 generated by PDBSum (http://www.ebi.ac.uk/pdbsum/) (Laskowski, 2009).

The total interface as detailed by the PISA server (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) differs between the two dimers in the asymmetric unit: the calculated interface area between A and B is 2841 Å (44% of residues) whilst for C and D, it is 2174 Å (37% of residues). These values are quite different and superposition of chains A and B (dimer 1) onto chains C and D (dimer 2) give a root mean square deviation (RMSD) of 1.0 Å although the conformations of the two are

not obviously very different (Figure 3.16). For the superpositioning of structural homologues onto Sco3914, chains A and B were used for the dimer superposition whilst chain A1 was used for the monomer. The results for the secondary structure matching (SSM) are shown in Table 3.4 and the superpositions of some of the models are shown in Figure 3.17. Crystal contacts are reported between the two dimers: Chains B and C (from different dimers) make salt bridge and H-bond contacts through Asp115 (B) and Arg79 (C). This is also observed between chains A and D.
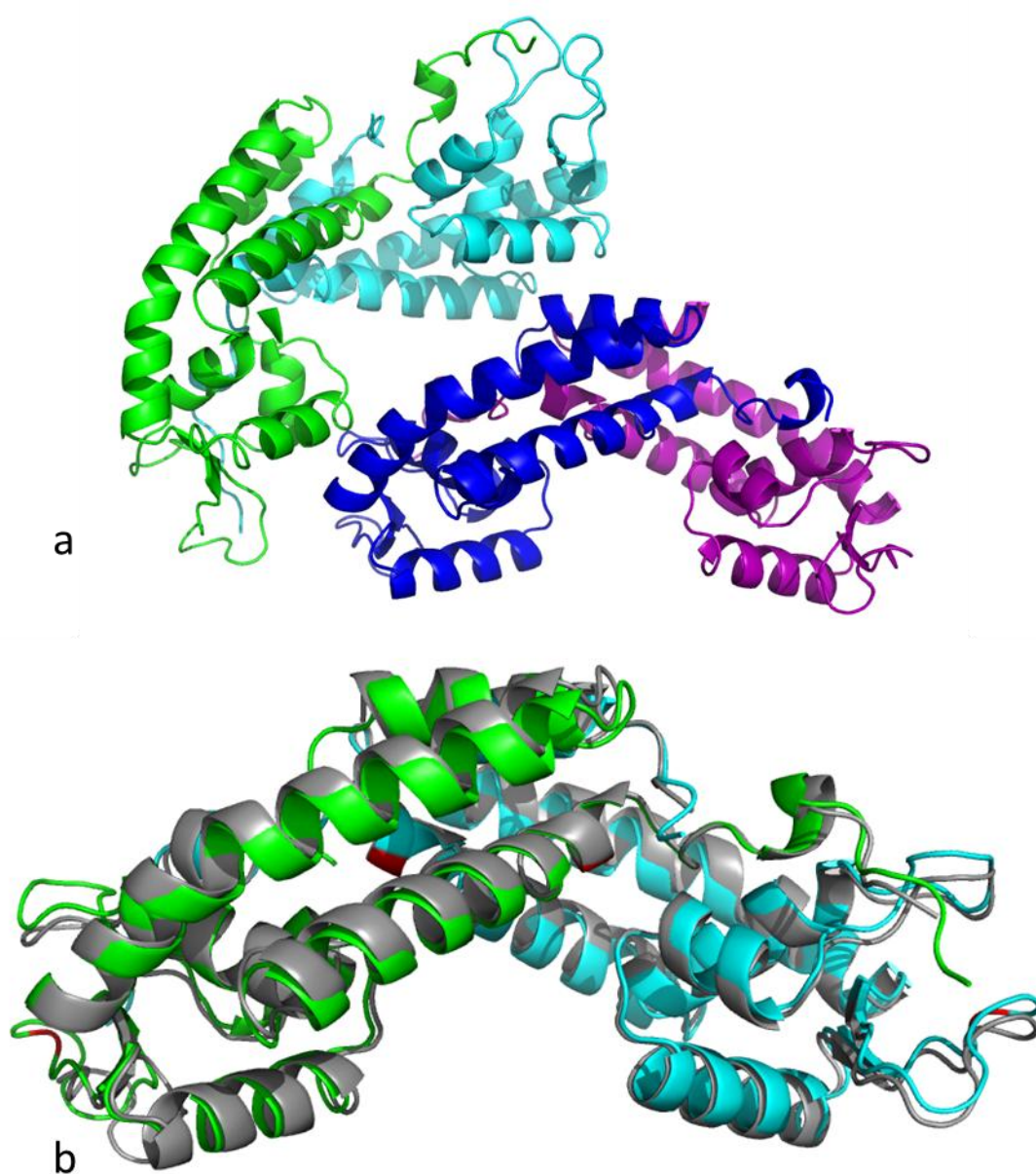


**Figure 3.16** (a) Two Sco3914 dimers in the asymmetric unit; (b) Chains C and D (grey) superimposed onto chains A and B (green and cyan). The red colour indicates the position of the cysteines.

Sco3914 has a shape which is relatively flat compared to other MFRs. However, as the superpositions show (Figure 3.17), it could still potentially be in a conformation which is compatible with DNA binding. The twist of the DNA recognition helices of Sco3914 is slightly different to the SlyA- and OhrR-DNA complexes, but the distance between the helices appears very similar, suggesting that it may be in a conformation for binding DNA. Although HypR. also has a dimeric shape which is relatively "flat" compared to other MFRs, superposition of the dimer was not possible due to its more unusual shape. The α4-α4' helices in HypR shift only a small distance when comparing the reduced and oxidized forms of this protein: from 30 Å to 27 Å respectively. The superposition of the monomer can be seen in Figure 3.17 and the locations of the two cysteines on each protein molecule are quite different.

**Table 3.4** Structural homologues of Sco3914

| Protein | Source | DNA or ligand | PDB code | Dali | | | R.m.s. deviation (Å)/ aligned residues | | Reference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Z-score | Alignment length | Seq. ID (%) | Subunit | Dimer | |
| Transcriptional regulator | *Rhodococcus jostii RHA1* | - | 3FM5 | 8.6 | 110 | 32 | 1.5/109 | 2.0/226 | To be published |
| ST1710 | *Sulfolobus tokodaii* | salicylate | 3GF2 | 7.8 | 107 | 24 | 1.8/105 | 3.4/204 | (Kumarevel *et al.*, 2009) |
| ST1710 | *Sulfolobus tokodaii* | DNA | 3GFI | 7.7 | 108 | 24 | 2.3/106 | 5.5/195 | (Kumarevel *et al.*, 2009) |
| Probable regulator | *Pseudomonas aeruginosa* | - | 2NNN | 7.2 | 114 | 22 | 1.8/111 | 3.0/221 | To be published |
| SarZ | *Staphylococcus aureus* | Sulfenic acid form | 3HRM | 6.8 | 100 | 17 | 2.5/93 | 3.4/189 | (Poor *et al.*, 2009) |
| SlyA | *Salmonella enterica* | DNA | 3Q5F | 6.8 | 112 | 24 | 2.0/110 | 2.4/220 | (Dolan *et al.*, 2011) |
| SarZ | *Staphylococcus aureus* | Mixed disulfide | 3HSR | 6.7 | 99 | 17 | 2.4/97 | 3.4/189 | (Poor *et al.*, 2009) |
| MarR | *E. coli* | salicylate | 1JGS | 6.5 | 114 | 17 | 2.2/112 | 2.9/145 | (Alekshun *et al.*, 2001) |
| OhrR | *Bacillus subtilis* | DNA | 1Z9C | 6.4 | 116 | 18 | 2.2/98 | 3.2/190 | (Hong *et al.*, 2005) |
| BldR | *Sulfolobus Solfataricus* | - | 3F3X | 6.2 | 108 | 21 | 2.7/105 | 2.3/215 | (Di Fiore *et al.*, 2009) |
| SarZ | *Staphylococcus aureus* | Reduced form | 3HSE | 5.8 | 85 | 16 | 2.8/87 | 3.3/167 | (Poor *et al.*, 2009) |
| MgrA | *Staphylococcus aureus* | - | 2BV6 | 5.5 | 113 | 13 | 2.7/92 | 4.0/203 | Chen *et al.* |
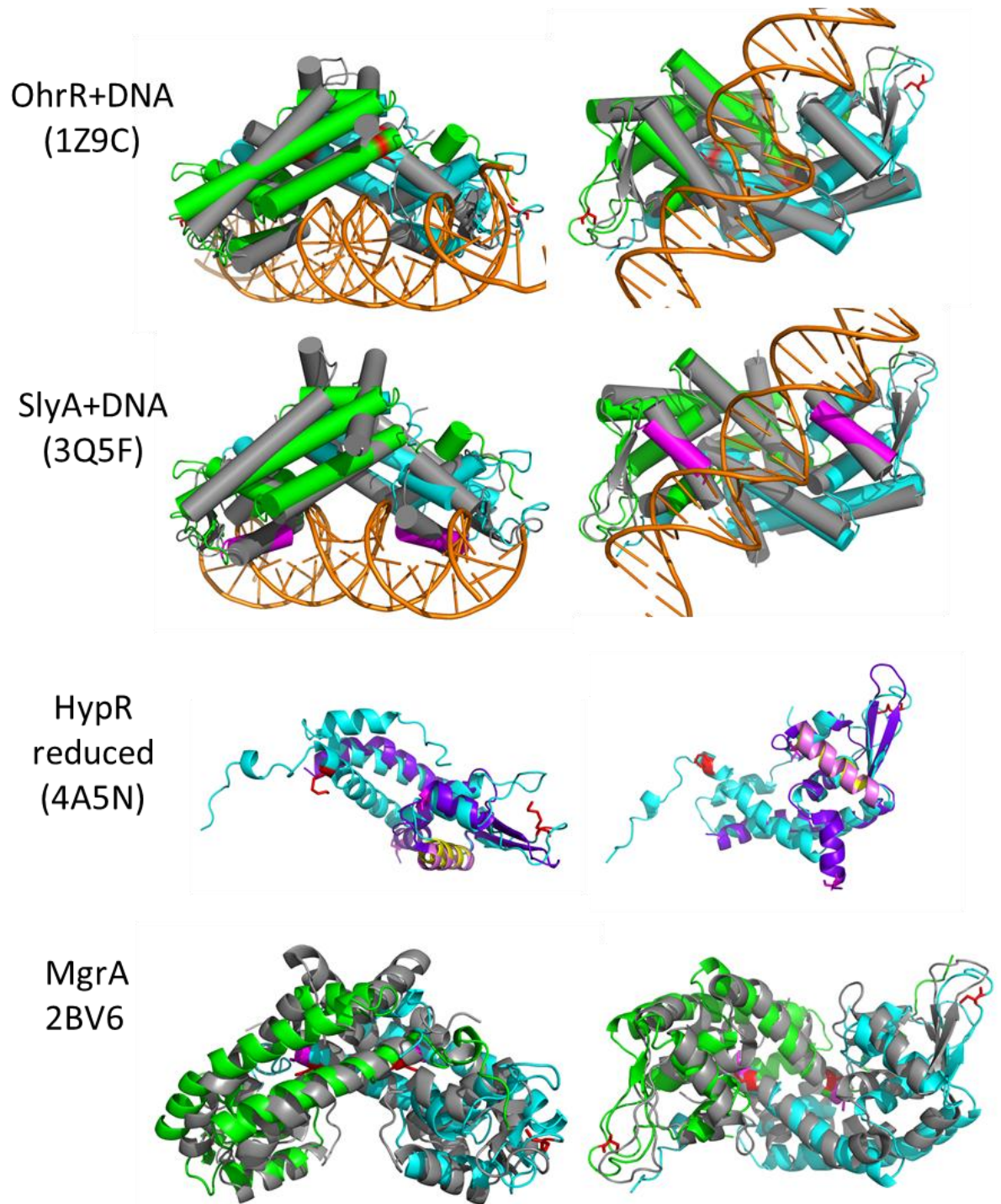
**Figure 3.17** Superimpositions of structures of MFRs from other organisms (in grey) on Sco3914 (green and cyan). Views are from the front and, rotating by 90°, from underneath. The cysteines are in red (Sco3914) or magenta (HypR and MgrA).
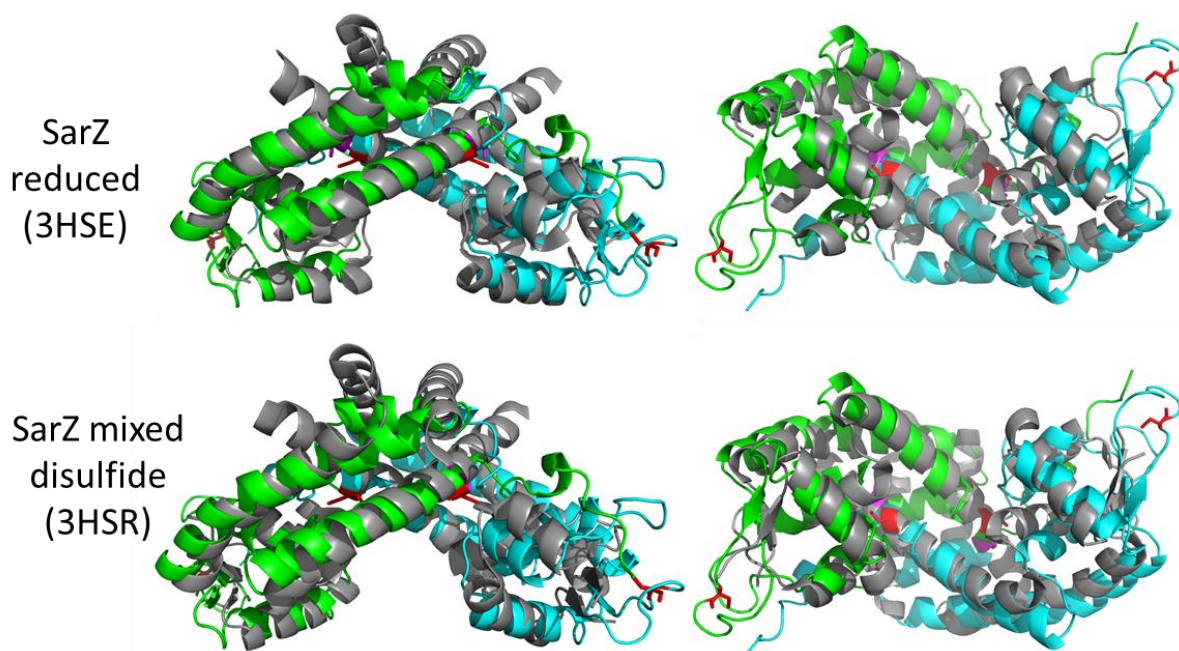
**Figure 3.18** Superposition of Sco3914 onto the crystal structures of SarZ reduced and SarZ mixed disulphide (Poor *et al.*, 2009). The Sco3914 cysteines are in red; the SarZ cysteine is in magenta.

SarZ is a global transcriptional regulator from *Staphylococcus aureus* which uses a single cysteine, Cys13, to sense peroxide stress (Poor *et al.*, 2009). The superimposition of Sco3914 onto the three different SarZ forms reveals that, at the dimer level, Sco3914 superimposes onto the reduced form better than the sulfenic acid or mixed disulphide forms. However, at the subunit level, superimposition upon the reduced SarZ monomer has the highest RMSD and the α4 helices appear to be in quite different positions compared to the superimposition upon the mixed disulphide form (Figure 3.18). These superimpositions also show that the cysteines (Cys16 on Sco3914 and Cys13 on SarZ) are located in a similar position, towards the N-terminus of the first α-helix on each monomer. Similarly, the cysteine in MgrA (Cys12), also a key global regulator in *S. aureus*, is located in this position. Single cysteine induction of transcriptional regulators is a mechanism used by many bacteria in defence against peroxide stress (see Chapter 1). Both MgrA and SarZ use this oxide sensing cysteine mechanism and so it is interesting that Sco3914 also appears to have a similarly conserved cysteine (even though it is not a single cysteine). The environment of Cys16 in Sco3914 (Figure 3.19) suggests that it could

hydrogen bond to the carbonyl oxygen of Gln13. The proximity of Trp15 could also stabilize this residue. In SarZ, the Cys13 is reported to be located in a pocket and surrounded by an extensive hydrogen bond network which can stabilize the sulfenic acid intermediate.



**Figure 3.19** The environment of cysteine 16 in Sco3914.

However, taking a sequence alignment view of Cys16 compared to other MFRs containing an N-terminal cysteine which have been shown to be redox responsive regulators, it becomes apparent that Cys16 does not have the similarly conserved surrounding residues.

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

```
                .........10.........20.........30.........40.........50
SCO3914        ------------  --GPTATDPA  LTALSQGWCA  LSLLHGRIEA  HIERALQTGH
SCO2987_OhrR   MTTPATEPAA  EPAAEPATDW  LRLDQQICFS  LSAASRAFGS  VY-RVVLKDL
XcOhrR         MDTTTATTAR  T------DTL  LQLDNQLCFA  LYSANLAMHK  LY-RGLLKAL
BsOhrR         MENK------  -------FDH  MKLENQLCFL  LYASSREMTK  QY-KPLLDKL
MgrA           ----------  -------GSH  MNLKEQLCFS  LYNAQRQVNR  YYSNKVFKKY
SarZ           ----------  -------GSH  MYLSKQLCFL  FYVSSKEIIK  KY-TNYLKEY
SarA           ----------  --------M   AITKINDCFE  LLSMVTYADK  LK-SLIKKEF
Consistency    1000000000  0000000233  6364484774  8445444526  3605264644

                .........60.........70.........80.........90........100
SCO3914        DLSVREYSLL  DVLSRQHDGD  GGHLQMKQVA  DAVVLSQSAT  TRLVTRLEDR
SCO2987_OhrR   GLTYSQYLVL  LVLWEH---G  E--LPVKRLG  EHLRLDSGTL  SPLLKRLEAA
XcOhrR         DLTYPQYLVM  LVLWET---D  E--RSVSEIG  ERLYLDSATL  TPLLKRLQAA
BsOhrR         NITYPQYLAL  LLLWEH---E  T--LTVKKMG  EQLYLDSGTL  TPMLKRMEQQ
MgrA           NLTYPQFLVL  TILWDE---S  P--VNVKKVV  TELALDTGTV  SPLLKRMEQV
SarZ           DLTYTGYIVL  MAIEND---E  K--LNIKKLG  ERVFLDSGTL  TPLLKKLEKK
SarA           SISFEEFAVL  TYISEN---K  EKEYYFKDII  NHLNYKQPQV  VKAVKILSQE
Consistency    5887468679  4584640004  4005368574  6483866566  6678868754

                .........110.........120.........130.........140.........150
SCO3914        GLLERYLCPT  DRRGIYTNVT  PAGLQLLDEA  RPTNDTALRE  ALDDAAANPD
SCO2987_OhrR   GLVRRERSTR  DERSVQVRLT  EEGAALRGRA  VEVPRRIAAA  TGL---DLAE
XcOhrR         GLVTRTRAAS  DERQVIIALT  ETGRALRSKA  GAVPEQVFCA  SAS---SLDE
BsOhrR         GLITRKRSEE  DERSVLISLT  EDGALLKEKA  VDIPGTILGL  SKQ---SGED
MgrA           DLIKRERSEV  DQREVFIHLT  DKSETIRPEL  SNASDKVASA  SSL---SQDE
SarZ           DYVVRTREEK  DERNLQISLT  EQGKAIKSPL  AEISVKVFNE  FNI---SERE
SarA           DYFDKKRNEH  DERTVLILVN  AQQRKKIESL  LSRVNKRITE  AN----NEIE
Consistency    6673947453  *7*4847388  5464465446  3454355435  5320006338

                .........160.........170.......
SCO3914        LAPLVRAVET  LKAPVPA---  -------
SCO2987_OhrR   VQDLRARLDR  LTAALDAYGT  APAPEPS
XcOhrR         LRQLKQELEK  LRSSLGAG--  -------
BsOhrR         LKQLKSALYT  LLETLHQKN-  -------
MgrA           VKELNRLLGK  VIHAFDE---  -------
SarZ           ASDIINNLRN  FVSKNF----  -------
SarA           L---------  ----------  -------
Consistency    6336232623  5233312000  0000000
```

**Figure 3.20** Sequence alignment of Sco3914 with other MFRs that utilize redox active cysteines. Sequence alignment was carried out using the PRALINE server (Simossis & Heringa, 2005).

All of the sequences in the alignment, except Sco3914, show conservation of residues surrounding the N-terminal cysteine, suggesting that this is perhaps not a well conserved cysteine in Sco3914. Nevertheless, it is obviously reactive and worthy of further investigation. Cysteines in MFRs have been shown to initiate derepression by i) formation of intersubunit disulphide bonds resulting in a dramatic conformational rearrangement or ii) modification of a single cysteine residue by an effector molecule which is

communicated to the DNA binding domain. Most recently, a new mechanism of phosphorylation of these conserved cysteines in SarA, MgrA and SarZ has been shown to occur (Sun *et al.*, 2012) and this again presents an interesting point for future work with Sco3914.

## 3.4   Site-directed mutagenesis

The modifications of the two cysteines suggested that these may be particularly reactive and so could potentially play a role in binding or recognition of DNA and/or of a ligand. Mutagenesis of these two cysteines was the logical next step to determining their roles. Initially, the decision was made to produce two mutants of both cysteines: a serine, to generate a structurally similar mutant, and an alanine. As described in Chapter 2, the oligos were designed to mutagenize these cysteines by PCR, using the Quikchange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies).  However, this strategy proved unsuccessful after many attempts and variations of the method. To further progress with this, the Sco3914C16S gene cloned into the pOPINF expression vector was ordered from GenScript. Interestingly, GenScript also appeared to have difficulties in cloning this gene. The turnaround time for production was much longer than expected as they reported problems with mutations when they sequenced the gene.

This was also the case with the C95S mutation, which was cloned into pUC19 (again by GenScript) but was problematic to subclone into pOPINF. It was decided to use pOPINF as the expression vector in an effort to keep as much of the overexpression and purification methods as similar as possible to the native protein. The gene was amplified with the pOPINF oligos using the pUC19 vector as a template and the PCR product checked by agarose electrophoresis. This step proved to be difficult as during several attempts, no bands of the correct size could be detected. The PCR reaction was optimized but the results were often inconsistent. A gradient PCR was also attempted but no product observed on agarose. Further optimization of the PCR method included a 20 times dilution of the primers which did result in a product (although this again was inconsistent

in further attempts, but it did give more product than had previously been achieved). Obtaining the PCR product was by no means the end of this problem: the next step was the InFusion reaction to clone the gene into pOPINF. Blue-white screening was carried out as described previously to detect successful transformants and these white colonies were checked by colony PCR to see if they had a band of the correct size. Subsequent sequencing performed by The Genome Analysis Centre (TGAC) returned many sequences which had insertions or deletions of random base pairs. These were predominantly at the beginning or the end of the sequence (where the primers anneal) although several were also detected in the middle of the sequence. It is still unclear as to why this gene has proved difficult to clone into an expression vector, but suggestions have included the possibility that the gene may be toxic to *E. coli,* which is why random mutations have been favoured. After screening many colonies by PCR and sequencing the purified vector from approximately 40 colonies, 6 of them proved to have the correct sequence and so it was possible to proceed with the relatively straightforward task of protein expression and purification.

### 3.4.1    Overexpression and purification of Sco3914C16S and Sco3914C95S

The mutants were overexpressed and purified as per the native protein: immobilized metal affinity chromatography followed by size exclusion gel filtration chromatography. The elution profiles and the SDS-PAGE analyses appeared to be much the same as for the native protein (Figure 3.21).
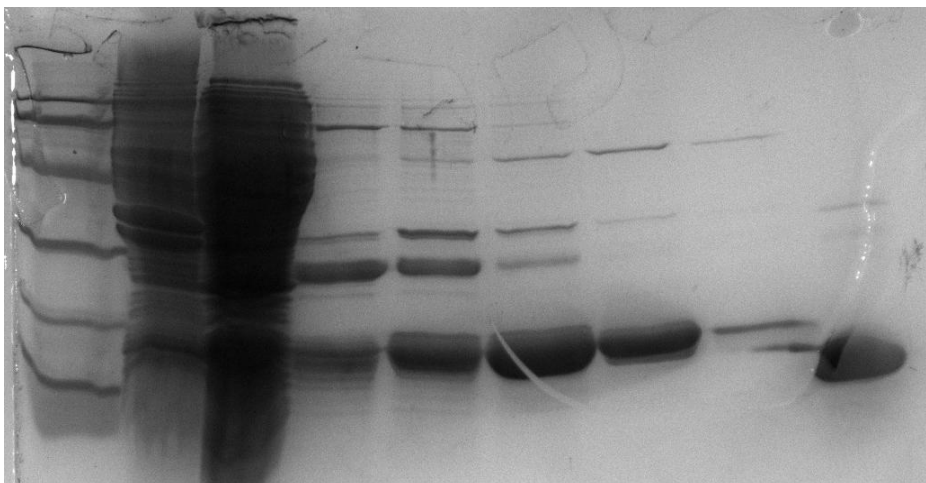
**Figure 3.21** 15% SDS-PAGE following nickel affinity purification of Sco3914C95S. The lanes contain: (1) Lyzed pellet; (2) Supernatant; (3) Fraction A4; (4) A7; (5) A8; (6) A9;( 7) A10; (8) A11.
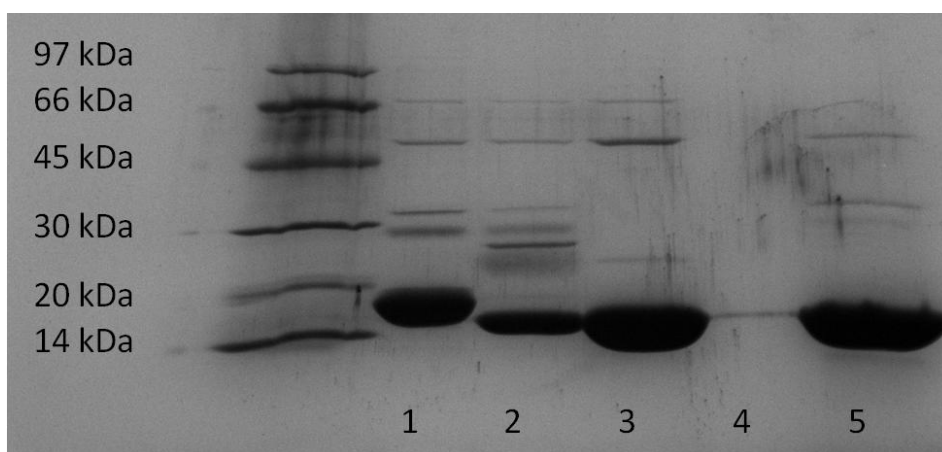


**Figure 3.22** 15% SDS-PAGE following tag cleavage with 3CP protease. Lane (1) Pre-reaction sample; (2) Post-cleavage sample; (3) Flowthrough from nickel column; (4) Sample eluted with 500 mM imidazole; (5) Previous preparation of cleaved wild type Sco3914.

### 3.4.2 Structural solution of Sco3914C16S

Sco3914C16S crystallized in the same conditions as Sco3914. Morphologically, the crystals appeared similar (Figure 3.23) although this time, the space group was $P2_12_12_1$ with cell dimensions of $a$ = 58.4 Å, $b$ = 68.6 Å, $c$ = 77.1 Å.
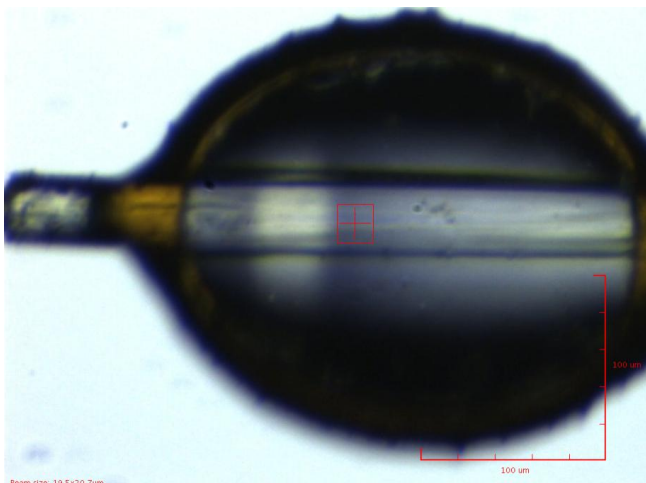
**Figure 3.23** Sco3914C16S crystal in the cryoloop at Diamond Light Source.

Data were collected at station i04 at Diamond Light Source. Several initial attempts to collect data were aborted as the spots appeared to be split. A more successful crystal was also aborted as the diffraction was severely affected by radiation damage. Translating the beam approximately 50 µm along the crystal enabled the data collection to be resumed and 200 images were collected successfully. The images from both runs were combined (using SortMTZ from the CCP4 suite) following processing and prior to scaling and merging. The structure was solved by molecular replacement, using the native Sco3914 model as the template. The data were processed using MOSFLM and scaled and merged using SCALA. Chain A of the Sco3914 model was used as the search model in PHASER, searching for two copies of the monomer in the asymmetric unit (giving a calculated solvent content of 47%).

**Table 3.5** Summary of X-ray data for Sco3914 reduced form and Sco3914C16S

| Dataset | Sco3914 reduced | Sco3914 C16S |
|---|---|---|
| Number of crystals | 1 | 1 |
| Beamline | i02, Diamond Light Source, UK | i04 |
| Wavelength (Å) | 0.91 | 0.9763 |
| Detector | Pilatus 6M | ADSC Q315r |
| Crystal-to-detector distance (mm) | 260.8 | 270.5 |
| Rotation range per image (°) | 0.4 | 0.5 |
| Exposure time per image (s) | 1.2 | 1.0 |
| Total rotation range (°) | 280 | 250 |
| Resolution range (Å) | 68.15-1.9 | 68.62-2.05 |
| Space Group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell parameters (Å) | $a$ = 60.28, $b$ = 68.15, $c$ = 76.37 | $a$ = 58.32, $b$ = 68.54, $c$ = 77.01 |
| Estimated mosaicity (°) | 0.4 | 0.6 |
| Total no. of measured intensities | 275576 (20377) | 75186(5694) |
| Unique reflections | 25038 (1801) | 18793(1363) |
| Multiplicity | 11.0 (11.3) | 4.0(4.2) |
| Mean $I/\sigma(I)$ | 13.9 (4.2) | 11.0(2.5) |
| Completeness (%) | 99.9 (100.0) | 97.0(96.4) |
| $R_{merge}$[†] | 0.121 (1.07) | 0.081(0.683) |

[†]$R_{merge} = S_{hkl} S_i |I_i(hkl) - (I(hkl))| / S_{hkl} S_i I_i(hkl)$, where $I_i(hkl)$ is the $i$th observation of reflection $hkl$ and $<I(hkl)>$ is the weighted average intensity for all observations of reflection $hkl$.

### 3.4.3   Analysis of the structures of Sco3914C16S and reduced form of native Sco3914

Overall, the structure of the mutant protein, Sco3914C16S is closely similar to that of native Sco3914, with an RMSD value for the superposition of the two of 0.82 for the dimer and 0.65 for the monomer. It exhibits almost the exact same conformation: no apparent reorientation of any loops or helices is evident. However, there remains a modification of Cys95, which was expected as the protein was prepared in the same way as the native protein. Ser16 lacks this modification but this has not altered the positioning of any surrounding residues (Figure 3.24).
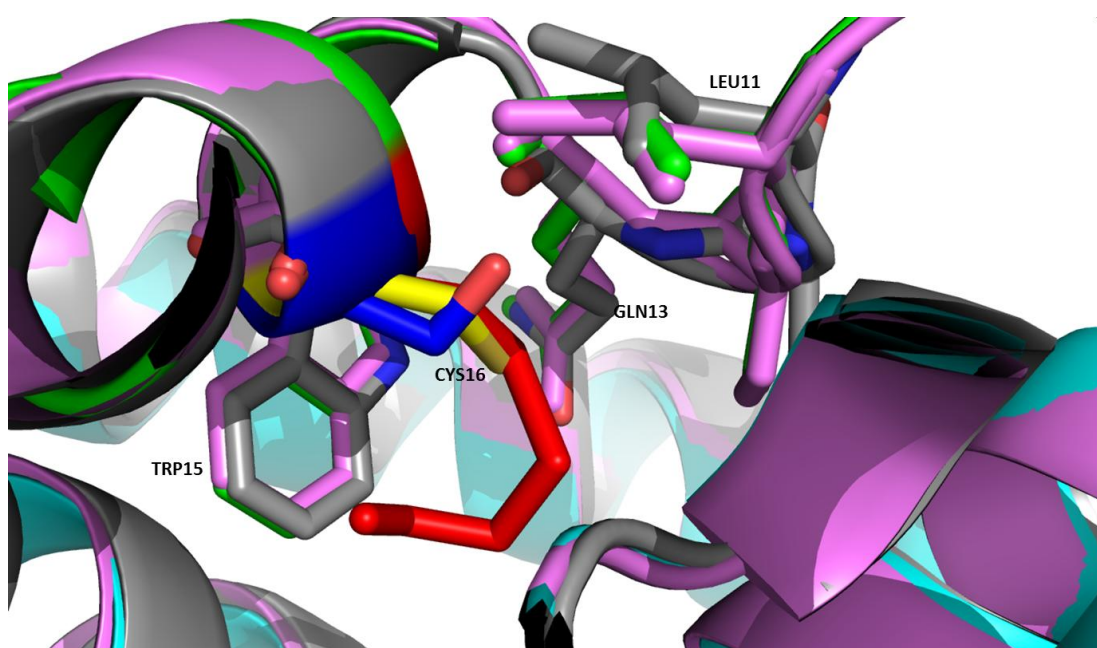


**Figure 3.24** Superimposition of Sco3914 (green and cyan) with Cys16-βME modification (red); Sco3914 reduced form (magenta) with Cys16 (yellow); Sco3914C16S (grey) with Ser16 (blue and red).

The reduced form of the protein was also crystallized under similar conditions (with the addition of 5 mM tris(2-carboxyethyl)phosphine (TCEP) as the reductant) and the structure solved using the native structure as the molecular replacement model. The reduced Sco3914 crystallized in space group $P2_12_12_1$ with cell dimensions of $a$ = 60.3 Å, $b$

= 68.2 Å, *c* = 76.4 Å. This too exhibited a closely similar conformation to the native protein in the crystal structure and the RMSD values for superposition of the dimers and monomers are 0.73 Å and 0.52 Å respectively. As with Sco3914C16S, there is certainly no significant reorientation of any of the helices and no dramatic difference in the position of the α4 recognition helices.
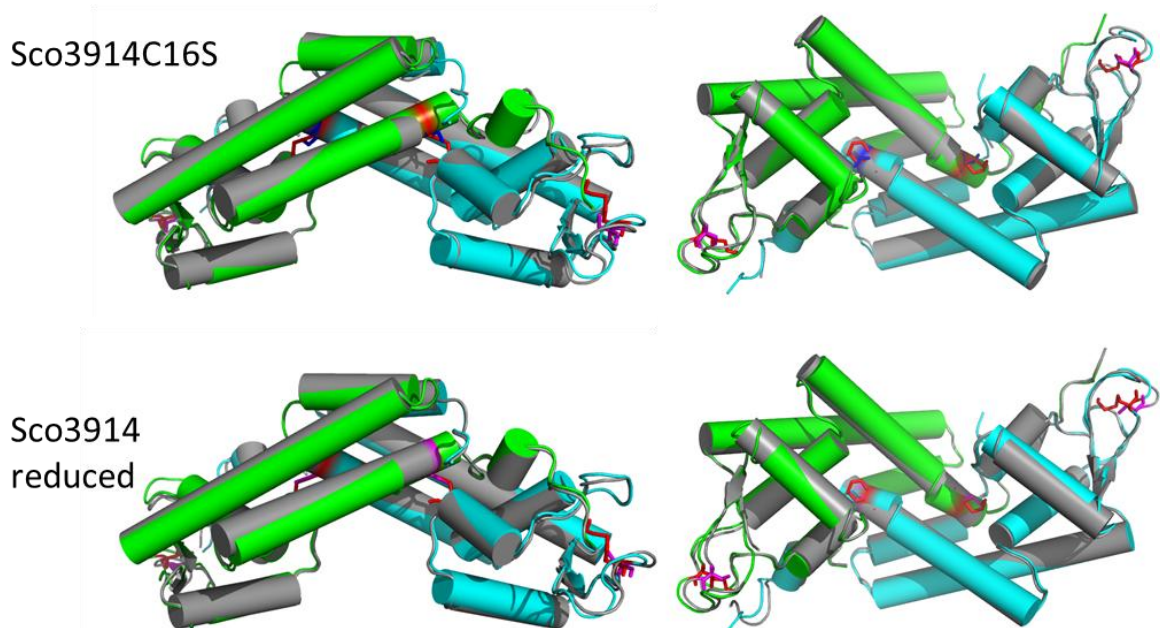


**Figure 3.25**  Superimposition of the C16S mutant and reduced Sco3914 (both structures in grey with the cysteines/serine in magenta) onto the original, cysteine modified model of Sco3914 (green and cyan with the cysteines in red).

There is no obvious conclusion that can be drawn from this result but it is likely that the protein in solution exists in a state of constant flux but prefers to crystallize only in a certain conformation, despite giving different crystal forms. Perhaps the full picture cannot be seen in this structure alone.

### 3.4.4    Confirmation of cysteine modification by mass spectrometry

The Sco3914 sample at a concentration of 10 mg ml$^{-1}$ in 25 mM HEPES, 0.5 M NaCl pH 7.5 was submitted for intact mass analysis. The molecular weights of the proteins were

determined by MALDI-ToF MS (Bruker Reflex III MALDI-ToF mass spectrometer) and the spectrum is displayed in Figure 3.26.

The results for the native protein indicate an intact mass of 16873.6 ±0.01 Da. The expected theoretical mass of the cleaved protein is 16721.8 ±0.03 Da: a difference of 151.8 Da, which suggests a covalent modification of the protein. The initial hypothesis that the two cysteines in Sco3914 are highly reactive fits well with this mass increase as it could be explained by these two residues having a covalently attached β-mercaptoethanol adduct. β-mercaptoethanol was used in the tag-cleavage buffer at a concentration of 2 mM and the individual molecules have a mass of 76 Da, corresponding to modification of both cysteines in each monomer.
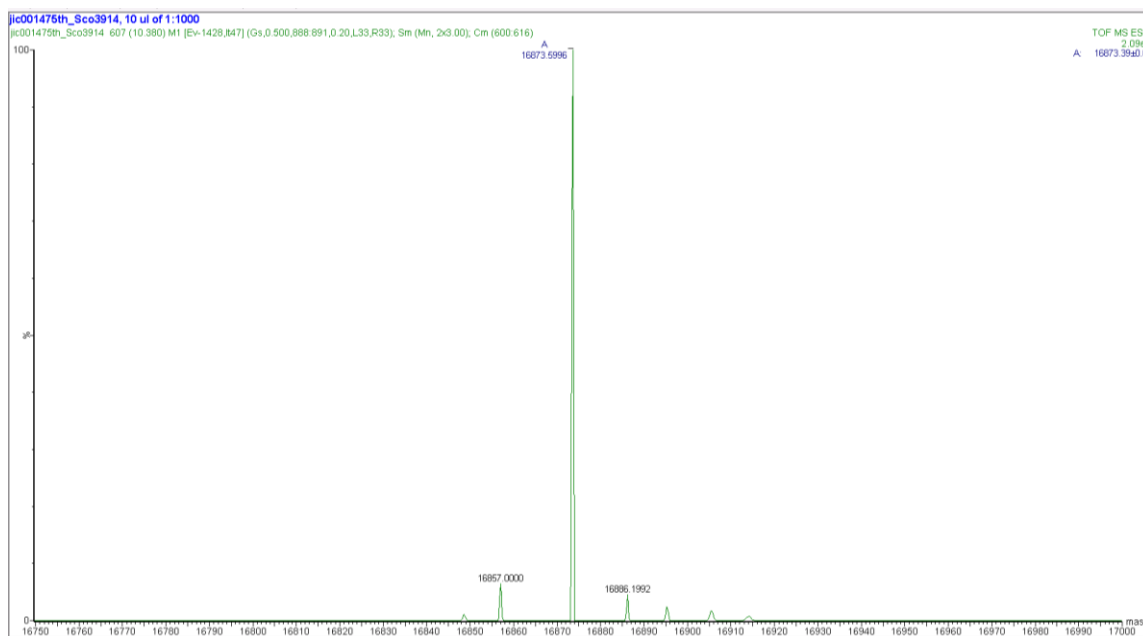


**Figure 3.26** Intact mass analysis of Sco3914. The main peak indicates that the protein has a molecular weight of 16873.6 Da ±0.01. This corresponds to a difference of +151.8 Da from the expected molecular weight of the protein which is 16721.8 Da.

Mass spectrometric analysis of Sco3914C16S was carried out to investigate this further, the expectation being that the single cysteine mutation should reduce the amount of modification to the protein by β-mercaptoethanol. The result displayed in Figure 3.28

shows that the protein has a mass difference of 76 Da from the expected value of 16705 Da for the cleaved protein. This corresponds exactly to a single βME adduct on the solitary cysteine 95.
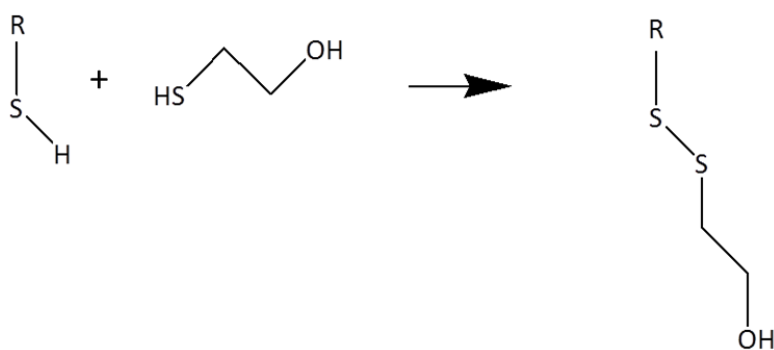


**Figure 3.27** Modification of cysteine by β-mercaptoethanol increases the mass by 76 Da.



**Figure 3.28** Intact mass analysis of Sco3914C16S. The main peak indicates that the protein has a molecular weight of 16781 Da ±0.03. This corresponds to a difference of +76 Da from the expected molecular weight of the protein which is 16705 Da.

## 3.5    Protein-DNA Interactions

### 3.5.1    Surface Plasmon Resonance

The preliminary stage of this experiment involved determining whether the protein would bind the entire intergenic sequence. The *sco3914-3915* intergenic region (Figure 3.29) was amplified by PCR (using the primers shown in Table 3.6) to incorporate a single-stranded overhanging sequence which is complementary to the biotinylated ssDNA annealed to the streptavidin chip (details in Chapter 2). As described in Chapter 2, because of the high GC content of *S. coelicolor*, designing specific primers can be difficult. So, in order to minimize the possibility of non-specific PCR product, a PCR template was initially amplified from the *S. coelicolor* gDNA to incorporate ~150 base pairs flanking sequence either side of the sequence of interest. The PCR product from this reaction was purified and used as a template for the subsequent PCR reaction, the purpose of which was to introduce the "nested" sequence to the intergenic sequence.

### *Sco3914-3915* Intergenic sequence

CATGGAGGGGTACACCTTCCCAAACGATTGCAGGAGCGCTTTGCGTTCGCTCACAATAATTGCAC
ACGCGGGCTATATGCAAGCGCTGGCAATCGCGTGCCTGGTCTACCCTGGTCTCAGCCACTCCTGG
ACGGAGGATCCACATG

**Figure 3.29**  The DNA sequence between *sco3915* (divergently transcribed gene, the proposed start codon of which is highlighted in yellow) and *sco3914*, the start codon of which is highlighted in green.

**Table 3.6**  Oligos used for the 3 step PCR amplification of the intergenic region. The nested sequence is in green and the sequence which is complementary to the biotinylated oligo is in blue. The reverse oligo from step 2 was used again in step 3.

| Oligo name | Sequence |
|---|---|
| Step 1 Forward | CGAGCATCACACCGAGCGCGT |
| Step 1 Reverse | GACGTCGAGCAGGGAGTACTCGC |
| Step 2 Forward | TTATCAAAAAGAGTATTGACGGAGGGGTACACCTTCCCAAA |
| Step 2 Reverse | TCATGTGGATCCTCCGTCCAGG |
| Step 3 Forward | TTAGCGTTTGCGTATGCG/dSpacer/TTATCAAAAAGAGTATTGAC |

Initial results (Figure 3.30) indicated that there is a significant interaction between the protein and the DNA, but only with the addition of 10 mM DTT. This was attempted under both "as prepared" (i.e. with β-mercaptoethanol) and reducing conditions (10 mM final concentration DTT in the running buffer). However, it was also noted that the protein had been exposed to β-mercaptoethanol in the tag cleavage buffer which, as shown in the crystal structure, forms adducts with cysteines 16 and 95. This modification appears to hinder DNA-binding, since the protein preparation which has not been exposed to β-mercaptoethanol binds DNA both in the presence and absence of 10 mM DTT (Figure 3.30). Although these results do not provide any strong evidence for the precise role of either cysteine in protein:DNA interactions, they do suggest that a βME modification of the cysteines hinders these interactions.
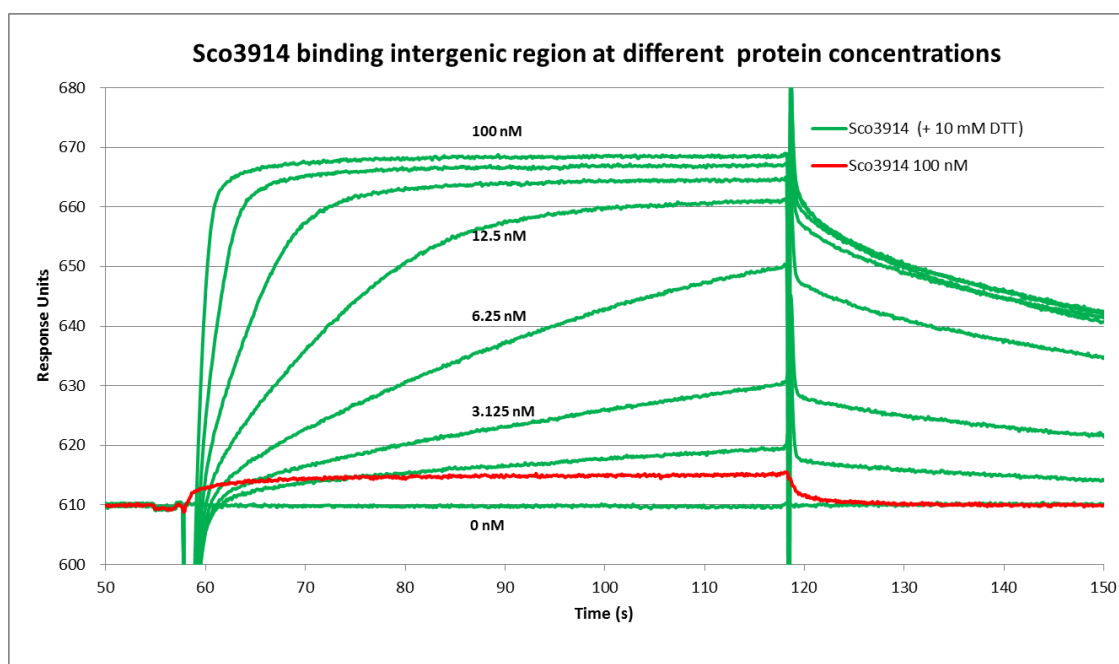


**Figure 3.30** Sensorgram showing Sco3914, at a range of concentrations (100 nM; 50 nM; 25 nM; 12.5 nM 6.25 nM; 3.125 nM; 1.5625 nM and 0 nM), interacting with the PCR amplified intergenic DNA sequence. For this experiment, the sample of Sco3914 had been prepared with 2 mM β-mercaptoethanol in the buffer. 10 mM DTT was added to this sample which enabled DNA-binding. A concentration dependent interaction may be observed in this plot which indicates that there is a specific interaction between the protein and this DNA sequence.

### 3.5.2   Mutant binding

Preliminary studies on the binding of Sco3914C16S and Sco3914C95S to a fragment of the intergenic sequence were also carried out. The initial hypothesis was that as the cysteines are reactive (based on the observations of the βME modifications in the crystal structure), then it was possible that they may participate in DNA recognition and interactions. Mutating each cysteine to a serine could prevent modification and thus affect DNA interactions or allosterically modulate DNA binding.

A fragment of DNA which contained the putative MEME motif (and had been shown to elicit binding of Sco3914 in preliminary experiments not shown here) was ordered from Sigma and annealed to the chip surface. The sequence ordered was intergenic base pairs 45-75:  CGCTCACAATAATTGCACACGCGGGCTATAT. This DNA was stripped between each cycle using 1 M NaCl and 50 mM NaOH. The results in Figure 3.31 show that samples prepared in the presence of β-mercaptoethanol have only a relatively small and transient interaction with the DNA, whereas the samples prepared in the absence of β-mercaptoethanol have a stronger interaction. These interactions were not precisely quantified at this stage, and there is a difference in that the non-βME treated samples retain the 17 amino acid His-tag, but the concentrations of all three samples were equivalent (100 nM). From this experiment, it would appear that the C95S mutant has a faster association and dissociation than do the native protein and the C16S mutant. The binding profile of the C16S mutant resembles that of the native protein but the binding level of the mutant is approximately 40% more than that of the native protein. This could partially be due to inaccuracies in protein concentration determination, or it could suggest that in the reduced form, Cys16 enhances the binding of the protein. Clearly more work on this is required to reach a more definitive conclusion.
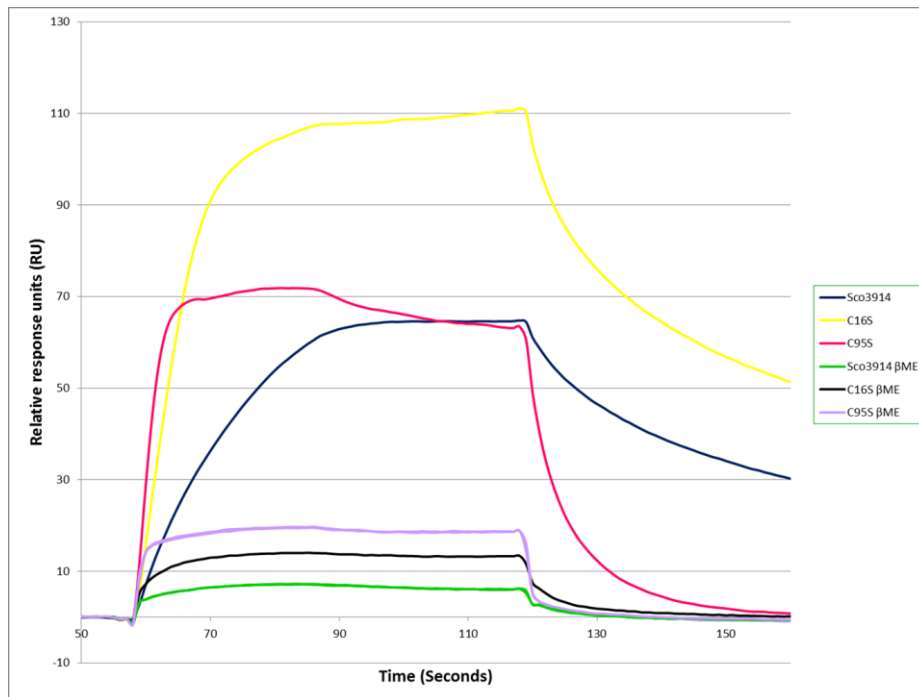
**Figure 3.31** Sensorgram showing the binding of Sco3914; Sco3914C16S and Sco3914C95S (all at 100 nM concentration) to DNA (base pairs 45-75 in the upstream sequence). The samples were prepared with and without β-mercaptoethanol.

### 3.5.3 DNA Footprinting – Stage 1

The next step involved "fragmenting" the intergenic region in order to define the cognate binding sequence. The intergenic region of 140 base pairs was fragmented into 30 base pair sequences, incorporating a 22 bp overlap between each sequence as shown below (Figure 3.32). Oligos of this size were chosen as previous crystal structures of DNA-bound MFRs (Dolan *et al.,* 2011) suggested that 22 base pairs was a reasonable number of base pairs to expect for a footprint and so the overlapping region needs to be long enough to ensure that the full binding site was not missed between adjacent fragments. The maximum length of 30 base pairs was chosen as DNA synthesis becomes expensive and less accurate above 50 bases (accounting for the extra 20 base pairs included on the reverse strand for annealing to the linker). The oligos were ordered from Sigma. The forward oligo included the 20 bp ssDNA sequence (Chapter 2) which is complementary to the ssDNA annealed to the streptavidin chip. In this experiment, the complementary ssDNA sequences were annealed by mixing 60 µl of the reverse oligo with 50 µl of the

forward oligo to give a 45 µM stock solution (PCR program used for this is detailed in Chapter 2). A larger volume of the forward oligo was used in order to maximize the potential for all of the linker strand (reverse) to form dsDNA and to reduce the opportunities for ssDNA to anneal to the chip. This was diluted to the working concentration of 1 µM with HBS-EP+ buffer. Each dsDNA sequence was then captured on the chip by running it across the analytical flow cell containing the ssDNA linker (Chapter 2). The full-length protein Sco3914 (including the His tag sequence) was then run across both the control and analytical flow cells and the results displayed as a percentage of the theoretical $R_{max}$ for each cycle (Figure 3.33).

```
30bp oligos with 22bp overlap:

GGAGGGGTACACCTTCCCAAACGATTGCAG:1
    ACACCTTCCCAAACGATTGCAGGAGCGCTT:2
        CCAAACGATTGCAGGAGCGCTTTGCGTTCG:3
            TTGCAGGAGCGCTTTGCGTTCGCTCACAAT:4
                GCGCTTTGCGTTCGCTCACAATAATTGCAC:5
                    CGTTCGCTCACAATAATTGCACACGCGGGC:6
                        CACAATAATTGCACACGCGGGCTATATGCA:7
                            TTGCACACGCGGGCTATATGCAAGCGCTGG:8
                            GCGGGCTATATGCAAGCGCTGGCAATCGCG:9
                                TATGCAAGCGCTGGCAATCGCGTGCCTGGT:10
                                    CGCTGGCAATCGCGTGCCTGGTCTACCCTG:11
                                        ATCGCGTGCCTGGTCTACCCTGGTCTCAGC:12
                                            CCTGGTCTACCCTGGTCTCAGCCACTCCTG:13
                                                ACCCTGGTCTCAGCCACTCCTGGACGGAGG:14
                                                    GTCTCAGCCACTCCTGGACGGAGGATCCAC:15
```

**Figure 3.32** The *sco3914-3915* intergenic region overlapping oligos as output by POOP (Le, unpublished work). The 30 base pair oligos include a 22 base pair overlap. The MEME motif is highlighted in pink.

Prior to this experiment, a MEME analysis of the upstream intergenic regions of 33 putative Sco3914 orthologues identified in reciprocal BLAST searches was performed and the following MEME motif found:-
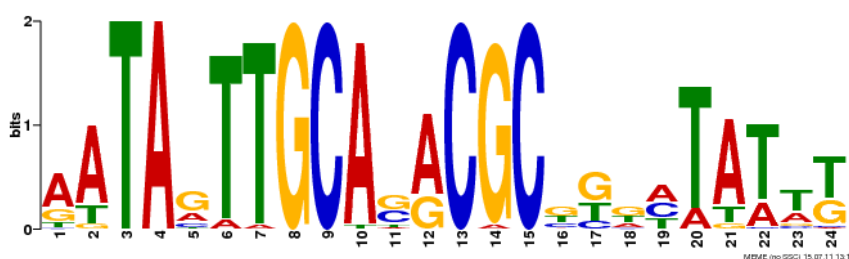
This motif is contained entirely within sequence 7, and partially in sequences 4, 5, 6, 8, 9 and 10. As the results in Figure 3.33 demonstrate, there is a distinctive interaction between the protein and sequence 7, a much weaker interaction with sequence 8 and very small interactions with sequences 5 and 6. These data support the bioinformatics analysis using MEME.
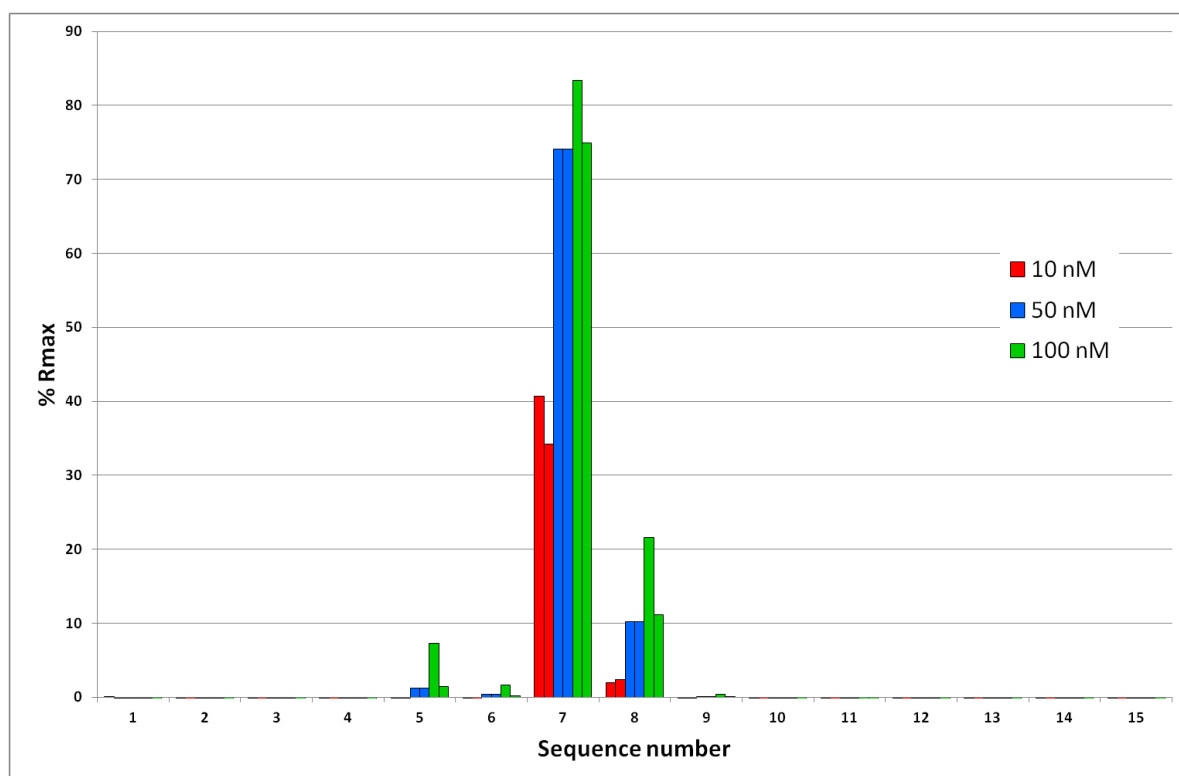


**Figure 3.33** Histogram plot of DNA sequence against the percentage of the calculated $R_{max}$ value for 3 different concentrations of Sco3914 (replicates are shown). There is a small interaction between Sco3914 and sequences 5 and 6, whilst there is a stronger and more significant interaction with sequence 7. This interaction drops off sharply with sequence 8.

120

**Table 3.7** Raw data from the SPR experiment showing the amount of DNA captured at each cycle and the calculated R$_{max}$ for each cycle. The protein interaction was measured in response units which were then calculated as a percentage of the R$_{max}$. There are replicates for each protein concentration.

| DNA Sequence | | Run 1 | | | | Run 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| 10 nM | MW dsDNA | DNA Capture | Theoretical Rmax | Protein bound | % Rmax | DNA Capture | Theoretical Rmax | Protein bound | % Rmax |
| 1 | 24410 | 385.4 | 460.8 | 0.2 | 0.0 | 408.5 | 488.4 | -0.8 | -0.2 |
| 2 | 24409 | 414.2 | 495.2 | -0.4 | -0.1 | 424.7 | 507.8 | -1.0 | -0.2 |
| 3 | 24410 | 426.1 | 509.4 | -0.6 | -0.1 | 434.9 | 520.0 | -1.2 | -0.2 |
| 4 | 24409 | 344.1 | 411.4 | -0.3 | -0.1 | 353.1 | 422.2 | -1.0 | -0.2 |
| 5 | 24408 | 396.2 | 473.7 | -0.1 | 0.0 | 404.7 | 483.9 | -0.8 | -0.2 |
| 6 | 24410 | 429.2 | 513.1 | -0.2 | 0.0 | 438.0 | 523.7 | -1.0 | -0.2 |
| 7 | 24407 | 422.2 | 504.8 | 205.6 | 40.7 | 429.2 | 513.2 | 175.5 | 34.2 |
| 8 | 24411 | 406.5 | 486.0 | 9.5 | 2.0 | 414.8 | 495.9 | 11.6 | 2.3 |
| 9 | 24412 | 411.2 | 491.6 | -0.2 | 0.0 | 418.4 | 500.2 | -0.6 | -0.1 |
| 10 | 24410 | 401.1 | 479.6 | -0.2 | 0.0 | 408.7 | 488.6 | -1.2 | -0.2 |
| 11 | 24412 | 442.6 | 529.1 | -0.7 | -0.1 | 450.7 | 538.8 | -1.6 | -0.3 |
| 12 | 24411 | 411.9 | 492.4 | -0.5 | -0.1 | 418.7 | 500.6 | -1.5 | -0.3 |
| 13 | 24411 | 435.4 | 520.5 | -0.7 | -0.1 | 441.6 | 528.0 | -1.6 | -0.3 |
| 14 | 24412 | 411.2 | 491.6 | -0.6 | -0.1 | 427.6 | 511.2 | -1.4 | -0.3 |
| 15 | 24411 | 440.2 | 526.3 | -0.6 | -0.1 | 445.9 | 533.1 | -1.5 | -0.3 |
| 50 nM | | | | | | | | | |
| 1 | 24410 | 401.0 | 479.4 | -0.3 | -0.1 | 408.8 | 488.8 | -1.1 | -0.2 |
| 2 | 24409 | 422.1 | 504.7 | -0.5 | -0.1 | 423.9 | 506.8 | -1.1 | -0.2 |
| 3 | 24410 | 431.5 | 515.9 | -0.6 | -0.1 | 437.0 | 522.5 | -1.4 | -0.3 |
| 4 | 24409 | 349.8 | 418.2 | -0.5 | -0.1 | 353.9 | 423.1 | -1.1 | -0.3 |
| 5 | 24408 | 401.9 | 480.5 | 5.9 | 1.2 | 405.5 | 484.9 | 7.2 | 1.5 |
| 6 | 24410 | 436.1 | 521.4 | 2.0 | 0.4 | 438.6 | 524.4 | 1.1 | 0.2 |
| 7 | 24407 | 426.5 | 510.0 | 377.7 | 74.1 | 430.5 | 514.8 | 386.3 | 75.0 |
| 8 | 24411 | 411.2 | 491.6 | 50.1 | 10.2 | 416.3 | 497.7 | 55.7 | 11.2 |
| 9 | 24412 | 414.8 | 495.9 | 0.5 | 0.1 | 417.7 | 499.4 | 0.4 | 0.1 |
| 10 | 24410 | 406.1 | 485.5 | -0.1 | 0.0 | 409.1 | 489.1 | -1.2 | -0.2 |
| 11 | 24412 | 446.4 | 533.7 | -0.9 | -0.2 | 449.7 | 537.6 | -1.7 | -0.3 |
| 12 | 24411 | 416.1 | 497.5 | -0.8 | -0.2 | 418.4 | 500.2 | -1.4 | -0.3 |
| 13 | 24411 | 438.4 | 524.1 | -0.8 | -0.2 | 442.9 | 529.5 | -1.5 | -0.3 |
| 14 | 24412 | 418.6 | 500.4 | -0.8 | -0.2 | 428.0 | 511.7 | -1.5 | -0.3 |
| 15 | 24411 | 444.1 | 530.9 | -0.9 | -0.2 | 447.4 | 534.9 | -1.6 | -0.3 |
| 100 nM | | | | | | | | | |
| 1 | 24410 | 404.0 | 483.0 | -0.3 | -0.1 | 410.6 | 490.9 | -1.1 | -0.2 |
| 2 | 24409 | 422.2 | 504.8 | -0.4 | -0.1 | 426.0 | 509.3 | -1.1 | -0.2 |
| 3 | 24410 | 433.8 | 518.6 | -0.5 | -0.1 | 438.0 | 523.7 | -1.4 | -0.3 |
| 4 | 24409 | 350.8 | 419.4 | -0.4 | -0.1 | 356.1 | 425.8 | -1.1 | -0.3 |
| 5 | 24408 | 402.8 | 481.6 | 35.1 | 7.3 | 405.3 | 484.6 | 7.2 | 1.5 |
| 6 | 24410 | 434.5 | 519.5 | 8.4 | 1.6 | 438.9 | 524.7 | 1.1 | 0.2 |

| 7 | 24407 | 427.2 | 510.8 | 425.8 | 83.4 | 431.0 | 515.4 | 386.3 | 75.0 |
| 8 | 24411 | 413.9 | 494.8 | 106.6 | 21.5 | 416.2 | 497.6 | 55.7 | 11.2 |
| 9 | 24412 | 415.7 | 497.0 | 1.8 | 0.4 | 419.8 | 501.9 | 0.4 | 0.1 |
| 10 | 24410 | 406.6 | 486.1 | -0.2 | 0.0 | 410.8 | 491.2 | -1.2 | -0.2 |
| 11 | 24412 | 445.9 | 533.1 | -1.0 | -0.2 | 451.3 | 539.5 | -1.7 | -0.3 |
| 12 | 24411 | 417.1 | 498.7 | -0.8 | -0.2 | 419.8 | 501.9 | -1.4 | -0.3 |
| 13 | 24411 | 440.1 | 526.2 | -0.9 | -0.2 | 443.1 | 529.7 | -1.5 | -0.3 |
| 14 | 24412 | 420.1 | 502.2 | -0.9 | -0.2 | 429.1 | 513.0 | -1.5 | -0.3 |
| 15 | 24411 | 443.5 | 530.2 | -0.9 | -0.2 | 449.0 | 536.8 | -1.6 | -0.3 |

The results shown here indicate that there is a significant interaction between the protein and the sequence number 7, and partially with sequence 8. As the interaction decreases sharply with sequence 8, the assumption is that within the eight base pairs lost from the left hand side of sequence 7, there are bases that are crucial for the binding of the protein. It is also fair to conclude from this that the extra bases added to the right hand side of the sequence in sequence 8 contribute nothing to the stability of the interaction either, and so this sequence is of no further interest. The next stage of the analysis involved probing this sequence further in order to establish the precise recognition sequence, or at least the minimum sequence required for the protein to bind. This next stage was carried out by taking the combined sequences of 6 and 7 and three base pairs from 8 as shown in Figure 3.34.

### 3.5.4 DNA Footprinting Stage 2

CGTTCGCTCACAATAATTGCACACGCGGGC:6

CACAATAATTGCACACGCGGGCTATATGCA:7

TTGCACACGCGGGCTATATGCAAGCGCTGG:8

CGTTCGCTCACAATAATTGCACACGCGGGCTATATGCAAGC          41 bp

**Figure 3.34** The 3 DNA sequences from the sco3914-3915 intergenic region which elicited the highest response in the SPR experiment. Sequences 6 and 7 were combined with 3 base pairs from sequence 8 to arrive at the starting sequence for the Stage 2 base pair elimination round. The MEME motif is highlighted in pink.

The strategy was to eliminate 2 base pairs from each end in turn to narrow down the binding site. In order to account for any effects from the linker sequence (such as stabilization of binding due to the extra bases), the linker was subsequently attached to the 3' end of the forward strand sequence for deletions from the left hand side (see Chapter 2). Two base pairs at a time were first removed from the right hand side of the sequence, followed by two base pairs at a time from the left. The sequences can be seen in Table 3.8.

**Table 3.8** Stage 2 DNA sequences for testing the binding of Sco3914 to its upstream DNA sequence. (a) Sequences to test the limits of the right hand boundary; (b) sequences to test the limits of the left hand boundary. Sequences highlighted in yellow gave the most significant interactions with the protein in SPR.

### (a) Sco3914 RH boundary

| Name | Sequence |
|---|---|
| 3914_RH_F | CGTTCGCTCACAATAATTGCACACGCGGGCTATATGCAAGC |
| 3914_RH_R | GCTTGCATATAGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus2_F | CGTTCGCTCACAATAATTGCACACGCGGGCTATATGCAA |
| 3914_RH_minus2_R | TTGCATATAGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus4_F | CGTTCGCTCACAATAATTGCACACGCGGGCTATATGC |
| 3914_RH_minus4_R | GCATATAGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus6_F | CGTTCGCTCACAATAATTGCACACGCGGGCTATAT |
| 3914_RH_minus6_R | ATATAGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus8_F | CGTTCGCTCACAATAATTGCACACGCGGGCTAT |
| 3914_RH_minus8_R | ATAGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus10_F | CGTTCGCTCACAATAATTGCACACGCGGGCT |
| 3914_RH_minus10_R | AGCCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus12_F | CGTTCGCTCACAATAATTGCACACGCGGG |
| 3914_RH_minus12_R | CCCGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus14_F | CGTTCGCTCACAATAATTGCACACGCG |
| 3914_RH_minus14_R | CGCGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |
| 3914_RH_minus16_F | CGTTCGCTCACAATAATTGCACACG |
| 3914_RH_minus16_R | CGTGTGCAATTATTGTGAGCGAACGcctaccctacgtcctcctgc |

### (b) Sco3914 LH boundary

| Name | Sequence |
|---|---|
| 3914_LH_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTGTGAGCGAACG |
| 3914_LH_R, | CGTTCGCTCACAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus2_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTGTGAGCGAA |
| 3914_LH_minus2_R, | TTCGCTCACAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |

| 3914_LH_minus4_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTGTGAGCG |
|---|---|
| 3914_LH_minus4_R, | CGCTCACAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus6_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTGTGAG |
| 3914_LH_minus6_R, | CTCACAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus8_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTGTG |
| 3914_LH_minus8_R, | CACAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus10_F, | GCTTGCATATAGCCCGCGTGTGCAATTATTG |
| 3914_LH_minus10_R, | CAATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus12_F, | GCTTGCATATAGCCCGCGTGTGCAATTAT |
| 3914_LH_minus12_R, | ATAATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus14_F, | GCTTGCATATAGCCCGCGTGTGCAATT |
| 3914_LH_minus14_R, | AATTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |
| 3914_LH_minus16_F, | GCTTGCATATAGCCCGCGTGTGCAA |
| 3914_LH_minus16_R, | TTGCACACGCGGGCTATATGCAAGCcctaccctacgtcctcctgc |

These sequences were annealed and the 1 µM working solution prepared as described in Chapter 2. Again, three concentrations of the protein were run and the interactions plotted as a percentage of the calculated $R_{max}$ (Figure 3.35).
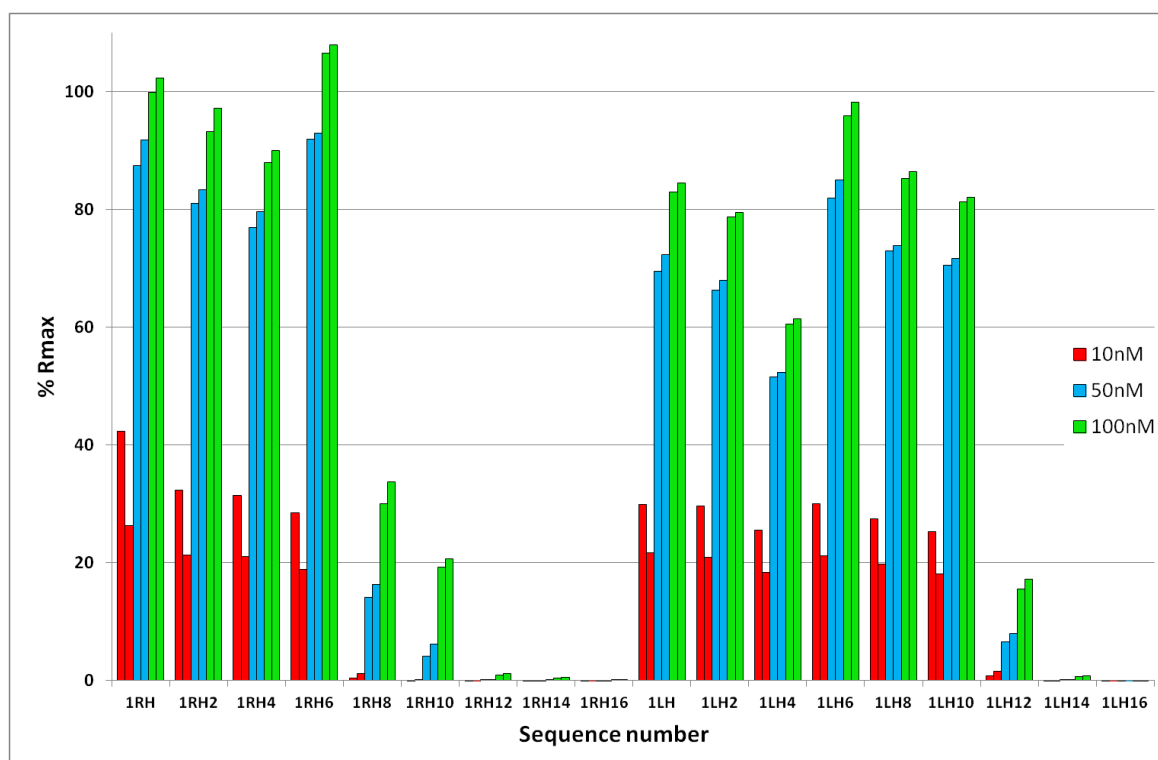


**Figure 3.35** DNA Footprinting Stage 2 Results. A sharp drop in binding may be observed when 8 base pairs are eliminated from the right hand side of the sequence and when 10 base pairs are eliminated from the left hand side.

There is a clear decrease in the affinity of the protein for the sequence when 8 base pairs are removed from the right hand side and 10 base pairs from the left hand side:-

**RH**

CGTTCGCTCA<mark>CAATAATTGCACACGCGGGCTATAT</mark>~~GCAAGC~~ 25 bp

**LH**

GCTTGC<mark>ATATAGCCCGCGTGTGCAATTATTG</mark>~~TGAGCGAACG~~

Sco3914 will tolerate RH-6 and LH-10, which would give the footprint as:-

CAATAATTGCACACGCGGGCTATAT          25 bp

This is one base pair longer than the motif suggested by MEME:-

AATAATTGCACACGCGGGCTATAT          24 bp

Stage 3 of the footprinting assay involved quantifying the affinity of the protein for this sequence. Anticipating that the 25mer may not elicit the strongest response (as the LH-6 sequence had a stronger response than the LH-10), the following DNA sequences were ordered from Sigma:-

**Table 3.9**  The oligos that were used for Stage 3 of the footprinting assay.

| 25mer_F | CAATAATTGCACACGCGGGCTATAT |
|---------|---------------------------|
| 25mer_R | ATATAGCCCGCGTGTGCAATTATTGcctaccctacgtcctcctgc |
| 29mer_F | CTCACAATAATTGCACACGCGGGCTATAT |
| 29mer_R | ATATAGCCCGCGTGTGCAATTATTGTGAGcctaccctacgtcctcctgc |

The sequences were annealed as described previously to give a final concentration of 1 μM.

### 3.5.5    DNA Footprinting Stage 3

The affinity of Sco3914 for this particular sequence was investigated using a similar approach to the method described in stages 1 and 2, but this time less of the ssDNA-linker was loaded onto the flow cells so that the affinities could be measured more accurately (52 RU on flow cell 3 and 150 RU on flow cell 4). A lower loading of ligand and high flow rate minimizes mass transport effects (Buckle, 2001). On flow cell 3 (the reference cell), a ssDNA sequence which was complementary to the ssDNA linker sequence was captured in order account for any non-specific interactions with the linker. The sequence being tested (the 25mer or the 29mer) was captured on flow cell 4 and the chip was regenerated after each cycle. This time the protein was injected across the DNA at a flow rate of 100 µl min$^{-1}$ for a period of 210 seconds to allow saturation of the DNA to occur. A dilution series was set up to include concentrations of protein from 200 mM to 0.391 mM and each concentration was tested in triplicate. Both DNA sequences were tested and it was observed that the protein had a greater affinity for the 29mer.



| KD (M) | SE(KD) | Rmax (RU) | SE(Rmax) |
|--------|--------|-----------|----------|
| 1.19E-07 | 8.50E-09 | 158.7 | 5.1 |
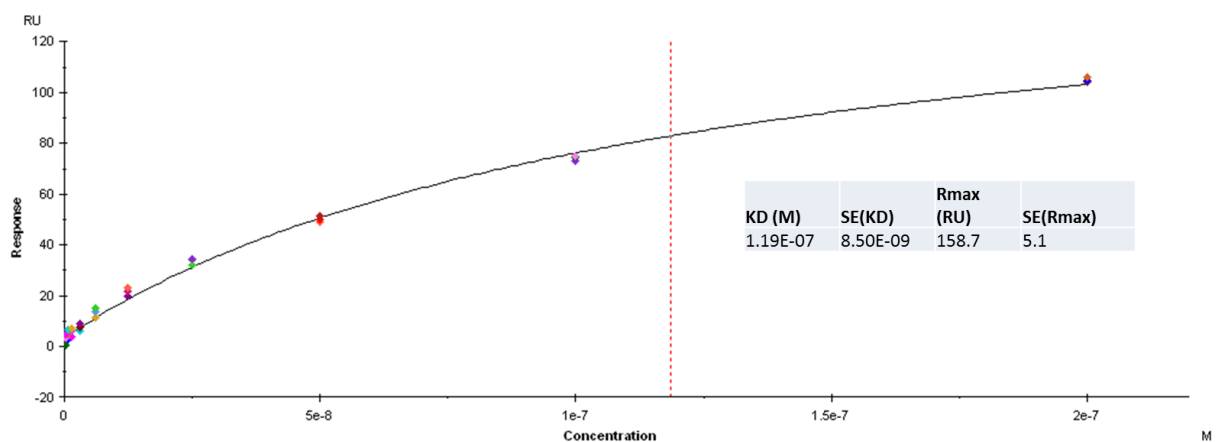
**Figure 3.36**  Stage 3 25mer. $K_D$ = 120 ± 9 nM
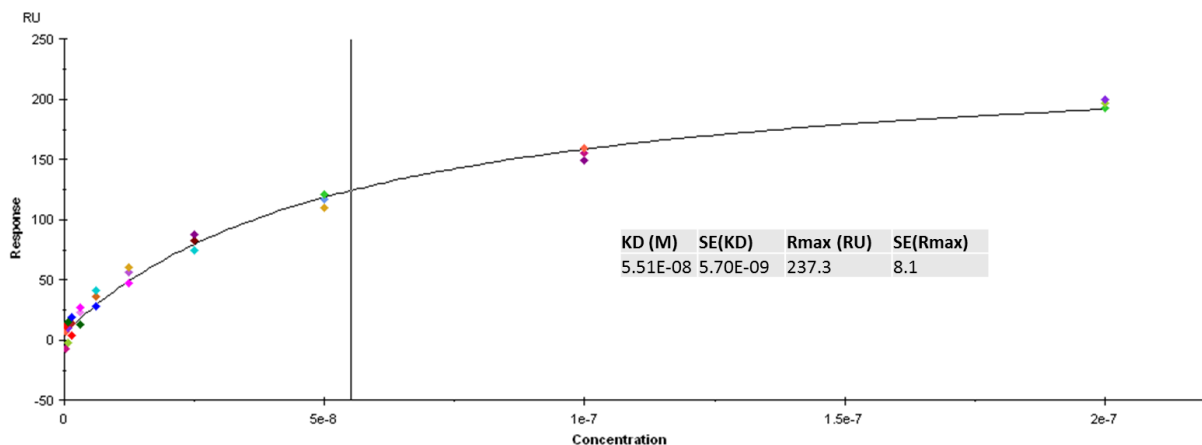
**Figure 3.37** Stage 3 29mer. $K_d$ = 55 ± 6 nM

The actual values determined for the affinity of Sco3914 for these sequences shows that this is a significant interaction. Whilst not quite into the nanomolar range, the $K_D$ value for the 29mer is 55 ± 6 nM which indicates that there is reasonably tight binding occurring between the protein and this specific sequence. These data lead to the conclusion that the binding site for Sco3914 is contained within the 29mer sequence. It is possible that the 25mer defines the actual footprint (and this agrees closely with the output from the MEME server) and that the extra bases in the 29mer lend the protein support to further stabilize binding. Further investigations could define this further, but for the purpose of this study, the footprint has been defined as CAATAATTGCACACGCGGGCTATAT. This sequence is not obviously palindromic and there are no apparent repeats, but it has a GC content of 44%, which is low compared to the average GC content of 72% for the *S. coelicolor* genome.

## 3.6 Determination of pK$_a$s of cysteines 16 and 95

Sco3914 has two cysteines: Cys16 and Cys95. Cys16 is located in what would appear to be a putative ligand binding site, as this is where salicylate has been observed in published MarR structures (Kumarevel *et al.*, 2009, Saridakis *et al.*, 2008) and also the region of conserved cysteines in several redox active MFRs (Chen *et al.*, 2008, Chen *et al.*, 2006, Chen *et al.*, 2009, Newberry *et al.*, 2007, Panmanee *et al.*, 2006, Soonsanga *et al.*, 2007, Hong *et al.*, 2005). Cys95 is located on the loop which could potentially make contacts

with DNA. In an attempt to understand their relative reactivities, the pK$_a$s of each of the cysteines in Sco3914 were determined using the DTNB assay (detailed in Chapter 2). This information, combined with knowledge about any effector molecules, could also suggest the relevance of these cysteines to ligand detection and DNA-binding.

Each cysteine reacts with the DTNB molecule to form a mixed thiol and releases the TNB$^{2-}$ anion. The absorbance of released TNB anion was measured at 412 nm and used to calculate the rates at each pH. The results from this experiment, along with their associated errors, are displayed in Figure 3.38 and Figure 3.39.
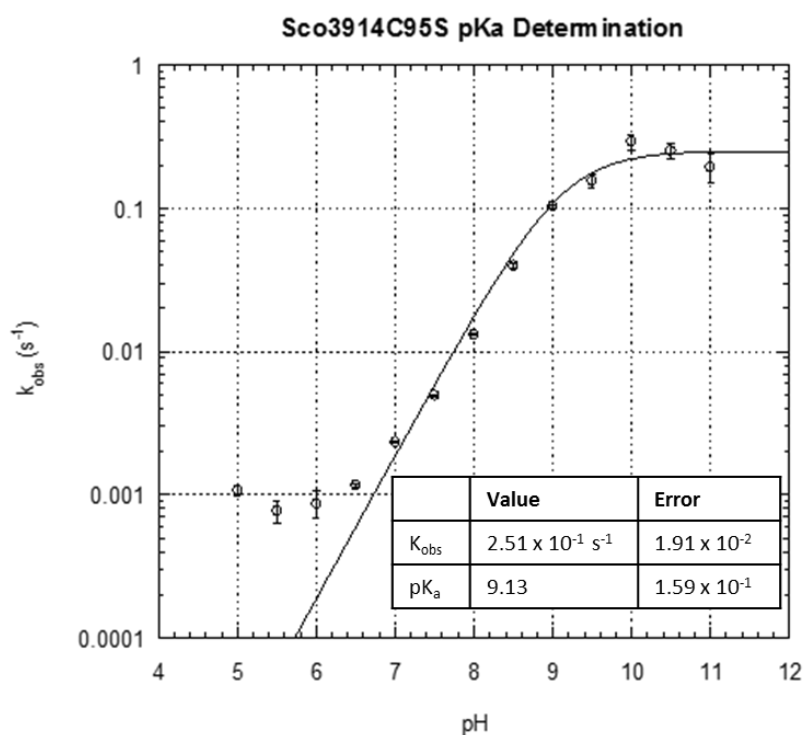


**Figure 3.38** Rates obtained from the titration of Sco3914C95S with buffers across a range of pH values containing 20 µM DTNB. The results shown here indicate that cysteine 16 has a pK$_a$ of 9.1.
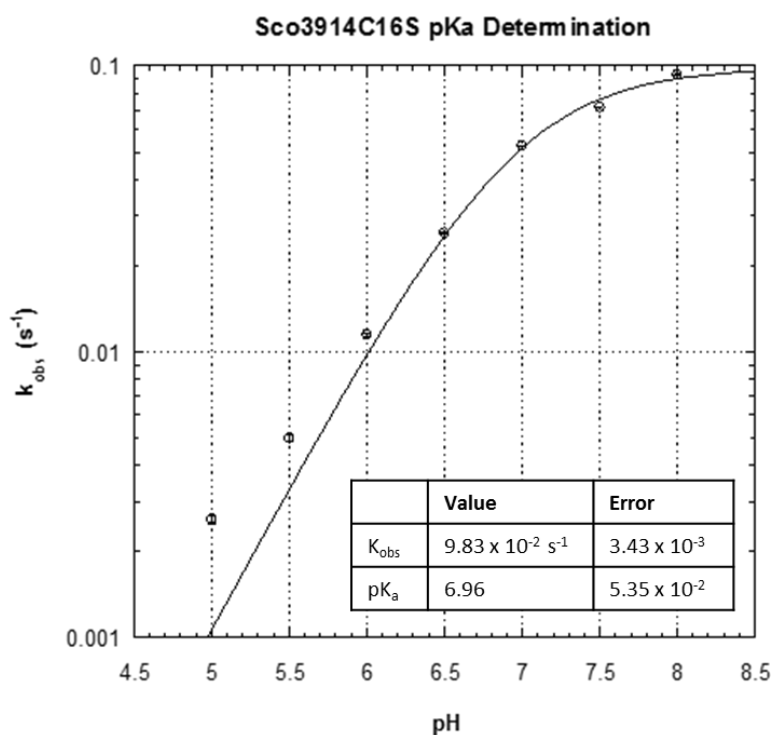
**Figure 3.39** Rates obtained from the titration of Sco3914C16S with buffers across a range of pH values containing 20 µM DTNB. The results shown here indicate that cysteine 95 has a $pK_a$ of 7.0.

The respective $pK_a$ values for cysteine 16 and cysteine 95 are 9.1 ± 0.2 and 7.0 ± 0.1 (free cysteine has a $pK_a$ of 8.3). These results support what is observed in the crystal structure in that Cys16 sits in a relatively hydrophobic pocket of the protein which would potentially destabilize the ionized form of this residue. The $pK_a$ of this cysteine is significantly higher than that determined for redox-sensing cysteines in other MFRs: the reactive Cys14 in HypR has a lower $pK_a$ value of 6.36 (Palm *et al.*, 2012) whilst for Cys15 in BsOhrR, the value is 5.2 (Hong *et al.*, 2005). The cysteines in these other structures are stabilized by a hydrogen bonding network provided by neighbouring tyrosines. Conversely, Cys95 is located on a solvent exposed loop (the wing region) and so it would be expected that the deprotonated form of the residue could be stabilized by solvent interactions and also by Tyr93 and Tyr103. This residue could potentially make contacts with the DNA.

## 3.7 PCR-Targeted Gene Disruption

The M145*sco3914* gene deletion strain was generated using the Redirect protocol (Gust *et al.*, 2003) detailed in chapter 2. The *sco3914* gene is contained on the StQ11 cosmid (base pairs 20697-21206 with a total cosmid length of 33873 bp). Cosmids were checked for rearrangements by restriction digest with BamHI and compared to the predicted digest fragment sizes by DNA gel electrophoresis (using the restriction tool on Streptomyces.org.uk). The primers used to amplify the disruption cassette from pIJ773 are displayed in Table 3.10. The gene deletion was confirmed by PCR using the P1 and P2 flanking primers (Chapter 2) as well as the gene specific checking primers (Table 3.11).

**Table 3.10** Primers used to amplify the pIJ773 disruption cassette. The sequence which is complementary to the disruption cassette is shown in red whilst the gene specific sequence is in black.

| Forward | ACCCTGGTCTCAGCCACTCCTGGACGGAGGATCCACATGATTCCGGGGATCCGTCGACC |
|---------|-------------------------------------------------------------|
| Reverse | TAGAACGGGCCGACGCCAGGTTGCCCGGACAGGCTCTATGTAGGCTGGAGCTG CTTC |

**Table 3.11** Primers used to verify positive transformants using PCR.

| Forward | CGTTCCGATCCCGAAGGCCC | 172 bp upstream of gene |
|---------|----------------------|-------------------------|
| Reverse | ACCCTGAGGCGTTCGAAGGC | 170 bp downstream of gene |

The mutagenized cosmid was then introduced into *S. coelicolor* by conjugation (detailed in Chapter 2) to undergo genetic recombination with the chromosomal DNA and so replace the *sco3914* gene with the disrupted gene. Successful mutants were detected by replica-plating on SFM agar plates supplemented with 1) apramycin and 2) apramycin + kanamycin. Those colonies which are kanamycin sensitive were then further verified by PCR (using both the gene specific checking primers and the F1 and F2 flanking primers.) Spore stocks of the mutant strains were then prepared as detailed in chapter 2.

### 3.7.1  Screening the successful mutants

The prepared spore stocks were screened for visibly detectable phenotypes by plating on a variety of media supplemented with different carbon sources (details in chapter 2). Twenty different conditions were tested in total and the observations for each mutant were monitored over the course of a week. Sco3914 did not exhibit an observable phenotype under any of the 20 conditions tested. It was decided not to investigate this further during this project.

## 3.8  Conclusions

It is still very much speculative that Sco3914 may also regulate the divergently transcribed Sco3915 (a putative efflux pump), but its operator is virtually in the middle of this intergenic sequence which could potentially hinder the transcription start sites of both genes. If this hypothesis were correct, then knocking out Sco3914 would allow the upregulation of Sco3915. Pushing these hypotheses further, if the role of Sco3915 were to remove toxic molecules from the cell, thus protecting it, then it would be difficult to observe a phenotype by deleting Sco3914 without knowledge of the "toxic" conditions. Also, removing the putative repressor would cause the pump to be upregulated and so a phenotype may not be seen even in the presence of a specific toxic effector. The next step would be to make a *sco3915* deletion mutant or, if more were known about a specific effector of Sco3914, to make a Sco3914 mutant which was unresponsive to the presence of this effector but was still capable of binding its cognate DNA. Further investigation of Sco3915 would be useful to uncover its function, and also to probe the role of Sco3914 on its regulation. The presence of four modifiable cysteines in the Sco3914 dimer also presents an interesting place to begin looking for potential ligands for this regulator, particularly as these residues are so well conserved in homologues of Sco3914. The p$K_a$ of Cys95, which is located in the wing motif, suggests that this residue would be reactive at a physiological pH whereas Cys16, which is buried, would not be reactive. This warrants further investigation to reveal the biological relevance of these cysteines. The crystal structures of the C16S mutant and the reduced form of Sco3914 did

not uncover any significant conformational differences to the βME modified protein. However, MFRs are typically very flexible and so perhaps the only way to determine any conformational rearrangements is to obtain the crystal structure of Sco3914 bound to the operator DNA or with a specific ligand. There are many avenues to explore with Sco3914 which could yield some very useful information about the nature of repressor-DNA interactions as well as further insights into its biological role in *S. coelicolor*.

# Chapter 4

# Sco5413

## 4.1 Introduction

Sco5413 is a putative transcriptional regulator from *S. coelicolor* belonging to the MarR family of proteins. It is predicted to be a mainly helical protein of 169 amino acids and has several orthologues in other species of streptomycete. This protein has been cloned, expressed and purified by nickel affinity and size exclusion chromatography. A truncated version of the protein has been crystallized and the structure solved to 1.25 Å resolution using the Single Isomorphous Replacement with Anomalous Scattering method (with mercury as the heavy atom providing the anomalous signal). A phenotype has been observed in the deletion mutant in the host organism and there is scope for more work in this area. However, several different experimental approaches have failed to detect any significant binding to the intergenic DNA sequence between *sco5413* and *sco5412*, which encodes an upstream hypothetical protein on the same coding strand. To date, whether Sco5413 requires a ligand/cofactor to bind to DNA and where it may bind in the whole of the genomic DNA sequence remains to be determined. The structure of the truncated version of this protein reveals no obvious ligand, although the pocket may be compromised due to the truncation of the N-terminal helix. However, when superimposed onto the OhrR-DNA complex (Hong *et al.*, 2005), it would appear to be in a DNA-binding conformation.



**Figure 4.1** A representation of the location of s*co5413* in the *Streptomyces coelicolor* genome (http://strepdb.streptomyces.org.uk).

The *sco5413* gene is 510 bp in length and has a GC content of 69%. It is an extremely well conserved gene amongst streptomycetes and has 68 orthologues amongst 112

actinobacterial genomes and *E. coli* and *Bacillus subtilis* (bioinformatics analysis done by Govind Chandra, John Innes Centre). Its close homology to orthologues in other published sequences of Streptomycetes can be seen in Table 4.1 below.

| | Amino acid sequence %ID | | | | DNA Sequence %ID | | | |
|---|---|---|---|---|---|---|---|---|
| **Sco** | **Sav** | **Ssc** | **Sgr** | **Sven** | **Sav** | **Ssc** | **Sgr** | **Sven** |
| Sco5413 | Sav2834 90% | Scab28331 90% | Sgr2122 89% | Sven5067 91% | Sav2834 87% | Scab28331 88% | Sgr2122 88% | Sven5067 89% |

**Table 4.1**  The sequence identities shared between the orthologues from other Streptomycetes and *Streptomyces coelicolor* Sco5413. Sav – *Streptomyces avermitilis*; Ssc – *Streptomyces scabius*; Sgr – *Streptomyces griseus*: Sven – *Streptomyces venezuelae*.

Interestingly, the *Streptomyces avermitilis* orthologue, *sav2834*, is 25 amino acids shorter than sco5413. However, this could be a misannotation of the start codon in the sequencing of *Streptomyces avermitilis* as this sequence may be extended by 75 base pairs to give a very similar alternative starting point. This includes a potential helix at the N-terminus which features later in this chapter, relating to the crystallization of Sco5413.

The results obtained for the biophysical, genetic and structural characterization of this protein will be discussed in this chapter.

## 4.2  PCR-Targeted Gene Disruption

The M145*sco5413* gene deletion strain was generated using the REDIRECT[©] technology protocol (Gust *et al.*, 2003), detailed in Chapter 2. The *sco5413* gene is contained on the St8F4 cosmid (Redenbach *et al.*, 1996)(base pairs 20697-21206 with a total cosmid length of 33873 bp). Cosmids were checked for rearrangements by restriction digest and compared to the predicted digest fragment sizes by DNA gel electrophoresis (using the restriction tool on Streptomyces.org.uk). St8F4 was digested with BamHI. The primers used to amplify the disruption cassette from pIJ773 are displayed in Table 4.2. The pIJ773 (*aac(3)IV*, *oriT*) cassette contains an apramycin resistance selectable marker and *oriT* which allows the PCR targeted cosmid DNA to be introduced into *S. coelicolor* through conjugation. The gene deletion was confirmed by PCR using the P1 and P2 flanking primers (Chapter 2) as well as the gene specific checking primers (Table 4.3). The size of the fragment expected from amplification of the successfully disrupted cosmid using the checking primers (Table 4.3) is 1643 base pairs, whereas the wild type should result in a fragment of 771 base pairs. PCR was used as the method of checking both the cosmids before conjugation and the genomic DNA following conjugation.

**Table 4.2**  Primers used to amplify the pIJ773 disruption cassette. The sequence which is complementary to the disruption cassette is shown in red; the *S. coelicolor* genomic specific sequence is in black.

| Forward | TGACCGGCATGTGACCGGCCGGTAAGGTCTCTTGCCGTGATTCCGGGGATCCGTCGACC |
|---|---|
| Reverse | GTCGGGCGACTCGCCGGGCCCGCCCGGCCCGCGCGCTCATGTAGGCTGGAGCTGCTTC |

**Table 4.3**  Primers used to verify positive transformants using PCR (Checking primers).

| **Forward** | GTGTGACGGACGGTCTCACG | 135 bp upstream of gene |
|---|---|---|
| **Reverse** | GGTGACGTACGCCATGACCC | 126 bp downstream of gene |
| **pIJ773_P1** | ATTCCGGGGATCCGTCGACC | Disruption cassette For |
| **pIJ773_P2** | TGTAGGCTGGAGCTGCTTC | Disruption cassette Rev |
| **5413F_39Flank** | TGACCGGCATGTGACCGGCCGGTAAGGTCTCTTGCCGTG | 39 upstream of start (including start codon) |
| **5413R_39Flank** | GTCGGGCGACTCGCCGGGCCCGCCCGGCCCGCGCGCTCA | 39 downstream of stop (including stop codon) |

The mutagenized cosmid was then introduced into *S. coelicolor* by conjugation (detailed in Chapter 2) to undergo genetic recombination with the chromosomal DNA and thus replace the *sco5413* gene with the disrupted gene. Successful mutants were detected by replica-plating on SFM agar plates supplemented with (1) apramycin and (2) apramycin + kanamycin. Those colonies which are kanamycin sensitive were then further verified by checking the gDNA by PCR (using both the gene specific checking primers and the F1 and F2 flanking primers.) Figure 4.2 shows that the fragments generated by the PCR amplification of the mutants are all ~1600 base pairs in length. The expected length for these fragments is 1643 base pairs. The length of the fragment from the wild type is ~780 base pairs and the expected length is 771 base pairs. These results in conjunction with the replica-plating for antibiotic resistance selection suggest that: i) the insertion of the disruption cassette into the cosmid was successful and ii) the recombination event to insert this disrupted gene into the M145 genomic DNA was also successful. Spore stocks of the mutant strains were then prepared as detailed in Chapter 2.
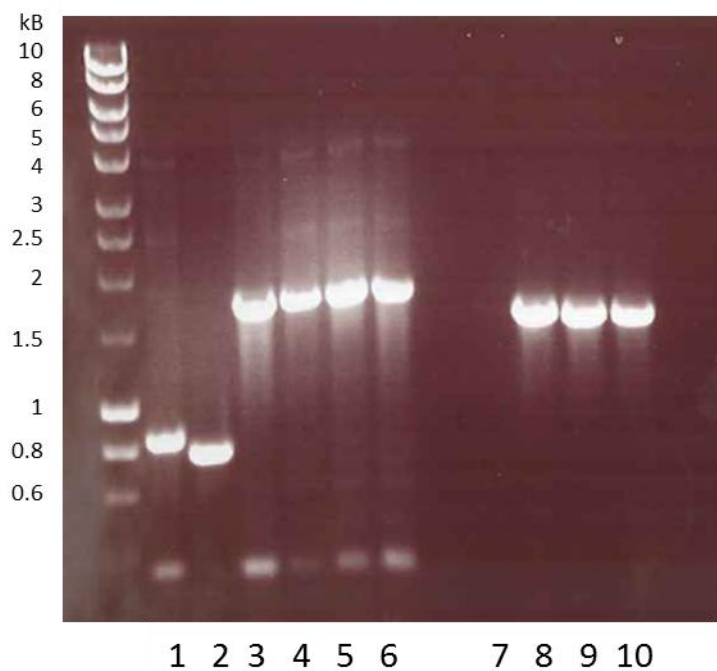


**Figure 4.2** 1% agarose DNA gel to check mutated genomic DNA against the wild type M145 gDNA. The first two lanes contain wild type gDNA amplified with (1) *sco3914* specific checking primers; (2) *sco5413* specific checking primers. Lanes 3-6 contain the M145Δ*sco3914*gDNA, amplified with the *sco3914* specific checking primers. Lanes 7-10 contain M145Δ*sco5413* gDNA, amplified with the *sco5413* specific checking primers.

### 4.2.1  Phenotypic screening of the successful mutants

The prepared spore stocks were screened for visibly detectable phenotypes by plating on a variety of solid media. All media were prepared according to Kieser *et al.* Spore stocks, which had been titred prior to this experiment, were streaked for single colonies onto Mannitol soya flour medium (SFM); Minimal medium (MM); R5 medium; Difco nutrient agar (DNA) and L agar, which were supplemented with different carbon sources (precise details in Chapter 2). Twenty different conditions were tested in total and the observations for each mutant were monitored over the course of a week. The *sco5413* knockout exhibited an interesting media-dependent phenotype on SFM supplemented with either 10 mM N-acetyl glucosamine (GlcNAc), 1% glucose or 1% glycine: it differed from the wild type in that it produced pigmented antibiotic prematurely and ectopically whilst the colonies appeared to be smaller in size than the wild type. Typically, the wild type single colonies, which were visible after two days, appeared to be very white at first, turning grey as they sporulated. The mutant single colonies were not visible until day 3, when they appeared very small and strikingly dark red in colour. Sporulation was also delayed in these mutants compared to the wild type, but did occur by day 5. These culture conditions were repeated in triplicate (the wild type in duplicate) and the same phenotype was observed in all replicates.

**a**    <u>**SFM + 1% Glucose**</u>

Day 3
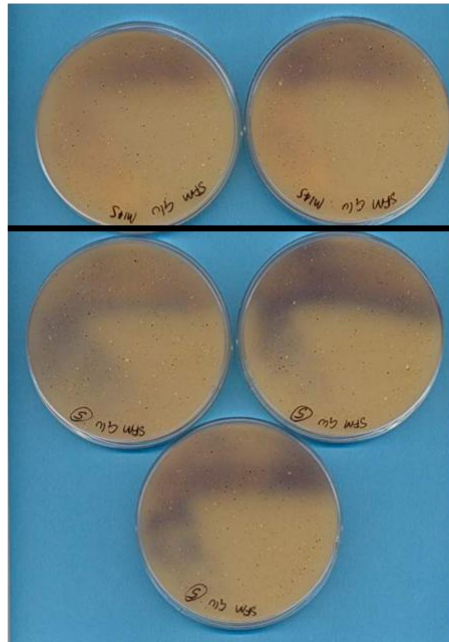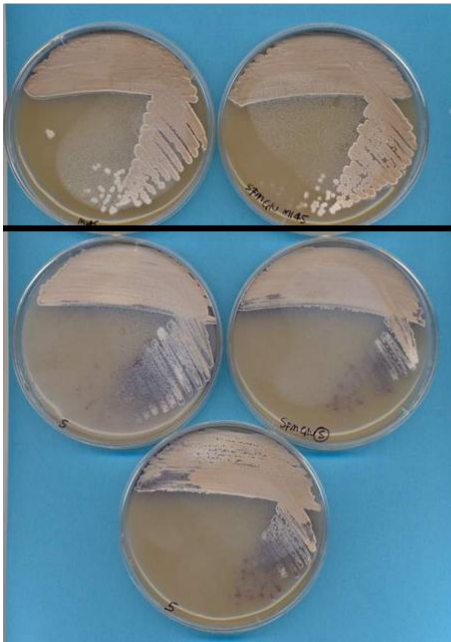


M145

M145Δ*sco5413*
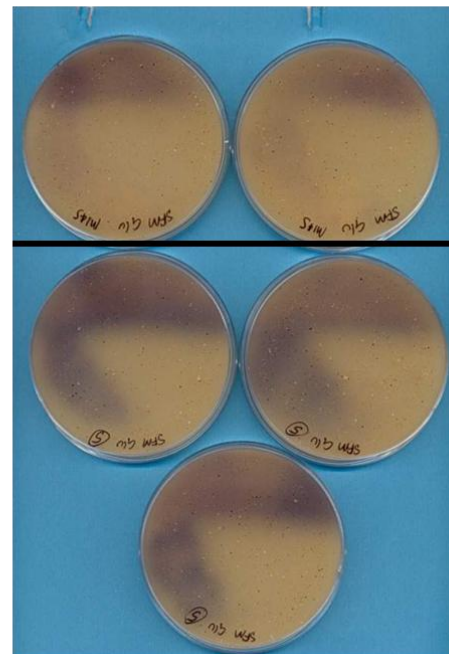
Day 4



M145

M145Δ*sco5413*

**b** **SFM + 1% Glycerol**

Day 3
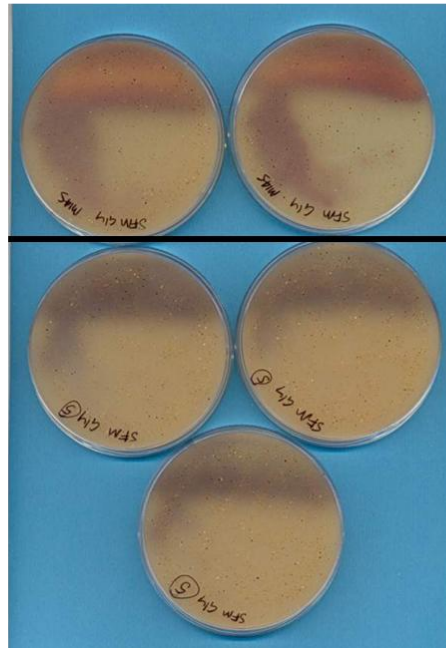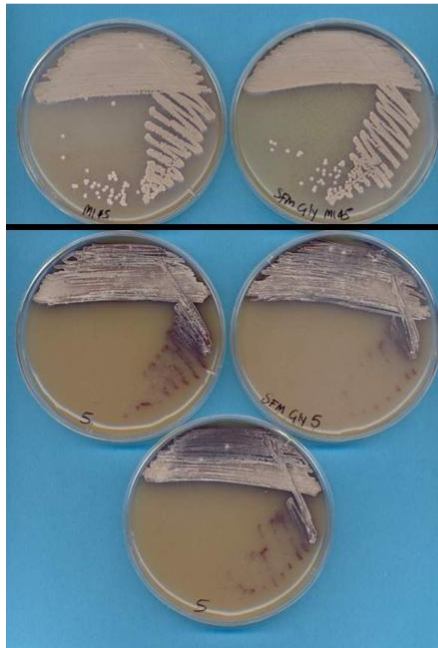


M145

M145Δ*sco5413*
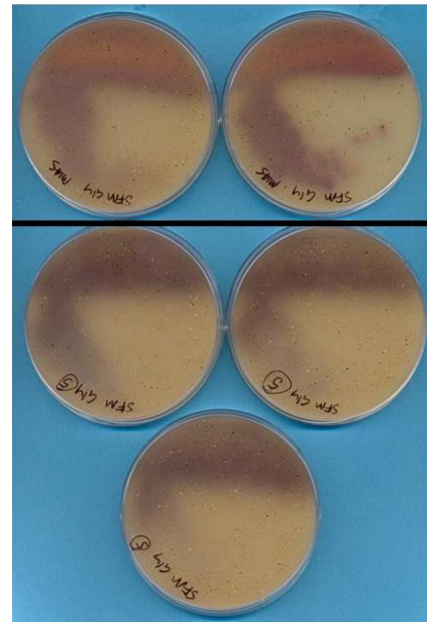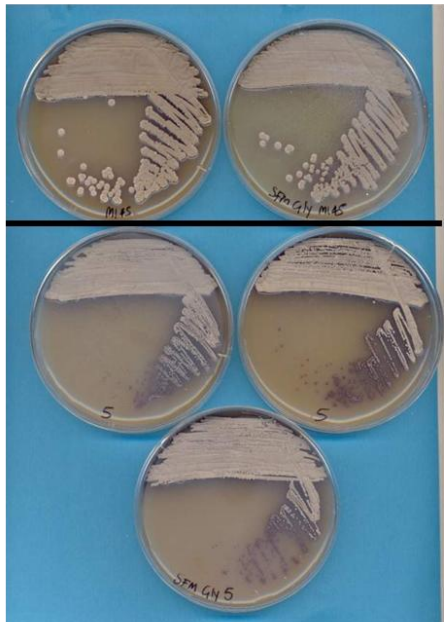
Day 4



M145

M145Δ*sco5413*

**c**    <u>**SFM + 10 mM GlcNAc**</u>

Day 3



M145

M145Δ*sco5413*
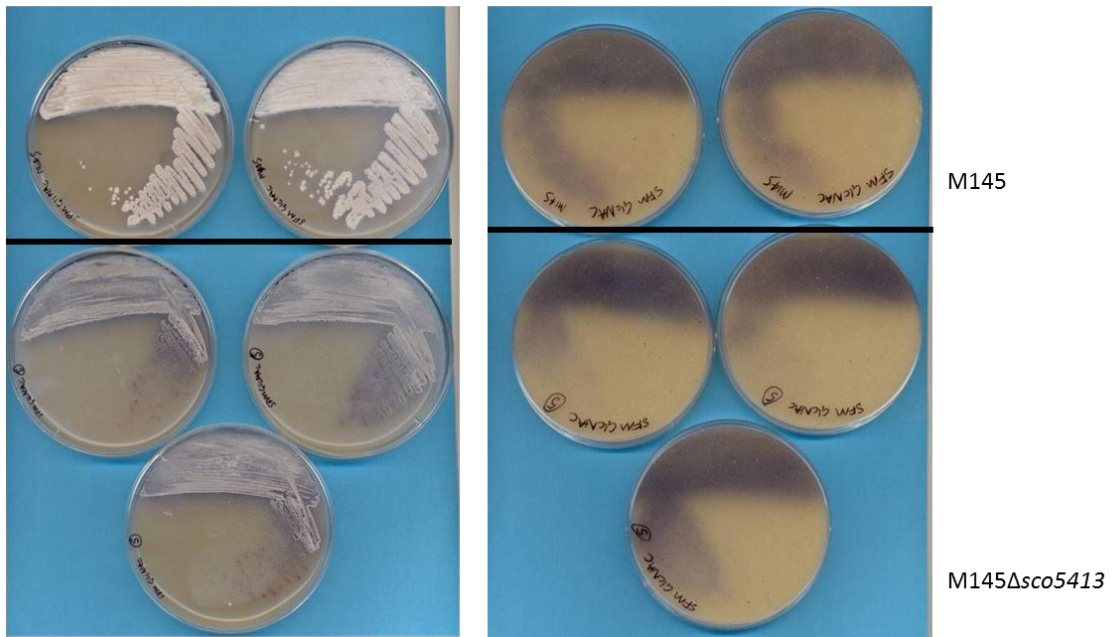
Day 4
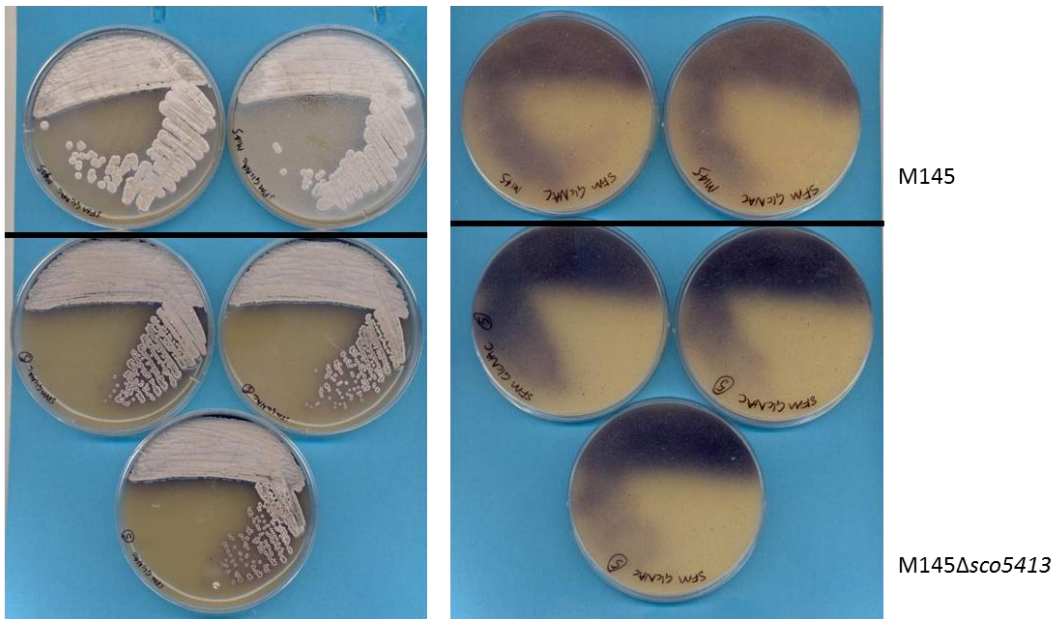


M145

M145Δ*sco5413*

**Figure 4.3**  M145 wild type and M145Δ*sco5413* cultivated on SFM (a) 1% Glucose; (b) 1% Glycerol; (c) 10 mM GlcNAc. Left hand plates – top view; right hand – view from underneath. The precocious and ectopic production of pigmented antibiotic in the mutant can be observed under all three conditions. The mutant colonies are also smaller than the Wild type and slower to grow.

The pictures of the plates in Figure 4.3 show how the growth of the mutant is affected three days after culturing. The dense lawn appears to be darker than that of the wild type in all conditions, most notably on SFM supplemented with 10 mM GlcNAc. The mutant seems to struggle with growth in this condition and the production of antibiotic is most pronounced here (as with the wild type), particularly looking at the underside of the plates where the dark blue in the vegetative mycelium can be seen. This is due to the antibiotic-inducing effect of GlcNAc (see later). When the medium is supplemented with 1% glucose or glycerol, the wild type lawn and colonies are a fairly grey in colour whilst the mutant is darker. In all conditions, the ectopic expression of antibiotic is evident in the mutant single colonies where its appearance on the surface of the colony may be observed.
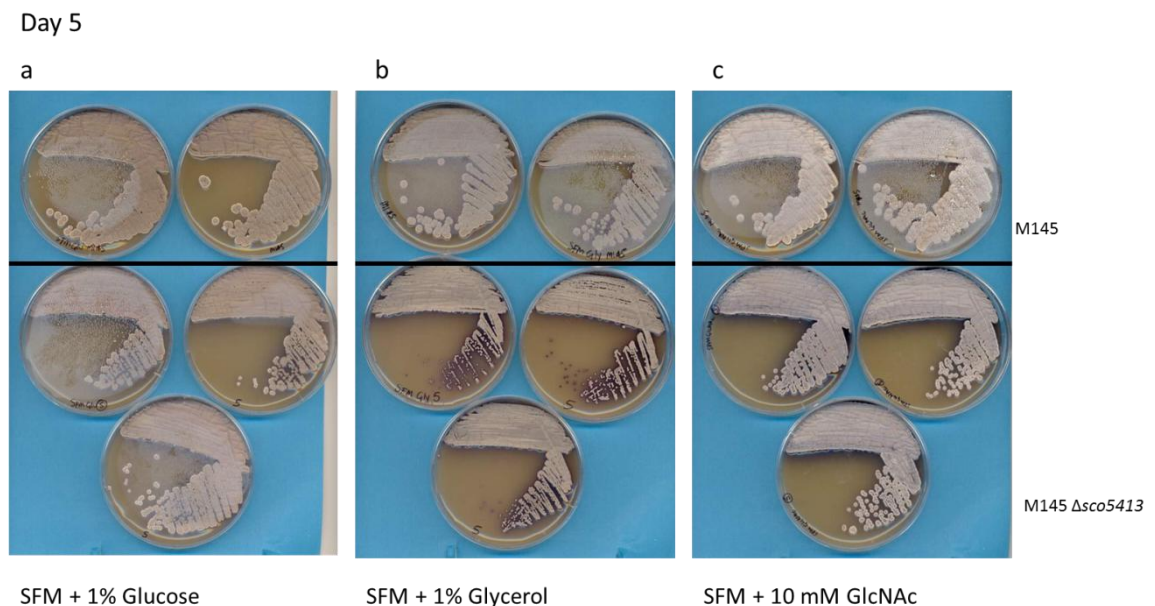


**Figure 4.4** M145 Wild type and M145Δ*sco5413* cultivated on SFM (a) 1% Glucose; (b) 1% Glycerol; (c) 10 mM GlcNAc. The differences between the mutant colonies and the wild type are still evident after 5 days growth.

Five days after plating, the differences between the wild type and mutant are less pronounced but still evident (Figure 4.4). The mutant single colonies are smaller and the overproduction of pigmented antibiotic can be seen surrounding the colonies. On SFM with 1% glucose, the differences at this stage are much less pronounced than on media

supplemented with 1% glycerol, where the mutant still exhibits delayed growth and excessive production of pigment. On media supplemented with 10 mM GlcNAc, M145 appears to be whiter than usual. The mutant colonies are slightly darker and again, the overproduction of pigment is particularly evident.

Interestingly, *N*-acetylglucosamine is a product of autolytic degradation of vegetative mycelium. At high concentrations, GlcNAc can act as an allosteric effector of DasR (Colson *et al.*, 2008) (a pleiotropic transcriptional repressor in *Streptomyces coelicolor*) which regulates pathway-specific activators involved in antibiotic synthesis and it is essential for development (Rigali *et al.*, 2006). GlcNAc has the effect of increasing antibiotic production and this is a common feature on streptomycetes (Rigali *et al.*, 2008). In *S. coelicolor*, during times of "feast", GlcNAc is present in the cell in its polymeric form, chitin (a component of fungi and insect cell walls). Chitin is metabolized by chitinases, which are part of the DasR regulon, and the presence of GlcNAc under these conditions actually blocks development and antibiotic production. However, during times of famine, DasR is inactivated by the presence of GlcNAc in its monomeric form, which is a sign of famine as it is derived from the hydrolysis of the bacterium's own cell walls. Studies conducted on the transport systems for both chitobiose (GlcNAc$_2$) and monomeric GlcNAc indicate that both transporters are DasR dependent and the model presented suggests that the cell is able to distinguish between these two forms of GlcNAc and so determine whether to switch on the developmental genes or not (Colson *et al.*, 2008). The M145Δ*sco5413* mutant also presented a phenotype different to the wild type on minimal media with 10 mM GlcNAc (Figure 4.5). Under this condition, the single mutant colonies appear to be producing pigmented antibiotic prematurely compared to the wild type single colonies which do not. This effect was not observed on minimal media which has been supplemented with other carbon sources.
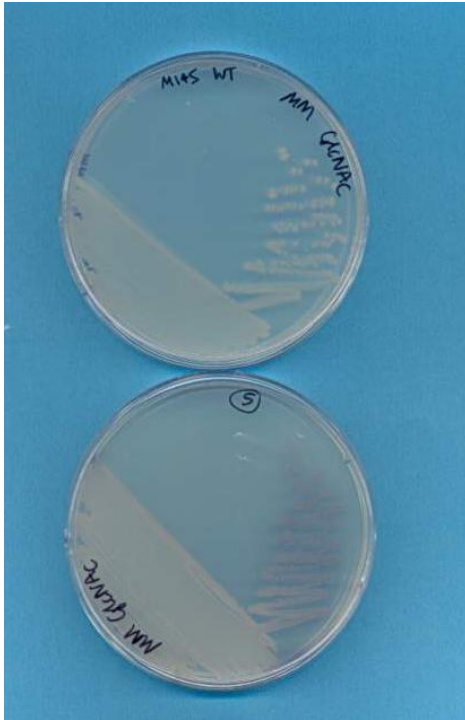
**Figure 4.5** M145 wild type and M145Δ*sco5413* cultured on minimal medium (MM) supplemented with 10 mM GlcNAc. After 3 days of growth, the single colonies of the mutant can be seen to be producing pigmented antibiotic whereas the wild type colonies do not.

The cultures were also grown in liquid media to monitor the growth of the strains over several days. This was done in minimal liquid medium which had been supplemented with 10 mM GlcNAc (again, liquid SFM posed a problem with the many aggregates which would interfere with the optical density measurements). The cells were pre-germinated according to the procedure described in Chapter 2 and used to inoculate three 50 ml replicates. Samples (1 ml) were then collected at 6 hour intervals over the course of 3 days and the optical density at 450 nm measured. A growth curve to compare the development of M145 wild type to the disruption mutant strain was generated from these observations.
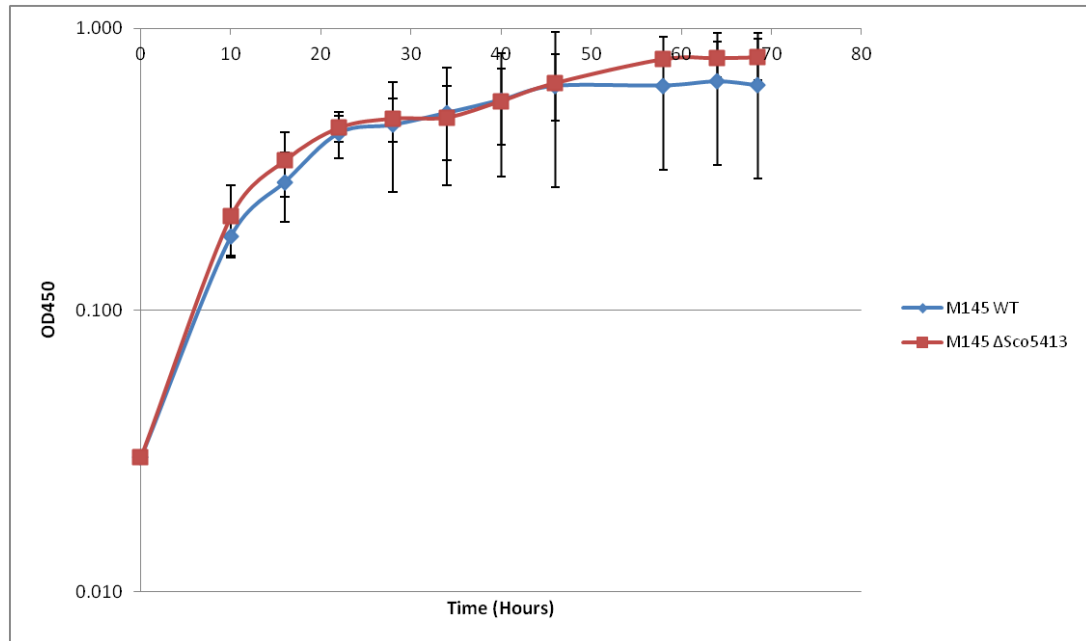
**Figure 4.6** Growth of M145 wild type and M145Δ*sco5413* in minimal liquid medium supplemented with 10 mM GlcNAc at 30°C. Three replicate cultures were set up and samples were extracted at 10, 16, 22, 28, 34, 40, 46, 58 and 64 hours following inoculation. The mean of the replicates and the standard deviation error bars were calculated (Excel).

The resultant growth curve (Figure 4.6) shows that the growth of the mutant under these conditions is comparable to that of the wild type. There are no apparent differences and so the conclusion is that there is no difference in visibly measurable phenotype under these growth conditions. The total protein content was also assessed by extracting the protein and measuring the absorbance of the supernatant at 750 nm (data not shown), but this reveals a similar trend.

Many studies have concluded that the production of secondary metabolites generally occurs in conjunction with the formation of aerial hyphae (Bibb, 1996, Kelemen *et al.*, 1998). Therefore, these results present an interesting observation of *Streptomyces* development, but an observation that is not easily interpreted. As a transcriptional regulator, Sco5413 could have one or many recognition sites within the *S. coelicolor* genome and to know where to begin looking for its possible site of regulation is one task.

To uncover the implications of deleting this gene on other regulatory networks is quite another and, unfortunately, this was not within the scope of this work.

## 4.3   Protein Expression and Purification

The protein was cloned into the pOPINF expression vector and expressed according to the method described in chapter 2. Typically, the protein elutes from the nickel affinity column with a double peak profile (Figure 4.7), the first of which contains several contaminating proteins. The fractions containing the contaminating proteins are disposed of and the second peak fractions pooled and applied to the size exclusion gel filtration column (GE Healthcare Superdex 75 16/60). The purity of the protein following the nickel affinity column purification and the size exclusion gel filtration chromatography column can be seen in Figure 4.8 and Figure 4.10.
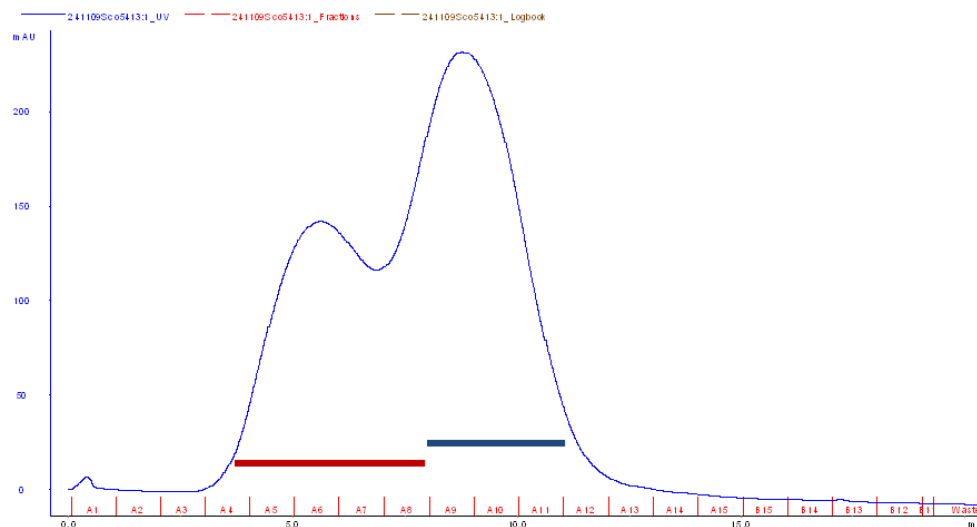


**Figure 4.7** Elution profile of Sco5413 from nickel affinity column chromatography. The double peak is characteristic of Sco5413 with the first peak containing contaminant proteins. The red line indicates the contaminated first peak whilst the blue line indicates the relatively pure second peak.
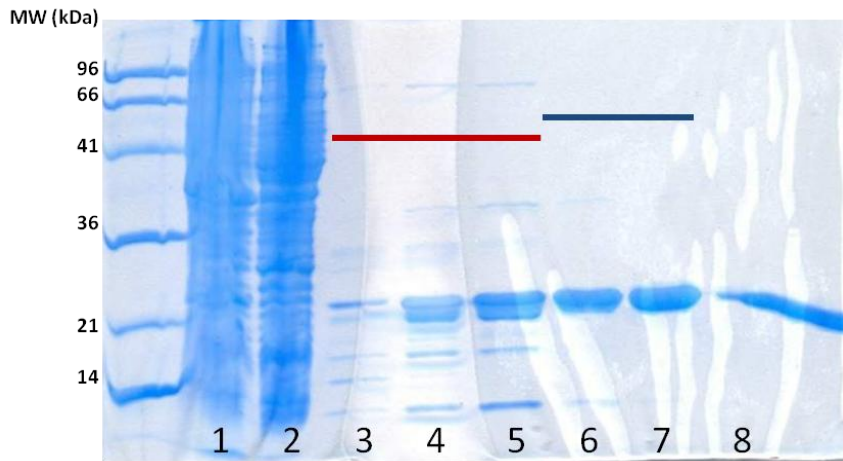
**Figure 4.8** 15% SDS-PAGE after nickel affinity column purification. The lanes contain: (1)Lyzed pellet; (2) Supernatant; (3) Fraction A4; (4) A7; (5) A8; (6) A9; (7) A10; (8) A11. The contaminated fractions from the first peak are demarcated by the red line whilst the relatively pure samples from the second peak are marked by the blue line.
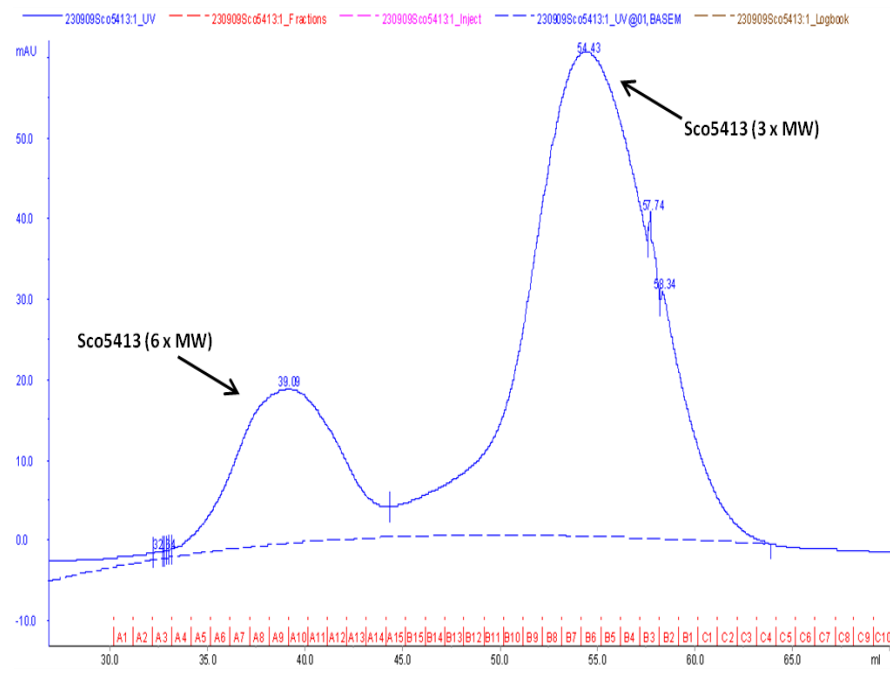


**Figure 4.9** Elution profile of Sco5413 from size exclusion gel filtration chromatography. The first peak contains the protein in a higher oligomeric state whilst the second peak also contains Sco5413.

The elution profile for the size exclusion gel filtration column typically displays two peaks. Both peaks contain Sco5413, although from the SDS-PAGE gel (Figure 4.10) it would appear that the first peak may contain a small amount of contaminant proteins whereas the second peak was judged to be more pure. Using the standard curve generated from

the column calibration, the elution volumes for both the first and second peaks correspond to molecular weights of 118 kDa and 59 kDa respectively. Sco5413 has a predicted molecular weight of 20.4 kDa and so these peaks would suggest that the protein elutes as a hexamer and a trimer. DLS analysis of the protein in solution was fairly inconclusive (data not shown) as it appears to be very polydisperse and so it is uncertain how the protein actually behaves in solution, although it is fair to conclude that it is certainly a dimer in the crystal form. As discussed later in this chapter, the dimer interface is quite substantial (34% of residues) and there is also a salt bridge connecting the monomers.
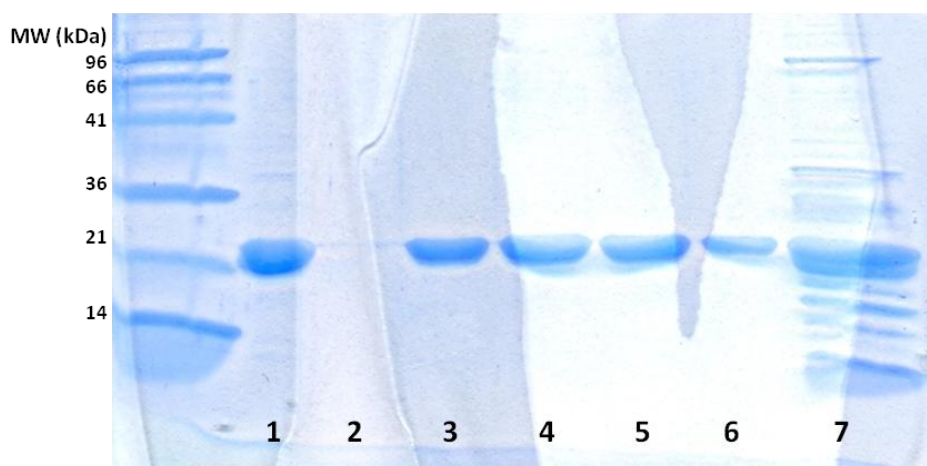


**Figure 4.10** 15% SDS-PAGE after size exclusion gel filtration column purification. The lanes contain: (1) Pooled fractions A9-A11 from nickel column; (2) Fraction A11; (3) B8; (4) B6; (5) B4; (6) B1; (7) Peak 1 (fractions A4-A8 pooled) from nickel column.

### 4.3.1    Truncation of Sco5413

As mentioned previously, Sco5413 has an orthologue in *Streptomyces avermitilis* (Sav2834; UniProtKB entry Q82JC1) which is predicted to be shorter (by the annotation in the StrepDB server) by 25 amino acids (Figure 4.11). As Sco5413 did not crystallize (even with the tag removed), the full-length sequence was analysed using the Disopred server (http://bioinf.cs.ucl.ac.uk/disopred/) (Ward *et al.*, 2004) and this predicted that this sequence had a high probability of disorder at the N-terminus (Figure 4.12).

```
SCO5413          1 VPKPLSLSFDPIARADELWAQRWGGVPSMAAITSIMRAQQILLGEVDAVV    50
                   ||||||||||||||.||||.||||||
SAV2834          1 -----------------------VPSMAAITSIMRAHQILLAEVDAVV    25

SCO5413         51 KPYGLTFARYEALVLLTFSKSGELPMSKIGERLMVHPTSVTNTVDRLVRS   100
                   |||||||||||||||||||:|||||||||||||||||||||||||||.|
SAV2834         26 KPYGLTFARYEALVLLTFSKAGELPMSKIGERLMVHPTSVTNTVDRLVTS    75

SCO5413        101 GLVAKRPNPNDGRGTLATITDKGREVVEAATRDLMAMDFGLGAYDAEECG   150
                   |||.||||||||||||:||||||||.:||||||||||||||||.|||.
SAV2834         76 GLVDKRPNPNDGRGTLASITDKGREVCDAATRDLMAMDFGLGAYDGEECA   125

SCO5413        151 EIFAMLRPLRVAAGDFDED     169
                   ||||||||||:||.||||:
SAV2834        126 EIFAMLRPLRIAAHDFDEE     144
```

**Figure 4.11** Amino acid sequences of Sco5413 and Sav2834 as annotated in the StrepDB server. Sequence alignment was carried out using Needleman-Wunsch global alignment by Emboss Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The former is longer by 25 amino acids at the N-terminus.

Therefore, the protein was truncated at the N-terminus in an attempt to improve crystallizability.  In choosing the transcriptional start site for Sco5413, removal of any predicted disordered regions, without disrupting any potential secondary structural elements, was the main objective. Bioinformatic analysis of the Sco5413 amino acid sequence was used to guide the choice of construct. The DISOPRED server (http://bioinf.cs.ucl.ac.uk/disopred/) (Ward *et al*., 2004)  predicted disorder for residues 1-12. The PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan *et al.*) predicted that residues 11-22 and 27-51 were α-helical whilst the Sco5413 orthologue in *Streptomyces avermitilis* (Sav2834), with which it is 90% identical, is annotated as having a transcriptional start site corresponding to residue 26 of Sco5413.
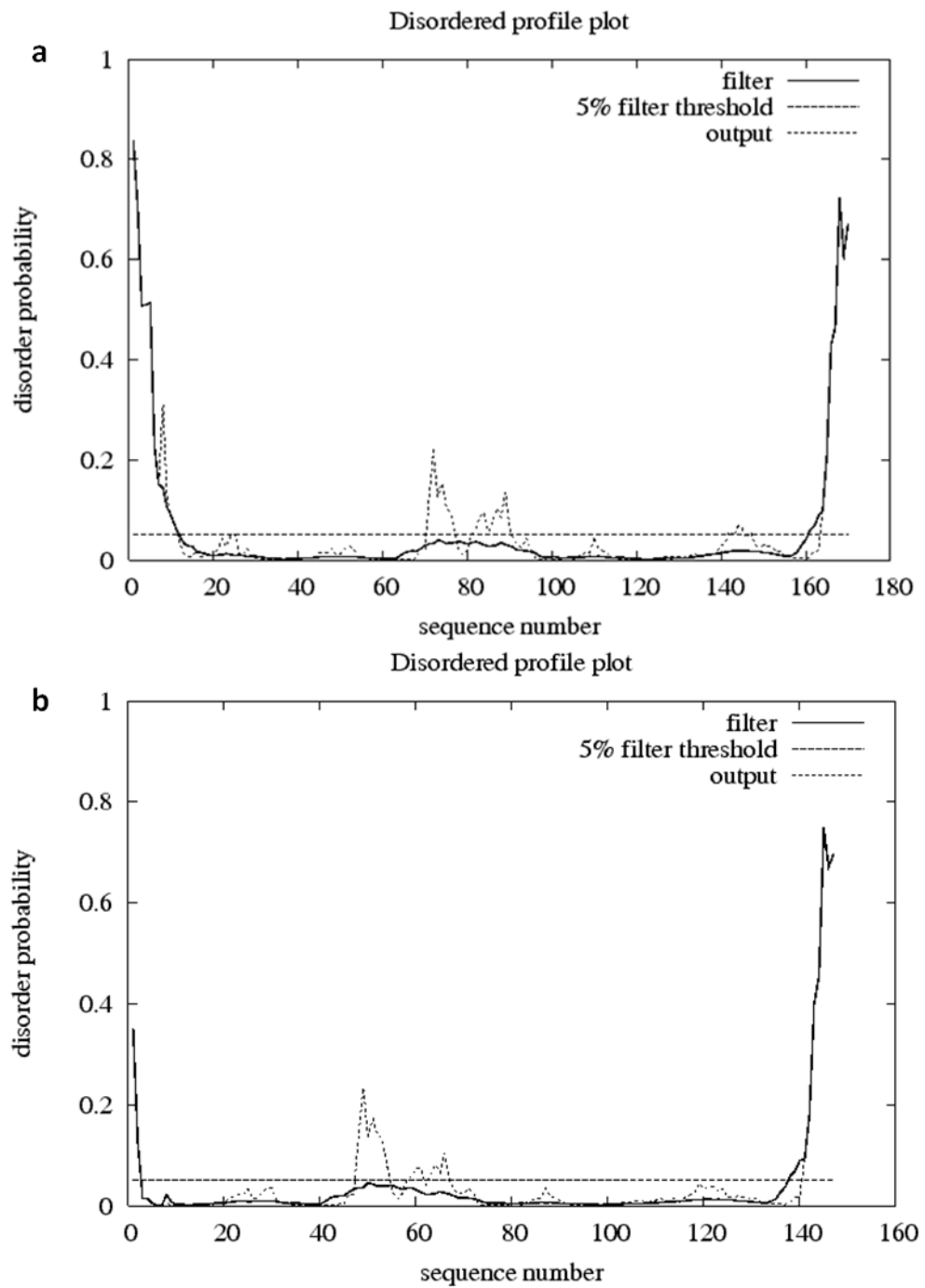
**Figure 4.12** Sco5413 (a) and Sco5413tr (b) disorder prediction (produced using DISOPRED (Ward *et al.*, 2004)). The results presented here suggest that removing the initial 12 amino acids may decrease the disorder.
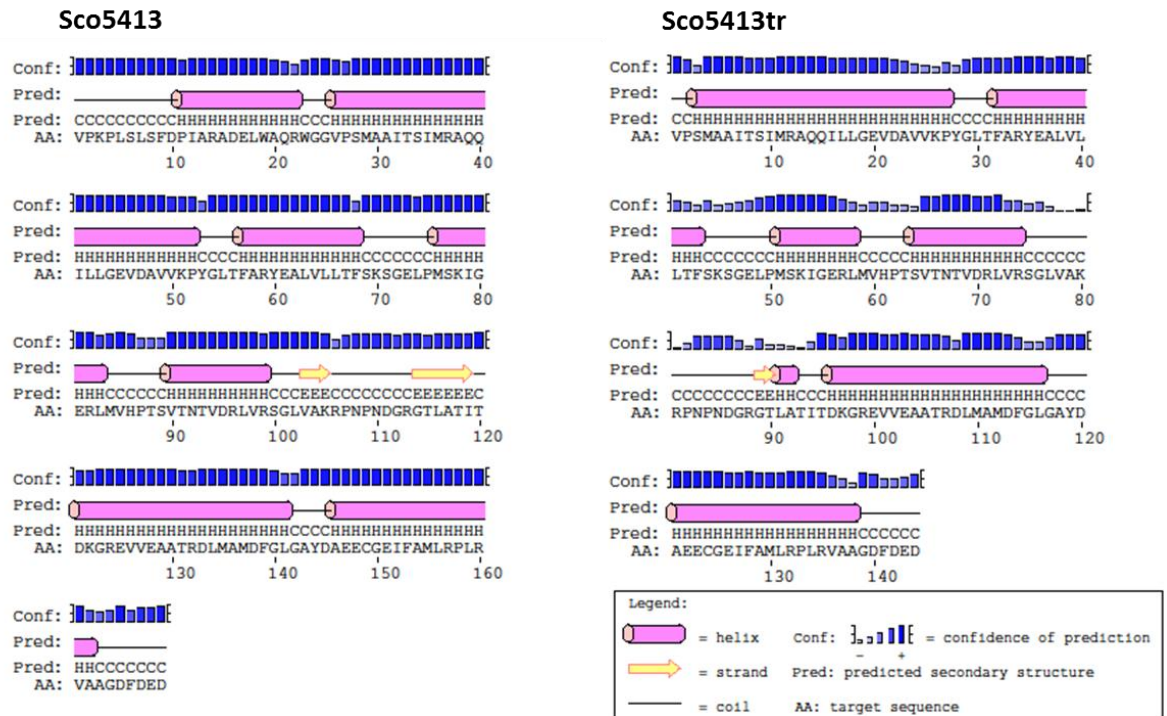
**Figure 4.13** Structure prediction for Sco5413 and Sco5413tr generated by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) (Murshudov *et al.*, 1997).

A start site was therefore chosen in the loop region between the two predicted N-terminal helices that corresponded to the Sav2834 N-terminus. Hereafter the truncated Sco5413 will be referred to as Sco5413tr. The truncation was achieved by amplifying the *sco5413* gene from *S. coelicolor* M145 genomic DNA using primers containing the pOPINF ligation independent cloning sequences (Forward AAGTTCTGTTTCAGGGCCCGGGTGTGCCCTCCATGG and reverse ATGGTCTAGAAAGCTTTAGTCCTCGTCGAAGTCGCCCGC). The sequence was cloned into the pOPINF expression vector (Berrow *et al.*, 2007) using the In-Fusion™ cloning system (Clontech Laboratories, Inc.). The pOPINFs*co5413tr* construct encoded residues 25-169 of the wild-type Sco5413 sequence with an N-terminal, 3C protease-cleavable, His$_6$-tag of sequence MAHHHHHHSSGLEVLFQGP. This construct was introduced into BL21(DE3)pLysS cells by transformation and the protein was expressed in the same way as full-length Sco5413.

## 4.4 Crystallography and Data Collection

Crystallization screening was carried out in 96-well sitting-drop MRC plates using a variety of commercial screens (Molecular Dimensions and Qiagen). The protein:well solution ratio was 1:1 (0.6 µl total volume). Many small rhombic dodecahedral crystals were grown from 0.2 M potassium nitrate and 25% PEG 3350 and optimizations of this condition were carried out in 24-well hanging drop plates. Further small crystals were generated. In order to improve these crystals, this condition was subjected to an additive screen (Hampton Research). Out of the 96 additives screened, it was observed that the addition of 0.1 M spermidine generated larger crystals. Further optimization trials revealed that the addition of 2-Methyl-2,4-pentanediol (MPD) also improved crystal size whilst the crystal form was not affected by the addition of 10% glycerol (used for cryoprotectant). The final condition consisted of 0.2 M potassium nitrate; 25% PEG 3350; 6% MPD; 10% glycerol and 1 mM spermidine. Hexagonal crystals with approximate dimensions of 200 µm x 200 µm x 150 µm grew from these conditions in less than 24 hours.
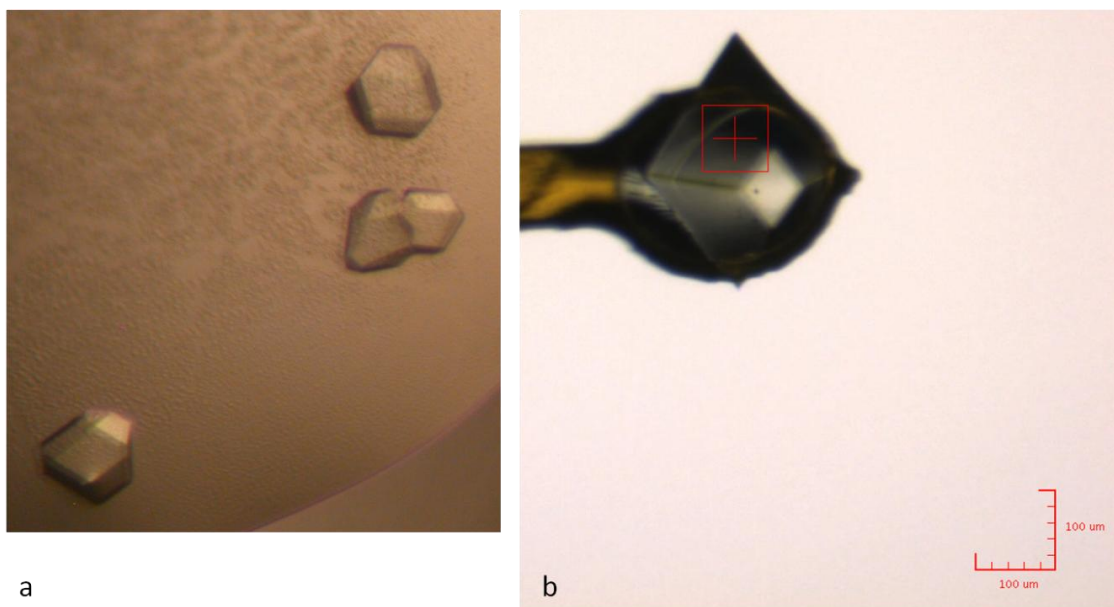


**Figure 4.14** Sco5413tr crystals. (a) In the hanging drop; (b) in the litholoop on beamline i02 at the Diamond Light Source synchrotron.

Mercury derivatized crystals were obtained by soaking the native crystals in a mother liquor solution containing a saturating amount of thimerosal for 4 hours. Crystals were backsoaked into the well-solution for 20 minutes prior to flash-freezing in liquid nitrogen. The crystals were transferred between solutions and frozen in liquid nitrogen for X-ray data collection using litholoops (Molecular Dimensions).

### 4.4.1   Native crystal data collection

The native data were collected at beamline i02 at Diamond Light Source synchrotron (Diamond) using an ADSC Quantum 315 CCD detector. It was established from test images that the crystal was orthorhombic with approximate cell parameters of $a$ = 63, $b$ = 66, $c$ = 70 Å, and the crystal diffracted X-rays to very high resolution. A low-dose medium resolution pass was collected before collecting to the maximum resolution, to ensure that the low resolution reflections were well measured (i.e. not saturated). For the medium resolution pass, 130 x 2.25° images were collected with the detector resolution set to 2.3 Å; for the high resolution pass, 500 x 0.3° images were collected with the detector resolution set to 1.20 Å. All data from both passes were integrated and merged together using the 3d option in Xia2 and, after inspection of the resultant statistics, the resolution was limited to 1.25 Å. The final data set was 100% complete at this resolution, with an overall $R_{merge}$ of 0.088 (data collection statistics are summarized in Table 4.4). An inspection of systematic absences indicated that the most likely space group was $P2_12_12_1$.
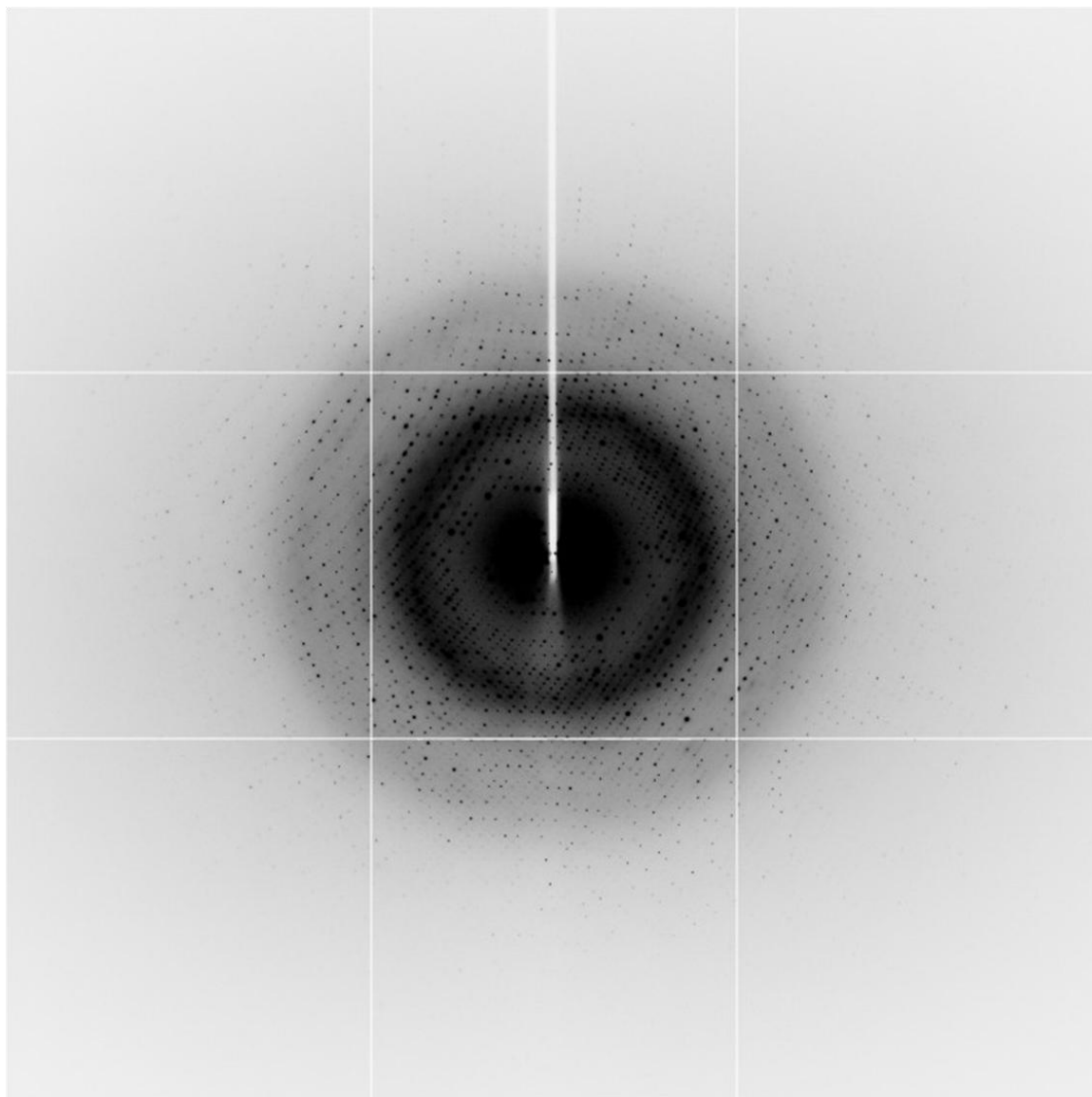
**Figure 4.15** Diffraction image for Sco5413tr showing diffraction to 1.2 Å resolution.

### 4.4.2 Heavy atom derivative crystal data collection

Derivative data were collected at beamline i24 at Diamond Light Source using a Pilatus 6M detector (Dectris). A fluorescence scan was taken around the mercury $L_{III}$ absorption edge to confirm the presence of Hg in this crystal (Figure 4.16). 5000 x 0.2° oscillation diffraction images were collected in a continuous sweep at a single wavelength of 1.0072 Å, corresponding to the $L_{III}$ edge of Hg, to a maximum resolution of 2.32 Å. The derivative dataset indexed as spacegroup $P2_12_12_1$ with cell dimensions $a$ = 62.66, $b$ = 63.10, $c$ =

68.71, using the 3d option in Xia2. The statistics from this processing were good, yielding

a dataset that was 97.4% complete and 32-fold redundant at 2.32 Å resolution.
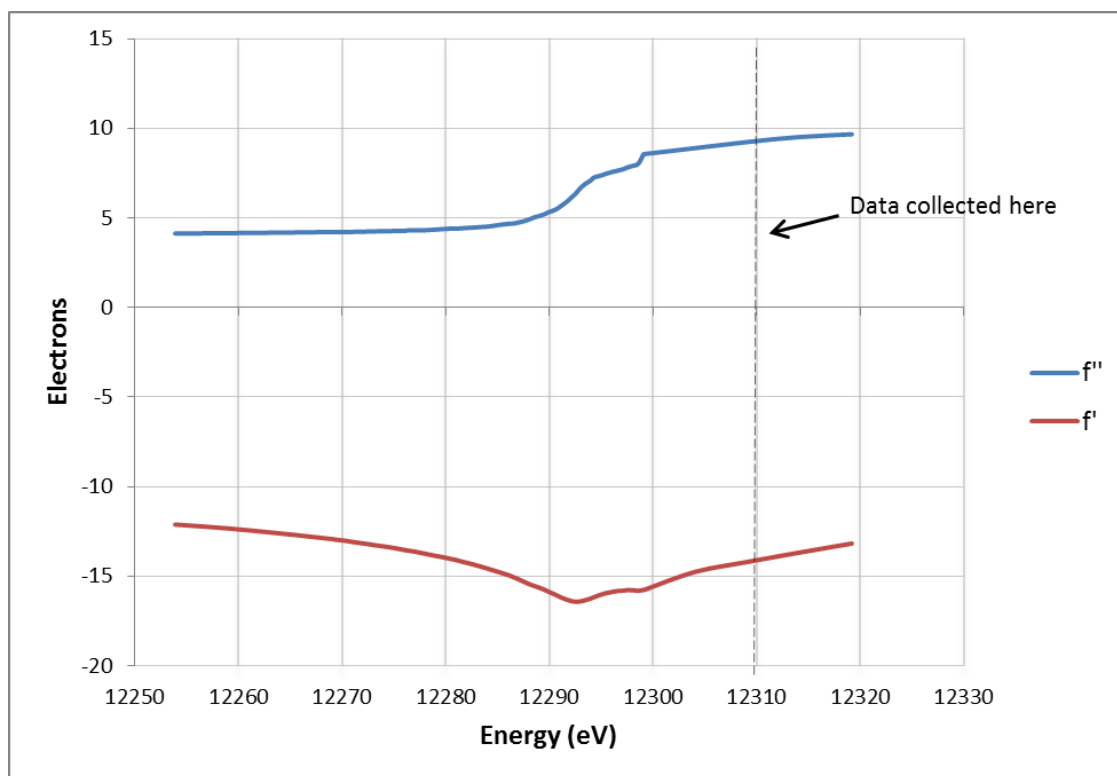


**Figure 4.16** Fluorescence scan of mercury derivative crystal taken around the mercury $L_{III}$ absorption edge.

**Table 4.4** Summary of X-ray data for Sco5413tr

| Dataset | Native | Thimerosal derivative |
|---|---|---|
| **Number of crystals** | 1 | 1 |
| **Beamline** | I02, Diamond Light Source, UK | I24, Diamond Light Source, UK |
| **Wavelength (Å)** | 0.9100 | 1.0072 |
| **Detector** | Quantum 315 CCD | Pilatus 6M |
| **Crystal-to-detector distance (mm)** | 375.2/159.8 | 501.1 |
| **Rotation range per image (°)** | 2.25/0.3 | 0.2 |
| **Exposure time per image (s)** | 0.25/1.00 | 0.2 |
| **Beam transmission (%)** | 76/100 | 23 |
| **Total rotation range (°)** | 292.5/150 | 1000 |
| **Resolution range (Å)** | 69.72 - 1.25 (1.28 - 1.25) | 31.55 - 2.26 (2.32 - 2.26) |
| **Space Group** | $P2_12_12_1$ | $P2_12_12_1$ |
| **Cell parameters (Å)** | $a = 63.40$, $b = 65.65$, $c = 69.72$ | $a = 62.66$, $b = 63.10$, $c = 68.71$ |
| **Estimated mosaicity (°)** | 0.2 | 0.3 |
| **Total no. of measured intensities** | 669686 (35277) | 415840 (17449) |
| **Unique reflections** | 81107 (5923) | 12870 (724) |
| **Multiplicity** | 8.3 (6.0) | 32.3 (24.1) |
| **Mean $I/\sigma(I)$** | 11.5 (2.6) | 26.3 (5.4) |
| **Completeness (%)** | 100.0 (100.0) | 97.4 (77.5) |
| **$R_{merge}$**[†] | 0.088 (0.709) | 0.095 (0.943) |
| **$R_{meas}$**[‡] | 0.096 (0.868) | 0.098 (0.983) |
| **Wilson $B$ value (Å$^2$)** | 14.6 | 49.6 |

[†] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. [‡] $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the $i$th observation of reflection $hkl$, $\langle I(hkl) \rangle$ is the weighted average intensity for all observations $i$ of reflection $hkl$ and $N$ is the number of observations of reflection $hkl$. [#] Collected in two passes at medium/high resolution.

### 4.4.3  Structural determination

The native and derivative data were combined to solve the structure by the Single Isomorphous Replacement with Anomalous Scattering (SIRAS) method using the Autosol wizard in PHENIX (Adams *et al.*, 2010) which used SOLVE for phasing and RESOLVE (Terwilliger, 2000, Terwilliger & Berendzen, 1999) for density modification. Autosol found two clear mercury sites and this solution gave a figure of merit (FOM) of 0.318 to 2.29 Å which increased to 0.671 following density modification.  The solution assumed two copies of the Sco5413tr monomer in the asymmetric unit (ASU) with a solvent content of 46%. These phases were then fed into Autobuild (Terwilliger *et al.*, 2008) which produced a partial model comprised of 238 residues with $R_{work}$ and $R_{free}$ values of 0.283 and 0.345, respectively, and a FOM of 0.616 at 2.29 Å resolution. Although this partial model comprised of two subunits, these did not resemble the typical MarR structure but this was amended by applying the appropriate symmetry operator to one of the subunits. Restrained refinement using REFMAC5 (Murshudov *et al.*, 1997)  against the 1.25Å  native dataset gave $R_{work}$ and $R_{free}$ values of 0.306 and 0.323 respectively, and a FOM of 0.553. The phases calculated from this model were density modified in PARROT (Cowtan, 2010) with two-fold averaging using the NCS operators derived from the dimer model. This gave a FOM of 0.755 to 1.25 Å resolution. The model was then rebuilt from scratch using BUCCANEER (Cowtan, 2006), yielding a model consisting of 286 residues with $R_{work}$ and $R_{free}$ values of 0.279 and 0.299 respectively, and a FOM of 0.802. Water molecules were fitted with Arp/WARP (Perrakis *et al.*, 1999), giving a model with $R_{work}$ and $R_{free}$ values of 0.206 and 0.236 respectively, and a FOM of 0.863. This model then underwent several iterations of rebuilding in COOT (Emsley *et al.*) and restrained refinement in REFMAC5 (with anisotropic temperature factors) resulting in a final model with 279 residues (96% of annotated sequence), 372 water molecules and 2 chloride ions. The final $R_{work}$ and $R_{free}$ values are 0.147 and 0.175 respectively, and a FOM of 0.920 at 1.25 Å resolution (final statistics in Table 4.5).
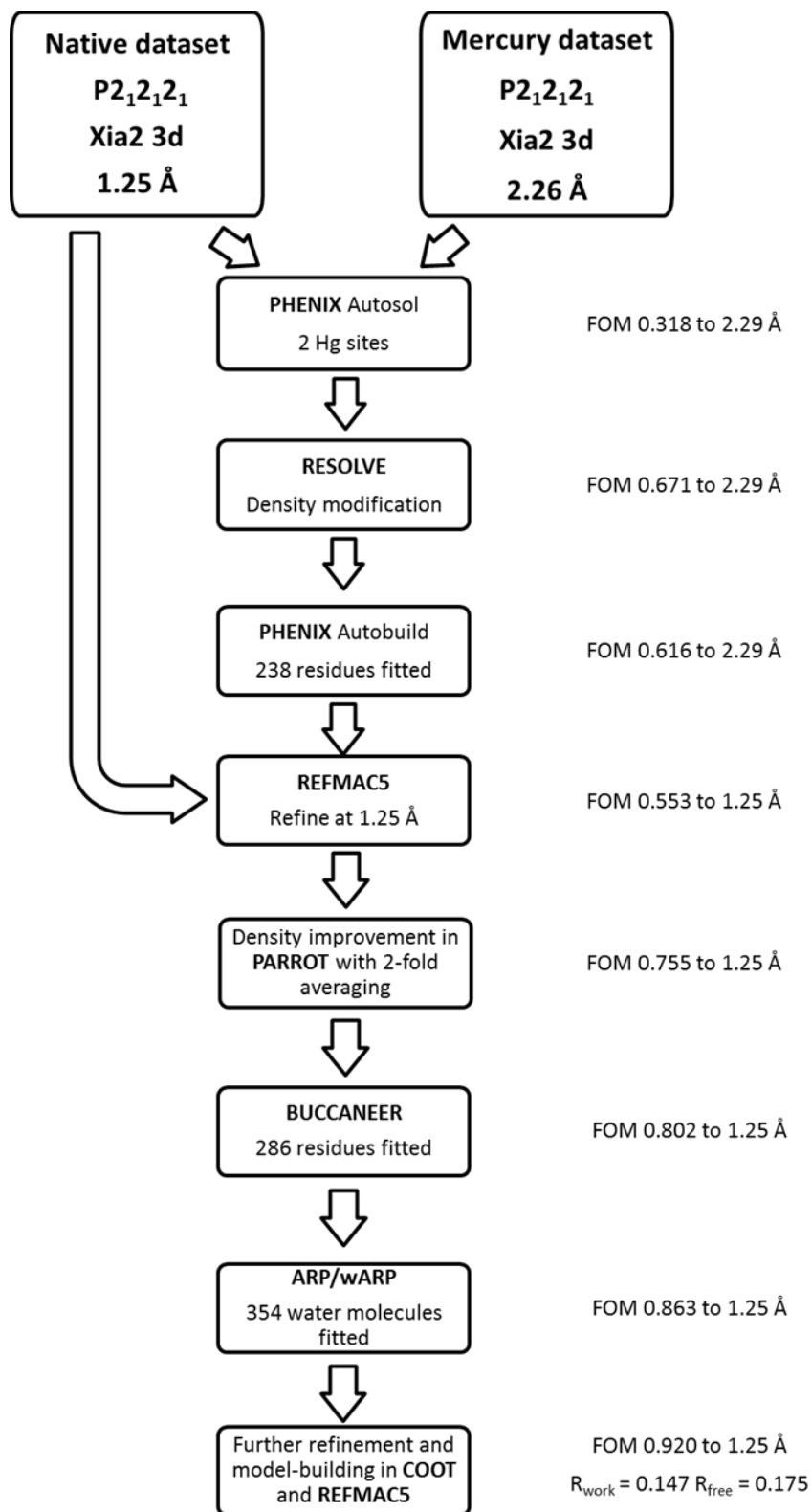
**Figure 4.17** Structure solution steps for Sco5413tr

**Table 4.5** Summary of refined model parameters for Sco5413tr

| Resolution range (Å) | 69.72 - 1.25 |
|---|---|
| $R_{work}$/ $R_{free}$[†] | 0.147/0.175 |
| Ramachandran plot: favoured/allowed/disallowed[‡] (%) | 100/0/0 |
| R.m.s. bond distance deviation (Å) | 0.01 |
| R.m.s. bond angle deviation (º) | 1.41 |
| No. of protein residues: chain A/chain B | 140/139 |
| No. of water molecules/chloride ions | 372/2 |
| Mean $B$ factors: protein/water/overall ($Å^2$) | 17.7/31.5/19.8 |

[†] The R-factors $R_{work}$ and $R_{free}$ are calculated as follows: $R = \sum(| F_{obs} - F_{calc} |)/\sum| F_{obs} | \times 100$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitudes, respectively. [‡] As calculated using *MOLPROBITY* (Davis *et al.*, 2007, Chen *et al.*, 2010).

### 4.4.4  Structural Validation using Molprobity

The model was evaluated using Molprobity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010) and 100% of the fitted residues were in the favoured region of the Ramachandran plot (Figure 4.18).

## MolProbity Ramachandran analysis

sco5413nat_2012xia2_refmac18FH.pdb, model 1
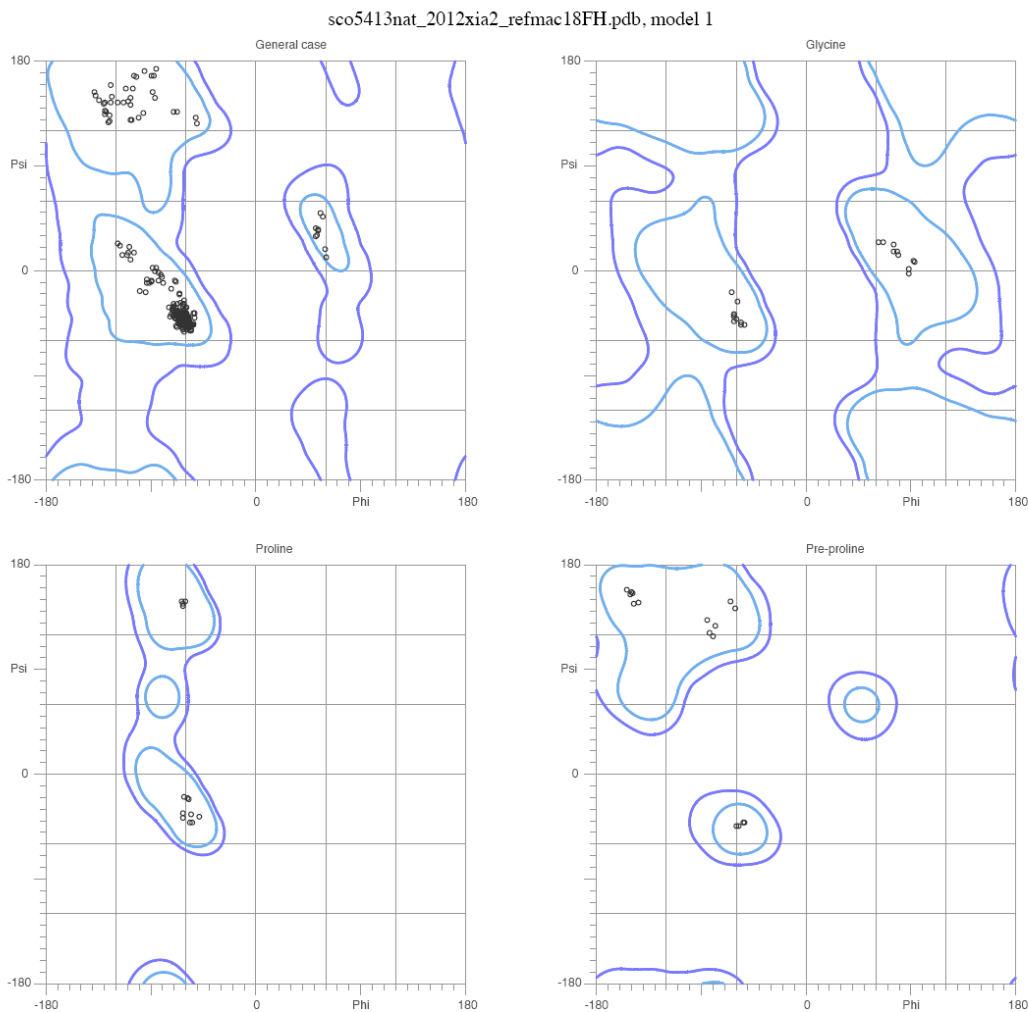


100.0% (288/288) of all residues were in favored (98%) regions.
100.0% (288/288) of all residues were in allowed (>99.8%) regions.

There were no outliers.

http://kinemage.biochem.duke.edu                    Lovell, Davis, et al. Proteins 50:437 (2003)

**Figure 4.18** Ramachandran plot for Sco5413tr.

### 4.4.5    Structural Analysis of Sco5413tr

The structure of Sco5413tr has the typical MarR features: α2 and α3 constitute the helix-turn-helix domain whilst the wing consists of two antiparallel β-strands, β2 and β3, which are connected by a loop (Figure 4.19). These are the DNA-binding domains in most MarR homologues. The wing is usually not well-resolved in other MFR structures which are not complexed with DNA, due to its innate flexibility. It is fairly well-resolved in this structure, although there are still a few residues from this region which are missing (86-89 in chain A and 87-89 in chain B). The wing is important for DNA interactions and it is better resolved in proteins which have been solved in complex with their cognate DNA sequences. Sco5413tr has the triangular shape typical of MarR family members and the dimer interface that is comprised of the N- and C- terminal helices, α1, α5 and α6. The PISA server (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) indicates that 48 residues (34% of total) of each monomer are located at the dimer interface: a total area of 2015 Å$^2$. There is also a salt bridge between Arg13 and Asp141' on each chain which connect the two monomers at the interface via the N- and C-terminal helices (although the N-terminal helix in this model is probably the second helix in the full-length protein). These could act as hinges around which the individual monomers could rotate whilst maintaining the connection at the dimer interface and contributing to the stability of the dimer. This salt-bridge could be disrupted by binding of an effector molecule. Again, this is speculative and worthy of further investigation once more is known about cognate ligands for this protein.

Other significant features of this structure are the α4 DNA recognition helices which, when superimposed onto the α4 recognition helices of the OhrR-DNA complex structure using Secondary Structure Matching (Krissinel & Henrick, 2004), have an RMSD of 0.9 Å (this value is for the superimposition of the α-recognition helices only).  This would suggest that in this structure, Sco5413tr is in a conformation that may allow DNA binding. However, superimposing the Sco5413tr α4 recognition helices onto the SlyA-DNA complex α4 recognition helices, the RMSD is 2.1 Å.  Superimposing the monomers onto each other (Sco5413tr onto both the SlyA- and OhrR-DNA complexes) produces similar

results, whereas superimposing the dimers it becomes apparent that the fit is much better with OhrR-DNA complex. A closer inspection of the conformation of the DNA in both structures shows that it differs slightly and so it is difficult to speculate about the positions of the Sco5413tr recognition helices as they relate to DNA binding. However, they do appear to be appropriately disposed relative to one another to engage with adjacent major grooves of the DNA helix.
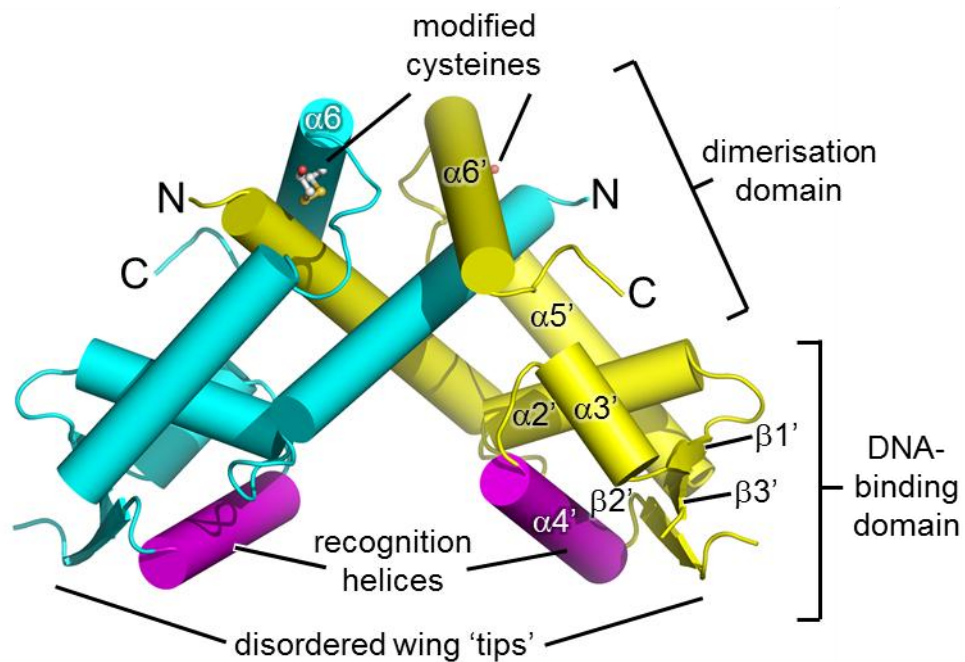


**Figure 4.19** Structure of Sco5413tr dimer. The intertwining of the N- and C- terminal helices can be seen in this model.

BldR
(3F3X)

SarZ
(3HRM)

OhrR+DNA
(1Z9C)

SlyA+DNA
(3Q5F)

HucR
(2FBK)
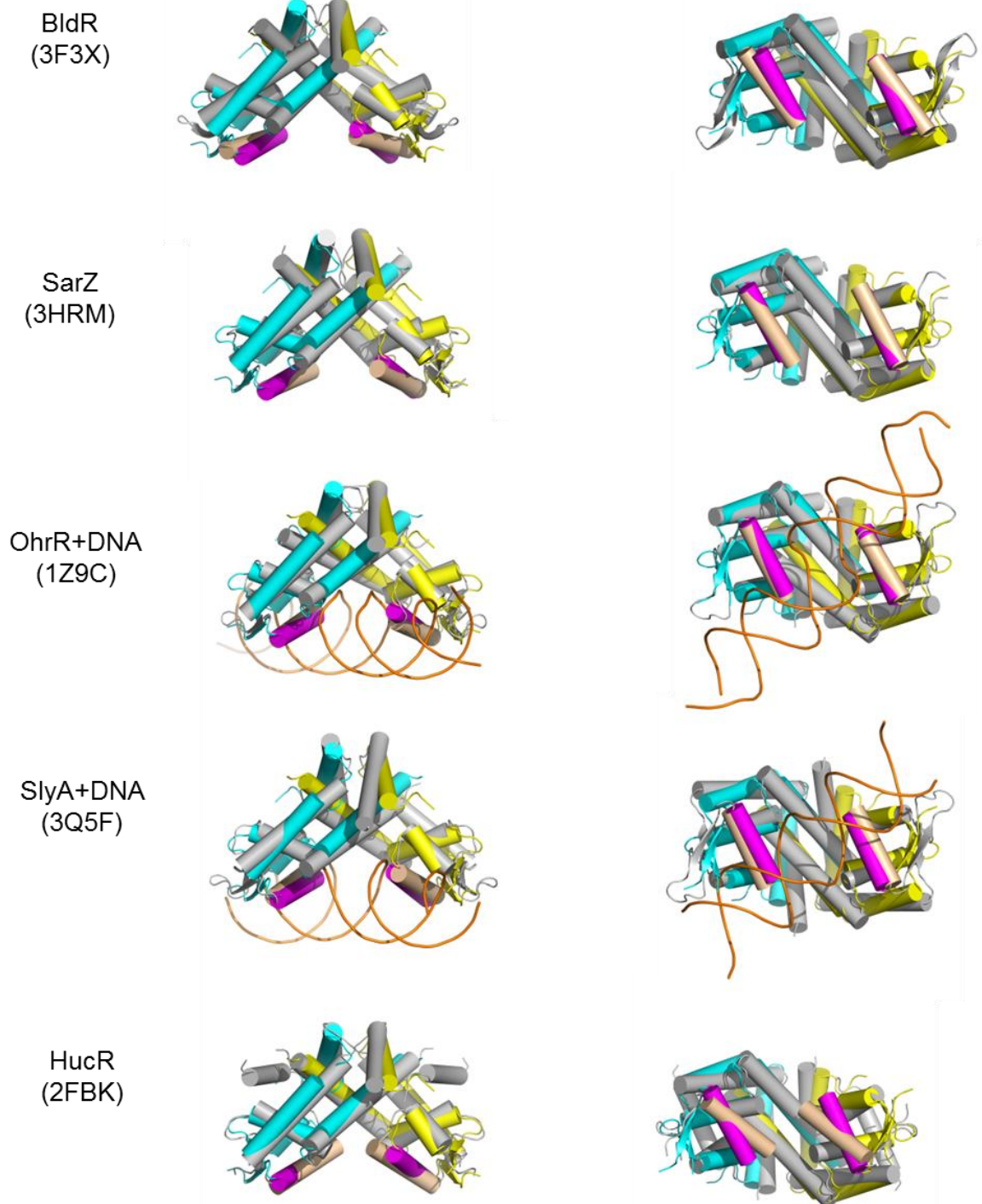


**Figure 4.20** Structural homologues superimposed onto Sco5413tr (viewed head on and from underneath). Sco5413tr is coloured in yellow and cyan and the comparison structures are in grey with pale brown recognition helices.

**Table 4.6** Structural homologues of Sco5413

| Protein | Source | DNA or ligand | PDB code | Dali | | | R.m.s. deviation (Å)/ aligned residues | | Reference |
| | | | | Z-score | Alignment length | Seq. ID (%) | Subunit | Dimer | |
|---------|--------|---------------|----------|---------|------------------|-------------|---------|-------|-----------|
| BldR | *Sulfolobus Solfataricus* | - | 3F3X | 16.2 | 129 | 19 | 1.9/126 | 2.7/246 | (Di Fiore *et al.*, 2009) |
| SarZ | *Staphylococcus aureus* | Sulfenic acid form | 3HRM | 15.9 | 129 | 16 | 1.5/125 | 1.9/249 | (Poor *et al.*, 2009) |
| OhrR | *Bacillus subtilis* | DNA | 1Z9C | 15.6 | 132 | 17 | 1.8/114 | 1.9/226 | (Hong *et al.*, 2005b) |
| SlyA | *Salmonella enterica* | DNA | 3Q5F | 15.4 | 130 | 15 | 1.7/126 | 2.8/254 | (Dolan *et al.*, 2011) |
| ST1710 | *Sulfolobus tokodaii* | salicylate | 3GF2 | 16.6 | 135 | 16 | 2.2/126 | 2.6/245 | (Kumarevel *et al.*, 2009) |
| MTH313 | *Methanobacterium thermoautotrophicum* | salicylate | 3BPX | 15.4 | 132 | 15 | 2.3/124 | 2.5/244 | (Saridakis *et al.*, 2008) |
| MarR | *E. coli* | salicylate | 1JGS | 14.5 | 127 | 20 | 2.3/109 | 3.8/202 | (Alekshun *et al.*, 2001) |
| SlyA | | salicylate | 3DEU | 13.8 | 129 | 16 | 2.5/124 | 2.8/248 | unpub |
| HucR | *Deinococcus radiodurans* | - | 2FBK | 13.2 | 132 | 20 | 2.1/1.25 | 2.2/256 | (Bordelon *et al.*, 2006) |

Also unknown is whether Sco5413tr has a ligand binding site. However, as the protein was truncated, it lacks the primary N-terminal helix. Using a HucR homology model generated by the PHYRE server, the N-terminal helix of this model was superimposed onto the Sco5413tr model to see where it may be in the structure. In HucR, an MFR from *Deinococcus radiodurans*, this N-terminal helix is referred to as an "extra" helix (Bordelon *et al*., 2006). MarR structures typically comprise six α-helices, whereas HucR has seven and Sco5413 is predicted to have seven. As discussed in Chapter 6, Sco4122 also falls into this subcategory of MarR homologues which have an extra N-terminal helix, the purpose of which is still unknown. Recent work has proposed the existence of a subfamily of MarR homologues, the urate responsive transcriptional regulator (UrtR) (Perera & Grove, 2011), which have a common mechanism of urate-mediated attenuation of DNA-binding. The study reports that all of these UrtRs have an N-terminal extension as a "salient" feature, along with the conservation of several residues involved in urate interactions, particularly the conserved tryptophan in the N-terminal helix. This residue also appears to be conserved in Sco5413 in the alignment generated by the PRALINE server (Simossis & Heringa, 2005) shown in Figure 4.21. This alignment also shows how the full-length

Sco5413 may be expected to adopt the "extra" N-terminal helix present in the UrtR homologues and Sco4122. Whether or not this is of any biological significance remains to be determined, but it is nevertheless interesting that the presence of an extra N-terminal helix has been reported as a feature in this subfamily of MFRs.
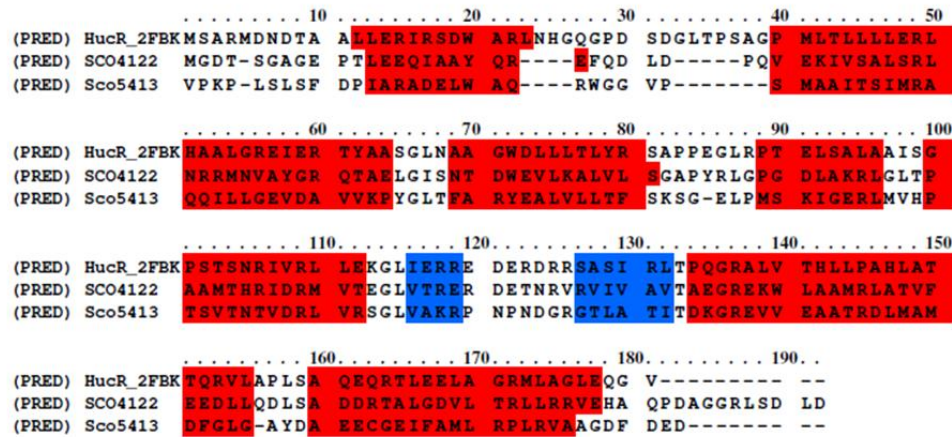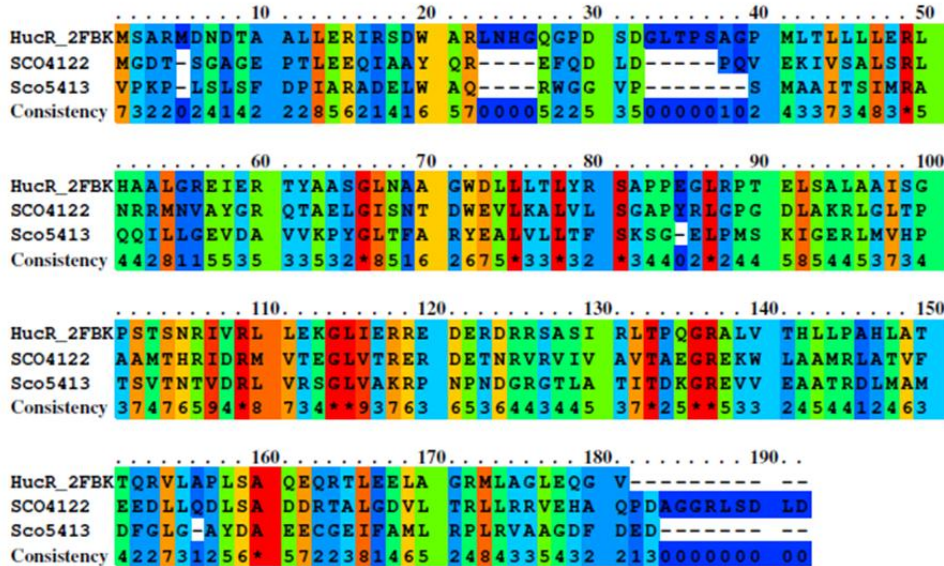


**Figure 4.21** Sequence alignments generated by the PRALINE server (Simossis & Heringa, 2005). (a) secondary structure prediction; (b) conservation of amino acids.

As with Sco3914, this protein underwent His-tag cleavage with 3CP protease which included 2 mM β-mercaptoethanol in the cleavage buffer. Again, a cysteine modification may be observed in the native structure, although this is a surface cysteine which is not located within a pocket or on the wing, as was the case with Sco3914. It is not a well conserved cysteine either and so is not considered particularly important in terms of the function of this protein.

An electrostatic surface map (Figure 4.23, generated in PyMOL) indicates a cleft located close to the dimer interface on the "shoulder" on each monomer. This putative pocket is lined mainly with neutral residues, with charged residues nearer the surface. Further structural analysis performed by PDBSum (Laskowski, 2001) suggests that there are two substantial clefts (one on each monomer in roughly equivalent positions) which extend from the one observed in Pymol through to the DNA interaction domains. However, no ligand is reported to be bound and there is no unexplained density in the structure and so it remains uncertain as to whether this is a ligand binding pocket. Superpositioning of the HucR model onto Sco5413tr presents an interesting theory: the additional N-terminal α-helix 1 sits over the cleft in Sco5413tr. In HucR, this helix contains Trp20 which, as discussed previously, is important for interactions with the substrate urate and this residue is conserved in full-length Sco5413 (Figure 4.21).
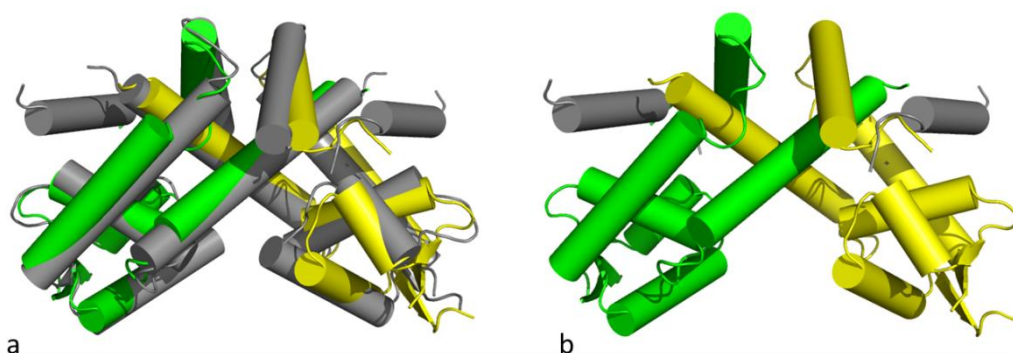


**Figure 4.22** (a) HucR (grey) superimposed onto Sco5413tr; (b) the extra N-terminal helix from HucR superimposed onto Sco5413tr.
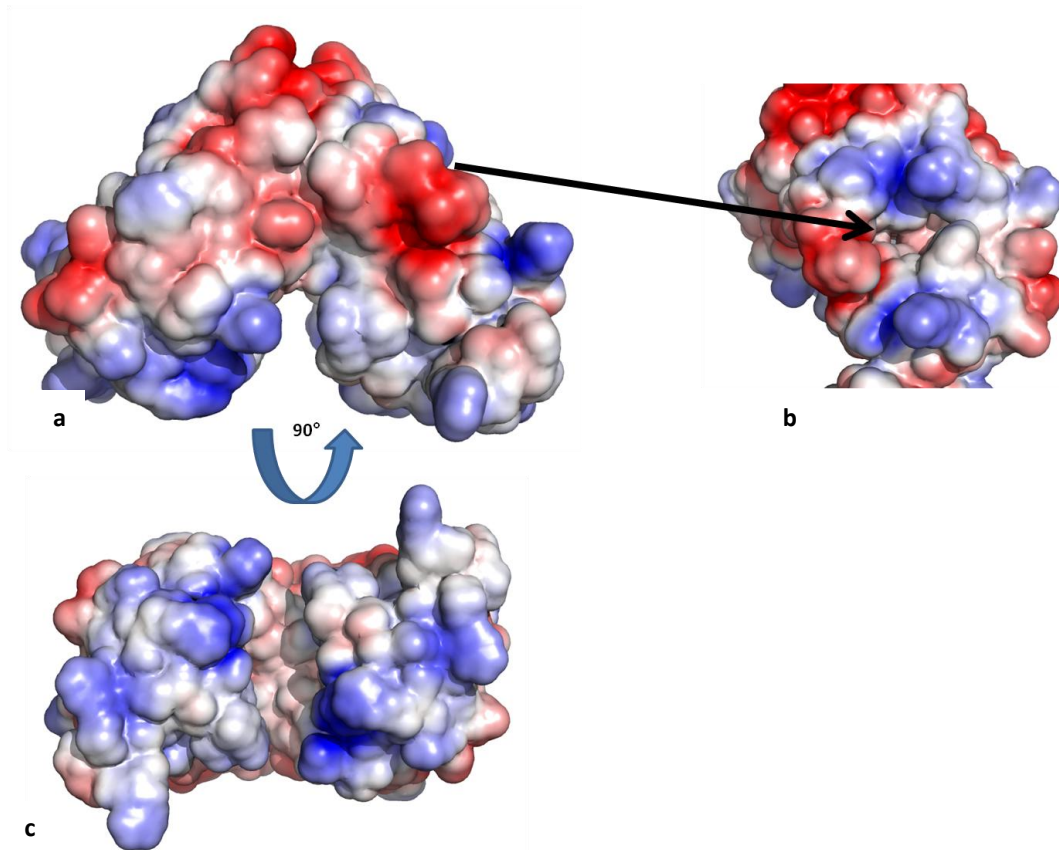
**Figure 4.23** Electrostatic surface representation of Sco5413tr dimer (negative charge coloured in red; positive charge in blue). Figure created using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). (a) shows the front view of Sco5413tr; the putative ligand pocket is magnified in (b). The DNA binding domains can be visualised in (c) and here the residues are evidently more positively charged.
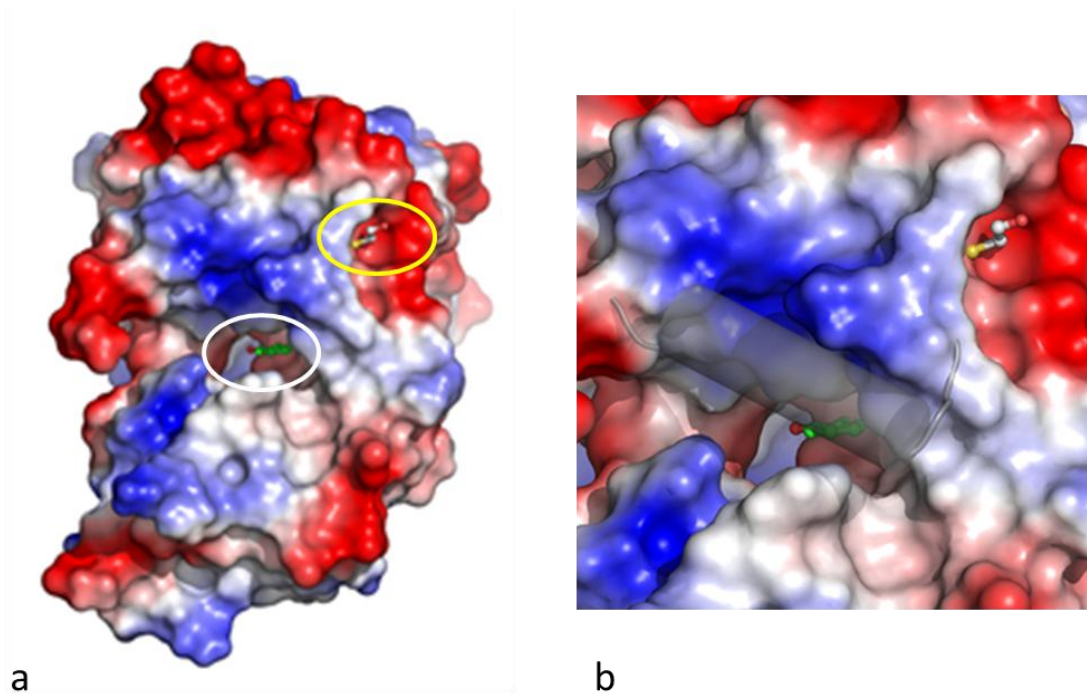
**Figure 4.24** (a) View of the putative binding site on Sco5413 with salicylate from the MTH313 structure superimposed (circled in white); the modified cysteine may also be seen (circled in yellow); (b) magnified view, including superimposed extra helix from a homology model of full-length Sco5413 based on HucR (generated by the PHYRE server (http://www.sbg.bio.ic.ac.uk/phyre2).

Relatively little is known about effector molecules of MFRs, yet much of the work has focused on salicylate as *E. coli* MarR was crystallized with salicylate bound and it has been shown to induce the multiple antibiotic resistance phenotype (Alekshun & Levy, 1999, Alekshun *et al.*, 2001, Martin & Rosner, 1995). MTH313 from *Methanobacterium thermoautotrophicum* (Saridakis *et al.*, 2008) and ST1710 from *Sulfolobus tokodaii* (Kumarevel *et al.*, 2009) are just two of the MFRs which have been co-crystallized with salicylate but whether or not this is the natural ligand for any of these proteins remains to be determined. However, these structures have shed light upon some of the features of the binding pockets and the superposition of one of the bound salicylates (SAL1) from MTH313 on Sco5413tr (Figure 4.24) shows that such an amphipathic molecule could possibly be accommodated in this cleft. There is some evidence, therefore, to suggest that this is a putative ligand binding pocket and crystallization of the full-length protein

and knowledge of the effector molecule would be useful to further investigate this hypothesis.

## 4.5  Western Blot to detect expression of Sco5413 *in vivo*

The expression of Sco5413 *in vivo* was monitored across a series of selected time points during several days of growth in liquid cultures (minimal liquid medium + 1% glucose). The proteins were incubated with an antibody against Sco5413 derived from rabbit sera then incubated with anti-rabbit IgG secondary antibody. A protein of ~17 kDa can be observed in the wild type but not in the mutant. Sco5413 is expected to be 18413 Da based on the sequence obtained from StrepDB. However, as mentioned in Section 5.1, Sco5413 has an orthologue in *S. avermitilis* (Sav2834) which has an expected size of 15669 Da and if Sco5413 were to be translated from the second potential start codon (base pair 76), then it would have a theoretical molecular weight of 15589 Da. Another explanation is that the protein runs anomalously on the SDS-PAGE gel. It is of course also possible that this is not Sco5413 and maybe a downstream protein which is affected by the mutation. For example, Sco5413 could negatively regulate another transcriptional repressor and, in the absence of Sco5413, this repressor is upregulated and so its target gene is downregulated. It is therefore very difficult to speculate what effects the removal of a putative transcriptional repressor might have.

The antibody is evidently not specific for Sco5413 and it is therefore significant that the other proteins which have an affinity for this antibody are present in what appear to be equal amounts in both samples. Also labelled are purified, tagged Sco5413 and Sco4961. This shows that the antibody has some affinity for Sco4961 (33% sequence identity to Sco5413) and it is therefore plausible that it may also have an affinity for many other proteins being expressed at these time points, hence the many bands on the Western blot (Figure 4.25). Also apparent are several bands for the purified Sco5413 which is the result of overloading the sample which also accentuates the contaminant proteins present in the sample. These may also be breakdown products of Sco5413.
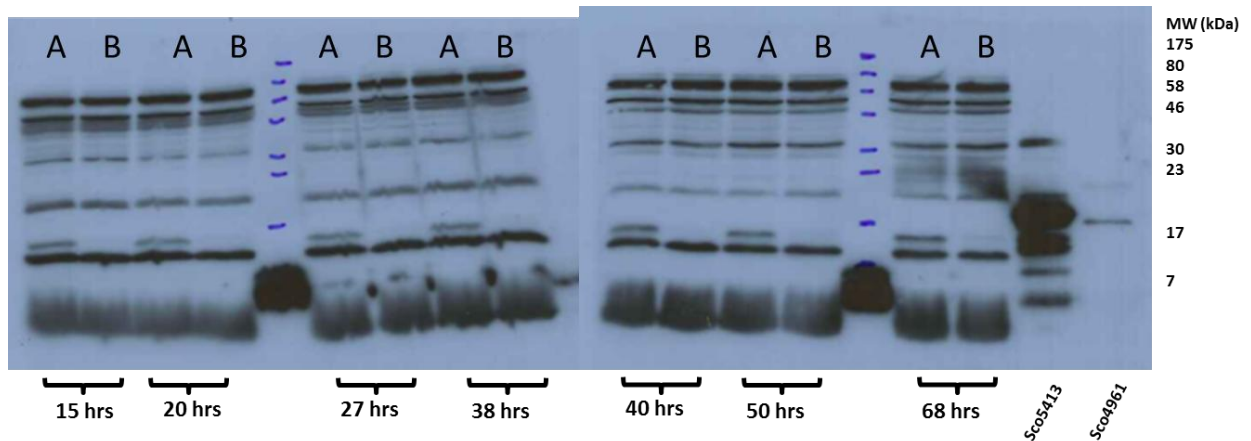
**Figure 4.25** Time course analysis of Sco5413 expression using Western Blot. A series of samples were taken at the times indicated on the blot. Sample **A** – M145 WT; Sample **B** – M145Δ*sco5413*. A band which corresponds to a protein of size ~17 kDa can be observed in the wild type but not in the mutant. It is possible that this could be Sco5413 or maybe a downstream protein, the expression of which is regulated either directly or indirectly by Sco5413.

## 4.6 DNA Interactions

### 4.6.1 Electrophoretic Mobility Shift Assay

As many prokaryotic transcription factors are self-regulating, it was necessary to check if Sco5413 would bind its own intergenic region. The first experimental approach employed was an electrophoretic mobility shift assay in which the protein is incubated with radiolabelled DNA fragments in order to detect any retardation of the DNA fragments on a non-denaturing polyacrylamide gel. The whole of the intergenic regions between *sco5412-5413* and *sco4961-4962* were amplified using the oligos shown in Table 4.7. The *sco4961-4962* intergenic region was used as a control.

**Table 4.7** Oligos used to amplify the intergenic regions between sco5412-5413 and sco4961-4962.

| | |
|---|---|
| CGCCACCTCGCCCCCTGACCGTGCCGGTGG | Int54125413_F |
| ACTGAGGGGCTTCGGCACGGCAAGAGACCT | Int54125413_R |
| GGCGGTCGTCATGTCCATCGACATTTCAGG | Int49614962_F |
| CGCTGTGCTGTAGCTCATAGCTCCAGCATC | Int49614962_R |

5 µl of the PCR products (~ 100 ng) were labelled with $[^{32}P\text{-}\gamma]$-ATP and incubated with differing concentrations of protein for the binding reaction (detailed in Chapter 2). Sco5413 was in a 25 mM Tris pH8, 100 mM NaCl buffer. Figure 4.26 shows that there was no retardation of the fragments on the gel when incubated with the protein, suggesting that Sco5413 does not bind either of these intergenic regions under the experimental conditions.
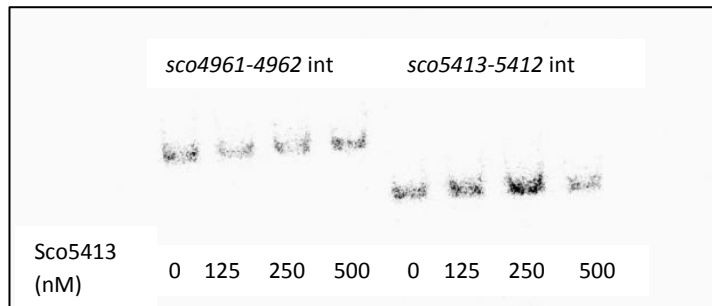


**Figure 4.26** Electrophoretic Mobility Shift Assay using purified (tagged) Sco5413 and the amplified intergenic regions of *sco4961-4962* and *sco5413-5414*. The concentrations of each protein sample are indicated above each lane. The image shows that Sco5413 did not bind to either of the intergenic fragments under these conditions.

It is not certain whether the protein required incubation in a different buffer in or the addition of a specific cofactor in order to bind these sequences, or whether it is as simple as Sco5413 does not bind these intergenic sequences. However, a single EMSA was not conclusive enough to rule out the possibility that Sco5413 was autoregulatory and so further experiments were pursued to test this hypothesis.

### 4.6.2   ChIP on Chip

ChIP on chip was used to identify potential operator sites for Sco5413 within the *Streptomyces coelicolor* genome and thereby define its regulon. Sco5413 was chosen for this experiment as it was the only protein in this project that exhibited a visible phenotype in the deletion mutant strain. The samples sent to Oxford Gene Technology (OGT) for this experiment were the wild type M145 genomic DNA and the M145Δ*sco5413* genomic DNA. All the samples were prepared as detailed in Chapter 2 by growing up the cells and cross-linking any proteins attached to the genomic DNA using formaldehyde.

The cells were lysed and the genomic DNA was fragmented by sonication into fragments of ~500 bp. The Sco5413 specific antibody (prepared by Cambridge Research Biochemicals) was used to immunoprecipitate this protein from the cell lysate and the DNA was purified from this extraction. The purified DNA fragments were then labelled and hybridized to an OGT ChIP microarray. (This stage was carried out by OGT).

Two time points were selected for the experiment: a sample was cross-linked at 16 hours and another sample at 90 hours. The selection of these time points was based on the mutant phenotypes as there was no apparent difference between the wild type and the mutant until at least 48 hours. It was therefore hoped that these time points would allow a distinct difference between the two strains to be detected in the microarray data. A screenshot of some of the results as displayed in the OGT Ltd ChIP Browser is shown in Figure 4.27.
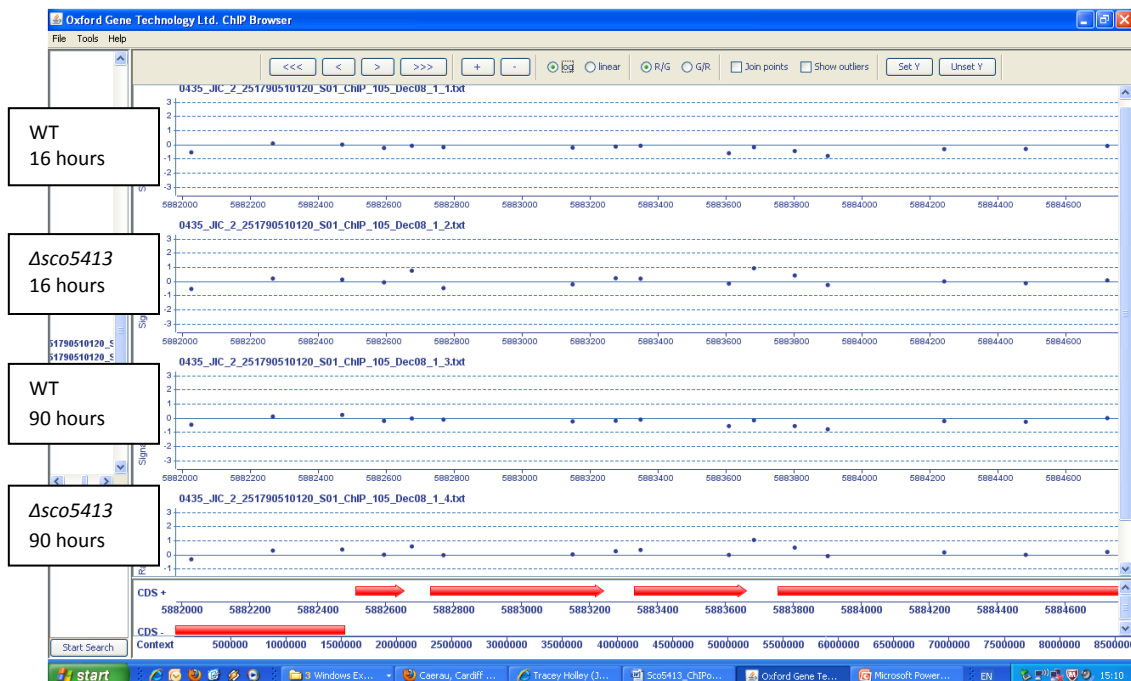


**Figure 4.27** A screenshot of the OGT ChIP Browser with the microarrays for each of the four samples (Wild type and mutant at two different time points) displayed. This window shows the region of Sco5413 and there are no obvious peaks detected between the microarrays. Each spot refers to the DNA probe which is attached to the chip surface.

Yet, as the results show, there was no discernable difference between the two: no obvious peaks in the microarray data and so what conclusions can be drawn from this? Firstly, perhaps the wrong time points were selected. As a transcriptional regulator, it is possible that Sco5413 does not necessarily need to bind its cognate sequence at these specific time points and so it was unfortunate that the appropriate window may have been missed. Secondly, there may have been impurities with the antibody. Although the protein was purified before it was sent to CRB for the antibody generation, it is of course possible that it contained undetected amounts of contaminating proteins from the expression *E. coli* strain. Another possibility is that there is enough similarity between Sco5413 and other *S. coelicolor* proteins for the antibody to be able to bind several other proteins in the immunoprecipitation step. A Western Blot (see section 4.5) demonstrated that in the native producer, the antibody targeted several other proteins and so it is conceivable that this could have been an issue. This would introduce a lot of noise into the experiment and so potentially mask any otherwise significant signal.

Another potential problem with this experiment was that the phenotype was manifested on SFM (mannitol soya flour) solid medium which had been enriched with 1% glucose/1% glycerol or 10 mM GlcNAc. It was considered too difficult to attempt to extract the samples from liquid SFM as this medium has many large, aggregated particles which could hinder the extraction of protein and/or DNA. The liquid medium used was 50:50 TSB:YEME (Chapter 2). Under these conditions it is possible that Sco5413 was not expressed (or at least not in any quantity sufficient for the success of this experiment).

### 4.6.3   Surface Plasmon Resonance

The purpose of the experiment was to determine whether the protein binds its own intergenic region and the precise DNA sequence to which the protein binds. This would then hopefully give some indication as to whether or not the protein is self-regulating and opens up the potential to uncover other putative binding sites in the genome. Although the EMSA experiment had given a negative result, it was important to ensure that this

was not due to the conditions of the experiment. The advantage of using this approach over the EMSA method was that this technology requires very little protein and also eliminates the need for radioactive labelling. SPR also offers the potential to explore the kinetics of the interactions.

The preliminary stage of this experiment involved determining whether the protein would bind the whole intergenic sequence. The *sco5413-5412* intergenic region was amplified by PCR (using the primers shown in Table 4.8) to incorporate a single-stranded overhanging sequence which is complementary to the biotinylated ssDNA annealed to the streptavidin chip (details in Chapter 2). Unlike with Sco3914 and Sco4122, the first step to generate a template was not required as non-specific binding was not as much of an issue. This whole intergenic sequence was used as a preliminary check to see if the protein would bind. Initial results (Figure 4.28) indicate that there is a negligible interaction between the protein and the DNA. The response unit values are all less than 10 RU and the reponses drop off immediately at the end of the protein injection, signifying a negligible interaction. This was attempted under both normal and reducing conditions (10 mM final concentration dithiothreitol (DTT) in the running buffer). The increased response observed for the 10 mM DTT sample is most likely an effect of the DTT rather than an enhanced protein interaction.

**Table 4.8** Oligos used for the 2 step PCR amplification of the intergenic region. The nested sequence is in green and the sequence which is complementary to the biotinylated oligo is in blue. The reverse oligo from step 2 was used again in step 3.

| Oligo name | Sequence |
|---|---|
| Step 2 Forward | TTATCAAAAAGAGTATTGACCCGTGCCGGTGGGCGG |
| Step 2 Reverse | CACGGCAAGAGACCTTACCGGC |
| Step 3 Forward | TTAGCGTTTGCGTATGCG/dSpacer/TTATCAAAAAGAGTATTGAC |
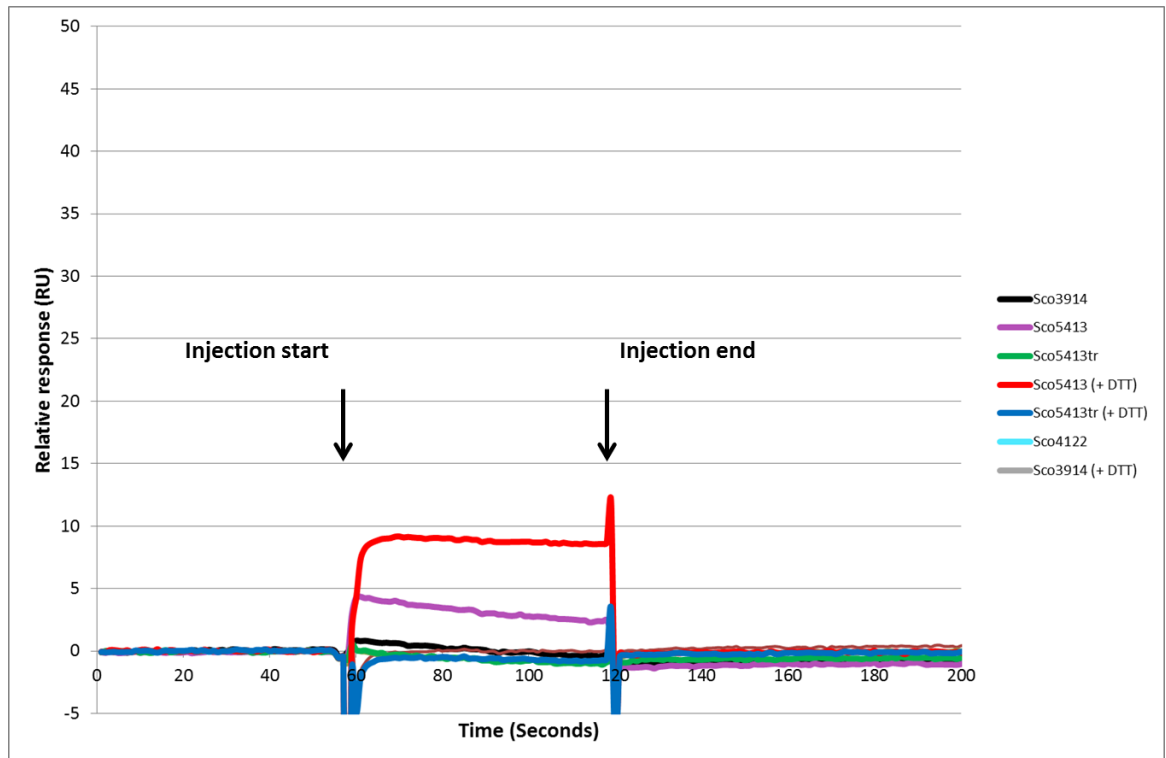
**Figure 4.28** Sensorgram showing Sco5413 (and Sco5413tr) interacting with the PCR amplified intergenic DNA sequence. Sco3914 (as prepared and with 10 mM DTT) and Sco4122 were used as controls to indicate that there may be some small interaction between Sco5413 and this specific sequence. However, it is likely that this is transient and perhaps not significant given the immediate dissociation at the end of the injection.

The next step involved "fragmenting" the intergenic region in order to see if there is any specific interaction (or indeed any interaction). The intergenic region of 76 base pairs was fragmented to incorporate a 25 bp overlap between each sequence as shown below and the oligos were ordered from Sigma. The forward oligo included the 20 bp ssDNA overhang which is complementary to the ssDNA annealed to the streptavidin chip. In this experiment, the complementary ssDNA sequences were annealed by mixing 60 µl of the reverse oligo with 50 µl of the forward oligo to give a 45 µM stock solution (PCR program used for this is detailed in the materials and methods chapter). This was diluted to the working concentration of 1 µM with HBS-EP+ buffer. Each dsDNA sequence was then captured on the chip by running it across the analytical flow cell containing the ssDNA linker. This is detailed in Chapter 2. The full-length protein Sco5413 (including the His-tag

sequence) was then run across both the control and analytical flow cells and the results for this are shown in Table 4.9.

Prior to this experiment, a MEME analysis of the upstream intergenic regions of several top orthologues in reciprocal BLAST searches was performed and the following MEME motif found:-
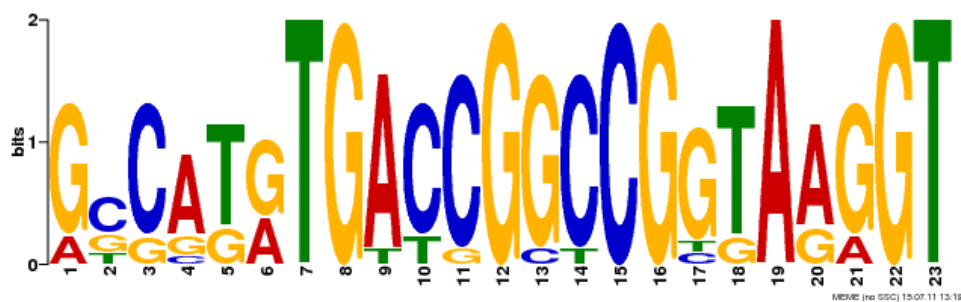


### *sco5413-5414* Intergenic sequences

TGACCGTGCCGGTGGGCGGTGGTATCGAAGGGCGAGACGGAGCTGACCGGCATGTGACCGGCCGGTAAGGTCTCTTGCCGTG

0-40 CCGTGCCGGTGGGCGGTGGTATCGAAGGGCGAGACGGAGC

16-55    GTGGTATCGAAGGGCGAGACGGAGCTGACCGGCATGTGAC

31-70 GAGACGGAGCTGACCGGCATGTGACCGGCCGGTAAGGTCT

46-76 GGCATGTGACCGGCCGGTAAGGTCTCTTGCC

**Figure 4.29** The 76 base pair intergenic region between *sco5413* and *sco5414* and the four fragments used in the experiment. The MEME motif is highlighted in pink.

This motif is contained within sequences 3 and 4 and so the expectation was that if there was to be any binding, it would be within these sequences. However, as can be seen in the results shown in Table 4.9, no apparent preferential binding of either sequence was

observed. In fact, there was no apparent binding of any of the sequences. The interaction of the protein with each sequence was calculated as a percentage of the $R_{max}$ value for each DNA sequence captured and plotted as a histogram. The plot quite clearly shows that there was no protein:DNA interaction with any of the sequences or any of the concentrations of protein. The figures shown in the table are the calculated percentages of the $R_{max}$ and the negative values would suggest that there is in fact slightly more interaction of the protein with the control flow cell (fc1) which contains the linker sequence alone.

**Table 4.9** Raw data from the SPR experiment showing the amount of DNA captured at each cycle and the calculated R$_{max}$ for each cycle. The protein interaction was measured in response units which were then calculated as a percentage of the R$_{max}$. There are two replicates for each protein concentration.

| Run 1 | | | | | | Run 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| **100 nM** | **MW** | **DNA Capture** | **Theoretical Rmax** | **Protein bound** | **%Rmax** | **DNA capture** | **Theoretical Rmax** | **Protein bound** | **%Rmax** |
| **1** | 30590 | 444.0 | 462.4 | -1.0 | -0.2 | 446.3 | 464.8 | -1.1 | -0.2 |
| **2** | 30586 | 460.6 | 479.8 | -1.0 | -0.2 | 462.3 | 481.5 | -1.3 | -0.3 |
| **3** | 30587 | 438.2 | 456.4 | -0.9 | -0.2 | 441.6 | 460.0 | -1.2 | -0.3 |
| **4** | 25024 | 463.9 | 590.6 | -0.9 | -0.2 | 465.2 | 592.3 | -1.3 | -0.2 |
| **50 nM** | | | | | | | | | |
| **1** | 30590 | 446.3 | 464.8 | -0.9 | -0.2 | 447.1 | 465.6 | -1.1 | -0.2 |
| **2** | 30586 | 461.3 | 480.5 | -0.9 | -0.2 | 461.9 | 481.1 | -1.2 | -0.2 |
| **3** | 30587 | 440.1 | 458.4 | -0.9 | -0.2 | 441.5 | 459.9 | -1.1 | -0.2 |
| **4** | 25024 | 465.2 | 592.3 | -1.0 | -0.2 | 466.6 | 594.0 | -1.3 | -0.2 |
| **10 nM** | | | | | | | | | |
| **1** | 30590 | 445.5 | 464.0 | -0.9 | -0.2 | 448.6 | 467.2 | -1.0 | -0.2 |
| **2** | 30586 | 462.2 | 481.4 | -1.0 | -0.2 | 462.7 | 482.0 | -1.2 | -0.2 |
| **3** | 30587 | 440.6 | 458.9 | -0.8 | -0.2 | 442 | 460.4 | -1.2 | -0.3 |
| **4** | 25024 | 465.5 | 592.6 | -1.0 | -0.2 | 466.9 | 594.4 | -1.2 | -0.2 |

The results shown here indicate that there is no significant interaction between the protein and the DNA and therefore further interrogation of this intergenic region will not be pursued.

## 4.7 Conclusions

Sco5413 has been shown in this study to possess the physical characteristics of a MarR family regulator (MFR): it is a dimer with a winged-helix DNA-binding domain and the dimerization domain consists of the N- and C- terminal helices interacting at the dimer interface. However, as a putative transcriptional regulator, it has not been shown that it is capable of binding DNA and there is more work to be done in this area.

There are many potential reasons as to why the ChIP on Chip experiment did not yield any positive results and, although frustrating in terms of trying to find a function for Sco5413, it does support the previous findings in the EMSA and SPR experiments, which also failed to detect any significant DNA interactions. Perhaps Sco5413 requires a ligand or cofactor to bind to DNA. Perhaps it only binds under specific conditions which have yet to be assayed. It could be that this protein is not expressed at the time points selected, but as it does not self-regulate, it cannot be detected at all. This experiment was carried out prior to the SPR experiment and so, apart from the negative result from the EMSA, there was no reason to suspect that this experiment would not yield useful results. Nevertheless, in the interests of presenting all of the work done on this protein as a confirmation of its lack of DNA-binding capability, these negative results are included.

The structure of truncated Sco5413, which was solved to 1.25 Å, is the highest resolution structure of a MarR homologue to date. This structure reveals a distinct channel which is lined by well conserved residues and this is possibly a binding site for ligands which may regulate the affinity of the protein for its cognate DNA sequence. Interestingly, a structure prediction for the full-length protein indicates that it would likely have the extra N-terminal helix seen in HucR and Sco4122 (see Chapter 5). It also has a tryptophan in this putative helix which appears to be conserved in HucR and interacts with its substrate urate. The modified cysteine does not appear to be well conserved and so it was not investigated further. Additionally, comparisons with other MFR structures indicate that Sco5413tr is in a conformation which is compatible with DNA-binding.

One final point of speculation about Sco5413 is that despite the lack of evidence for it binding DNA, it is still a protein worth more investigation. It is highly represented in the genomes of other bacteria and as such, it is most likely to have some important function. The distinct phenotype in the gene deletion strain is also a further indication of its potential to yield some very useful information relating to actinomycete development.

# Chapter 5

# Sco4122

## 5.1 Introduction

Sco4122 is predicted to be a 22 kDa protein which is divergently transcribed to Sco4121, a possible integral membrane protein (Figure 5.1). As another putative member of the MarR family of transcriptional regulators (MFRs), this protein shares a sequence identity of 22% with the MarR homologue from *Deinococcus radiodurans*, HucR (Bordelon *et al.*, 2006). Biophysical analysis of its DNA-binding capabilities revealed that it is capable of specific binding to its own promoter region and so the assumption is that this protein is also autoregulatory.

The *sco4122* gene is 549 bp in length and has a GC content of 71%. It has 15 orthologues amongst 112 actinobacterial genomes and *E. coli* and *Bacillus subtilis* (bioinformatics analysis done by Govind Chandra, John Innes Centre). Its close homology to orthologues in other published sequences of streptomycetes can be seen in Table 5.1 below.

**Table 5.1**  The sequence identities shared between the orthologues from other streptomycetes and *Streptomyces coelicolor* Sco4122. Sav – *Streptomyces avermitilis*; Ssc – *Streptomyces scabius*; Sgr – *Streptomyces griseus*: Sven – *Streptomyces venezuelae*.

| | Amino acid sequence %ID | | | | DNA Sequence %ID | | | |
|---|---|---|---|---|---|---|---|---|
| **Sco** | **Sav** | **Ssc** | **Sgr** | **Sven** | **Sav** | **Ssc** | **Sgr** | **Sven** |
| Sco4122 | Sav4105 91% | Scab49171 87% | Sgr3909 80% | Sven38764 80% | Sav4105 87% | Scab49171 86% | Sgr3909 81% | Sven38764 83% |
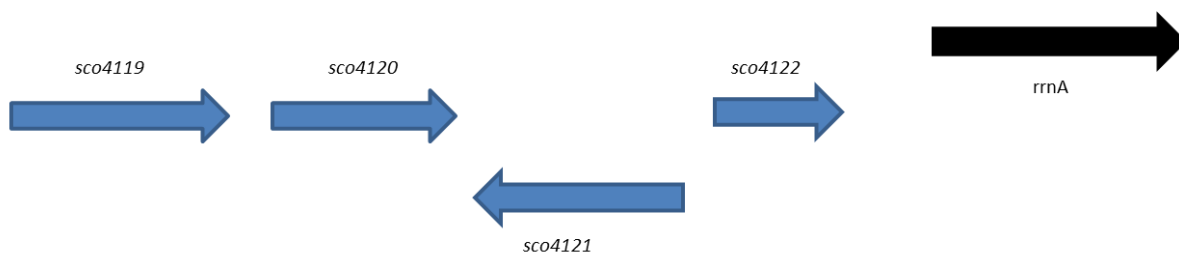
Figure 5.1    A representation of the location of s*co4122* in the *Streptomyces coelicolor* genome (http://strepdb.streptomyces.org.uk).

Sco4122 has been cloned, expressed, purified and crystallized and the structure solved to 2.35 Å resolution. A gene disruption mutant was generated in the host organism but no visible phenotype observed under the conditions tested.

Bioinformatic analysis of this intergenic region also revealed several putative MEME motifs, the first of which corresponded to a significant response in SPR experiments. The DNA footprint of Sco4122 has been determined in this work using Surface Plasmon Resonance. There is the opportunity for further studies to be carried out regarding this particular footprint as there is a suggestion that there may be some form of cooperative binding occurring in this region. Cooperative binding of transcriptional repressors has been shown in other cases: OhrR in *Streptomyces coelicolor* binds cooperatively to the *ohrR-ohrA* intergenic region under reducing conditions (Oh *et al.*, 2007), repressing the transcription of the divergently transcribed OhrA (an organic hydroperoxide resistance protein). Sco4122 is transcribed divergently to Sco4121, a possible integral membrane protein which shares a strong sequence identity with putative integral membrane proteins from other streptomycetes as well as various other actinobacteria.

There is also the potential for Sco4122 to be crystallized with its cognate sequence in order to identify the precise protein-DNA interactions.

The results obtained for the biophysical, genetic and structural characterization of this protein will be discussed in this chapter.

## 5.2   Protein Expression and Purification

The N-terminally His-tagged protein was cloned into the expression vector pOPINF (Berrow *et al.*, 2007) and expressed in BL21 (DE3) pLysS cells according to the method described in chapter 2. Sco4122 elutes from the nickel affinity column with a shouldered-peak profile (Figure 5.2), the first part of which contains several contaminating proteins. The fractions were pooled and applied to the size exclusion gel filtration column (GE Healthcare Superdex 75 16/60). The purity of the protein following the nickel affinity column purification and the size exclusion gel filtration chromatography column can be seen in the SDS-PAGE analyses (Figure 5.3 and Figure 5.5).
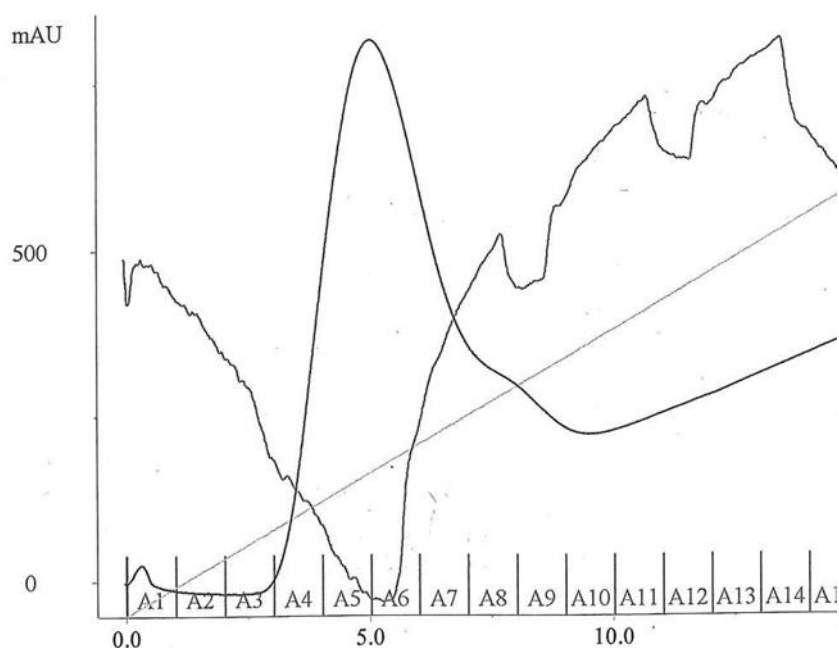


**Figure 5.2** Elution profile of Sco4122 from nickel affinity column chromatography. The peak has a slight shoulder as the protein tends to elute with some contaminating proteins. (The UV absorbance continues to increase after the protein has eluted as the imidazole used absorbs UV light at 280 nm).
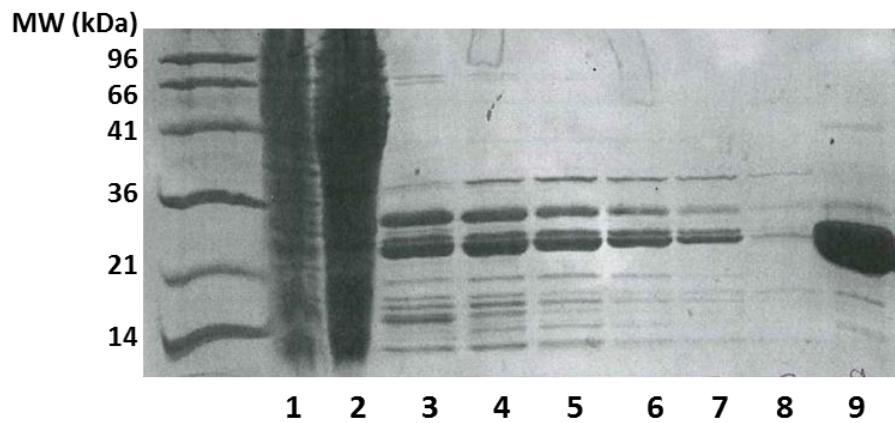
**Figure 5.3** 15% SDS-PAGE after nickel affinity column purification. The lanes contain: (1) Lyzed pellet; (2) Supernatant; (3) Fraction A4; (4) A5; (5) A7; (6) A8; (7) A9; (8) A11; (9) previous preparation of Sco4122.
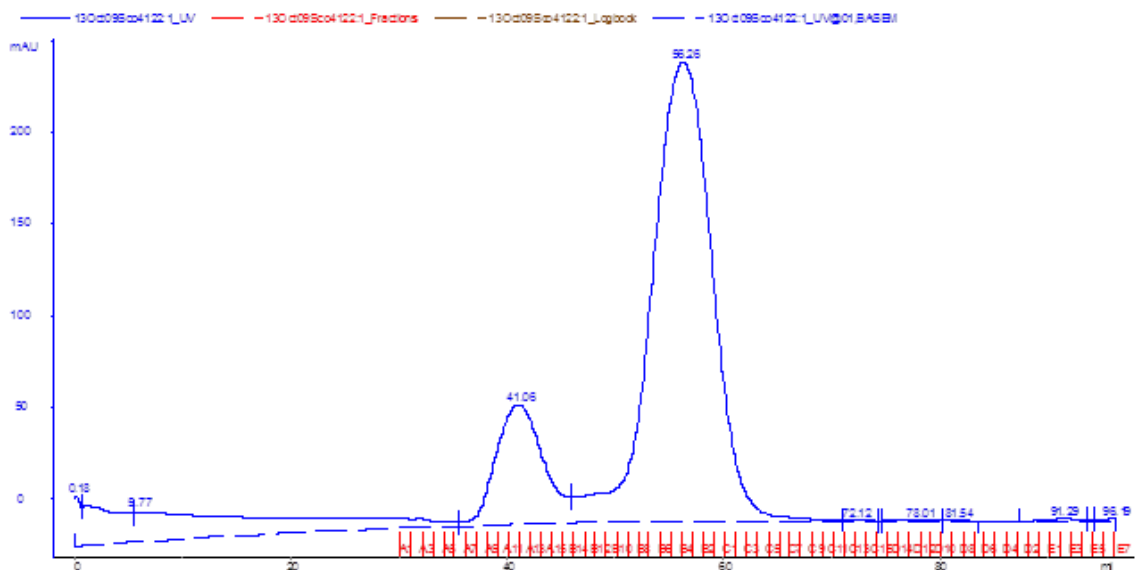


**Figure 5.4** Elution profile of Sco4122 from size exclusion gel filtration chromatography. The first peak contains the protein in a higher oligomeric state whilst the second peak also contains Sco4122 (calculated to be slightly larger than a dimer at 52 kDa).
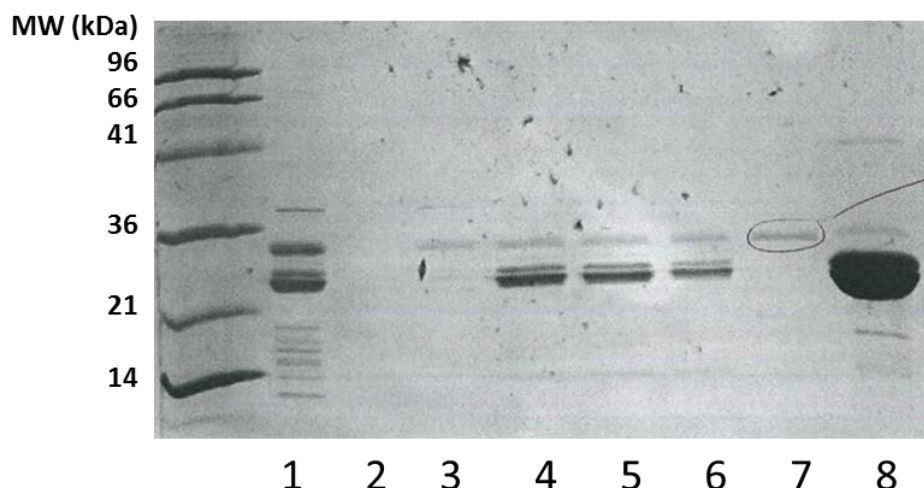
**Figure 5.5** 15% SDS-PAGE after size exclusion gel filtration column purification. The lanes contain: (1) Pooled fractions A9-A11 from nickel column; (2) Fraction A13; (3) B12; (4) B5; (5) B4; (6) B2; (7) C8; (8) previous preparation of Sco4122.

The standard curve generated by the calibration of the gel filtration column was used to calculate the estimated molecular weight of the protein eluted from the column. The largest peak (eluted at 56.26 ml) corresponds to a molecular weight of 52 kDa. The predicted molecular weight of the Sco4122 dimer is 44.326 kDa. This includes the 17 amino acid 6xHis tag, highlighted in yellow on the full-length sequence below:-

MAHHHHHHSSGLEVLFQGPGDTSGAGEPTLEEQIAAYQREFQDLDPQVEKIVSALSRLNRRMNVAY
GRQTAELGISNTDWEVLKALVLSGAPYRLGPGDLAKRLGLTPAAMTHRIDRMVTEGLVTRERDETNR
VRVIVAVTAEGREKWLAAMRLATVFEEDLLQDLSADDRTALGDVLTRLLRRVEHAQPDAGGRLSDLD

As with Sco3914 and Sco5413, this value is slightly larger than the predicted molecular weight but both of these were shown to be dimers in the crystal structures. Nevertheless, DLS was performed on this protein and demonstrates that it is most likely a dimer in solution. The experiment recorded a peak indicating that there is a predominant species with a radius corresponding to molecular mass of 35 kDa. This is closely similar to the predicted 44 kDa and therefore it is assumed that Sco4122 is a dimer.
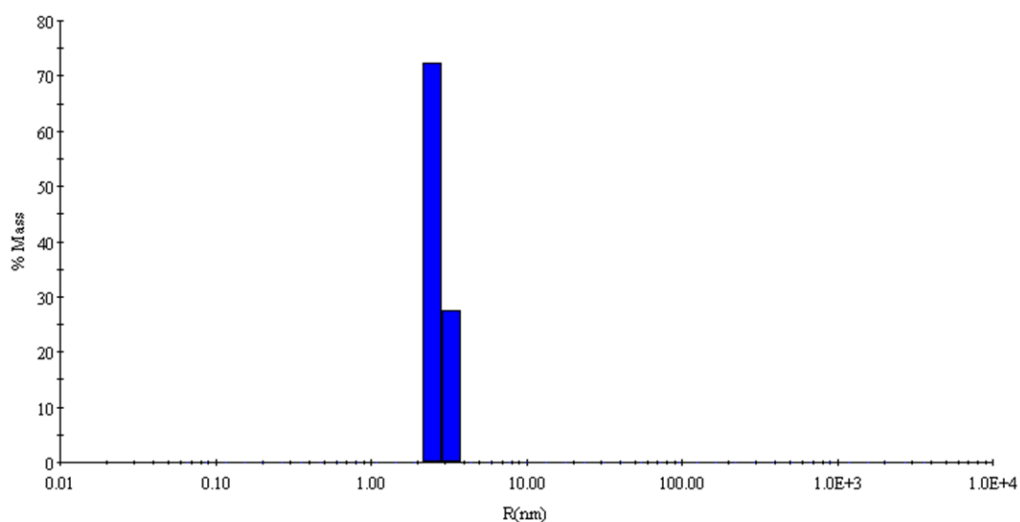
| Item | R (nm) | %Pd | MW-R (kDa) | %Int | %Mass |
|---|---|---|---|---|---|
| **Peak 1** | 2.7 | 13.2 | 35 | 61.7 | 100 |
| **Peak 2** | 37 | 36.6 | 15697 | 35.7 | 0 |
| **Peak 3** | 220.7 | 0 | 1026160 | 2.6 | 0 |

**Figure 5.6** Dynamic Light Scattering profile of Sco4122. The peaks represent the species of a particular size (radius) as a percentage of the total mass present in solution. Sco4122 is present in solution as a dimer.

### 5.2.1 Expression and Purification of Selenomethionine substituted protein

For the preparation of selenomethionine substituted protein, the growth cultures were subjected to metabolic inhibition and supplemented with the appropriate stocks of amino acids and selenomethioine prior to induction with 1 mM IPTG (details in Chapter 2). The protein was purified using the same methods as with the native protein, but the yield was much lower (as would be expected): approximately 0.4 mg l$^{-1}$. However, this protein eluted with less contaminants than the native protein and so gel filtration was not performed to avoid further loss of sample.
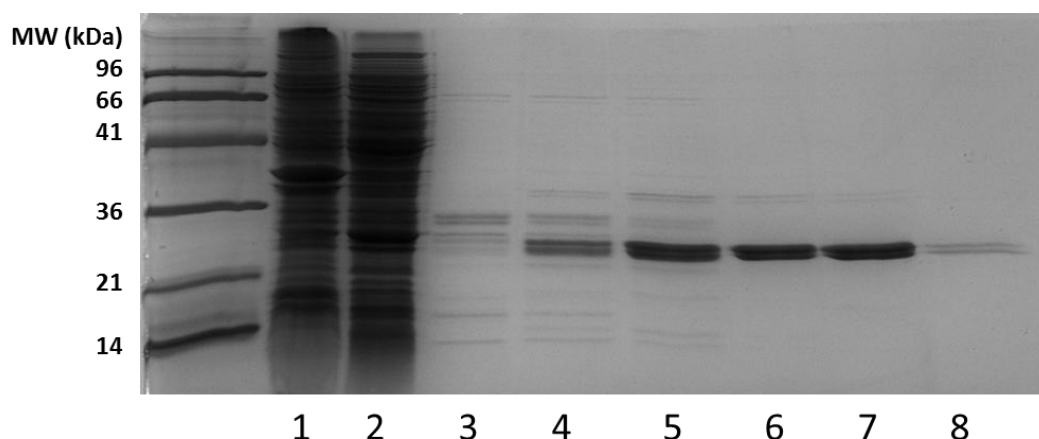
**Figure 5.7** 15% SDS-PAGE after nickel affinity column purification of selenomethionine substituted Sco4122. The lanes contain: (1) uninduced pellet; (2) induced pellet; (3) to (8) fractions containing Sco4122 (all of which were pooled).

Sco4122 has five methionines in its sequence and these are highlighted in cyan the sequence below (including the N-terminal methionine of the His-tagged sequence shown below in yellow).

MAHHHHHHSSGLEVLFQGPGDTSGAGEPTLEEQIAAYQREFQDLDPQVEKIVSALSRLNRRMNVAY
GRQTAELGISNTDWEVLKALVLSGAPYRLGPGDLAKRLGLTPAAMTHRIDRMVTEGLVTRERDETNR
VRVIVAVTAEGREKWLAAMRLATVFEEDLLQDLSADDRTALGDVLTRLLRRVEHAQPDAGGRLSDLD

Successful incorporation of selenomethionine in Sco4122 was confirmed by mass spectrometry. Electrospray ionization (Micromass Q-ToF 2 mass spectrometer) was used to determine the molecular weight of the proteins (native and selenomethionine-substituted). The spectra, shown in Figure 5.8, indicate that several main components existed in the selenomethionine-substituted protein preparation. The molecular masses of these components correspond to different amounts of selenomethionine in the sequence. The native protein has a molecular mass of 22031 Da. This is 132 Da less than the theoretical 22163 Da for the Sco4122 monomer, but this may be explained by the possible absence of the N-terminal methionine which would leave 4 methionines in the sequence. The spectra for the SeMet derivative protein has 4 different species present, the first of which corresponds to the native protein (22031 Da), suggesting that there is

some native protein present in the sample. The most abundant species in the sample have molecular weights of 22219 Da and 22397 Da; a mass increase on the native protein of 188 Da and 366 Da respectively. There is a mass difference of 47 Da between sulphur and selenium. Taking the first species of 22219 Da, the extra 188 Da corresponds to 4 x 47 Da, suggesting that all four methionines are in fact selenomethionines. The larger species could have the N-terminal (seleno)methionine intact which would account for the extra 366 Da: (4 x 47 Da) + 178 Da = 366 Da (selenomethionine has a molecular weight of 178 Da).
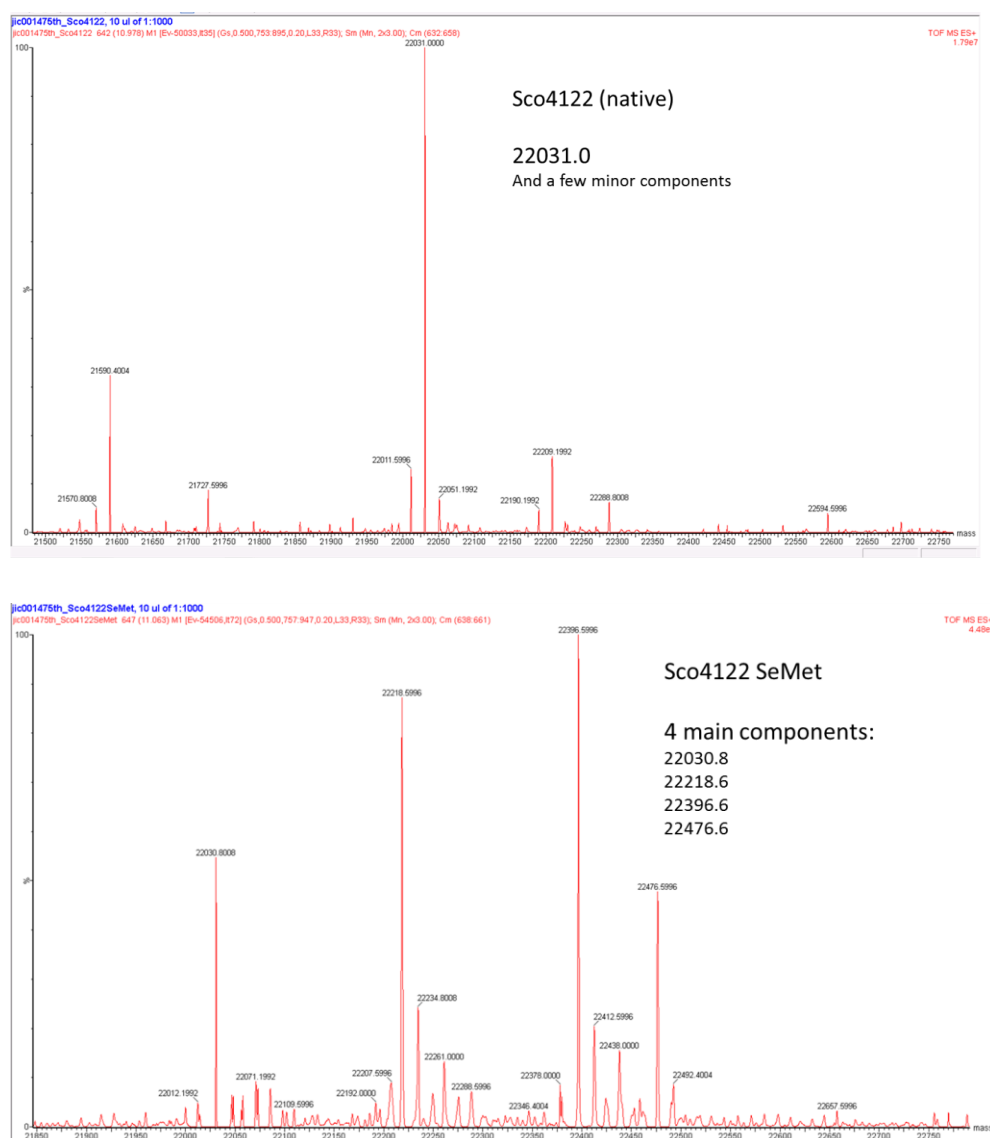


**Figure 5.8** Spectra showing the molecular weights of the proteins as determined using electrospray mass spectrometry (ESMS). The top figure shows data for native Sco4122 and the lower figure shows the data for selenomethionine-substituted protein (Sco4122 SeMet).

## 5.3 Crystallography and Data Collection

Crystallization screening was carried out in 96-well sitting-drop MRC plates using a variety of commercial screens (details in Chapter 2). The protein:well solution ratio was 1:1 (0.6 µl total volume). Many small needle-like crystals grew from 0.2 M ammonium nitrate and 25% PEG 3350 (JCSG-plus™ Molecular Dimensions) after 5 days. Crystal growth usually began with precipitate which then formed oily-looking globules in the drops before forming microcrystals, some of which then developed into needle-like clusters (Figure 5.9). Optimizations of this condition were carried out in 24-well hanging drop plates. Further small crystals were generated (Figure 5.10), although crystal growth was inconsistent and could take up to 7-10 days. Sometimes the crystals did not grow well in conditions that had already generated crystals, perhaps an indication of how sensitive this protein is to the smallest perturbations. Further optimization trials revealed that the addition of 0.4% (v/v) 2-Methyl-2,4-pentanediol (MPD) slightly improved crystal size, whilst crystal growth was negatively affected by the addition of 10-15% glycerol. Some crystals were grown in the presence of 10% ethylene glycol, but the resultant diffraction was not good enough for structural solution. The final condition consisted of 0.2 M ammonium bromide; 15% PEG 3350; 0.4% (v/v) MPD. This condition was arrived at following unsuccessful attempts to obtain a halide derivative crystal by cocrystallization. Needle-like crystals with approximate dimensions of 200 µm x  15 µm x 15 µm grew from these conditions in approximately 7 days. The crystals were soaked briefly (~ 10 seconds) in this condition supplemented with 15% (v/v) ethylene glycol prior to flash-freezing in liquid nitrogen.
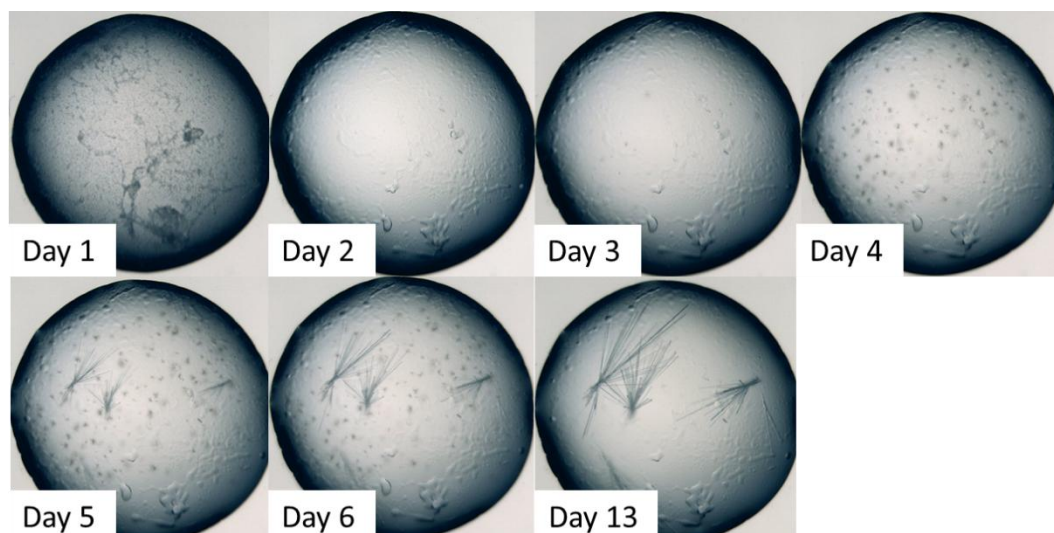
**Figure 5.9** The growth of Sco4122 crystal in the sitting drop crystallization screen plate over the course of 13 days. The crystal growth begins with precipitation leading to phase separation, formation of microcrystals and culminating in growth of very delicate needle-like crystals. These thin crystals extend and thicken slightly after 2 weeks.
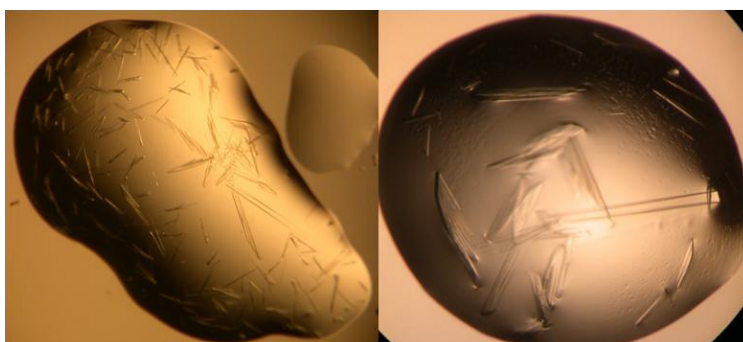


**Figure 5.10** Sco4122 crystals obtained by optimization of the original conditions from the crystallization screen (0.2M Ammonium nitrate, 20% w/v PEG 3350). It was possible to obtain more disperse and thicker crystals with slight variations of this condition.

Structural determination was attempted using molecular replacement with a template model based on HucR (2FBK) generated by the PHYRE server (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley & Sternberg, 2009), but this was unsuccessful, although not surprising given the low identity (23%). Attempts were made to generate derivative crystals using co-crystallization (ammonium bromide) and heavy atom soaking techniques (details in Chapter 2). These soaks often resulted in the cracking

or complete destruction of the crystal, but it was possible to collect data from a few of the survivors. Data were collected for a variety of these crystal soaks (platinum, gold and bromide) close to their respective x-ray absorption edges, but no significant anomalous signal was detected. The native data were collected on station i02 at the Diamond Light Source Synchrotron (Oxfordshire) to a resolution of 2.35 Å. The unit cell dimensions of $a$ = 44.91, $b$ = 92.64, $c$ = 93.33, $\alpha$ = 90.00°, $\beta$ = 95.67°, $\lambda$ = 90.00° were consistent with the space group, $P2_1$. Statistics for these data are shown in Table 5.2. For the anomalous dataset, selenomethionine substituted protein crystals were obtained by metabolic inhibition (Chapter 2). Crystals were cryoprotected by soaking briefly (~10 seconds) in the well solution containing 15% ethylene glycol and then immediately the crystals were flash-frozen in liquid nitrogen. The crystals were transferred between solutions and frozen in liquid nitrogen for X-ray data collection using litho-loops (Hampton Research).  A fluorescence scan was taken around the selenium $K$ absorption edge to confirm the presence of selenium in the protein (Figure 5.12). The data were collected to 2.8 Å resolution at a wavelength of 0.9796 Å corresponding to the absorption peak. A first pass consisted of 500 images and a second pass in a different part of the crystal consisted of 600 images. These images were combined and processed using Xia2 to 3.0 Å resolution. The cell parameters for the SeMet crystal were  = 44.86 Å, b = 103.37  Å, c = 81.78 Å, $\alpha$ = 90.00°, $\beta$ = 96.21°, $\lambda$ = 90.00°, consistent with the space group $P2_1$ but with a different cell to the native dataset. Examples of the diffraction images from each data collection are shown in Figure 5.11.
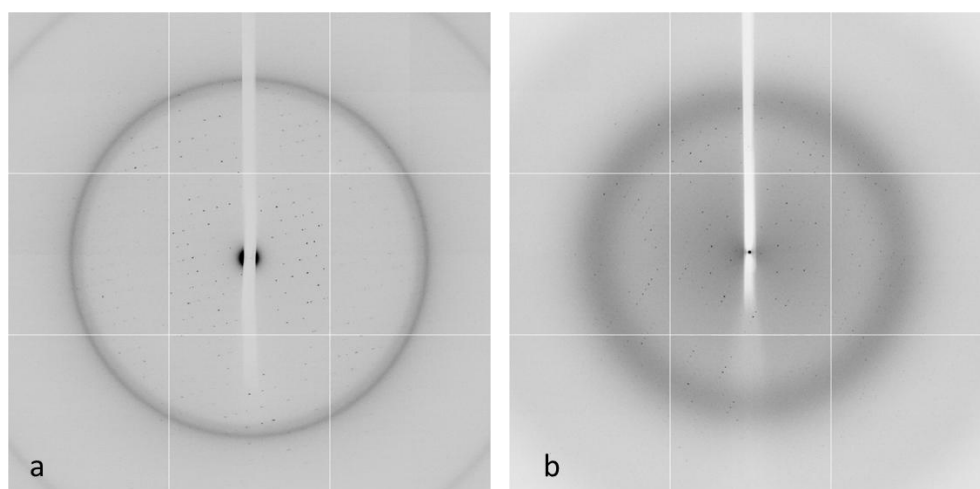


**Figure 5.11** Diffraction images from Sco4122 (a) SeMet crystal and (b) native crystal.

**Table 5.2** Data collection summary for Sco4122. Values in parentheses are for the outer resolution shell.

| Dataset | Native | Selenomethionine derivative |
|---|---|---|
| Number of crystals | 1 | 1 |
| Beamline | i02, Diamond Light Source, UK | i04, Diamond Light Source, UK |
| Wavelength (Å) | 0.97 | 0.9796 |
| Detector | ADSC Q315r detector | ADSC Q315r detector |
| Crystal-to-detector distance (mm) | 345.5 | 422.0 |
| Rotation range per image (°) | 1.0 | 1.0 |
| Exposure time per image (s) | 1.5 | 0.6 |
| Beam transmission (%) | 44.73 | 100 |
| Total rotation range (°) | 300 | 660 |
| Resolution range (Å) | 92.87-2.35 (2.41-2.35) | 81.30-3.00 (3.08-3.00) |
| Space Group | $P2_1$ | $P2_1$ |
| Cell parameters (Å) | $a$ = 44.91, $b$ = 92.64, $c$ = 93.33 | $a$ = 44.86, $b$ = 103.37, $c$ = 81.78 |
|  | $\beta$ = 95.67 | $\beta$ = 96.21 |
| Total no. of measured intensities | 200232 (14242) | 302458 (21418) |
| Unique reflections | 31769 (2310) | 14947 (1084) |
| Multiplicity | 6.3 (6.2) | 20.2 (19.8) |
| Mean $I/\sigma(I)$ | 13.5 (2.4) | 13.1 (3.4) |
| Completeness (%) | 99.9 (99.8) | 100 (100) |
| $R_{\text{merge}}$[†] | 0.108 (0.811) | 0.215 (1.087) |
| $R_{\text{meas}}$[‡] | 0.129 (0.977) | 0.227 (1.146) |

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. [‡] $R_{\text{meas}} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the $i$th observation of reflection $hkl$, $\langle I(hkl) \rangle$ is the weighted average intensity for all observations $i$ of reflection $hkl$ and $N$ is the number of observations of reflection $hkl$.
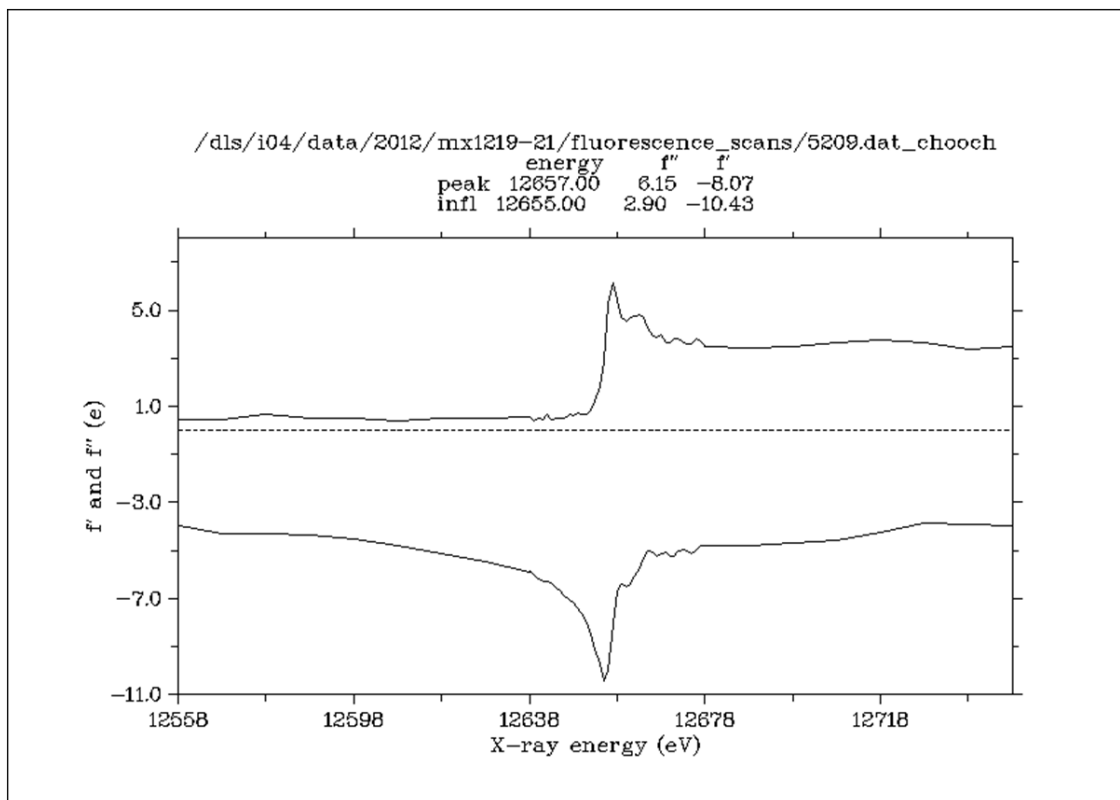
**Figure 5.12** A fluorescence scan taken at the selenium *K*-edge at the Diamond Light Source Synchrotron, station i04. The energy at the peak is 12657 eV, where f'' = 6.15 e and f' = -8.07 e.

### 5.3.1 Structural determination

The diffraction images were initially processed using the 3dii option in Xia2. The SeMet datafiles were fed into PHENIX (Adams *et al.*) which used the AutoSol wizard (Terwilliger *et al.*, 2009) to solve the structure by the Single Wavelength Anomalous Dispersion (SAD) method. This wizard starts with the datafiles containing the amplitudes of structure factors and identifies the heavy atom sites (using HySS) and calculates the associated phases (using SOLVE). Fifteen good selenium sites were identified, giving a figure of merit of 0.34 to 3.0 Å resolution following density modification (using RESOLVE). There are 4 copies of the monomer in the asymmetric unit (728 residues in total expected) with an estimated solvent content of 48%. These phases were then fed into AutoBuild which generated a partial model. This partial model was very incomplete: 376 residues had been built, of which 53 residues had been sequenced. Yet, as with Sco3914, the predominantly α-helical structure characteristic of MarR family proteins was apparent. Four copies of a

template MarR model based on HucR (PDB code 2FBK) (generated by the PHYRE server) were docked onto this partial model in COOT (Emsley *et al.*, 2010), using a combination of SSM superpose (Krissinel & Henrick, 2004) and manual fitting. This gave two dimers in the asymmetric unit with 432 residues in total and this model was refined (restrained refinement) in REFMAC5. The phases from this model were then combined with the unmodified experimental phases using SIGMAA from the CCP4 Suite (Winn *et al.*, 2011). The combined phases were then four-fold averaged in PARROT (Cowtan, 2010). The output from PARROT was used by BUCCANEER (Cowtan, 2006) to rebuild the model from scratch. 640 residues were sequenced with a figure of merit (FOM) of 0.692 at this stage. This model was improved in COOT, rebuilding the most complete chain and copying this to the other chains, before refinement in REFMAC5 resulting in a FOM of 0.724. This seemed a reasonable model to use as a molecular replacement template to solve the structure with the native dataset. PHASER used the monomer from this model to search for four copies in the native dataset at 4.5 Å resolution, yielding a model which consisted of two dimers in the ASU. After 20 cycles of restrained refinement with automatic NCS restraints, the FOM had increased to 0.785 with an $R_{work}$ of 0.275 and $R_{free}$ of 0.293. Further iterations of rebuilding, refinement and solvent addition using ARP/wARP (Perrakis *et al.*, 1999) gave a model with 616 residues and 104 waters fitted with an $R_{work}$ of 0.209, $R_{free}$ of 0.245 and FOM of 0.849.
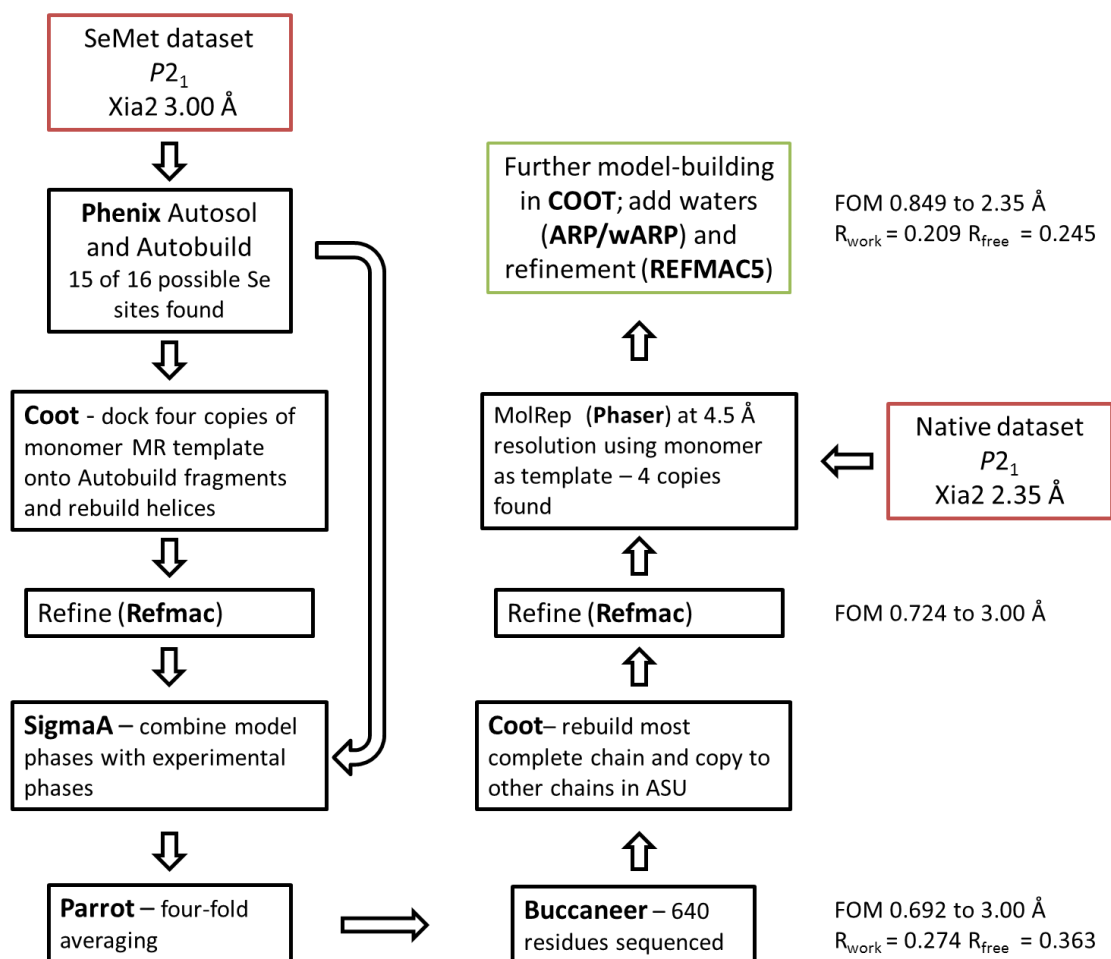
**Figure 5.13** Structure solution steps for Sco4122

**Table 5.3** Summary of refined model parameters for Sco4122

| | |
|---|---|
| Resolution range (Å) | 92.87-2.35 |
| $R_{work}$/ $R_{free}$[†] | 0.20/0.25 |
| Ramachandran plot: favoured/allowed/disallowed[‡] (%) | 98.2/100/0 |
| R.m.s. bond distance deviation (Å) | 0.01 |
| R.m.s. bond angle deviation (º) | 1.40 |
| No. of protein residues: chain A/chain B/chain C/chain D | 160/153/152/152 |
| No. of water molecules | 105 |
| Mean $B$ factors: protein/water/overall (Å$^2$) | 36.7/35.5/36.7 |

[†] The R-factors $R_{work}$ and $R_{free}$ are calculated as follows: $R = \sum(|\,F_{obs} - F_{calc}\,|)/\sum|\,F_{obs}\,| \times 100$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factor amplitudes, respectively. [‡] As calculated using *MOLPROBITY* (Davis *et al.*, 2007, Chen *et al.*, 2010).

Structural validation was carried out using Molprobity (Chen *et al.*, 2010, Davis *et al.*, 2007). The Molprobity score is respectable (in the 97th percentile) and the Ramachandran plot (Figure 5.14) shows that there are no outliers in the structure.
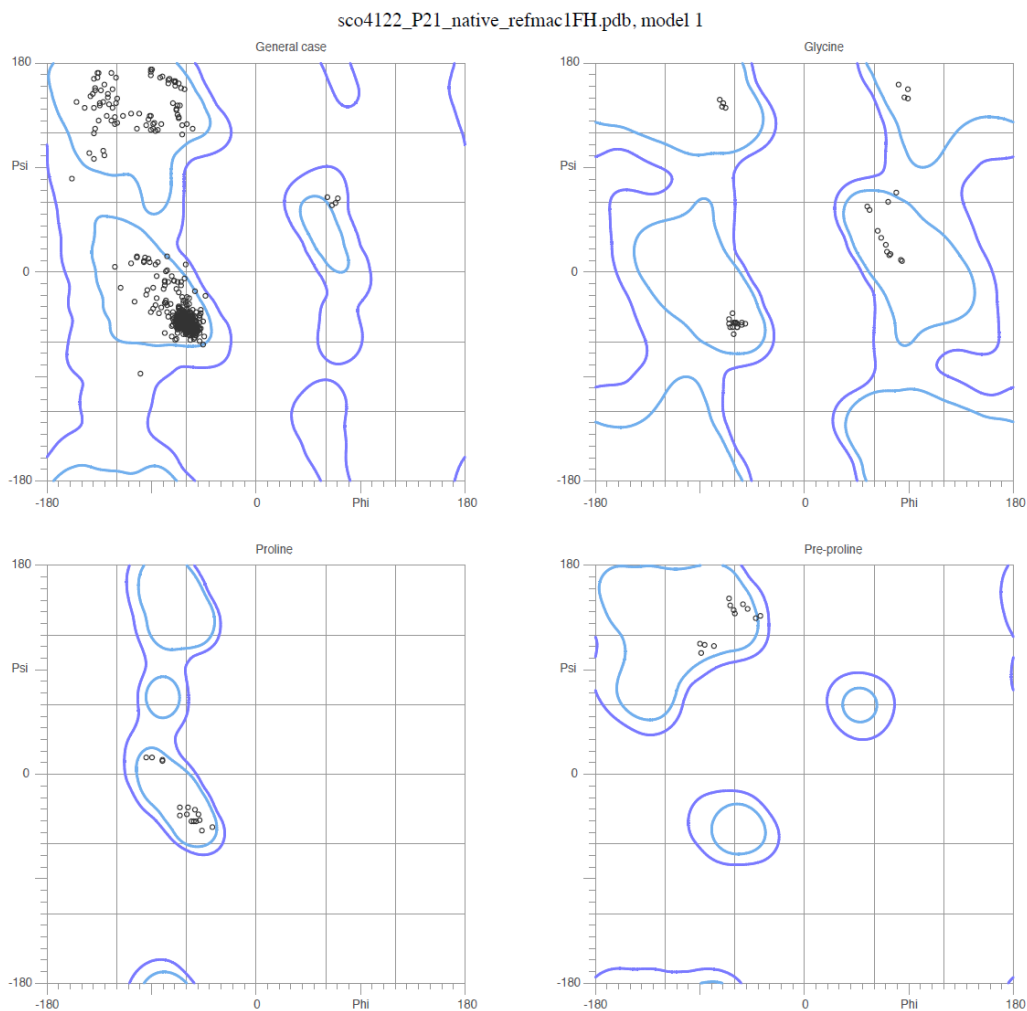


98.2% (592/603) of all residues were in favored (98%) regions.
100.0% (603/603) of all residues were in allowed (>99.8%) regions.

There were no outliers.

http://kinemage.biochem.duke.edu                    Lovell, Davis, et al. Proteins 50:437 (2003)

**Figure 5.14**  Ramachandran plot for Sco4122

**5.3.2 Structural analysis**
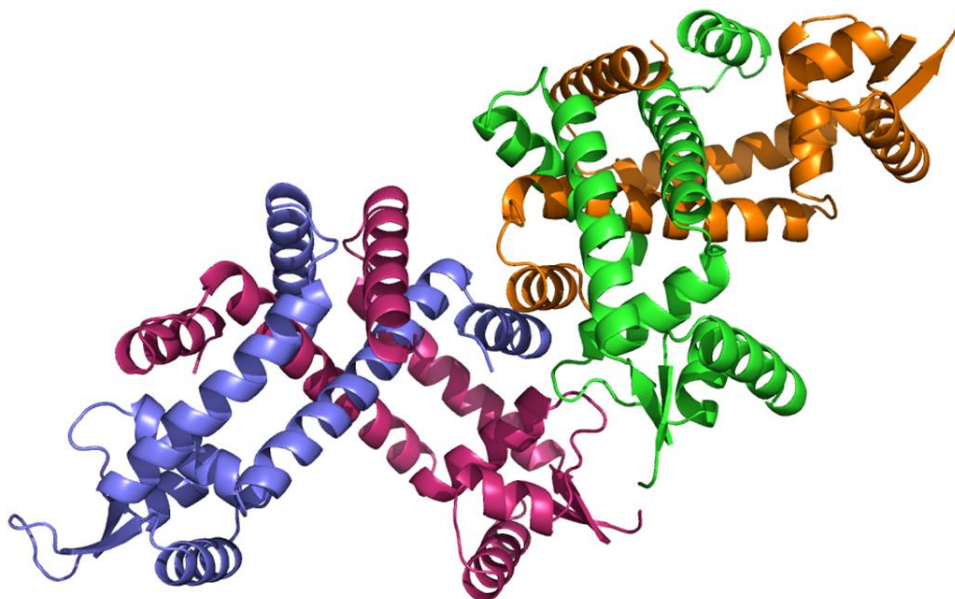


**Figure 5.15** The two Sco4122 dimers in the ASU. Chains A and B (coloured blue and pink) were used for the superimpositions with other MFRs.

The structure has been refined to 2.35 Å resolution and displays the features which are characteristic of MFRs, being mainly α-helical with a winged helix-turn-helix motif. It resembles the structure of HucR (2FBK) in that it possesses the additional α1 helix which braces the amino-terminal end of α2 helix. The purpose of this "extra" helix in Sco4122 is unknown and it is difficult to speculate without the protein-DNA complex or having a ligand bound, but it is apparent that there are quite a few contacts between this helix and other helices in the protein which would likely serve to stabilize the structure. An alignment of the sequences of 2FBK and Sco4122 was performed by the PRALINE server (Simossis & Heringa, 2005) and can be seen in Figure 5.16. The secondary structure prediction alignment suggests that the conserved Trp20 in HucR has been substituted for Tyr19 in Sco4122
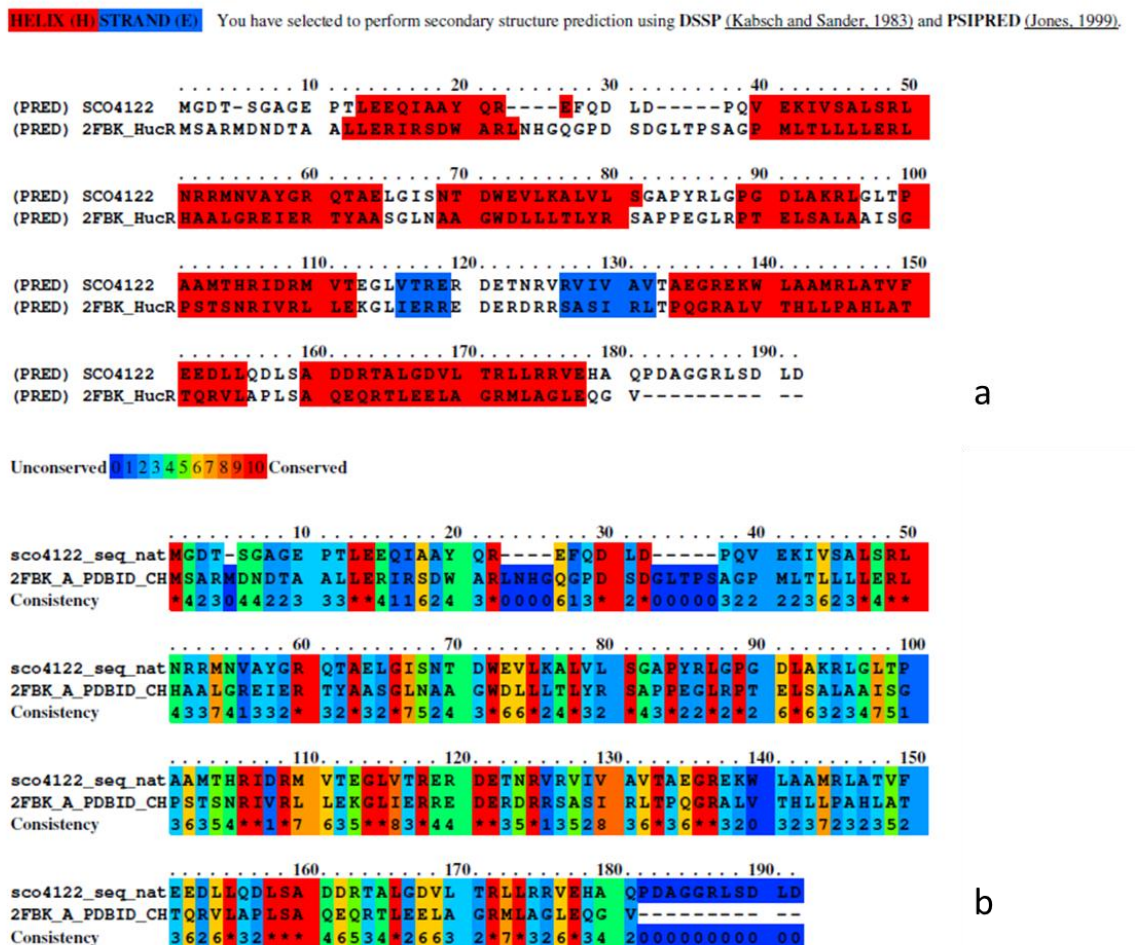
**Figure 5.16** Sequence alignment of 2FBK (HucR) with Sco4122. Alignment performed by the PRALINE server (Simossis & Heringa, 2005). (a) Secondary structure prediction; (b) conservation of amino acids.

This alignment shows that there are some well conserved residues between the two sequences and some perhaps interesting substitutions, but it is difficult to specify if these have any functional significance in Sco4122. For example, Trp20 in HucR is important for interactions with the substrate urate and the equivalent residue in Sco4122 is Tyr19. Its position in the first N-terminal helix appears to be similar as well, but again, this is speculation about its significance. There are evidently some well conserved residues between the two structures and, despite having an overall low sequence identity, there also seems to be good conservation of structural features too, which can be observed in the superposition of 2FBK onto Sco4122 (superposed using SSM superpose in COOT). The RMSD for this superposition is 2.5 Å and the α5 recognition helices appear to be in quite

different orientations (Figure 5.17). Other structural homologues which were generated by PHYRE2 (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver) were superimposed onto Sco4122 and the results are shown in Figure 5.18.
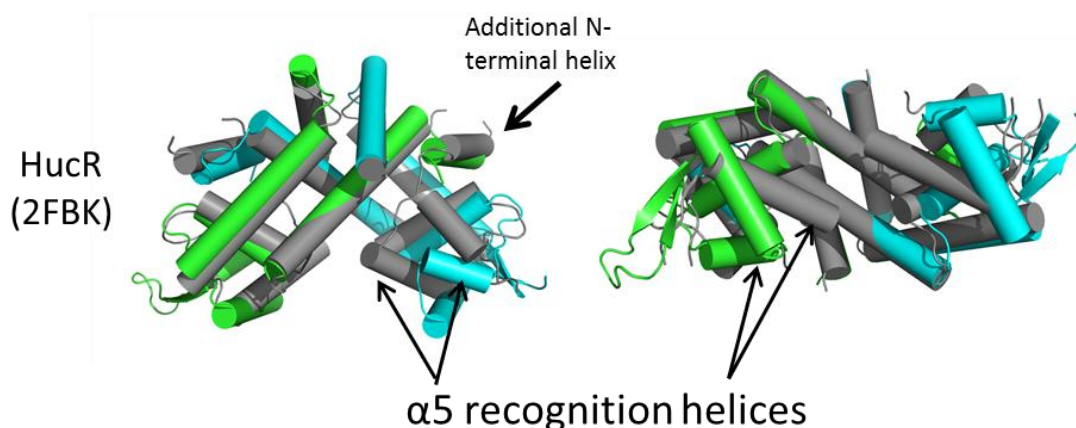


**Figure 5.17** HucR (grey) superimposed onto Sco4122 (green and cyan).

**Table 5.4** Structural homologues of Sco4122

| Protein | Source | DNA or ligand | PDB code | Dali | | | R.m.s. deviation (Å) | | Reference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Z-score | Alignment length | Seq. ID (%) | Subunit | Dimer | |
| MepR | *Staphylococcus aureus* | - | 3ECO | 9.6 | 133 | 19 | 1.9 | 2.9 | (Kumaraswami *et al.*, 2009) |
| YusO | *Bacillus subtilis* | - | 1S3J | 9.3 | 137 | 18 | 1.7 | 2.8 | unpub. |
| MexR | *Pseudomonas aeruginosa* | ArmR | 3ECH | 9 | 124 | 21 | 2.1 | 2.7 | (Wilke *et al.*, 2008) |
| ST1710 | *Sulfolubus tokodaii* | salicylate | 3GF2 | 6 | 131 | 27 | 2.6 | 2.5 | (Kumarevel *et al.*, 2009) |
| SlyA | *Salmonella enterica* | DNA | 3Q5F | 5.5 | 133 | 16 | 3.3 | 3.5 | (Dolan *et al.*, 2011) |
| TcaR | *Staphylococcus epidermis* | penicillin | 3KP2 | 5 | 116 | 22 | 2.6 | 3.7 | (Chang *et al.*, 2010) |

The superimposition with SlyA in its DNA-bound configuration has a relatively high RMSD at both the monomer and the dimer level, suggesting that Sco4122 is not in a suitable conformation for DNA binding. At a dimer level, it superimposes best with salicylate-bound ST1710 and MexR in complex with the peptide effector, ArmR, both of which are

assumed to be in a conformation which is also incompatible with DNA binding. However, the current lack of structures in the Protein Data Base (PDB) of MFRs in complex with DNA or effector molecules makes it particularly difficult at this point to speculate about the predisposition of Sco4122 to DNA-binding in this conformation.

MepR 3ECO

MexR 3ECH

ST1710 3GF2

SlyA 3Q5F

TcaR 3KP2
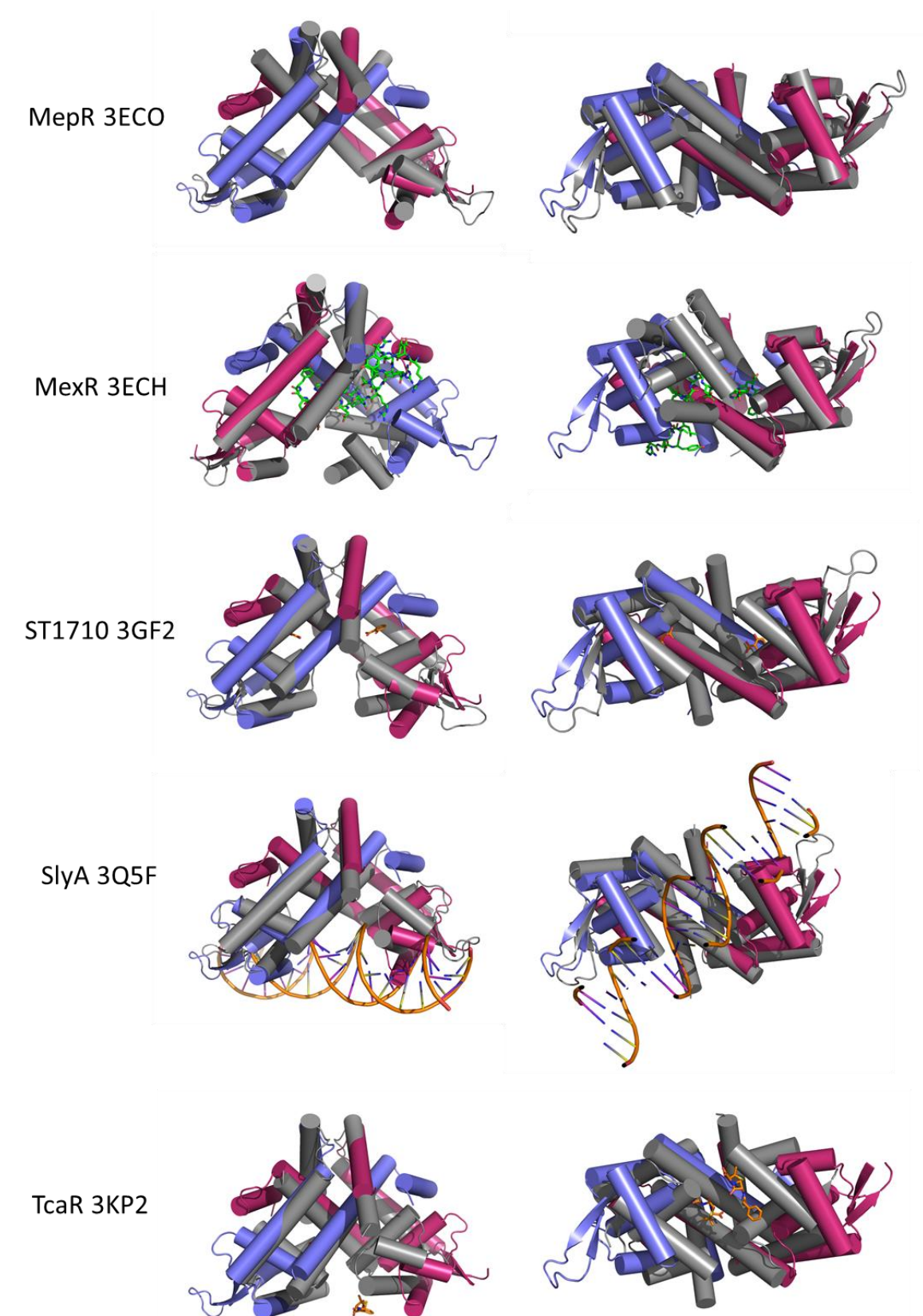
**Figure 5.18** Superpositions other MFRs (grey), which are structural homologues of Sco4122, onto Sco4122 chains A and B (blue and pink).

## 5.4 DNA Interactions

As with Sco3914 and Sco5413, the upstream intergenic sequence was selected as a starting point to investigate protein:DNA interactions, autoregulation being a key feature of over half of all characterized MarR homologues to date (Wilkinson & Grove, 2006).

### 5.4.1 Surface Plasmon Resonance

In order to determine whether the protein would bind the whole *sco4121-4122* intergenic sequence (Figure 5.19), this region was amplified by 3 PCR steps (using the primers shown in Table 5.5) to incorporate a single-stranded overhanging sequence which is complementary to the biotinylated ssDNA annealed to the streptavidin chip (details in Chapter 2). The final PCR product with the single-stranded overhang was annealed to the biotinylated linker using the wizard for immobilization on the Biacore T100.

*sco4121-4122* Intergenic sequence

<mark>CAT</mark>TGCCCTGGTCACGTTCACCCTCTCTTGATCTCTCGGATCGTCCGCTCCAGCCGTTAGAGCTGA
AGACTTAGAAGCTAAAGTTCGAAGGTGAAGAGTACATATCGAAGGACTTCAAAGCAAAGGAGTT
GCGTGCCATACTGCGGCC<mark>ATG</mark>

**Figure 5.19** The DNA sequence between *sco4121* (divergently transcribed gene, the proposed start codon of which is highlighted in yellow) and *sco4122*, the start codon of which is highlighted in green.

GCCACATAGACGTACAGGTAGGGGACCGTGAAGCCGAGTCCGAACGCGCTGAGTGCG
TTGCCCACGTGGATCCGGCGCATCGCCGCGCCCATTGCCCTGGTCACGTTCACCCTCTCTTGATCT
CTCGGATCGTCCGCTCCAGCCGTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTGAAGAG
TACATATCGAAGGACTTCAAAGCAAAGGAGTTGCGTGCCATACTGCGGCCATGGGCGACACCTCC
GGCGCCGGCGAGCCGACACTCGAAGAGCAGATCGCC

**Figure 5.20** The initial template (288 bp) which was amplified from the genomic DNA. The intergenic region is highlighted in yellow whilst the oligo sequences are highlighted in cyan (the reverse complement was taken for the 3' end).

**Table 5.5** Oligos used for the 3 step PCR amplification of the intergenic region. The nested sequence is in green and the sequence which is complementary to the biotinylated oligo is in blue. The reverse oligo from step 2 was used again in step 3.

| Oligo name | Sequence |
| --- | --- |
| Step 1 Forward | GCCACATAGACGTACAGGTAGG |
| Step 1 Reverse | GGCGATCTGCTCTTCGAGTG |
| Step 2 Forward | TTATCAAAAAGAGTATTGACCGCCCATTGCCCTGGTCA |
| Step 2 Reverse | CCGGAGGTGTCGCCCATG |
| Step 3 Forward | TTAGCGTTTGCGTATGCG/dSpacer/TTATCAAAAAGAGTATTGAC |

The protein was injected across the chip (the control flow cell contained a random DNA sequence of 150 base pairs – see Chapter 2) and binding was observed. Unlike Sco3914, the binding of Sco4122 to its intergenic sequence was unaffected by the addition of 10 mM DTT to the running buffer. Sco4122 contains no cysteines in its sequence and so it did not contain the β-mercaptoethanol adducts, thereby supporting the theory that these adducts affected the binding of Sco3914. Once binding had been confirmed, a series of dilutions were prepared in order to see if the interaction was concentration dependent. The sensorgram in Figure 5.21 shows that there is a concentration dependent interaction between Sco4122 and its intergenic DNA sequence which drops off sharply below 25 nM of protein. The next step was to determine the precise region of interaction in the sequence.
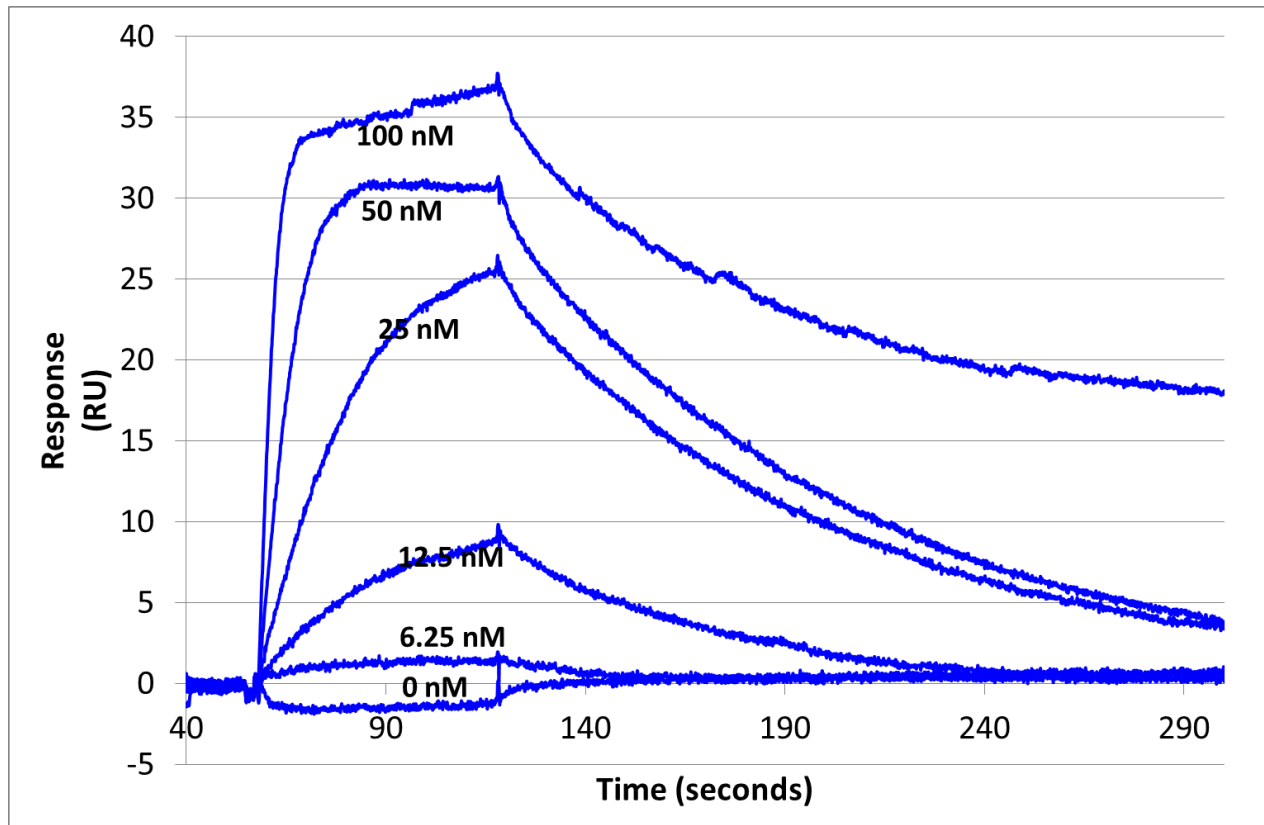
**Figure 5.21** Sensorgram showing Sco4122 interacting with the PCR amplified intergenic DNA sequence *sco4121-4122*.

### 5.4.2 DNA Footprinting – Stage 1

The intergenic region was "fragmented" in order to see if there is any specific interaction with a particular sequence. The intergenic region of 145 base pairs was fragmented to incorporate a 22 bp overlap between each 35 bp sequence as shown below and the oligos (forward and reverse) were ordered from Sigma. The forward oligo included the 20 bases ssDNA which is complementary to the ssDNA annealed to the streptavidin chip. In this experiment, the complementary ssDNA sequences were annealed by mixing 60 μl of the forward oligo with 50 μl of the reverse oligo to give a 45 μM stock solution (PCR program used for this is detailed in Chapter 2). This was diluted to the working concentration of 1 μM with HBS-EP$^+$ buffer. Each dsDNA sequence was then captured on the chip by running it across the analytical flow cell containing the ssDNA linker. The full-length protein Sco4122 (including the His tag sequence) was then run across both the control and analytical flow cells and the results for this are shown in Figure 5.23 and Table 5.6.

```
35bp oligos with 22bp overlap:

TGCCCTGGTCACGTTCACCCTCTCTTGATCTCTCG:1
                TTCACCCTCTCTTGATCTCTCGGATCGTCCGCTCC:2
                        GATCTCTCGGATCGTCCGCTCCAGCCGTTAGAGCT:3
                          GTCCGCTCCAGCCGTTAGAGCTGAAGACTTAGAAG:4
                            GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAG:5
                              ACTTAGAAGCTAAAGTTCGAAGGTGAAGAGTACAT:6
                                AGTTCGAAGGTGAAGAGTACATATCGAAGGACTTC:7
                                  AGAGTACATATCGAAGGACTTCAAAGCAAAGGAGT:8
                                    AAGGACTTCAAAGCAAAGGAGTTGCGTGCCATACT:9
                                      CAAAGGAGTTGCGTGCCATACTGCGGCCATGGGCG:10
```

**Figure 5.22** The *sco4121-4122* intergenic region overlapping oligos as output by POOP (Le, to be published). The 35 base pair oligos include a 22 base pair overlap. The MEME motif 1 is highlighted in yellow and MEME motif 2 in cyan.

Prior to this experiment, a MEME analysis of the intergenic regions of 15 top orthologues in reciprocal BLAST searches was performed and the following MEME motifs were found:-

MEME motif 1
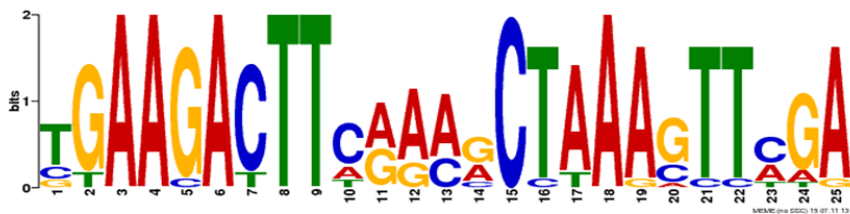


MEME motif 2
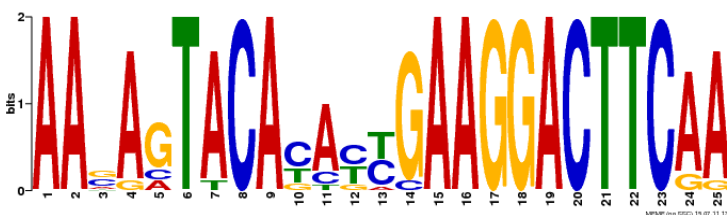


Motif 1 is contained in its entirety within sequence 5 and so the expectation was that if there was to be any binding, it would be within this sequence with perhaps lesser responses from the flanking sequences. The interaction of the protein with each sequence was calculated as a percentage of the theoretical $R_{max}$ value for each DNA sequence captured and plotted as a histogram. The results shown in Figure 5.23 and

Table 5.6 support this hypothesis and demonstrate that there is significant binding (almost 100% of the calculated $R_{max}$ at 100 mM of protein) to sequence 5 with a lesser response for the flanking sequences, 4 and 6. Surprisingly, there was a significant response (~ 60% of $R_{max}$ at 100 mM protein) with sequence number 7, which shares only 9 base pairs overlapping sequence number 5 (and only 7 of the MEME motif 1). This suggests a possible second binding site for Sco4122. However, this sequence does contain most of MEME motif 2, but as the response decreases dramatically with sequence 8, this implies that there is perhaps some element of the MEME motif 1 required for the interaction. This result raises important questions regarding the possible cooperativity that may exist between these two sites. It is notable that there is a significant drop-off in affinity for site 2 as the few base pairs from site 1 are omitted (sequence 8). (Note that sequence 8 also has the first base from the MEME motif 2 omitted). Investigating these two sites was the next logical step.
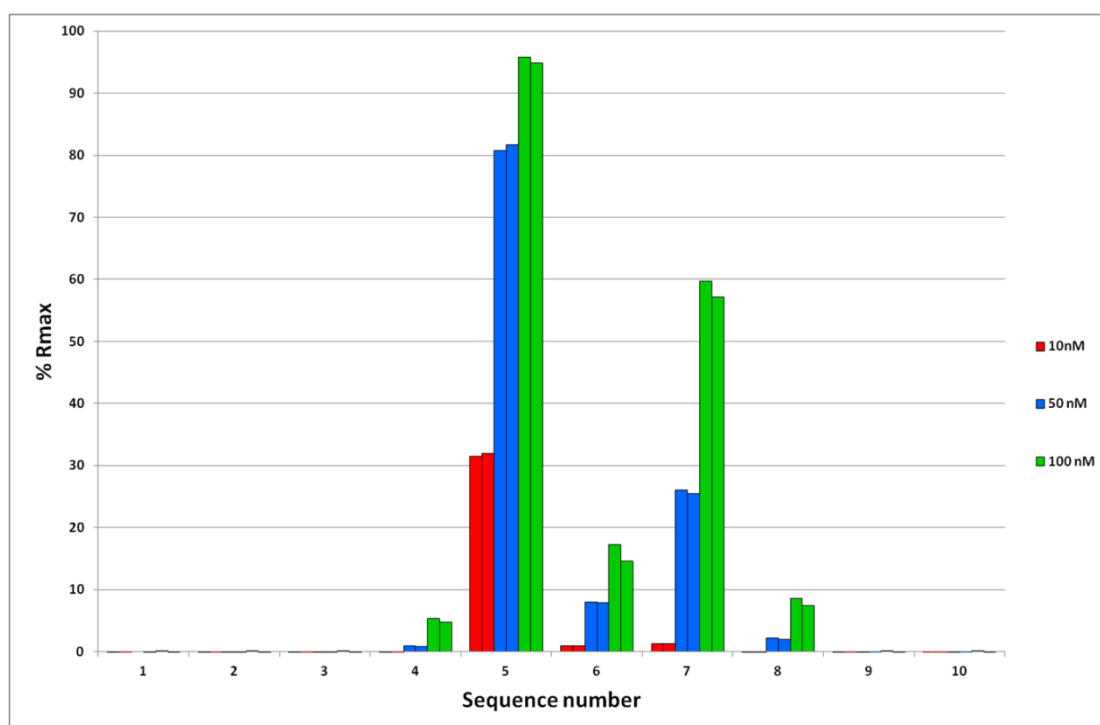


**Figure 5.23** Histogram plot of DNA sequence against the percentage of the calculated $R_{max}$ value. The bars represent the protein:DNA interaction (as a percentage of $R_{max}$) to the sequence at a given concentration.

**Table 5.6** Raw data from the SPR experiment showing the amount of DNA captured at each cycle and the calculated $R_{max}$ for each cycle. The protein interaction was measured in response units which were then calculated as a percentage of the $R_{max}$. There are replicates for each protein concentration.

| Protein conc. (nM) | DNA sequence | | Run1 | | | | Run2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | DNA MW | DNA capture | Theoretical Rmax | Protein bound | %Rmax | DNA capture | Theoretical Rmax | Protein bound | %Rmax |
| 10 nM | 1 | 27500 | 408.0 | 513.0 | -0.7 | -0.1 | 412.7 | 518.9 | -0.9 | -0.2 |
| | 2 | 27500 | 392.8 | 493.8 | -0.6 | -0.1 | 396.8 | 498.9 | -1.0 | -0.2 |
| | 3 | 27501 | 407.1 | 511.8 | -0.7 | -0.1 | 413.0 | 519.2 | -0.9 | -0.2 |
| | 4 | 27499 | 426.1 | 535.7 | -0.7 | -0.1 | 430.8 | 541.6 | -1 | -0.2 |
| | 5 | 27495 | 334.3 | 420.4 | 132.1 | 31.4 | 337.8 | 424.8 | 135.5 | 31.9 |
| | 6 | 27494 | 357.7 | 449.8 | 4.2 | 0.9 | 361.4 | 454.5 | 4.4 | 1.0 |
| | 7 | 27496 | 423.6 | 532.6 | 7.1 | 1.3 | 428.8 | 539.2 | 7.1 | 1.3 |
| | 8 | 27495 | 405.8 | 510.3 | -0.4 | -0.1 | 411.2 | 517.1 | -0.7 | -0.1 |
| | 9 | 27497 | 413.9 | 520.4 | -0.8 | -0.2 | 415.2 | 522.1 | -1.1 | -0.2 |
| | 10 | 27502 | 432.0 | 543.1 | -0.9 | -0.2 | 434.7 | 546.5 | -1.2 | -0.2 |
| 50 nM | 1 | 27500 | 408.5 | 513.6 | 0.0 | 0.0 | 414.6 | 521.3 | -0.6 | -0.1 |
| | 2 | 27500 | 396.1 | 498.0 | -0.1 | 0.0 | 397.5 | 499.8 | -0.7 | -0.1 |
| | 3 | 27501 | 407.6 | 512.4 | -0.3 | -0.1 | 410.5 | 516.1 | -0.7 | -0.1 |
| | 4 | 27499 | 427.3 | 537.2 | 4.9 | 0.9 | 430.1 | 540.8 | 4.4 | 0.8 |
| | 5 | 27495 | 334.8 | 421.0 | 339.9 | 80.7 | 338.7 | 425.9 | 347.9 | 81.7 |
| | 6 | 27494 | 358.1 | 450.3 | 35.8 | 7.9 | 361 | 454.0 | 35.8 | 7.9 |
| | 7 | 27496 | 424.8 | 534.2 | 139.1 | 26.0 | 427.9 | 538.1 | 137.2 | 25.5 |
| | 8 | 27495 | 408.1 | 513.2 | 11.2 | 2.2 | 410.2 | 515.8 | 10.1 | 2.0 |
| | 9 | 27497 | 414.8 | 521.6 | -0.2 | 0.0 | 416.7 | 523.9 | -1 | -0.2 |
| | 10 | 27502 | 432.1 | 543.2 | -0.4 | -0.1 | 436.7 | 549.0 | -1.2 | -0.2 |
| 100 nM | 1 | 27500 | 409.8 | 515.2 | 0.6 | 0.1 | 414.5 | 521.1 | -0.1 | 0.0 |
| | 2 | 27500 | 394.5 | 496.0 | 0.6 | 0.1 | 397.6 | 499.9 | -0.1 | 0.0 |
| | 3 | 27501 | 408.3 | 513.3 | 0.5 | 0.1 | 411.1 | 516.8 | -0.3 | -0.1 |
| | 4 | 27499 | 428.4 | 538.6 | 28.8 | 5.3 | 431.3 | 542.3 | 25.8 | 4.8 |
| | 5 | 27495 | 336.3 | 422.9 | 405.2 | 95.8 | 338.8 | 426.0 | 404 | 94.8 |
| | 6 | 27494 | 358.6 | 450.9 | 77.8 | 17.3 | 361.7 | 454.8 | 66.4 | 14.6 |
| | 7 | 27496 | 424.9 | 534.3 | 318.7 | 59.7 | 428.3 | 538.6 | 307.7 | 57.1 |
| | 8 | 27495 | 407.7 | 512.7 | 43.8 | 8.5 | 410.3 | 515.9 | 38 | 7.4 |
| | 9 | 27497 | 415.6 | 522.6 | 0.4 | 0.1 | 415.4 | 522.3 | -0.5 | -0.1 |
| | 10 | 27502 | 433.7 | 545.2 | 0.3 | 0.1 | 437.8 | 550.4 | -0.7 | -0.1 |

### 5.4.3    DNA Footprinting - Stage 2

It was decided to investigate both sites for stage 2 to ensure that the precise footprint could be determined. For site 1, sequence 5 was combined with 3 base pairs from sequence 6 as shown below in Figure 5.24. For site 2, sequences 6 and 7 were combined with 5 base pairs from sequence 8

**Site1**

Start from DNA 5 LH and 3 bp of DNA 6

        GTTAGAGC==TGAAGACTTAGAAGCTAAAGTTCGA==AG:5

                ==ACTTAGAAGCTAAAGTTCGA==AGGTG

        GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTG          38 bp

**Site2**

Start with DNA 7 and 3 bp from DNA 8.

        ==AGTTCGA==AGGTG==AAGAGTACATATCGAAGGACTTC==:7

                ==AGAGTACATATCGAAGGACTTCAA==A

        ==AGTTCGA==AGGTGAAGAGTACATATCGAAGGACTTCAAA          38 bp

**Figure 5.24**   The DNA sequences from the sco4121-4122 intergenic region which elicited the highest response in Stage 1 of the SPR experiment. In order to define the footprint, these sequences were combined with several base pairs from the subsequent sequence for Stage 2 of the experiment. MEME motif 1 is highlighted in yellow; MEME motif 2 in cyan.

As with Sco3914, the strategy was to eliminate 2 base pairs from each end in turn to define the recognition sequence. In order to account for any effects from the linker sequence (such as stabilization of binding due to the extra bases), the linker was also attached to the forward strand sequence for deletions from the left hand side. Two base

pairs at a time were first removed from the right hand side of the sequence, followed by two base pairs at a time from the left. The sequences can be seen in Table 5.7.

**Table 5.7** Stage 2 DNA sequences for testing the binding of Sco4122 to its upstream DNA sequence. (a) Sequences to test the limits of the site 1 right hand boundary; (b) sequences to test the limits of the site 1 left hand boundary; (c) sequences to test the limits of site 2 right hand boundary; (d) sequences to test the limits of site 2 left hand boundary. Sequences that complement the immobilized linker are in lower case.

**(a)** <u>**Sco4122 Site 1 RH boundary**</u>

| Name | Sequence |
|------|----------|
| Site1_RH_F | GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTG |
| Site1_RH_R | CACCTTCGAACTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus2_F | GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGG |
| Site1_RH_minus2_R | CCTTCGAACTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus4_F | GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAA |
| Site1_RH_minus4_R | TTCGAACTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus6_F | GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCG |
| Site1_RH_minus6_R | CGAACTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus8_F | GTTAGAGCTGAAGACTTAGAAGCTAAAGTT |
| Site1_RH_minus8_R | AACTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus10_F | GTTAGAGCTGAAGACTTAGAAGCTAAAG |
| Site1_RH_minus10_R | CTTTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus12_F | GTTAGAGCTGAAGACTTAGAAGCTAA |
| Site1_RH_minus12_R | TTAGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus14_F | GTTAGAGCTGAAGACTTAGAAGCT |
| Site1_RH_minus14_R | AGCTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |
| Site1_RH_minus16_F | GTTAGAGCTGAAGACTTAGAAG |
| Site1_RH_minus16_R | CTTCTAAGTCTTCAGCTCTAACcctaccctacgtcctcctgc |

**(b)** <u>**Sco4122 Site 1 LH boundary**</u>

| Name | Sequence |
|------|----------|
| Site1_LH_F | CACCTTCGAACTTTAGCTTCTAAGTCTTCAGCTCTAAC |
| Site1_LH_R | GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus2_F | CACCTTCGAACTTTAGCTTCTAAGTCTTCAGCTCTA |
| Site1_LH_minus2_R | TAGAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus4_F | CACCTTCGAACTTTAGCTTCTAAGTCTTCAGCTC |
| Site1_LH_minus4_R | GAGCTGAAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus6_F | CACCTTCGAACTTTAGCTTCTAAGTCTTCAGC |
| Site1_LH_minus6_R | GCTGAAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus8_F | CACCTTCGAACTTTAGCTTCTAAGTCTTCA |
| Site1_LH_minus8_R | TGAAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus10_F | CACCTTCGAACTTTAGCTTCTAAGTCTT |

| | |
|---|---|
| Site1_LH_minus10_R | AAGACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus12_F | CACCTTCGAACTTTAGCTTCTAAGTC |
| Site1_LH_minus12_R | GACTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus14_F | CACCTTCGAACTTTAGCTTCTAAG |
| Site1_LH_minus14_R | CTTAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus16_F | CACCTTCGAACTTTAGCTTCTA |
| Site1_LH_minus16_R | TAGAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |
| Site1_LH_minus18_F | CACCTTCGAACTTTAGCTTC |
| Site1_LH_minus18_R | GAAGCTAAAGTTCGAAGGTGcctaccctacgtcctcctgc |

**(c)  Sco4122 Site 2 RH boundary**

| Name | Sequence |
|---|---|
| Site2_RH_F | AGTTCGAAGGTGAAGAGTACATATCGAAGGACTTCAAA |
| Site2_RH_R | TTTGAAGTCCTTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus2_F | AGTTCGAAGGTGAAGAGTACATATCGAAGGACTTCA |
| Site2_RH_minus2_R | TGAAGTCCTTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus4_F | AGTTCGAAGGTGAAGAGTACATATCGAAGGACTT |
| Site2_RH_minus4_R | AAGTCCTTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus6_F | AGTTCGAAGGTGAAGAGTACATATCGAAGGAC |
| Site2_RH_minus6_R | GTCCTTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus8_F | AGTTCGAAGGTGAAGAGTACATATCGAAGG |
| Site2_RH_minus8_R | CCTTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus10_F | AGTTCGAAGGTGAAGAGTACATATCGAA |
| Site2_RH_minus10_R | TTCGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus12_F | AGTTCGAAGGTGAAGAGTACATATCG |
| Site2_RH_minus12_R | CGATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus14_F | AGTTCGAAGGTGAAGAGTACATAT |
| Site2_RH_minus14_R | ATATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |
| Site2_RH_minus16_F | AGTTCGAAGGTGAAGAGTACAT |
| Site2_RH_minus16_R | ATGTACTCTTCACCTTCGAACTcctaccctacgtcctcctgc |

**(d)  Sco4122 Site 2 LH boundary**

| Name | Sequence |
|---|---|
| Site2_LH_F | TTTGAAGTCCTTCGATATGTACTCTTCACCTTCGAACT |
| Site2_ LH_R | AGTTCGAAGGTGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_ LH_minus2_F | TTTGAAGTCCTTCGATATGTACTCTTCACCTTCGAA |
| Site2_ LH_minus2_R | TTCGAAGGTGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_ LH_minus4_F | TTTGAAGTCCTTCGATATGTACTCTTCACCTTCG |
| Site2_ LH_minus4_R | CGAAGGTGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_ LH_minus6_F | TTTGAAGTCCTTCGATATGTACTCTTCACCTT |
| Site2_ LH_minus6_R | AAGGTGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_ LH_minus8_F | TTTGAAGTCCTTCGATATGTACTCTTCACC |
| Site2_ LH_minus8_R | GGTGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_ LHminus10_F | TTTGAAGTCCTTCGATATGTACTCTTCA |
| Site2_LH_minus10_R | TGAAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |

| Site2_LH_minus12_F | TTTGAAGTCCTTCGATATGTACTCTT |
| Site2_LH_minus12_R | AAGAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_LH_minus14_F | TTTGAAGTCCTTCGATATGTACTC |
| Site2_LH_minus14_R | GAGTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_LH_minus16_F | TTTGAAGTCCTTCGATATGTAC |
| Site2_LH_minus16_R | GTACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |
| Site2_LH_minus18_F | TTTGAAGTCCTTCGATATGT |
| Site2_LH_minus18_R | ACATATCGAAGGACTTCAAAcctaccctacgtcctcctgc |

The experiment was run using a similar method to stage 1 (using replicates of 3 different protein concentrations) and the results are displayed in Figure 5.25.
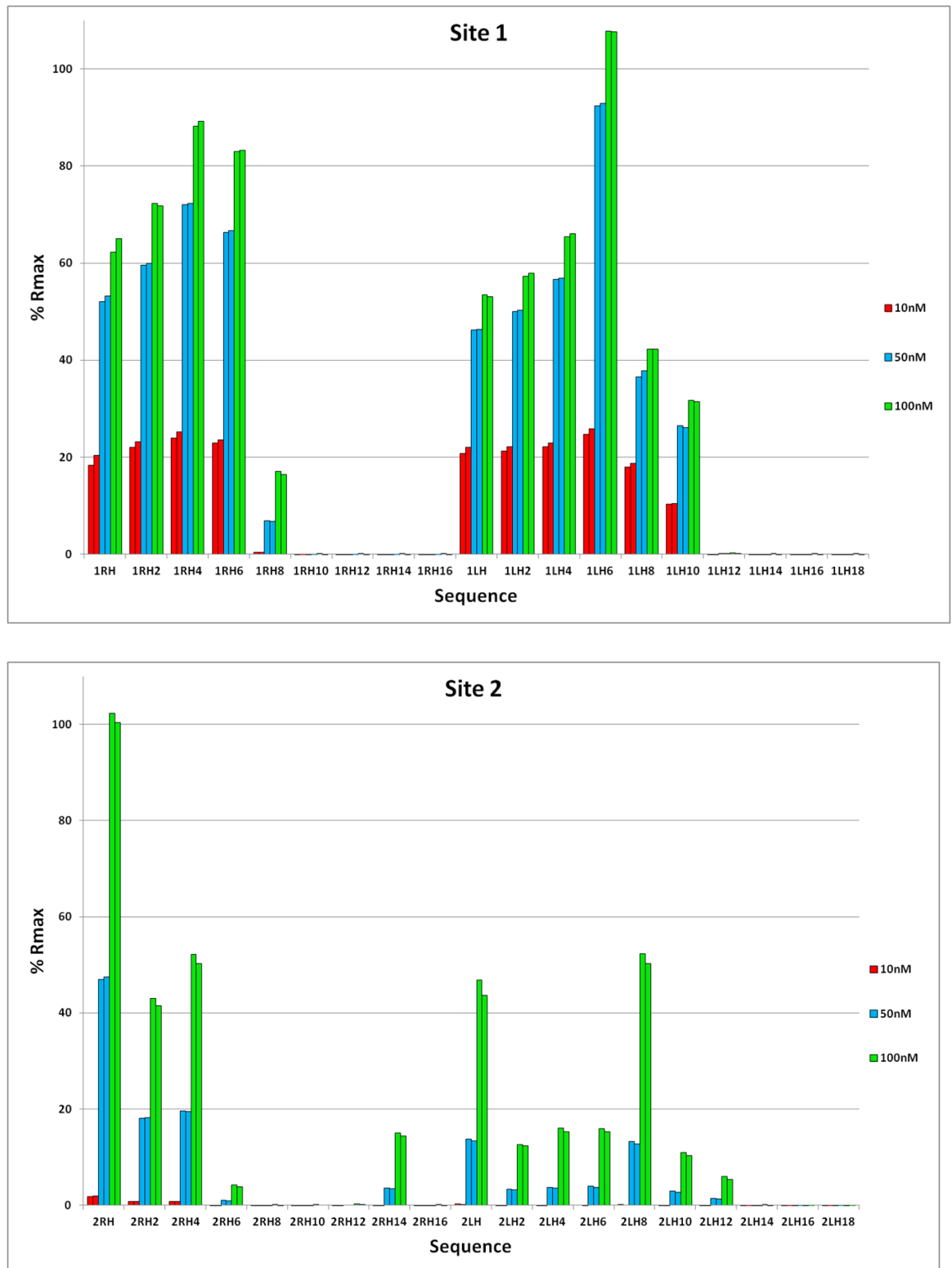
**Figure 5.25**  The results for Stage 2 of the footprinting experiment. Site 1 is shown in the top histogram and site 2 in the lower histogram.

For site 1, the results are clearer to interpret than for site 2. There is a pronounced decrease in binding of Sco4122 when 8 base pairs are removed from the right hand side of the sequence. The left hand side is slightly anomalous in that there is a marked increase in the binding when 6 base pairs are removed (> 100% of theoretical $R_{max}$) which decreases significantly when 8 base pairs are removed. The limit has therefore been determined as 6 base pairs from either side (although it is entirely possible that it could be 5 or 7).

**Site 1**

RH -6   GTTAGAGCTGAAGACTTAGAAGCTAAAGTTCG~~AAGGTG~~

LH -6   ~~GTTAGA~~GCTGAAGACTTAGAAGCTAAAGTTCGAAGGTG

This gives the footprint for site 1 as determined by this experiment as:-

GCTGAAGACTTAGAAGCTAAAGTTCG  (26 bp)


Site 2 is more complex: this was expected as the binding at stage 1 was ~50% of the theoretical $R_{max}$ at 100 nM and barely significant at 10 nM. Here, it is observed that there is virtually no binding at 10 nM and so the focus for this site shall be on 100 nM. Additionally, there is evidence for linker effects occurring as the binding of the full-length sequence when it is reversed (LH) is less than half of the $R_{max}$ obtained for the RH sequence.


**Site 2**

RH -0   AGTTCGAAGGTGAAGAGTACATATCGAAGGACTTCAAA

LH -8   ~~AGTTCGAA~~GGTGAAGAGTACATATCGAAGGACTTCAAA


RH-0 is evidently favoured, but Sco4122 will tolerate RH-4. Strangely, whilst LH-0 is favoured, so is LH-8. The binding for these two sequences is very similar.

So either the whole sequence, or:-

GGTGAAGAGTACATATCGAAGGACTTCAAA (30 bp)

or maybe even:-

GGTGAAGAGTACATATCGAAGGACTT  (26 bp)

There is also the possibility that there may be elements of cooperative binding occurring whereby the proteins may associate laterally via the DNA-binding domains (Perera & Grove, 2010). This is purely speculative at this stage and requires much more detailed investigation as these data are not clear enough to be able to deduce the stoichiometry with any confidence.

The main conclusion from this experiment is that site 1 contains the main footprint which has been defined as a 26 base pair sequence: GCTGAAGACTTAGAAGCTAAAGTTCG. This sequence is relatively AT rich at 58%: it has a GC content of just 42%, compared to 71% for the whole genome. A closer inspection of this sequence as it compares to site 1 reveals that there are similarities between the two sites which could account for the binding at this second site at higher concentrations of protein. Sequence alignments of either the full-length site 2 or the 26 bp site 2 with the defined footprint sequence from site 1 indicate that there are similarities:-

Taking the full length of Site 2 gives 42% identity for the following sections:-

G C T G A A G A C T T A G A A G C T A A A G T T C G  ( 2 6  b p )

A G T T C G A A G G T G A A G A G T A C A T A T C G A A G G A C T T C A A A (38 bp)

Or 50% for the 26 bp sequence omitting 8 base pairs from the LH and 4 from the RH:-

G C T G A A G A C T T A G A A G C T A A A G T T C G  ( 2 6  b p )

G G T G A A G A G T A C A T A T C G A A G G A C T T  ( 2 6  b p )

214

Site 2 also overlaps the second motif found by MEME:-
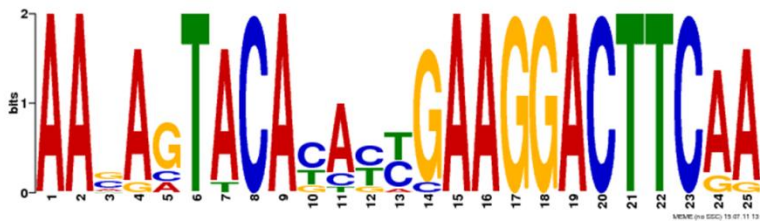


The location of these MEME motifs (of which there are 3) within the intergenic region is shown in Figure 5.26 and it is apparent that these two motifs are in close proximity (separated by only 5 base pairs). However, as the binding is very small at this second motif, it was not investigated further in this study. Another interesting feature of these motifs is that none of them appear to have any obvious palindromicity, whereas several MarR homologues which have already been characterised tend to bind palindromic or pseudopalindromic sequences. HucR binds a pseudopalindromic sequence: two 8 base pair half sites which are separated by 2 base pairs approximately in the middle of a 101 base pair intergenic sequence (Wilkinson & Grove, 2004). These are inverted imperfect repeats as opposed to direct repeats. This 2004 study concludes that the position of the binding site of HucR suggests that it may simultaneously repress the divergent *hucR* and uricase genes via a steric inhibition of RNA polymerase mechanism and so this has implications for the role of Sco4122. Like MEME motif 1 for Sco3914, motif 1 for Sco4122 is located towards the centre of this intergenic region and so the question still exists as to whether Sco4122 represses its own transcription along with that of the divergently transcribed gene, *sco4121*. This remains an interesting hypothesis as *sco4121* encodes a possible membrane protein which, according to the annotation in StrepDB, has a weak similarity to multidrug resistance proteins. This in itself does not imply a role for Sco4122 in cellular defence mechanisms, but it does offer opportunities for further investigation.



**Figure 5.26** The location of the MEME motifs within the *sco4121-sco4122* intergenic region.

## 5.5   PCR-Targeted Gene Disruption

The M145*sco4122* gene deletion strain was generated using the Redirect protocol (Gust *et al.*, 2003), detailed in Chapter 2. The *sco4122* gene is contained on the StD72a cosmid. Cosmids were checked for rearrangements by restriction digest and compared to the predicted digest fragment sizes by DNA gel electrophoresis (using the restriction tool on Streptomyces.org.uk). StD72a was digested with SacI.

The primers used to amplify the disruption cassette from pIJ773 are displayed in Table 5.8. The gene deletion was confirmed by PCR using the P1 and P2 flanking primers (Chapter 2) as well as the gene specific checking primers (Table 5.9).

**Table 5.8**  Primers used to amplify the pIJ773 disruption cassette. The sequence which is complementary to the disruption cassette is shown in red.

| | |
|---|---|
| Forward | CTTCAAAGCAAAGGAGTTGCGTGCCATACTGCGGCCATG<span style="color:red">ATTCCGGGGATCCGTCGACC</span> |
| Reverse | CCCGCAGAGGGCTGTCAAGCGTCGCTGTCAAGATCTTTA<span style="color:red">TGTAGGCTGGAGCTGCTTC</span> |

**Table 5.9**  Primers used to verify positive transformants using PCR.

| | | |
|---|---|---|
| **Forward** | CCCACGTGGATCCGGCGCA | 177 bp upstream of gene |
| **Reverse** | GTTACGGTCCCGTGGCAA | 73 bp downstream of gene |

The mutagenized cosmid was then introduced into *S. coelicolor* by conjugation (detailed in Chapter 2) to undergo genetic recombination with the chromosomal DNA and so replace the *sco4122* gene with the disrupted gene. Successful mutants were detected by replica-plating on SFM agar plates supplemented with 1) apramycin and 2) apramycin + kanamycin. Those colonies which are kanamycin sensitive were then further verified by PCR (using both the gene specific checking primers and the F1 and F2 flanking primers). Spore stocks of the mutant strains were then prepared as detailed in Chapter 2.

### 5.5.1   Screening the successful mutants

The prepared spore stocks were screened for visibly detectable phenotypes by plating on a variety of media supplemented with different carbon sources (details in Chapter 2). 20 different conditions were tested in total and the observations for each mutant were monitored over the course of a week. Sco4122 did not exhibit an observable phenotype which differed to that of the wild type on any of the conditions tested. This particular investigation was pursued no further.

## 5.6   Conclusions

The structure of Sco4122 shows distinct structural similarities to that of HucR, a urate responsive MarR homologue from *Deinococcus radiodurans*, despite their fairly divergent sequences. It is interesting that both structures exhibit the additional N-terminal α-helix, α1, the purpose of which remains to be determined, although the conservation of Trp20 in HucR which interacts with the substrate urate would suggest that the presence of this helix is necessary to the function of the protein. Interestingly, at the equivalent position in Sco4122 is Tyr19, which could also participate in π-stacking to orientate an aromatic ligand, if this were to be the mechanism at work in HucR. Also of note is the interaction of Asp61 with Arg99, which mirrors the interaction between the Asp in α-helix 3 (α3) and the Arg in α-helix 5 (α5) in HucR. The disruption of this salt-bridge interaction in HucR by binding of urate is crucial to the reorientation of the DNA recognition helix. In Sco4122, there are also clear interactions between Trp62 in α3 and Trp130 in α5, and Thr60 (α3) and Arg96 (α5), all of which could also potentially communicate the presence of a ligand to the DNA-binding domain. Although at this point this conjecture is hugely speculative in the absence of any ligand-bound structural information for either protein, it is worth noting for future investigations with Sco4122.  Compared to other structural homologues, it would appear that Sco4122 is not preconfigured for DNA binding. Obtaining a structure of the protein-DNA complex would be another challenge for this particular regulator.

This study has also made significant progress towards the determination of the DNA-footprint of Sco4122. One site has been shown to elicit strong binding and this site coincides with the first sequence reported by the MEME server, whilst there is the suggestion that there may be a second, lesser site, which may or may not have biological relevance. It does, however, correspond approximately to the second MEME motif and also has some sequence similarity to the first motif. The lack of an obvious palindromic repeat in this sequence is noteworthy as many MFRs have been shown to bind (pseudo)palindromic sequences. As with Sco3914, and HucR, the operator site for Sco4122 would appear to be located in the centre of the intergenic region and so could potentially block RNA polymerase at two operator sites. Thus the question remains as to whether Sco4122 regulates not only its own transcription but also that of the divergently transcribed *sco4121*, which encodes a possible transmembrane protein. Such an investigation would offer further insights into the role of this regulator and opportunites to uncover implications for other MFRs.

# Chapter 6

# General discussion and conclusions

The work presented in this thesis describes three MarR family regulators (MFRs) from *Streptomyces coelicolor* on biophysical, genetic and structural levels. These repressors are clearly well represented in the *S. coelicolor* genome and Sco5413 is particularly well represented across the genomes of other actinomycetes. Therefore, uncovering details regarding their respective structures and DNA footprints can inform future investigations into other MFRs, from streptomycetes and other organisms. In this study, the structures of all three MFRs have been determined using anomalous dispersion methods. The DNA footprints for Sco3914 and Sco4122 have been defined and a phenotype observed for the deletion mutant of Sco5413.

The structures presented in this thesis confirm that these MFRs are dimers and that they possess the characteristic features of MFRs: they have the typical topology α1-α2-α3-α4-(α5)-β1-β2-α5-α6, with α4 (or α5 in Sco4122) constituting the recognition helix in the helix-turn-helix motif, whilst the wing consists of the two antiparallel β-strands, β1 and β2 and the intervening loop (Sco5413 has three β-strands). These are the DNA-binding domains in most MarR homologues. The dimerization interface comprises the α1, α5 and α6 (or α1, α6 and α7 in Sco4122) helices from each monomer which intertwine with each other. The dimer interface is rich in hydrophobic residues in all three structures. Whilst there are the obvious structural similarities between them, they also exhibit features which are unique to themselves. Sco3914, for example, has the slightly flatter shape which is reminiscent of HypR, with the reactive cysteines which may or may not be functionally relevant. Sco4122 is interesting in that it has the "extra" N-terminal helix seen in HucR and potentially required for the stabilization of a ligand. Sco5413tr has been truncated but bioinformatic analysis of the full-length structure would suggest that this protein also has an extra N-terminal helix with a conserved tryptophan residue which corresponds to part of the urate binding signature of HucR (Perera & Grove, 2011).

The structure of truncated Sco5413 (Sco5413tr), which was solved to 1.25 Å, is the highest resolution structure of a MarR homologue to date. Due to inherent flexibility of these proteins, which it has been proposed is critical to their function, Sco5413 was truncated in order to crystallize this protein, but a structure prediction for the full-length protein indicates that it would likely have the extra N-terminal helix seen in HucR and Sco4122. It also has a tryptophan in this putative helix which appears to be conserved in HucR and interacts with its substrate urate. The modified cysteine in Sco5413tr, which at first was anticipated may have some functional significance, does not appear to be well conserved and so it was not investigated further.

The structures of all three MFRs were aligned (using Sco5413tr as the reference structure) with SSM Superpose (Krissinel & Henrick, 2004) and the results are displayed in Table 6.1.

**Table 6.1** Superimposition of Sco3914 and Sco4122 with Sco5413tr as the reference structure

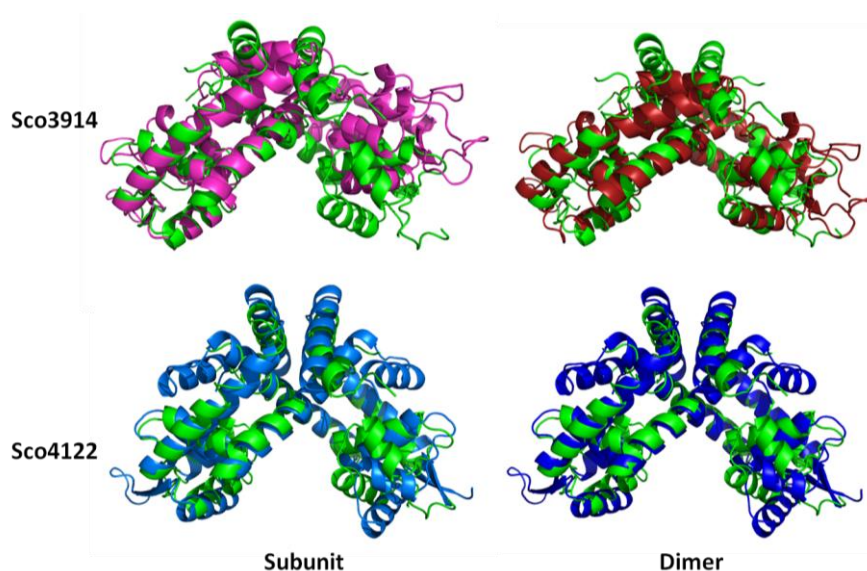| | R.m.s. deviation (Å)/aligned residues | |
|---|---|---|
| **Protein name** | **Subunit** | **Dimer** |
| **Sco3914** | 1.8/99 | 3.4/198 |
| **Sco4122** | 3.0/123 | 3.2/239 |



**Figure 6.1** Superimpositions of Sco3914 (pink and red) and Sco4122 (light blue and blue) onto Sco5413 (reference structure in green) at both the subunit and dimer levels.

The positions of the DNA recognition helices were also measured (from the corresponding C-alpha atoms of the approximate mid-point residue of the helix) in order to estimate whether the proteins may be in conformations which would enable DNA-binding. Sco3914 had a distance of 29.3 Å between the corresponding residues; Sco4122 had a distance of 24.6 Å and in Sco5413tr the distance was 24.6 Å. Interestingly, Sco5413tr and Sco4122 exhibited identical distances despite the slightly different orientations of their respective helices exhibited in the superimpositions. The distance between major grooves in B-DNA is ~34 Å, thus Sco3914 would appear to be in a conformation suitable for DNA-binding whereas Sco4122 and Sco5413tr are perhaps not in such optimal conformations for interacting with the major grooves. However, the superimposition of Sco5413tr on the OhrR-*ohrA* complex looks reasonable, but the superimposition onto the SlyA-DNA complex suggests that there would be clashes between this helix and the DNA. Again, without more structures of MFRs in complex with their cognate DNA sequences, it is difficult to reach a firm conclusion about the predisposition towards DNA binding of this particular MFR. The positioning of the recognition helices in Sco4122 with respect to the DNA of the SlyA complex is much more dramatic: clashes are visually apparent and so the conclusion is that Sco4122 is not in a conformation which is compatible with DNA-binding.

None of the structures present any indication of a specific ligand, but taking all three structures together and superimposing them with HucR and the first salicylate site from MTH313 (Saridakis *et al.*, 2008), a putative pocket/ligand binding site may be hypothesized. The full-length Sco5413 has a tryptophan which may be the equivalent of Trp20 in HucR, which in HucR is required for the interaction with its effector ligand, urate. However, there the similarity appears to end as Sco5413 does not possess the other conserved residues which have been proposed to constitute a urate binding signature. The structure of Sco5413tr reveals a distinct channel which is lined by well conserved residues and this is possibly a binding site for ligands which may regulate the affinity of the protein for its cognate DNA sequence. This putative binding site may also be accessed from "underneath", where there also appears to be a cleft. The extra N-terminal helix in

both HucR and Sco4122 appear to be well-anchored to the structure via interactions with residues from the other subunit and so perhaps this helix would be unlikely to move to accommodate a ligand.



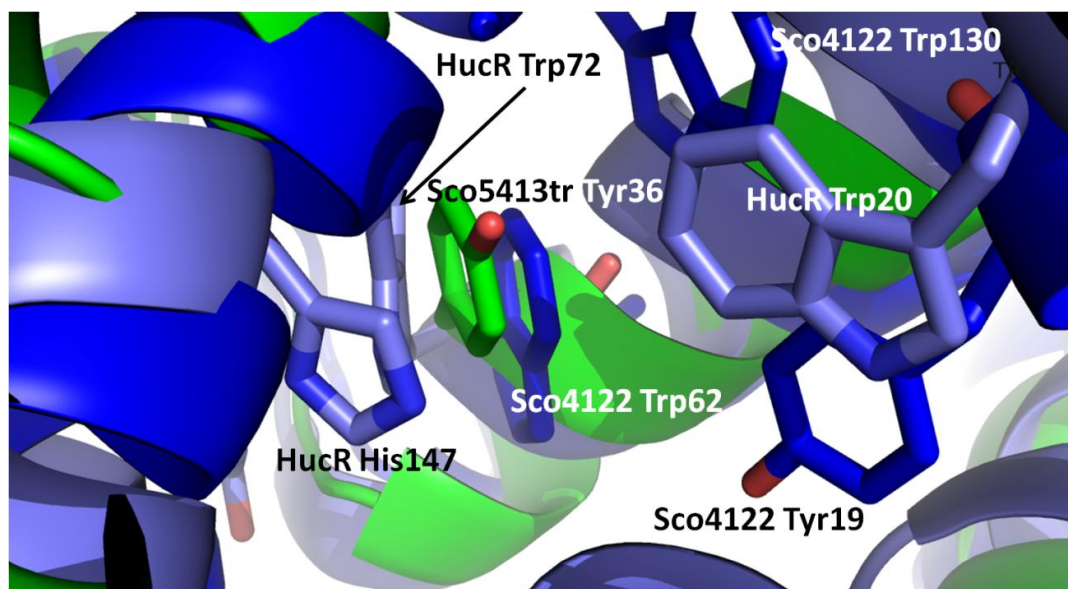**Figure 6.2** Putative ligand binding pocket of Sco5413tr and Sco4122, based on superimposition with HucR. Sco5413tr is in green; Sco4122 is in dark blue and HucR is in slate blue.

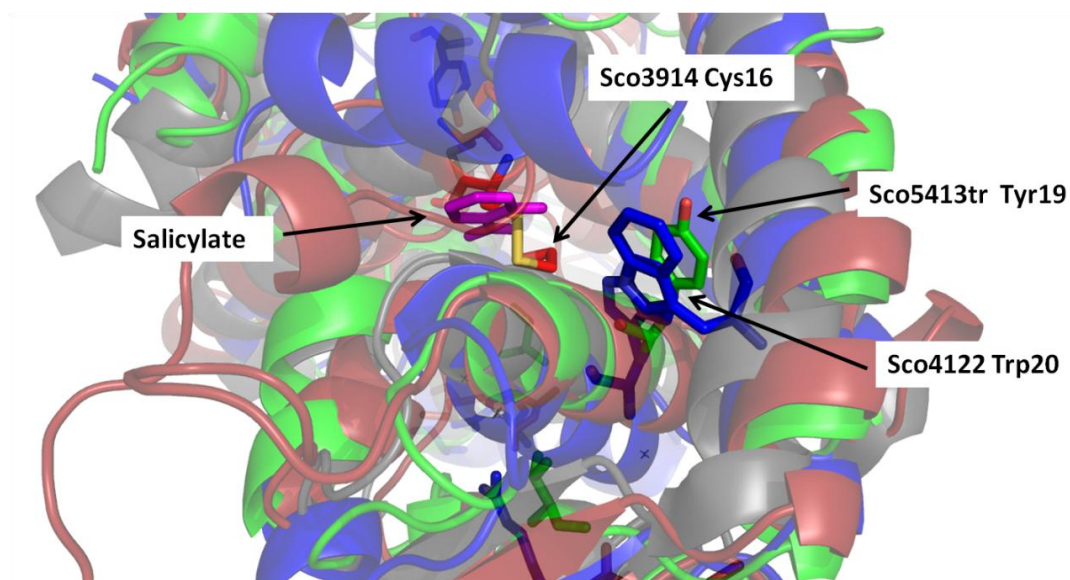

**Figure 6.3** View of the putative conserved binding site for a ligand. Salicylate (magenta) from the MTH313-salicylate complex (PDB code 3GF2) is superimposed onto Sco5413tr (green). Sco4122 is in blue; Sco3914 is in red and MTH313 is in grey.

The presence of Cys16 from Sco3914 in this putative binding site may also still bear some significance, despite the findings in this work that it has a relatively high $pK_a$ (9.13) and would likely be protonated at a physiological pH. However, its formation of an adduct with β-mercaptoethanol does suggest that this residue could be readily modified by low molecular weight thiols and so this raises questions for future research into the function of Sco3914. The wing cysteine, Cys95, may have a more physical role in protein-DNA interactions which remains to be determined. Clearly, the modification affects DNA-binding in the wild type and both C16S and C95S mutants. These cysteines are probably too far apart (~25 Å) to be involved in an intramolecular disulfide bond and these values are closely similar for the structures of the C16S mutant and the reduced form of the protein. Thus, at this point, the functional importance of these residues remains uncertain.

The DNA footprints of Sco3914 and Sco4122 have been determined in this work and are as follows:-

GCTGAAGACTTAGAAGCTAAAGTTCG        Sco4122        26 bp

CAATAATTGCACACGCGGGCTATAT        Sco3914        25 bp

The lengths of these two footprints seem to correspond to the typical MFR footprints shown in Table 6.2.

The apparent lack of palindromicity in the DNA footprints determined for Sco3914 and Sco4122 is also quite unique in terms of MFRs which have been characterized to date, most of which have been shown to bind (pseudo)palindromic repeats. If the suggestion by Dolan *et al.* (2011) holds true and that an equivalent of the R86 in SlyA could potentially participate in an indirect readout of the DNA and is therefore able to recognize several slightly different sequences, then perhaps this lack of palindromicity could feature more prevalently as other transcriptional regulators are characterized.

**Table 6.2** DNA footprints of several MFRs. (* correspond to multiple binding sites).

| Protein | Organism | Footprint (base pairs) | Reference |
|---|---|---|---|
| MepR | *Staphylococcus aureus* | 43*, 27 | (Kaatz *et al.*, 2006, Kumaraswami *et al.*, 2009) |
| MarR | *Escherichia coli* | 21 | (Martin & Rosner, 1995) |
| OhrR | *Xanthomonas campestris* | 44* | (Mongkolsuk *et al.*, 2002) |
| SlyA | *Salmonella typhurium* | 25 | (Stapleton *et al.*, 2002) |
| HucR | *Deinococcus radiodurans* | 21 | (Wilkinson & Grove, 2005) |
| MexR | *Pseudomonas aeruginosa* | 28 | (Evans *et al.*, 2001) |
| Sco3914 | *Streptomyces coelicolor* | 25 | This work |
| Sco4122 | *Streptomyces coelicolor* | 26 | This work |

The presence of MFR structures in the PDB is increasing rapidly and, with the addition of more protein-DNA or protein-ligand complexes, significant advances could be made towards using these data to develop therapeutics aimed at these repressors. This would be extremely useful in the manipulation of the biosynthetic gene clusters in actinomycetes to increase the production of secondary metabolites. The data presented in this work will hopefully contribute to the growing wealth of knowledge in this area and follow-up work could include determination of the specific ligand as well as crystallization of the proteins with their cognate DNA sequences. These would allow us to structurally define the repression-derepression cycles of these proteins. On a more general level, these structures will contribute to our understanding of protein-ligand and protein-DNA recognition.

# References

## References

Aceti, D. J. & W. C. Champness, (1998) Transcriptional Regulation of Streptomyces coelicolor Pathway-Specific Antibiotic Regulators by the absA and absB Loci. *Journal of Bacteriology* 180: 3100-3106.

Adamidis, T. & W. Champness, (1992) Genetic analysis of absB, a Streptomyces coelicolor locus involved in global antibiotic regulation. *Journal of Bacteriology* 174: 4622-4628.

Adamidis, T., P. Riggle & W. Champness, (1990) Mutations in a new Streptomyces coelicolor locus which globally block antibiotic biosynthesis but not sporulation. *J Bacteriol* 172: 2962-2969.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger & P. H. Zwart, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-221.

Alekshun, M. N. & S. B. Levy, (1999) The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol* 7: 410-413.

Alekshun, M. N., S. B. Levy, T. R. Mealy, B. A. Seaton & J. F. Head, (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. *Nat Struct Mol Biol* 8: 710-714.

Alekshun, M. N., Y. S. Kim & S. B. Levy, (2000) Mutational analysis of MarR, the negative regulator of marRAB expression in Escherichia coli, suggests the presence of two regions required for DNA binding. *Mol Microbiol* 35: 1394-1404.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li & W. S. Noble, (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202-208.

# References

Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill & D. A. Hopwood, (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417: 141-147.

Berrow, N. S., D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart & R. J. Owens, (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35: e45.

Bibb, M. J., (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* 8: 208-215.

Bibb, M., (1996) 1995 Colworth Prize Lecture. The regulation of antibiotic production in Streptomyces coelicolor A3(2). *Microbiology* 142 ( Pt 6): 1335-1344.

Bordelon, T., S. P. Wilkinson, A. Grove & M. E. Newcomer, (2006) The crystal structure of the transcriptional regulator HucR from Deinococcus radiodurans reveals a repressor preconfigured for DNA binding. *J Mol Biol* 360: 168-177.

Bradford, M. M., (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.

Brandenberger, M., M. Tschierske, P. Giachino, A. Wada & B. Berger-Bachi, (2000) Inactivation of a novel three-cistronic operon tcaR-tcaA-tcaB increases teicoplanin resistance in Staphylococcus aureus. *Biochim Biophys Acta* 1523: 135-139.

Brian, P., P. J. Riggle, R. A. Santos & W. C. Champness, (1996) Global negative regulation of Streptomyces coelicolor antibiotic synthesis mediated by an absA-encoded putative signal transduction system. *J Bacteriol* 178: 3221-3231.

Brunger, A. T., (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* 49: 24-36.

Buchan, D. W., S. M. Ward, A. E. Lobley, T. C. Nugent, K. Bryson & D. T. Jones, Protein annotation and modelling servers at University College London. *Nucleic Acids Res* 38: W563-568.

Buckle, M., (2001) Surface plasmon resonance applied to DNA-protein complexes. *Methods Mol Biol* 148: 535-546.

Buckle, M., R. M. Williams, M. Negroni & H. Buc, (1996) Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance. *Proc Natl Acad Sci U S A* 93: 889-894.

Cathelyn, J. S., S. D. Crosby, W. W. Lathem, W. E. Goldman & V. L. Miller, (2006) RovA, a global regulator of Yersinia pestis, specifically required for bubonic plague. *Proc Natl Acad Sci U S A* 103: 13514-13519.

Chang, Y. M., W. Y. Jeng, T. P. Ko, Y. J. Yeh, C. K. Chen & A. H. Wang, (2010) Structural study of TcaR and its complexes with multiple antibiotics from Staphylococcus epidermidis. *Proc Natl Acad Sci U S A* 107: 8617-8622.

Chen, H., C. Yi, J. Zhang, W. Zhang, Z. Ge, C. G. Yang & C. He, (2010) Structural insight into the oxidation-sensing mechanism of the antibiotic resistance of regulator MexR. *EMBO Rep* 11: 685-690.

Chen, H., J. Hu, P. R. Chen, L. Lan, Z. Li, L. M. Hicks, A. R. Dinner & C. He, (2008) The Pseudomonas aeruginosa multidrug efflux regulator MexR uses an oxidation-sensing mechanism. *Proceedings of the National Academy of Sciences* 105: 13586-13591.

Chen, P. R., S. Nishida, C. B. Poor, A. Cheng, T. Bae, L. Kuechenmeister, P. M. Dunman, D. Missiakas & C. He, (2009) A new oxidative sensing and regulation pathway mediated by the MgrA homologue SarZ in Staphylococcus aureus. *Mol Microbiol* 71: 198-211.

Chen, P. R., T. Bae, W. A. Williams, E. M. Duguid, P. A. Rice, O. Schneewind & C. He, (2006) An oxidation-sensing mechanism is used by the global regulator MgrA in

Staphylococcus aureus. *Nat Chem Biol* 2: 591-595.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson & D. C. Richardson, (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12-21.

Cheung, A. L., K. A. Nishina, M. P. Trotonda & S. Tamber, (2008) The SarA protein family of Staphylococcus aureus. *Int J Biochem Cell Biol* 40: 355-361.

Cohen, S. P., H. Hachler & S. B. Levy, (1993a) Genetic and functional analysis of the multiple antibiotic resistance (mar) locus in Escherichia coli. *J Bacteriol* 175: 1484-1492.

Cohen, S. P., L. M. McMurry, D. C. Hooper, J. S. Wolfson & S. B. Levy, (1989) Cross-resistance to fluoroquinolones in multiple-antibiotic-resistant (Mar) Escherichia coli selected by tetracycline or chloramphenicol: decreased drug accumulation associated with membrane changes in addition to OmpF reduction. *Antimicrob. Agents Chemother.* 33: 1318-1325.

Cohen, S. P., S. B. Levy, J. Foulds & J. L. Rosner, (1993b) Salicylate Induction of Antibiotic-resistance in Escerichia-coli - Activation of the Mar Operon and a Mar-independent Pathway. *Journal of Bacteriology* 175: 7856-7862.

Colson, S., G. P. van Wezel, M. Craig, E. E. E. Noens, H. Nothaft, A. M. Mommaas, F. Titgemeyer, B. Joris & S. Rigali, (2008) The chitobiose-binding protein, DasA, acts as a link between chitin utilization and morphogenesis in Streptomyces coelicolor. *Microbiology* 154: 373-382.

Cowtan, K., (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62: 1002-1011.

Cowtan, K., (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66: 470-478.

Crouch, R. J., M. Wakasa & M. Haruki, (1999) Detection of nucleic acid interactions using

surface plasmon resonance. *Methods Mol Biol* 118: 143-160.

D'Arcy, A., (1994) Crystallizing proteins - a rational approach? *Acta Crystallogr D Biol Crystallogr* 50: 469-471.

Dauter, Z., M. Dauter & K. R. Rajashankar, (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr D Biol Crystallogr* 56: 232-237.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson & D. C. Richardson, (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: W375-383.

Davis, J. R. & J. K. Sello, (2010) Regulation of genes in Streptomyces bacteria required for catabolism of lignin-derived aromatic compounds. *Appl Microbiol Biotechnol* 86: 921-929.

Di Fiore, A., G. Fiorentino, R. M. Vitale, R. Ronca, P. Amodeo, C. Pedone, S. Bartolucci & G. De Simone, (2009) Structural analysis of BldR from Sulfolobus solfataricus provides insights into the molecular basis of transcriptional activation in Archaea by MarR family proteins. *J Mol Biol* 388: 559-569.

Dolan, K. T., E. M. Duguid & C. He, (2011) Crystal structures of SlyA protein, a master virulence regulator of Salmonella, in free and DNA-bound states. *J Biol Chem* 286: 22178-22185.

Doublie, S., (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* 276: 523-530.

Ducruix, A. & R. Giege, (1992) *Crystallization of nucleic acids and proteins : a practical approach,* p. xxiv, 331 p. IRL Press at Oxford University Press, Oxford.

Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-2132.

Emsley, P., B. Lohkamp, W. G. Scott & K. Cowtan, (2010) Features and development of

Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501.

Evans, G. & G. Bricogne, (2002) Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu K[alpha] X-ray equipment. *Acta Crystallographica Section D* 58: 976-991.

Evans, K., L. Adewoye & K. Poole, (2001) MexR repressor of the mexAB-oprM multidrug efflux operon of Pseudomonas aeruginosa: identification of MexR binding sites in the mexA-mexR intergenic region. *J Bacteriol* 183: 807-812.

Evans, P., (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72-82.

Evolving Methods for Macromolecular Crystallography. In*.*: Springer Netherlands, pp. 41-51.

Fernandez-Moreno, M. A., A. J. Martin-Triana, E. Martinez, J. Niemi, H. M. Kieser, D. A. Hopwood & F. Malpartida, (1992) abaA, a new pleiotropic regulatory locus for antibiotic production in Streptomyces coelicolor. *Journal of Bacteriology* 174: 2958-2967.

Fuangthong, M. & J. D. Helmann, (2002) The OhrR repressor senses organic hydroperoxides by reversible formation of a cysteine-sulfenic acid derivative. *Proceedings of the National Academy of Sciences of the United States of America* 99: 6690-6695.

Fuangthong, M., S. Atichartpongkul, S. Mongkolsuk & J. D. Helmann, (2001) OhrR is a repressor of ohrA, a key organic hydroperoxide resistance determinant in Bacillus subtilis. *J Bacteriol* 183: 4134-4141.

Gajiwala, K. S. & S. K. Burley, (2000) Winged helix proteins. *Curr Opin Struct Biol* 10: 110-116.

George, A. M. & S. B. Levy, (1983) Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in Escherichia coli: involvement of a non-plasmid-determined efflux of tetracycline. *Journal of Bacteriology* 155: 531-540.

Gramajo, H. C., E. Takano & M. J. Bibb, (1993) Stationary-phase production of the

antibiotic actinorhodin in Streptomyces coelicolor A3(2) is transcriptionally regulated. *Mol Microbiol* 7: 837-845.

Grove, A., (2010) Urate-responsive MarR homologs from Burkholderia. *Mol Biosyst* 6: 2133-2142.

Guerra, A. J., C. E. Dann, 3rd & D. P. Giedroc, (2011) Crystal structure of the zinc-dependent MarR family transcriptional regulator AdcR in the Zn(II)-bound state. *J Am Chem Soc* 133: 19614-19617.

Gust, B., G. L. Challis, K. Fowler, T. Kieser & K. F. Chater, (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* 100: 1541-1546.

Hesketh, A., H. Kock, S. Mootien & M. Bibb, (2009) The role of absC, a novel regulatory gene for secondary metabolism, in zinc-dependent antibiotic production in Streptomyces coelicolor A3(2). *Mol Microbiol* 74: 1427-1444.

Hesketh, A., J. Sun & M. Bibb, (2001) Induction of ppGpp synthesis in Streptomyces coelicolor A3(2) grown under conditions of nutritional sufficiency elicits actII-ORF4 transcription and actinorhodin biosynthesis. *Mol Microbiol* 39: 136-144.

Hesketh, A., W. J. Chen, J. Ryding, S. Chang & M. Bibb, (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in Streptomyces coelicolor A3(2). *Genome Biol* 8: R161.

Hobbs, G., A. I. Obanye, J. Petty, J. C. Mason, E. Barratt, D. C. Gardner, F. Flett, C. P. Smith, P. Broda & S. G. Oliver, (1992) An integrated approach to studying regulation of production of the antibiotic methylenomycin by Streptomyces coelicolor A3(2). *J Bacteriol* 174: 1487-1494.

Hong, H. J., M. I. Hutchings, L. M. Hill & M. J. Buttner, (2005a) The role of the novel Fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *J Biol Chem* 280: 13055-13061.

Hong, M., M. Fuangthong, J. D. Helmann & R. G. Brennan, (2005b) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol Cell* 20: 131-141.

Horinouchi, S., (2003) AfsR as an integrator of signals that are sensed by multiple serine/threonine kinases in Streptomyces coelicolor A3(2). *J Ind Microbiol Biotechnol* 30: 462-467.

Kaatz, G. W., C. E. DeMarco & S. M. Seo, (2006) MepR, a Repressor of the Staphylococcus aureus MATE Family Multidrug Efflux Pump MepA, Is a Substrate-Responsive Regulatory Protein. *Antimicrob. Agents Chemother.* 50: 1276-1281.

Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* 66: 125-132.

Kallifidas, D., B. Pascoe, G. A. Owen, C. M. Strain-Damerell, H. J. Hong & M. S. Paget, (2010) The zinc-responsive regulator Zur controls expression of the coelibactin gene cluster in Streptomyces coelicolor. *J Bacteriol* 192: 608-611.

Kelemen, G. H., P. Brian, K. Flardh, L. Chamberlin, K. F. Chater & M. J. Buttner, (1998) Developmental regulation of transcription of whiE, a locus specifying the polyketide spore pigment in Streptomyces coelicolor A3 (2). *J Bacteriol* 180: 2515-2521.

Kelley, L. A. & M. J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4: 363-371.

Kieser, T. B., M.J.; Buttner, M.J.; Chater, K.; Hopwood, D.A., (2000) *Practical Streptomyces Genetics*.

Kleywegt, G. J. & A. T. Brunger, (1996) Checking your imagination: applications of the free R value. *Structure* 4: 897-904.

Krissinel, E. & K. Henrick, (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60: 2256-2268.

Krissinel, E., (2007) On the relationship between sequence and structure similarities in

proteomics. *Bioinformatics* 23: 717-723.

Kumaraswami, M., J. T. Schuman, S. M. Seo, G. W. Kaatz & R. G. Brennan, (2009) Structural and biochemical characterization of MepR, a multidrug binding transcription regulator of the Staphylococcus aureus multidrug efflux pump MepA. *Nucleic Acids Res* 37: 1211-1224.

Kumarevel, T., T. Tanaka, T. Umehara & S. Yokoyama, (2009) ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res* 37: 4723-4735.

Laemmli, U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.

Laskowski, R. A., (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* 29: 221-222.

Laskowski, R. A., (2009) PDBsum new things. *Nucleic Acids Res* 37: D355-359.

Leslie, A. G. W., H. R. Powell, R. J. Read & J. L. Sussman, (2007) Processing diffraction data with mosflm

Lim, D. C., K. Poole & N. Strynadka, (2002) Crystal structure of the MexR repressor of the mexRAB-oprM multi-drug efflux operon of Pseudomonas aeruginosa. *J. Biol. Chem.*: M111381200.

Liu, Y., A. C. Manna, C. H. Pan, I. A. Kriksunov, D. J. Thiel, A. L. Cheung & G. Zhang, (2006) Structural and function analyses of the global regulatory protein SarA from Staphylococcus aureus. *Proc Natl Acad Sci U S A* 103: 2392-2397.

Liu, Y., A. Manna, R. Li, W. E. Martin, R. C. Murphy, A. L. Cheung & G. Zhang, (2001) Crystal structure of the SarR protein from Staphylococcus aureus. *Proc Natl Acad Sci U S A* 98: 6877-6882.

Luong, T. T., P. M. Dunman, E. Murphy, S. J. Projan & C. Y. Lee, (2006) Transcription Profiling of the mgrA Regulon in Staphylococcus aureus. *J Bacteriol* 188: 1899-1910.

MacNeil, D. J., K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons & T. MacNeil, (1992) Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111: 61-68.

Mandel, M. & A. Higa, (1970) Calcium-dependent bacteriophage DNA infection. *J Mol Biol* 53: 159-162.

Maneewannakul, K. & S. B. Levy, (1996) Identification for mar mutants among quinolone-resistant clinical isolates of Escherichia coli. *Antimicrob Agents Chemother* 40: 1695-1698.

Martin, J. F. & A. L. Demain, (1980) Control of antibiotic biosynthesis. *Microbiol Rev* 44: 230-251.

Martin, R. G. & J. L. Rosner, (1995) Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proc Natl Acad Sci U S A* 92: 5456-5460.

Martin, R. G. & J. L. Rosner, (2004) Transcriptional and translational regulation of the marRAB multiple antibiotic resistance operon in Escherichia coli. *Mol Microbiol* 53: 183-191.

McCallum, N., M. Bischoff, H. Maki, A. Wada & B. Berger-Bachi, (2004) TcaR, a putative MarR-like regulator of sarS expression. *J Bacteriol* 186: 2966-2972.

McCoy, A. J., (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63: 32-41.

Miller, J. H., (1972) *Experiments in molecular genetics,* p. xvi, 466 p. Cold Spring Harbor Laboratory, [Cold Spring Harbor, N.Y.].

Mizuno, T., M. Y. Chou & M. Inouye, (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A* 81: 1966-1970.

Moffatt, B. A. & F. W. Studier, (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* 49: 221-227.

# References

Mongkolsuk, S., W. Panmanee, S. Atichartpongkul, P. Vattanaviboon, W. Whangsuk, M. Fuangthong, W. Eiamphungporn, R. Sukchawalit & S. Utamapongchai, (2002) The repressor for an organic peroxide-inducible operon is uniquely regulated at multiple levels. *Mol Microbiol* 44: 793-802.

Murshudov, G. N., A. A. Vagin & E. J. Dodson, (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240-255.

Newberry, K. J., M. Fuangthong, W. Panmanee, S. Mongkolsuk & R. G. Brennan, (2007) Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Mol Cell* 28: 652-664.

Nichols, C. E., S. Sainsbury, J. Ren, T. S. Walter, A. Verma, D. K. Stammers, N. J. Saunders & R. J. Owens, (2009) The structure of NMB1585, a MarR-family regulator from Neisseria meningitidis. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 65: 204-209.

Oh, S. Y., J. H. Shin & J. H. Roe, (2007) Dual role of OhrR as a repressor and an activator in response to organic hydroperoxides in Streptomyces coelicolor. *J Bacteriol* 189: 6284-6292.

Okusu, H., D. Ma & H. Nikaido, (1996) AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol* 178: 306-308.

Paget, M. S., L. Chamberlin, A. Atrih, S. J. Foster & M. J. Buttner, (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in Streptomyces coelicolor A3(2). *J Bacteriol* 181: 204-211.

Pagliai, F. A., C. L. Gardner, S. G. Pande & G. L. Lorca, (2010) LVIS553 transcriptional regulator specifically recognizes novobiocin as an effector molecule. *J Biol Chem* 285: 16921-16930.

Palm, G. J., B. Khanh Chi, P. Waack, K. Gronau, D. Becher, D. Albrecht, W. Hinrichs, R. J. Read & H. Antelmann, (2012) Structural insights into the redox-switch mechanism of the

MarR/DUF24-type regulator HypR. *Nucleic Acids Res* 40: 4178-4192.

Panmanee, W., P. Vattanaviboon, L. B. Poole & S. Mongkolsuk, (2006) Novel organic hydroperoxide-sensing and responding mechanisms for OhrR, a major bacterial sensor and regulator of organic hydroperoxide stress. *J Bacteriol* 188: 1389-1395.

Perera, I. C. & A. Grove, (2010a) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J Mol Cell Biol* 2: 243-254.

Perera, I. C. & A. Grove, (2010b) Urate is a ligand for the transcriptional regulator PecS. *J Mol Biol* 402: 539-551.

Perera, I. C. & A. Grove, (2011) MarR homologs with urate-binding signature. *Protein Sci* 20: 621-629.

Perera, I. C., Y.-H. Lee, S. P. Wilkinson & A. Grove, (2009) Mechanism for Attenuation of DNA Binding by MarR Family Transcriptional Regulators by Small Molecule Ligands. *Journal of Molecular Biology* 390: 1019-1029.

Perrakis, A., R. Morris & V. S. Lamzin, (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6: 458-463.

Poor, C. B., P. R. Chen, E. Duguid, P. A. Rice & C. He, (2009) Crystal structures of the reduced, sulfenic acid, and mixed disulfide forms of SarZ, a redox active global regulator in Staphylococcus aureus. *J Biol Chem* 284: 23517-23524.

Price, B., T. Adamidis, R. Kong & W. Champness, (1999) A Streptomyces coelicolor antibiotic regulatory gene, absB, encodes an RNase III homolog. *J Bacteriol* 181: 6142-6151.

Ramachandran, B. V., L. Engstrom & G. Agren, (1963) Fractionation of DF32P-binding proteins of rat-liver cell fractions by DEAE-cellulose chromatography. Distribution of esterase activity. *Biochem Pharmacol* 12: 167-172.

Redenbach, M., H. M. Kieser, D. Denapaite, A. Eichner, J. Cullum, H. Kinashi & D. A. Hopwood, (1996) A set of ordered cosmids and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome. *Mol Microbiol* 21: 77-96.

Reverchon, S., F. Van Gijsegem, G. Effantin, O. Zghidi-Abouzid & W. Nasser, (2010) Systematic targeted mutagenesis of the MarR/SlyA family members of Dickeya dadantii 3937 reveals a role for MfbR in the modulation of virulence gene expression in response to acidic pH. *Mol Microbiol* 78: 1018-1037.

Rigali, S., F. Titgemeyer, S. Barends, S. Mulder, A. W. Thomae, D. A. Hopwood & G. P. van Wezel, (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by Streptomyces. *EMBO Rep* 9: 670-675.

Rigali, S., H. Nothaft, E. E. Noens, M. Schlicht, S. Colson, M. Muller, B. Joris, H. K. Koerten, D. A. Hopwood, F. Titgemeyer & G. P. van Wezel, (2006) The sugar phosphotransferase system of Streptomyces coelicolor is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol Microbiol* 61: 1237-1251.

Rosenfeld, N., M. B. Elowitz & U. Alon, (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 323: 785-793.

Sambrook, J., P. MacCallum & D. Russell, (2001) Molecular Cloning: A Laboratory Manual, Third edition. *Cold Spring Harbor Laboratory Press*.

Saridakis, V., D. Shahinas, X. Xu & D. Christendat, (2008) Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from Methanobacterium thermoautotrophicum. *J Mol Biol* 377: 655-667.

Sello, J. K. & M. J. Buttner, (2008) The gene encoding RNase III in Streptomyces coelicolor is transcribed during exponential phase and is required for antibiotic production and for proper sporulation. *J Bacteriol* 190: 4079-4083.

Simossis, V. A. & J. Heringa, (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* 33: W289-294.

Soonsanga, S., M. Fuangthong & J. D. Helmann, (2007) Mutational analysis of active site residues essential for sensing of organic hydroperoxides by Bacillus subtilis OhrR. *J Bacteriol* 189: 7069-7076.

Stapleton, M. R., V. A. Norte, R. C. Read & J. Green, (2002) Interaction of the Salmonella typhimurium transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J Biol Chem* 277: 17630-17637.

Stevenson, C. E., H. Kock, S. Mootien, S. C. Davies, M. J. Bibb & D. M. Lawson, (2007) Crystallization and preliminary X-ray analysis of AbsC, a novel regulator of antibiotic production in Streptomyces coelicolor. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63: 233-235.

Stockley, P. G. & B. Persson, (2009) Surface plasmon resonance assays of DNA-protein interactions. *Methods Mol Biol* 543: 653-669.

Studier, F. W. & B. A. Moffatt, (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113-130.

Sun, F., Y. Ding, Q. Ji, Z. Liang, X. Deng, C. C. Wong, C. Yi, L. Zhang, S. Xie, S. Alvarez, L. M. Hicks, C. Luo, H. Jiang, L. Lan & C. He, (2012) Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc Natl Acad Sci U S A*.

Sun, J., A. Hesketh & M. Bibb, (2001) Functional analysis of relA and rshA, two relA/spoT homologues of Streptomyces coelicolor A3(2). *J Bacteriol* 183: 3488-3498.

Takano, E., H. C. Gramajo, E. Strauch, N. Andres, J. White & M. J. Bibb, (1992) Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2). *Mol Microbiol* 6: 2797-2804.

Terwilliger, T. C. & J. Berendzen, (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55: 849-861.

Terwilliger, T. C., (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol*

*Crystallogr* 56: 965-972.

Terwilliger, T. C., P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart & L. W. Hung, (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65: 582-601.

Terwilliger, T. C., R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read & P. D. Adams, (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64: 61-69.

Truong-Bolduc, Q. C., P. M. Dunman, J. Strahilevitz, S. J. Projan & D. C. Hooper, (2005) MgrA is a multiple regulator of two new efflux pumps in Staphylococcus aureus. *J Bacteriol* 187: 2395-2405.

Truong-Bolduc, Q. C., X. Zhang & D. C. Hooper, (2003) Characterization of NorR protein, a multifunctional regulator of norA expression in Staphylococcus aureus. *J Bacteriol* 185: 3127-3138.

Ward, J. J., J. S. Sodhi, L. J. McGuffin, B. F. Buxton & D. T. Jones, (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337: 635-645.

Wilke, M. S., M. Heller, A. L. Creagh, C. A. Haynes, L. P. McIntosh, K. Poole & N. C. Strynadka, (2008) The crystal structure of MexR from Pseudomonas aeruginosa in complex with its antirepressor ArmR. *Proc Natl Acad Sci U S A* 105: 14832-14837.

Wilkinson, S. P. & A. Grove, (2004) HucR, a Novel Uric Acid-responsive Member of the MarR Family of Transcriptional Regulators from Deinococcus radiodurans. *J. Biol. Chem.* 279: 51442-51450.

Wilkinson, S. P. & A. Grove, (2005) Negative cooperativity of uric acid binding to the transcriptional regulator HucR from Deinococcus radiodurans. *J Mol Biol* 350: 617-630.

# References

Wilkinson, S. P. & A. Grove, (2006) Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol* 8: 51-62.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin & K. S. Wilson, (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235-242.

Winter, G., (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43: 186-190.

Xu, W., J. Huang & S. N. Cohen, (2008) Autoregulation of AbsB (RNase III) expression in Streptomyces coelicolor by endoribonucleolytic cleavage of absB operon transcripts. *J Bacteriol* 190: 5526-5530.

Xu, W., J. Huang, R. Lin, J. Shi & S. N. Cohen, (2010) Regulation of morphological differentiation in S. coelicolor by RNase III (AbsB) cleavage of mRNA encoding the AdpA transcription factor. *Mol Microbiol* 75: 781-791.

# Publications