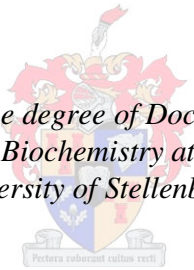
Regulators of dormancy/viability of *Mycobacterium tuberculosis* inside the human macrophages.

by
Maria Magdalena Botha



Dissertation presented for the degree of Doctor of Philosophy of Medical Biochemistry at
University of Stellenbosch

Promoter: Prof IJF Wiid
Co-promoter: Prof CP Kenyon
Faculty of Health Science
Division Molecular Biology and Human Genetics
Department of Biomedical Science

March 2012

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2012

Abstract

The investigation was aimed to improve the understanding of the binding interactions between DevS and DevR that are implicated in the regulation of the dormancy response in *Mycobacterium tuberculosis*. These binding interactions could provide good drug targets for the treatment of persistent tuberculosis, the mechanistic understanding of their binding interactions is important for the development of a validated inhibitor screen. A detailed *in silico* analysis of the amino acid residues that play a role in the binding of receptor DevR to both kinase DevS and the target DNA was undertaken. A reasonable approximation of the DevS structure was produced using homologous protein structures. *In silico* docking of DevS to DevR merely produced a set of probable candidate structures, since more than one conformation with similar docked energies was observed. The decision on which one is the more correct form can only be estimated by crystallization of this complex. Therefore, the functional expression and purification of the Dev TCS components were pursued. Denaturing HIS™-select nickel affinity gel purification in the form of matrix-assisted refolding led to the production of functional Dev TCS proteins. To understand the binding of DevR to DNA consensus sequences, as well as the nature of these interactions, a model was built of the full length DevR dimer binding to DNA consensus sequences. Based on this model, single mutations were made to DevR *in vitro* and their effects assessed in order to validate the model built. During Electrophoretic Mobility Shift Assay (EMSA) analysis, it was found that K179I and N183L mutants prevented the binding of DevR to the DNA consensus sequences. If DevR and DevS binding are to be used in a drug development program, it is essential to have the protocols to accurately measure their interaction, in addition to developing a fundamental understanding of how their interactions occur. The binding affinity of DevR to both DevS and the truncated soluble fragment of DevS (DevS201) were explored, using the BIAcore instrument, an SPR-based biosensor. For sufficiently strong binding between a histidine kinase and a response regulator, the KD needs to be in the nM range. The KD was calculated to be 255 nM for DevS201 and 184 nM for DevS. Therefore it can be concluded that DevS201 binds DevR strongly enough to be used in future studies, and that the BIAcore could be used to screen small-molecule inhibitors of DevR-DevS interactions.

Opsomming

Die Dev twee komponent sisteem (TKS) bestaan uit ‘n histidine kinase naamlik (DevS) en ‘n reaksie reguleerder DevR. DevS en DevR is betrokke by die regulering van die dormante stadium van *Mycobacterium tuberculosis*. Hierdie meganisme kan ‘n deurbraak dwelm teiken vir die behandeling van sluimerende tuberkulose wees. Die meganisme van hierdie bindings interaksies is van kritieke belang, tesame met die ontwikkeling van 'n erkende inhibeerder toets. ‘n Gedetaileerde *in silico* analise van die aminosuur volgordes wat 'n rol speel in die binding van die reseptor DevR aan beide DevS sowel as die teiken DNS is voltooi. ‘n Model van die DevS struktuur is saamgstel met behulp van homoloë proteïen strukture. *In silico* mering van DevS aan DevR het `n stel van die waarskynlike kandidaat strukture verskaf, aangesien meer as een konformasie met soortgelyke merings energieë waargeneem is. Die mees waarskynlike vorm kan alleenlik geïndentifiseer word na kristallisasie van hierdie kompleks. Die funksionele uitdrukking en suiwering van die Dev TKS proteine is gevolglik uitgevoer. Funksionele Dev TKS proteïene is verkry deur denaturerende HIS-select nikkel affiniteit jel suiwering, in die vorm van matriks-geassisteerde hervouing te gebruik.

Ten einde die binding te verstaan tussen DevR en DNS konsensus volgordes, sowel as die aard van hierdie interaksies, is 'n model gebou van die volle lengte DevR dimeer binding aan DNS konsensus volgordes. Hierdie model is gevalideer deur punt mutasies in DevR te skep en die gevolge daarvan te beoordeel met elektroforetiese mobiliteits verskuiwing reaksie analises. Dit is bevind dat K179I en N183L mutante, verhoed die binding van DevR aan die DNS konsensus volgordes.

Die gebruik van DevR en DevS bindings in ‘n dwelm ontwikkelingsprogram, benodig die fundamentele begrip van hoe die interaksies plaasvind, sowel as akkurate protokolle om die interaksies te meet. Die BIAcore instrument, ’n SPR-gebaseerde biosensor, is ingespan om die bindings affiniteit van DevR aan beide DevS en die fragment van DevS (DevS201) te ondersoek. Om voldoende sterk binding tussen DevS en die DevR te verseker, moet die KD in die nM omgewing wees. Die KD is bepaal as 255 nM en 184 nM vir DevS201 en DevS, onderskeidelik. Die afleiding kan dus gemaak word dat DevS201 sterk genoeg aan DevR bind om in verdere studies gebruik te kan word, en dat die BIAcore gebruik kan word om klein-molekule inhibeerders van DevR-DevS interaksies te toets.

Acknowledgments

Due to the contributions of many different people in my life and studies it has made it possible for me to write this thesis. I would like to extend my appreciation especially to the following people:

- It's an honour to thank Prof Colin Kenyon for his enthusiasm, his encouragement and his resolute dedication to me, this project and science in general. I would also like to show my gratitude to Prof I Wiid for his encouragement and guidance.
- I would like to thank Anjo Theron , whose patience and humour have rescued me from peril more times than I can recall, and Dr Robyn Roth, her painstaking effort in proof reading the drafts is highly appreciated.
- I am indebted to many of my colleagues especially Lyndon, Thandeka, Lionel and Stoyan for all their assistance, valued advice, insight and useful discussions.
- To my teachers, especially Mr. K de Wet, for awakening my love to wonder, which is the seed of science.
- This thesis would not have been possible unless, I had received funding from the CSIR.
- My parents for all their undaunted love, patience and support in all my failures and successes. My parents, whom throughout my childhood and career had always encouraged me to follow my heart and inquisitive mind in any direction it took me.
- I would like to show my gratitude towards my grandparents and family for their unwavering support and encouragement.
- Heartfelt appreciation to my friends for keeping me sane and never failing to lift my spirits.
- I am eternally grateful for the wisdom and perseverance that He has bestowed upon me during my research project and indeed throughout my life.

Table of Content

List of Figures

List of Tables

Abbreviations

3D	Three dimensional
BCG	Bacilli Calmette Guérin
BIA	Biomolecular Interaction Analysis
BLAST	Basic Local Alignment Search Tool
B. subtilis	*Bacillus subtilis*
CATH	Class Architecture Topology Homologous Superfamily
DOPE	Discrete Optimized Protein Energy
DNA	Deoxyribonucleic Acid
DS	Discovery Studio
DSC	Discrimination of protein Secondary structure Class
EC	Enzyme Commission
E. coli	*Escherichia coli*
EMBOSS	European Molecular Biology Open Software Suite
EMSA	Electrophoretic Mobility Shift Assays
GO	Gene Ontology
HIV	Human Immunodeficiency Virus
HK	Histidine Kinase
Ka	Association Constant
Kd	Dissociation Constant
KD	Equilibrium Constant
LB	Luria Bertani

MDR	Multi-Drug Resistant
MS	Mass Spectrometry
M. tb	*Mycobacterium tuberculosis*
NCBI	National Center for Biotechnology Information
NTA	Nitrilotriacetic Acid
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Database
PDF	Probability Density Function
PT	Phosphotransfer domain
ROI	Reactive Oxygen Intermediates
RMS	Root Mean Square
RMSD	Root Mean Square Deviation
RU	Response Units
SCOP	Structural Classification of Proteins
SDM	Site Directed Mutant
SDMS	Site Directed Mutagenesis
SRP	Surface Plasmon Resonance
TCS	Two-Component Systems
TB	Tuberculosis
TM	Transmembrane Regions

WHO	World Health Organization
XDR	Extensively Drug Resistant

Chemical Abbreviations

AMP	Adenosine Mono-Phosphate
ATP	Adenosine Tri-Phosphate
Ca	Calcium
$CaCl_2$	Calcium Chloride
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleosides
DTT	Dithiothreitol
EDC	1-ethyl-3(3-dimethylaminopropyl)-carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
K	Potassium
KCl	Potassium Chloride
KH_2PO_4	Potassium Dihydrogen Phosphate
Mg	Magnesium
$MgCl_2$	Magnesium Chloride
NaCl	Sodium Chloride
Na_2HPO_4	Disodium Hydrogen Phosphate

NHS	N-hydroxysuccinimide
NO	Nitric Oxide
O_2	Oxygen
PMSF	Phenylmethylsulphonyl Fluoride
SDS	Sodium Dodecyl Sulphate
TAE	Tris Acetate EDTA Buffer
TBE	Tris Borate EDTA Buffer
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
X-Gal	Bromo-chloro-indolyl-galactopyranoside

Buffer formulations

Binding Buffer	50mM Na_2HPO_4, 200mM NaCl, 10Mm Imidazole, pH 8.0
Buffer A	0.1M Tris-HCl, 0.1mM EDTA, 1mM DTT, 50mM NaCl, pH 8.0
Denaturation Buffer	20mM Tris-HCl, 500mM NaCl, 10% glycerol, 8M urea, pH 8.0
Dev Phosphorylation Buffer	50mM Tris-HCl, 50mM KCl, 10mM $MgCl_2$, pH 8.0
dNTP mix	Contains equal amounts of each dATP, dCTP, dGTP, dTTP in an aqueous solution neutralized to pH 7.0
Elution Buffer	50mM Na_2HPO_4, 300mM NaCl, 250mM Imidazole, pH 8.0
EMSA Binding Buffer	20mM Tris-HCl, 50mM KCl, 4mM $MgCl_2$, 5mM $CaCl_2$, 1% glycerol, pH 7.4
Equilibration Buffer	50mM Na_2HPO_4, 200mM NaCl, 10mM Imidazole, pH 8.0
DNA Loading Dye	10mM Tris-HCl, 0.15% Orange G, 0.03% Xylene cyanol FF, 60% Glycerol, 60mM EDTA, pH 7.6 (Fermentas)
Gel Storage Buffer	40% Methanol, 5% Sucrose, 5% Glycerol in dH_2O
HBS-N	10mM HEPES, 150mM NaCl, pH 7.4
Laemmli Loading Buffer	1.2% SDS, 30% Glycerol, 15% ß-mercaptoethanol, 0.18 mg/ml Bromophenol Blue, 0.15M Tris-HCl, pH 6.8
LB agar	1% Agar,1% Tryptone, 0.5% Yeast extract, 1% NaCl in dH_2O, pH 7.0
LB media	1% Tryptone, 0.5% Yeast extract, 1% NaCl in dH_2O, pH 7.0
MS Buffer A	0.5% Acetonitrile, 0.1% Formic acid
MS Buffer B	100% Acetonitrile, 0.1% Formic acid
MS Phosphorylation Buffer	20mM Tris-HCl, 50mM KCl, 4mM $MgCl_2$, and 5mM $CaCl_2$, pH 7.4
Native Running Gel 5%	5% Acrylamide-Bisacrylamide mix, 0.06% Ammonium Persulphate, 0.1% TEMED, 5.2% TBE, pH 8.0

Native Stacking Gel 4%	4% Acrylamide-Bisacrylamide mix, 0.05%Ammonium Persulphate, 0.1% TEMED, 2.5% TBE, pH 8.0
PAGE Destaining Solution	40% Methanol, 10% Glacial Acetic Acid in Water
PAGE Staining Solution	0.1 g Coomassie Blue G250, 40% Methanol, 10% Glacial Acetic Acid in Water
PBS	137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4, and 1.4mM KH_2PO_4, pH 7.5
Renaturation Buffer	20mM Tris-HCl, 500mM NaCl, 10% Glycerol, pH 8.0
Running Gel 14%	14% Acrylamide-Bisacrylamide mix, 0.1% SDS, 0.05% Ammonium Persulphate, 0.1% TEMED, 375mM Tris-HCl, pH 8.8
SpoII Phorylation Buffer	50mM Tris-HCl, 50mM KCl, 10mM $MgCl_2.6H_2O$, 1mM DTT , 0.1mM EDTA, pH 7.6
Stacking Gel 4%	4% Acrylamide-Bisacrylamide mix, 0.1% SDS, 0.05% Ammonium Persulphate, 0.1% TEMED, 50mM Tris-HCl, pH 6.8
TAE	0.04M Tris-Acetate, 1mM EDTA, pH 8.0
TBE	890mM Tris-borate, 890mM boric acid, 20mM EDTA, pH 8.0
TE Buffer	10mM Tris, 1mM EDTA, pH 8.0
Tris-Glycine Buffer	25mM Tris-HCl, 0.2M Glycine, pH 8.3
Wash Buffer	50mM Na_2HPO_4, 200mM NaCl, 10mM Imidazole, pH 8.0

Chapter 1 Literature Review

"Today we are calling on the world to recognize that we can't fight HIV unless we do much more to fight TB as well."

- President Nelson Mandela

1.1 *Mycobacterium* genus.

The genus *Mycobacterium* includes several medically-important species that cause a high degree of human morbidity and mortality. *Mycobacterium tuberculosis* (*M. tb)* is estimated by the World Health Organization (www.who.org) to affect almost one third of the world population. The genus is divided into slow growers and fast growers, taking either more or less than 7 days to form colonies on selected media, respectively. Based on a high degree of genetic similarity, a number of these species can be grouped into complexes that will also cause similar disease syndromes (Saviola and Bishai, 2006). The *Mycobacterium tuberculosis* complex is comprised of *M. tb, M. bovis, M. microti* and *M. africanum.* and *M. bovis* are most prevalent in humans, while *M. africanum* is found in patient isolates collected from Africa (Shinnick and Good, 1994). *M. bovis* is known for its attenuated vaccine strain Bacilli Calmette Guerin (BCG) (Saviola and Bishai, 2006). The *Mycobacterium avium* complex consists of *M. avium, M. intracellulare* and *M. paratuberculosis*, commonly found in the environment and isolated from soil, water and food (Havlir and Ellner, 2000). Chickens, pigeons and other birds can show tuberculosis-like symptoms caused by an infection of *M. avium*, whereas *M. paratuberculosis* infects ruminants to cause related symptoms called

Johne's disease (Scanu *et al*., 2007). *M. leprae*, another example of a slow growing *Mycobacterium* species, causes leprosy, recognized by the bodily deformities that are symptomatic to the disease (Gelber *et al*., 2009). *M. marinum*, closely related to the *M. tb* complex, produces cutaneous lesions such as swimming pool granuloma or fish handler's nodule. Since the reservoir for *M. marinium* is water, humans become infected via environmental water, handling fish or during the cleaning and maintenance of fish tanks (Toosi and Ellner, 1998).

1.2 Tuberculosis

Tuberculosis (TB) is caused by *M. tb* - one of the most dreaded human pathogens. TB is still a primary cause of death in low- and middle-income countries, particularly in sub-Saharan Africa where people have an increased vulnerability conferred by Human Immunodeficiency Virus (HIV) infection and the acceleration in the emergence of drug resistant *M. tb* strains (World Health Organization report, 2006). Multidrug resistant (MDR) tuberculosis is resistant to rifampicin and isoniazid (frontline drugs), with or without other resistance (Maartens and Wilkinson, 2007). Extensively drug resistant (XDR) tuberculosis is resistant to at least rifampicin and isoniazid, in addition to any quinolone and at least one injectable second line agent such as capreomycin, amikacin or kanamycin (Maartens and Wilkinson, 2007). In 2007, the World Health Organization estimated that there were approximately 9 million new TB cases, which was higher than the estimated 8.24 million new cases in 2006. India and China had the highest incidence rates, as can be seen in Figure 1.1, followed by South Africa, Nigeria and Indonesia. In terms of total number of cases, India ranked first with 2.0 million, before China (1.3 million), Indonesia (0.53 million), Nigeria (0.49 million) and South Africa (0.46 million). Countries like Russia, Ethiopia, Kenya, Zaire, Uganda, Tanzania and Zimbabwe also fall in this high prevalence group, as seen in Figure 1.1.

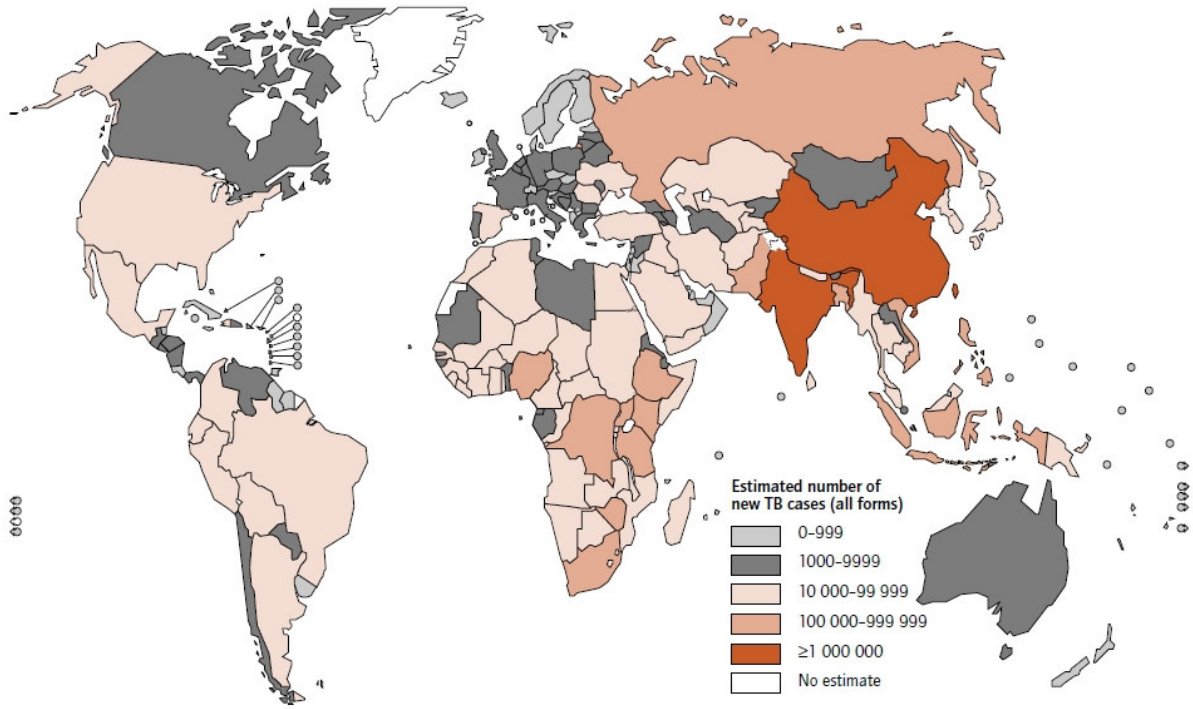


Figure 2.1: The estimated new TB infections (incidence) of all forms per country. The countries with a low incidence are indicated in shades of grey, countries with high incidence are indicated in shades of red. India and Indonesia has the highest incidence. WHO estimates that 9.27 million new cases of TB occurred in 2007 (139 per 100,000 population), this is higher compared with 8.24 million new cases (140 per 100,000 population) in 2006. (World Health Organization report, 2007).

It was estimated that 23% of total deaths of people infected with *M. tb* can be attributed to co-infection with HIV, equivalent to 465,000 deaths amongst HIV-positive people and a total of 1.3 million deaths among HIV-negative people in 2007 (WHO Report, 2009). The increasing numbers of new cases can be ascribed to drug resistance, co-infection with HIV, urban crowding, international travel and migration, as well as poverty, poor infection control policies and inadequate drug treatment regimens. Contributing to these factors are the difficulties encountered during the diagnosis and treatment of children, especially in third world countries (Fätkenheuer *et al*., 1999). It has become evident that only an attempt to fight TB and HIV simultaneously will have an impact on disease management (Corbett *et al*., 2003).

1.3 *Mycobacterium tuberculosis* infection

When a person with an active *M. tb* infection coughs, the bacteria spread via the aerosols produced. After these infectious bacilli are localized in the lungs (primary site of infection) and various disease outcomes are then possible (Cosma *et al*., 2003). Most people who are exposed to the tuberculosis bacilli have an innate protection and will never show any sign of infection (Parish *et al*., 1999). In a small proportion of people the bacilli will actively multiply and the host will develop the active disease. Following infection, bacilli reside inside the macrophages in the lung. If the host's immune system does not eradicate the bacterium, *M. tb* will adapt to the hostile environment within the macrophage and start to replicate. Typically, during most other bacterial infections, the invading bacterium will be destroyed within the phagosome via acidification of the phagosome after fusion with the lysosome (Fisher *et al*., 2002). Destruction of the bacilli should occur because of the lytic enzymes present in the lysosome, as well as the sterilizing effect of the production of reactive oxygen intermediates (ROI) (Clemens and Horwitz, 1995).

In contrast, *M. tb*-containing phagosomes resist acidification as well as lysosomal fusion. *M. tb* arrests the maturation of its phagosome at the stage where the phagosome has the characteristics of early and late endosomes but not the characteristics of a lysosome (Schnappinger *et al*., 2003). This leads to the evasion of the first line of defence of the host immune system by the *M. tb* bacilli.

For this reason, a portion of people will harbour a dormant infection that is a constant source of reactivation disease (Corbett *et al*., 2003). During latent infection, the bacilli are suppressed, the host does not develop the active disease and consequently the bacteria avoid being eliminated. When the host immune system fails to keep the bacilli in check, the infection will be reactivated and cause active disease (Flynn and Chan, 2001). The threat of *M.tb* lies in its ability to revert to asymptomatic latent infections that serves as a reservoir for future infections.

Current treatment of latent tuberculosis is not as effective as active TB treatment because of the latent bacilli's low metabolic state. Failure to treat both active and latent tuberculosis will hamper the progress in global eradication of the disease (Voskuil *et al.,* 2004)

1.4 *Mycobacterium tuberculosis*: The dormant stage and dormancy regulon

The non-proliferating stage of *M. tb* infection is referred to as the stationery phase, latency, dormancy or persistence (Zahrt and Deretic, 2000; Rhen *et al*., 2003). The metabolic and replicative status during this dormant phase is not clear. *M. tb* does not form true spores like *B. subtilis* nor does it go into a non–differentiating stage like *E. coli*. Furthermore, *M. tb* dormancy is not driven by starvation conditions but rather by hypoxic conditions and low nitric oxide (NO) concentrations (Voskuil *et al*., 2004). *M. tb* bacilli in this dormant state show a phenotypical resistance to anti-mycobacterial drugs, mainly because these drugs target growth-related functions (Boon and Dick, 2002). In order to replicate dormancy, experimental models are used and include exposure to low concentrations of oxygen and increased NO (Saini *et al*., 2004a). Voskuil *et al*. (2003) characterized a gene set of 48 genes that are up-regulated when *M. tb* enters into a low oxygen and increased NO dormant state. These genes are in close proximity, indicating that this type of arrangement might support a fast, co-ordinated transcriptional response when dormancy is entered (Voskuil *et al*., 2003). It was demonstrated that this gene set failed to induce when the devR gene was disrupted, and the set was labelled the "dormancy regulon" (Park *et al*., 2003). Most of the genes identified were conserved hypothetical proteins, but this dormancy regulon also included enzymes involved with glutamine and nitrate transport, as well as acyl-CoA dehydrogenase and a resuscitation promoting factor (Kendall *et al*., 2004).

1.5 Two-component signal transduction systems in Prokaryotes

Pathogen survival is based on the capacity of pathogens to express no more than those gene products necessary for the optimal growth in their ever-changing environment. Prokaryotes have developed complex adaptation machineries to survive in various ecological niches, one of which is the two-component signal transduction system (Beier and Gross, 2006). Currently approximately 4,000 bacterial two-component systems (TCSs) have been identified in 145 sequenced genomes (Ulrich *et al*., 2005). The TCSs, which function by means of histidine phosphorylation, are a critical component of prokaryotic cell signalling and are also found in lower eukaryotes. TCSs consist typically of a membrane-bound sensor with histidine kinase (HK) activity and a cytoplasmic response regulator (West and Stock, 2001). Bacterial and

archaeal species make use of this signalling process. Response regulators make up approximately 23% of all encoded proteins of 200 completely sequenced genomes (Galperin, 2006). The basic biochemistry of the TCS starts with the sensor kinase's capability to recognize the change in the environment, autophosphorylation and subsequent phosphorylation of a specific response regulator, as seen in Figure 1.2. The HK protein is phosphorylated at a conserved histidine located in the kinase domain of the protein. Phosphotransfer occurs between the phosphorylated-histidine on the HK to an aspartate residue on the response regulator protein (Figure 1.2). The regulatory domain of the response regulator changes conformation based on this phosphorylation, which alters the effector domain's DNA binding properties (Beier and Gross, 2006; Puttick *et al*., 2008). Response regulators then activate the transcription of a number of genes that are specifically associated with survival (Grigoroudis *et al*., 2007). Despite their global presence and importance in bacteria, surprisingly little is known about the external stimuli that activate different systems and the subsets of genes that are affected. Examples of signals that can activate the kinase sensor include ions, pH, temperature, oxygen pressure, redox state, osmolarity and others (Beier and Gross, 2006).

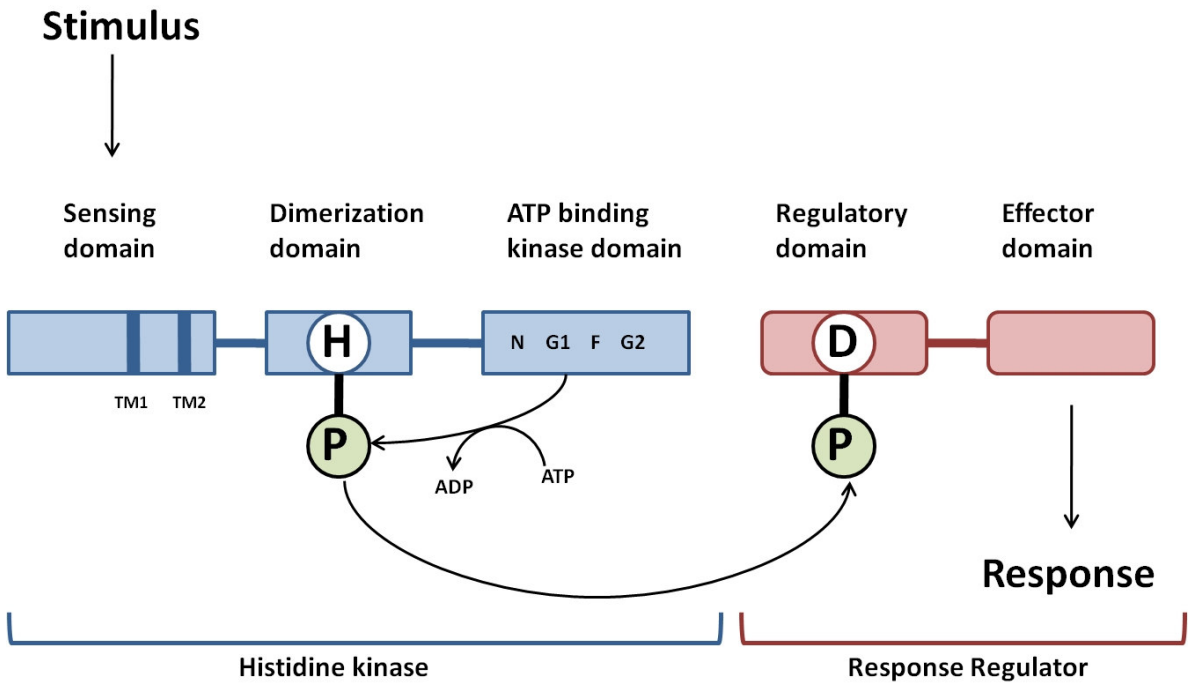


Figure 1.2: The two-component phosphotransfer scheme (Klumpp and Krieglstein 2002). A TCS typically consists of a dimeric transmembrane sensor histidine kinase and a cytoplasmic response regulator. The ATP-binding domain of the HK consists of approximately 250 amino acids and contains four conserved blocks of amino-acid sequences: N, G1 (or D), F and G2 (or G). Similarly, the Regulatory domain of the response regulator can be identified by the number and spacing of conserved aspartate, lysine, and hydrophobic residues.

1.6 The sensor histidine kinases

Most members of the histidine kinase (HK) family have two functionally and structurally different segments: an N-terminal sensor domain and a C-terminal conserved domain (Wolanin *et al*., 2002). The sensing domains are variable in sequence, reflective of the many different environmental signals to which HKs are responsive, whereas the kinase core is more conserved (Galperin, 2006). The kinase core includes dimerization and ATP binding domains and can be identified by five conserved sequence motifs present in prokaryotic HKs (Stock *et al*., 1988). These motifs have been termed the H, N, G1 (or D), F and G2 (or G) boxes (Bilwes *et al*., 2001). The conserved histidine that gets phosphorylated is the central feature in the H box, whereas the N, D, F and G boxes define the ATP binding cleft (Marina *et al*., 2001). In most HKs, the H box is part of the dimerization domain. However, for some proteins, like *T. maritima* CheA, the conserved histidine is located at the far N-terminus of the protein in a separate Histidine-containing phosphotransfer (PT) domain (Bilwes *et al*., 2001). The N, G1, F and G2 boxes are usually contiguous, but the spacing between these motifs can be somewhat varied, as seen in Figure 1.3 and Figure 1.4.

The work by Grebe and Stock (1999) identified 11 distinct sub-families of HKs, based in the outcome of multiple protein sequence alignments. Nevertheless, the sub-families all have the representative homology boxes. Figure 1.3 shows the alignment of a representative of each of the sub-families, indicating the presence of the relevant boxes. Kim and Forst (2001) also took a genomic approach to analyse the HK family in bacteria and archae, identifying the HK groups based on phylogenetic analysis and sequence similarity. They concluded that the HKs could be grouped into 5 different classes (Figure 1.4).

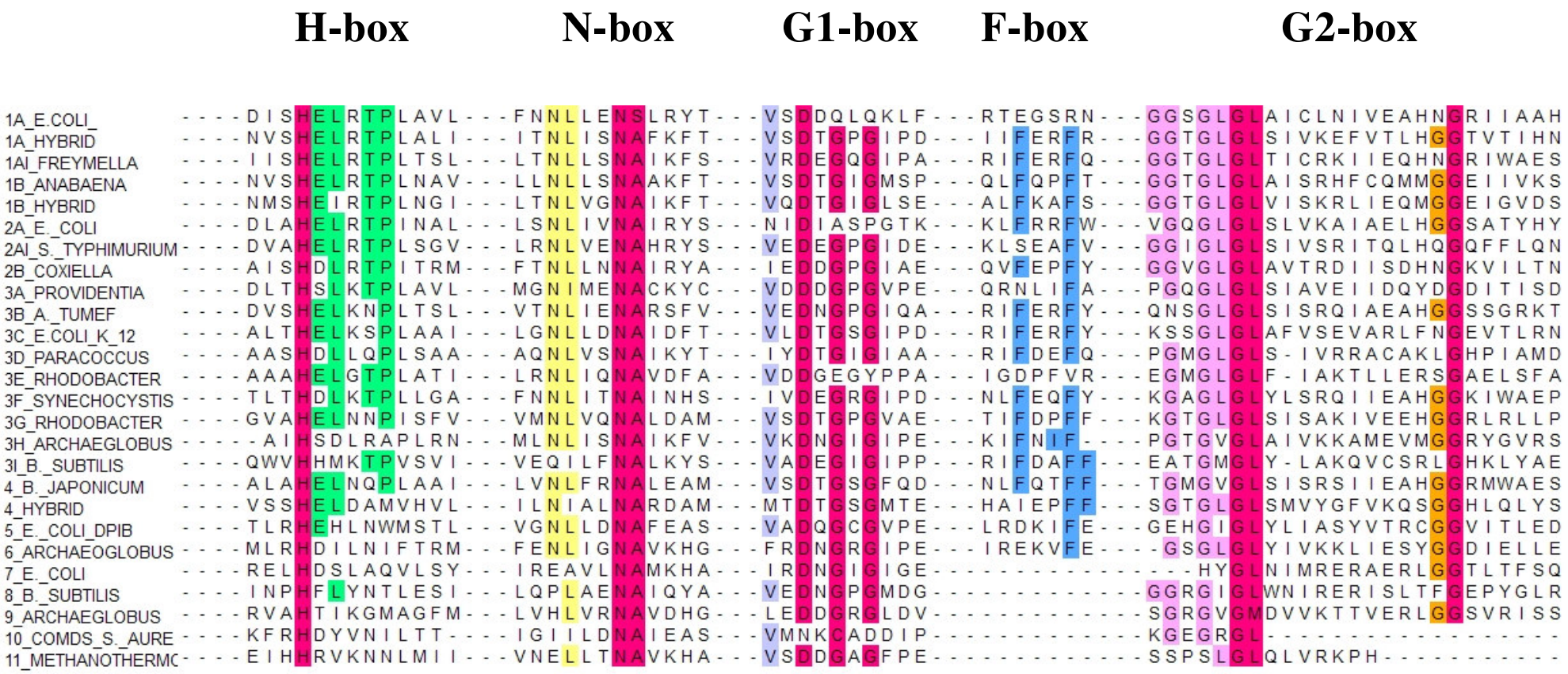


Figure 1.3: Protein sequence alignment of representatives of the 11 histidine kinase sub-families. The conserved boxes and the motives they comprise are shaded. The dark pink indicates highly conserved residues. The other colours indicate residues that are more than 50% conserved in HK's. The alignment was adapted from Grebe and Stock, 1999. The numerals indicate the group to which the sequence belongs. Some groups have several sub-groups and this is indicated by the alphabetic character. Abbreviations of the species is used and the full names are found Appendix A.

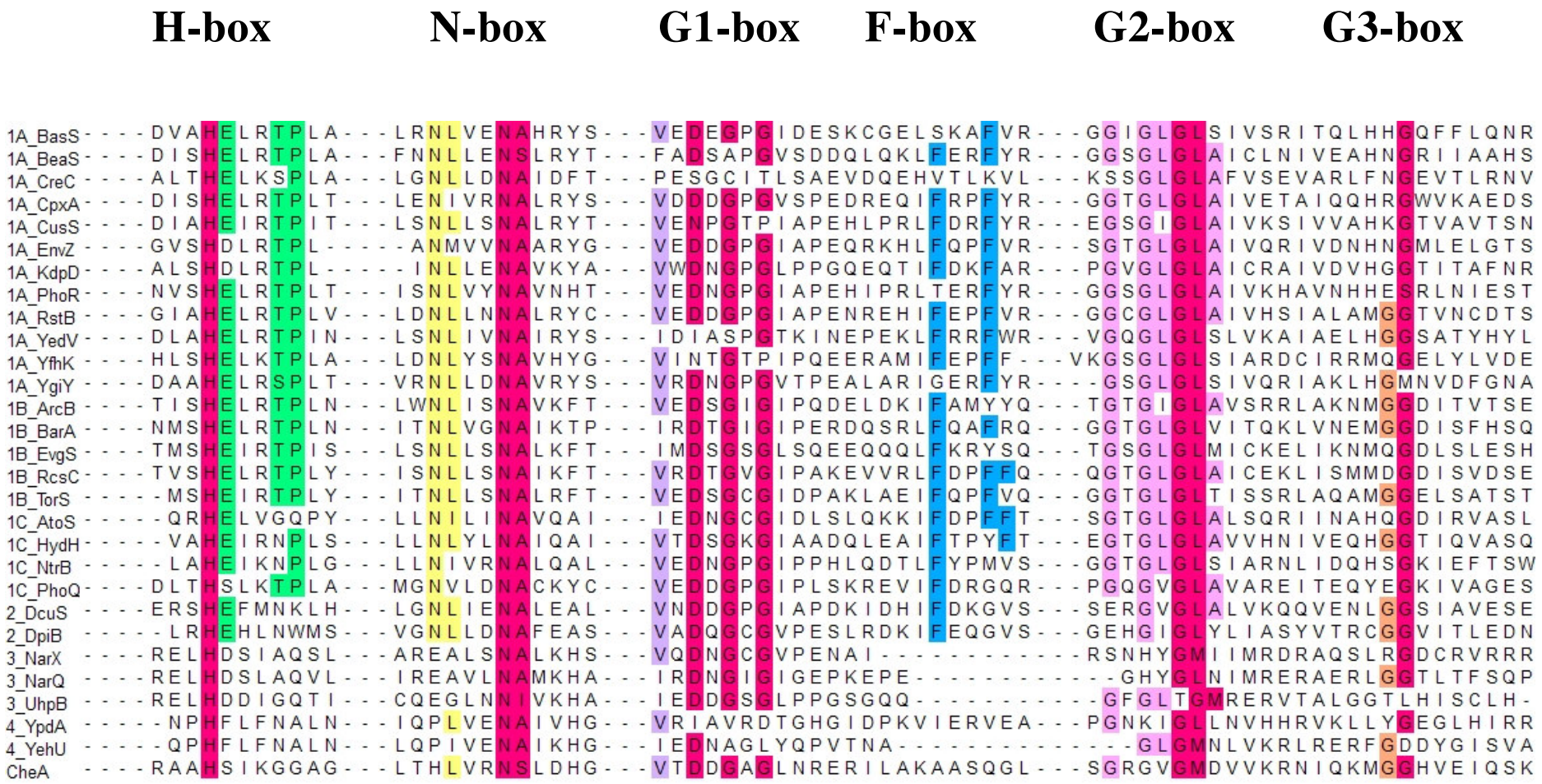


Figure 1.4: Alignment of the homology boxes of the HKs found in *E. coli*. The conserved motifs are shaded. The dark pink indicates highly conserved residues. The other colours indicate residues that are more than 50% conserved in HK's. The classification of each HK found in *E. coli,* is indicated by the numeral identifier and the sub-class is indicated by the alphabetic character. CheA is the only member of class 5. The alignment was adapted from Kim and Forst, 2001.

1.7 *Mycobacterium tuberculosis*: two-component systems

Tubercle bacilli survive and multiply in severe, hostile environments within the human host. To ensure survival, it is important for the bacilli to express different sets of genes at different stages of infection (Parish *et al*., 2003). *M. tb* contains 11 paired two-component HK signal transduction systems, as well as a few unpaired sensor kinases and orphan regulators (Haydel and Clark-Curtiss, 2004). A summary of these TCS is shown in Table 1.1. Two-component regulatory systems are usually encoded in operons that include both the sensor kinase and response regulator genes for co-expression of the corresponding proteins. This also decreases the chances of cross-talk between non-cognate sensors and regulators (Laub and Goulian, 2007). In contrast, bacterial genomes often code for unpaired two-component system proteins where a sensor kinase gene is not accompanied by a response regulator gene and orphan regulators not accompanied by a sensor kinase (Raghavan and Groisman, 2010). Very little is known about the global regulatory mechanisms of *M. tb* that co-ordinate the gene expression after phagocytosis by the human macrophage, but *M. tb* initiates a latent infection that overcomes the defence reactions launched by the host cell (Rhen *et al*., 2003; Haydel and Clark-Curtiss, 2004). TCSs allow bacteria to sense and adapt to different environmental factors, and must assist *M. tb* to survive inside the macrophage in the presence of low oxygen, high NO, presence of reactive oxygen intermediates (ROIs), low iron concentration, pH differences and nutrient deprivation. Compared to bacteria like *E.coli*, *B.subtilis* and *Streptococcus pneumonia, M. tb* has relatively few TCSs (Fontan *et al*., 2004). The DevR-DevS TCS, also referred to as DosR-DosS, is a two-component system of *M. tb* identified to play a significant role in the dormancy regulon (Park *et al*., 2003). DevS is the sensor kinase and DevR the response regulator. The Dev TCS was found to express at higher levels in a virulent strain compared to the avirulent strains (Dasgupta *et al*., 2000). Dev was named as "**d**ifferentially **e**xpressed in **v**irulence strain" and is also known as Dos for "**do**rmancy **s**urvival regulator" (Gupta *et al*., 2009).

Table 1.1 The 11 paired two-component systems found in *M. tuberculosis*

Kinase Sensor	**Response Regulator**	**Output function**
PhoQ	PhoP	Perez *et al.* (2001) implicated the Pho TCS in virulence. In *Salmonella typhimurium* the PhoP-PhoQ system functions to detect cellular concentrations of Mg^{2+}. PhoQ detects low Mg^{2+} concentration and PhoP will initiate expression of high affinity Mg^{2+} transport systems to assist the bacteria to overcome the Mg^{2+} starvation (Vescovi *et al*., 1997). The *M. tb* PhoP shows sequence similarity to *S. typhimurium* PhoP and a similar function can be inferred (Groisman, 2001). The specific genes under PhoP control have not been identified (Fontan *et al*., 2004).
KdpD	KdpE	The K^{+} uptake machinery in *M. tb* is believed to be under the control of the Kdp TCS (Altendorf *et al*., 1994).
PrrB	PrrA	PrrA-PrrB assists the adaption of the bacilli during the early growth phase (Ewann *et al*., 2004). The PrrA–PrrB system belongs to a wide subfamily of TCSs, of which OmpR–EnvZ is the prototype (Ewann *et al*., 2004).
MprB	MprA	This TCS plays a role in the establishment and maintenance of the persistent infection (Zahrt and Deretic, 2000).
NarX	NarL	*M. tb* NarL was classified as a putative nitrate response regulator involved in the control of anaerobic metabolism (Schnell *et al*., 2008). Protein sequence homology revealed that the closest sequence homologue to *M. tb* NarL was from *E. coli.* However, it displayed significant differences in three-dimensional structure (Schnell *et al*., 2008).
MtrB	MtrA	Zahrt and Dieretic (2000) reported, that although the function of this TCS is still unknown, it is essential to *M. tb* because it is differentially expressed between virulent and avirulent mycobacterium strains during growth in macrophages and MtrA is induced in iron deficient conditions (Rodriguez *et al*., 2002).
SenX	RegX	Clover *et al*. (2007) have recognized the SenX3-RegX3 TCS of *M. smegmatis* as a phosphate-dependent regulator of genes implicated in phosphate acquisition. The TCS is one of the five conserved *Mycobacterium* TCSs, suggesting that it plays a similar and important role in the survival of the mycobacterium in both virulent strains and avirulent strains (King-Scott *et al*., 2007).
TrcS	TrcR	TrcS-TrcR mutations showed no phenotype in BALB/c mouse, but were hypervirulent in SCID mouse models (Parish *et al.*, 2003; Ewann *et al*., 2004). The role this TCS plays in growth or virulence remains unknown. However, evidence was found that TrcR is involved in the repression of the Rv1057 β-propeller gene that is expressed during growth of *M. tb* within human macrophages (Haydel and Clark-Curtiss, 2006).
0600c/ 0601c/	*0602c* (TcrA)	Plays a role in the regulation of hypoxia in *M.tb*. However, the specific role is not known (Sherman *et al.,* 2001).
TcrY	TcrX	In a SCID mouse model, TcrY-TcrX mutants showed a hypervirulent phenotype (Parish *et al*., 2003).
DevS	DevR	Responds to hypoxia and is required for survival (Boon and Dick, 2002).

1.8 *Mycobacterium tuberculosis*: Dev two-component system

The *M. tb* Dev TCS is involved in the mechanism by which the bacterium enters dormancy when challenged with an increase of NO and hypoxia in the macrophage (Ioanoviciu *et al*., 2007). Hypoxia, a condition of decreased oxygen availability at the cellular level, is alleged to be connected with the induction and persistence of dormancy during the latent stage. Other known signals that activate the Dev regulon are low iron and ethanol (Schnappinger *et al*., 2003). The schematic of the organization of the Dev locus is shown in Figure 1.5. A stress or stimulus in the cellular environment leads to the autocatalytic phosphorylation of the histidine kinase (DevS), which in turn causes the phosphoryl group transfer between the HK and the response regulator, DevR. The activated DevR then induces the dormancy regulon (Park *et al*., 2003).

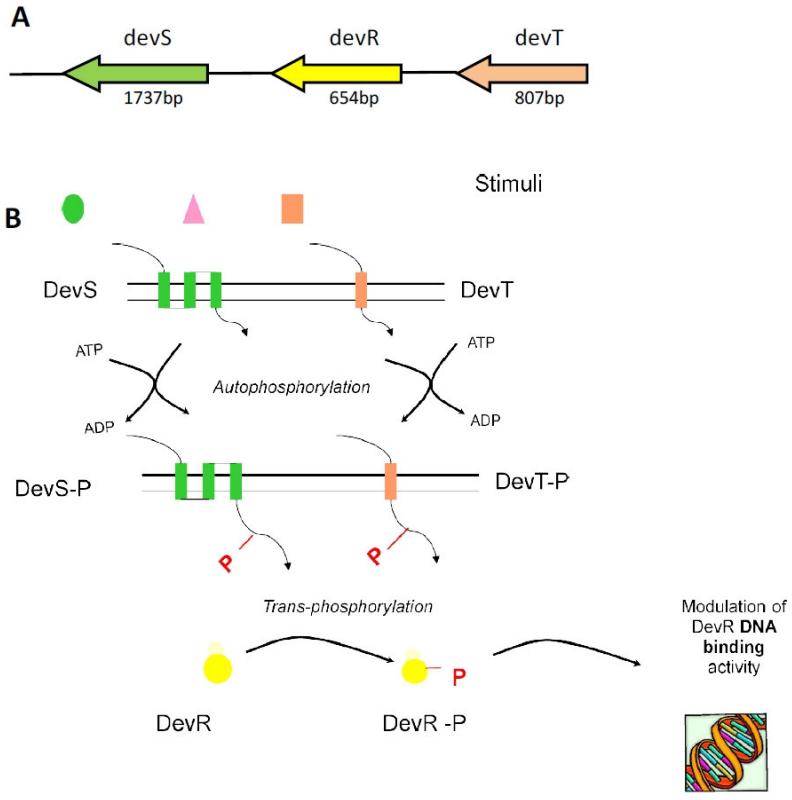


Figure 1.5: Schematic of the organization of the Dev locus. A) Schematic of the genome organization of the *M. tuberculosis* Dev locus (Chauhan and Tyagi, 2008a). B) The DevT-DevS-DevR phosphotransfer scheme (Roberts *et al*., 2004). Both DevS and DevT are sensor HKs, which get phosphorylated by ATP during autophosphorylation. The phosphate group then gets transferred to the response regulator DevR during trans-phosphorylation. The phosphorylated DevR can then bind DNA.

The two-step reaction takes place when the membrane-bound HK is autophosphorylated at conserved His395. During the second step, the response regulator DevR is phosphorylated from the histidine by transfer of the phosphoryl group to a conserved aspartic acid residue, Asp54. When the phosphorylation status of the DevR changes, it alters its affinity for DNA motifs present upstream of the target genes (Saini *et al*., 2004b). Roberts *et al*. (2004) hypothesized that DevR phosphorylation increases its affinity for a specific palindromic DNA sequence that precedes genes that are up-regulated during the hypoxic response (the dormancy regulon), switching on the transcription of these genes. DevT, an unpaired HK sensor, is co-transcribed with the DevR-DevS TCS, and has the ability to phosphorylate DevR as well. DevT is an alanine-valine-rich protein that belongs to the universal stress protein family found in various bacteria (Saini *et al*., 2002). DevS uses Mg^{2+} as divalent ion during autophosphorylation, whereas DevT utilizes both $Mg^{2+}$and Ca^{2+}. In contrast to *devS* transcription that is hypoxia responsive, the devT transcription levels were found not to be raised during hypoxia (Saini *et al*., 2004a). Malhotra *et al*. (2004) supported the principle that the function of the DevR-DevS TCS is linked to dormancy as it plays a less significant role in the early stages of the *M. tb* infection. The exact role DevT plays in *M. tb* infection remains unresolved (Malhorta *et al*., 2004; Saini *et al*., 2004b). It is not impossible to assume that DevT can also bind and phosphorylate DevR under the same or different environmental signals. Both DevT and DevS are heme-based O_2 sensors with K_d values for oxygen of 26 µM and 3 µM respectively (Sousa *et al*., 2007). This begs the question - why there are two kinase sensors for a single response regulator on the *M. tb* genome? Redundancy or a difference in physiological states could be a plausible explanation. However, the possibility exists that the difference in the affinity and kinetics of the sensors may be critical for an oxygen dose-dependent response needed under different conditions (Sousa *et al*., 2007). *M. tb* bacilli housing gene deletions for *devR* and *devS* are incapable of initiating a response to hypoxic conditions, making the DevR-DevS system an appealing drug target for persistent infection (Voskuil *et al*., 2003; Malhorta *et al*., 2004). The inhibition of proteins involved in signal transduction has an added downstream effect since this will interfere with the functionality of the complete regulatory network controlled by the response regulator. One inhibitor will therefore have a pleiotropic effect (Gupta *et al*., 2009). Another reason that the Dev TCS may be a good drug target is the low homology the Dev proteins have to other enzymes, as well as their absence in invertebrates and vertebrates - hence, inhibition will not affect cellular function of its eukaryotic host (Dasgupta *et al*., 2000).

1.9 Other *Mycobacterium* dormancy response regulators

DevR regulates universal stress protein expression and adaption during oxygen starvation in *M. smegmatis* (O'Toole *et al*., 2003). The *M. smegmatis* DevR exhibits a high sequence similarity and gene linkage to the response regulator found in *M. tb*. *M. smegmatis* DevR also supports adaptation and resistance to environmental stresses caused by oxygen starvation (O'Toole *et al*., 2003). To overcome experimental difficulties associated with *M. tb*, *M. bovis* BCG was used as a model organism to study whether DevR plays a role in the adaptation to hypoxic conditions (Boon and Dick, 2002). The DevS and DevR are co-transcribed in both *M. tb* and *M. bovis* BCG and the genes share a 99.8% sequence similarity (Cole *et al*., 1998; Dasgupta *et al*., 2000). The deletion of the DevR coding sequence resulted in the absence of other dormancy-induced proteins like alpha-crystallin, universal stress protein domain-containing Rv2623 and cystathionine β-synthase domain-containing Rv2626c. In contrast, the DevS HK deletion had a moderate effect on the response of the bacilli to the hypoxic conditions (Boon and Dick, 2002).

The design of inhibitors to the DevS-DevR TCS as part of a drug design initiative relies on an understanding of the molecular interactions and component binding of this system

1.10 Hypothesis

We hypothesize that an investigation into both the protein-protein and the protein-DNA interactions of the Dev TCS, can improve our understanding of the detailed mechanism associated with *M. tb* latency and provide a good drug target for persistent tuberculosis.

1.11 Research objectives

1. To present a detailed *in silico* analysis of the amino acid residues that play a role in the binding interactions of DevR and DevS. The *in silico* study will then be used to shed some light on the 3D structures of the components that have not yet been crystallized.
2. To functionally express and purify all the components of the *M. tb* Dev two-component system.
3. To analyze the binding interactions between DevS and DevR in order to use these interactions as drug targets and consequently develop inhibitor screening methodologies.
4. To identify the amino acids involved in the binding interaction between DevR and respective DNA fragments, to be used in future rational drug design strategies.

Chapter 2 Homology modelling and *In silico* binding analysis of the Dev two-component system

2.1 Introduction

2.1.1 Molecular modelling

The most common method of *in silico* protein modelling is knowledge-based modelling or homology modelling. This entails building a model based on known structures of proteins homologous to the target protein. These homologous template proteins are usually the same protein from a different organism or a protein that has the same function. The topology or fold of these proteins plays a vital part in structure prediction. Ultimately, the most important step in homology modelling is the amino acid sequence alignment between the target protein and the template proteins. This step is based in the assumption that the target protein and the template proteins have diverged from a common ancestor protein and will therefore have a similar 3D structure (Sander and Schneider, 1991). In other words, proteins with similar sequences will have similar structures (Marti-Renom *et al*., 2000).

There are various algorithms that can be used for homology modelling; the two best known algorithms are SWISS MODEL (Schwede *et al.,* 2003) and MODELLER (Sali and Blundell, 1993). The MODELLER program was used in this study. MODELLER uses a method called spatial restraint satisfaction modelling, during which various restraints can be used to guide model building. These restraints include bond angles, dihedral angles and stereochemistry. The MODELLER algorithm assumes that the aligned residues will have similar restraints and uses these restraints to build the model (Sali *et al*., 1993). Loop modelling is one of the most difficult tasks in homology modelling. A number of enzymes only differ by the length and

composition of their loops and consequently the loops are linked to the specificity of an enzyme (Marti-Renom *et al*., 2000).

Homology modelling has certain limitations. Choosing valid template proteins and assigning a fold to the target proteins is very important because errors occur in the models when the template model shows insertions or deletions with respect to the modelled proteins. Misalignment during the sequence alignment of proteins with low sequence homology leads to errors in the side chain packing and interaction analysis.

Model assessment is important when choosing the more correct form of the protein. Doing a fold comparison between the target model and the template model is one of the first steps. It is also important to check the basic stereochemistry of the model with a program like PROCHECK (Laskowski *et al*., 1993). Furthermore, molecular mechanics in the form of energy minimization can assist to discriminate between correctly and incorrectly folded proteins, for example the CHARMM force field for minimization.

Another limitation that homology models have is the resolution. Unlike crystal structures, the resolution cannot always go down to atom level; homology higher than 70% is needed for that to be possible.

2.1.2 Structural information of histidine kinases

Sensor histidine kinases (HKs) are strictly defined as those enzymes that catalyze the transfer of the terminal phosphate group from ATP to a conserved histidine on the HK. This kinase must have an associate response regulator to which this phosphate group can be transferred, in order to catalyze another reaction, usually involving DNA binding (Kim and Forst, 2001).

Literature shows that there are highly conserved sequence domains in kinase proteins, especially in the kinase domain (catalytic core) consisting of conserved H, N, G1 and G2 boxes, and a more variable F-box (Stock *et al*., 2000). Grebe and Stock (1999) were the first authors to suggest the idea that there are different subclasses of HKs that have evolved independently of each other. They used extensive sequence database searches to identify a HK dataset for their study. Multiple sequence alignment using the sequence homology boxes, found on the C-terminal of the HKs, were used to classify these kinases into distinct

subfamilies (see Figure 1.3). Kim and Forst (2001) used a BLAST analysis of the genomic sequences of the kinase domains of EnvZ, CheA, NarZ, YehU and DcuS to create a HK dataset for analysis (see Figure 1.4). A phylogenetic analysis using the unweighted pair-group method with arithmetic means algorithm (UPGMA) was used to classify the HK dataset into subfamilies.

The HK superfamily has been classified by various criteria, but none have been done using the three-dimensional structural data available. In a broader sense this study aimed to close the gap between sequence and structure for HKs. Therefore a summary of those two-component systems with structural data on both the sensor kinase and the response regulator in the Brookhaven Protein Databank (PDB) was compiled and is shown in Table 2.1.

More specifically the interaction analysis between DevS and DevR is important for this study. The number of available protein structures is far less than the number of known proteins. This shows that there are many more protein structures still to be elucidated. Computational modelling is used to fill the gap. The aim of this study was to perform *in silico* analysis of the interactions between homology-modelled DevS and DevR, in order to obtain models that may be used with confidence to guide initial experiments to probe the binding interactions of the Dev TCS. It is hoped that this knowledge may contribute to the identification of targets against latent tuberculosis.

Table 2.1: Two-component systems in bacteria with structural data available in the PDB database for both the histidine kinase and the corresponding response regulator

Kinase Sensor	**Response Regulator**	**Output Function**
ArcB	ArcA	*E. coli* ArcB phosphorylates ArcA in response to anaerobic conditions. A phosphoryl group is transferred from His292 in the kinase domain to Asp576 in the receiver domain and then His717 in the Hpt domain of ArcB. Asp54 of ArcA receives the phosphoryl group from the Hpt-domain (Mika en Hengge, 2005). (PDB ArcB: 1FR0, 1A0B, 2A0B) (PDB ArcA: 1XHE, 1XHF)
CheA	CheY CheB	Phosphorylated CheA phosphorylates response regulator CheY or CheB and phosphorylated CheY interacts with the flagellar motor switch complex to induce rotation of the flagella, resulting in smooth swimming behaviour. Clockwise rotation of the flagella takes place in the absence of the CheY-P interaction (Dons *et al.,* 2004). (PDB CheA: 1B3Q,1UOS,1I5N)(PDB CheY 3F7N, 3FFT)(PDB CheB 1A2O, 1CHD)
EnvZ	OmpR	EnvZ and OmpR are responsible for osmoregulation (Buckler *et al*., 2000). (PDB EnvZ 1BXD, 1JOY) (PDB OmpR 2JPB)
NarK	NarL	NarL was classified as a nitrate response regulator involved in the control of anaerobic metabolism (Schnell *et al*., 2008). (PDB NarK 1EZI)(PDB NarL 1ILH, 3EUL, 1JE8)
NtrB	NtrC	*E. coli* NtrB (NRII) phosphorylates the transcription factor nitrogen response regulator I (NtrC, NRI) for the control of nitrogen-regulated Ntr genes. (Ninfa et al., 1993) (PDB NtrB: 1RT2) (PDB NtrC: 1NTR).
LuxP LuxQ	LuxO LuxR	*Vibrio harveyi*'s autoinducer-1 (A-1) binds to the periplasmic receptor LuxP LuxQ then phosphorylates LuxO which indirectly represses the transcriptional activator LuxR. (Sun *et al.,* 2004) (PDB LuxP/LuxQ complex: 2HJ9, 1ZHH) (PDB LuxQ: 2HJE)
PhoQ	PhoP	*Salmonella typhimurium* PhoQ phosphorylates the transcription response regulator PhoP in response to stimuli upon phagocytosis, thereby promoting bacterial virulence. (Vescovi *et al.,* 1997; Groisman, 2001) (PDB PhoQ: 1ID0, 2C2A, 1YAX) (PDB PhoP: 1MVO, 2EU6, 2EUB).
PrrB	PrrA	*M. tuberculosis* two-component system of PrrB and PrrA has been shown to be induced upon macrophage phagocytosis and to be transiently required for the early stages of macrophage infection (Ewann *et al*., 2004).(PDB PrrA: 1YS6, 1YS7) (PDB PrrB: 1YSR, 1YS3)
FixL	FixJ	*Rhizobium* FixL detects oxygen tension. Oxygen-free (deoxy-) FixL transfers a phosphoryl group to FixJ at low-oxygen tension whereas the oxygen-bound form is inactivated. FixJ is a transcriptional factor (Wang *et al*., 2006). (PDB FixL: 1D06, 1DRM) (PDB FixL: 1X3U, 1DCM, 1DSW, 1DBW)
KinA KinB	Spo0F	*B. subtilis* 2 sensor histidine kinases, KinA and KinB autophosphorylate in response to nutrient deprivation or high cell density, and phosphorylate the first of two response regulators Spo0F. Spo0B, the second histidine kinase and phosphotransferase, receives a phosphate group from Spo0F and transfers it to the response regulator Spo0A (Min *et al.,* 1993; Bijlsma and Groisman, 2003)..(PDB Spo0F: 1FSP, 1PUX) (PDB Spo0F/Spo0B complex: 1F51, 1FTK) (PDB Spo0B: 1IXM)
Spo0B His30	Spo0A	

2.2 Methods

2.2.1 Identification of all structural data on two-component histidine kinases

It is believed that structure is more conserved than sequence (Illergard *et al*., 2009). The structural relationships within the previously identified HK groups were investigated using Discovery Studio (DS) software, to explore the relationship between structural motifs and the homology boxes found in HK sequences.

Extracting 3D structures from the PDB database for use in protein BLAST searches resulted in thousands of hits. The E-value obtained in these results describes the number of hits one can expect to observe by chance when searching a database. Hence the lower the E-value, or the closer it is to zero, the more significant the match. The E-value can also be used as a convenient way to create a significance threshold for reporting results. Most HK have similarity in their conserved motifs (boxes) but overall similarity is low. The E-value could therefore not be used to differentiate between true HKs and other proteins. Consequently, the protein sequences were extracted by using the Enzyme Classification (EC) numbering (http://au.expasy.org/enzyme/) of HKs and this was supplemented by using the known Gene Ontology (GO, http://www.geneontology.org/) terms for sensor HKs. The CATH (Class, Architecture, Topology and Homologous superfamily, http://www.cathdb.info) and SCOP (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/) databases were also used to identify close relatives in the HK family. The sequences found during these searches were then used as seed sequences for protein BLAST against the PDB database, to ensure all structures were searched exhaustively.

2.2.2 Structural modelling of full length and truncated DevS

The basic methodology used for homology modelling can be seen in Figure 2.1. It is a summary of the methods and results used in the *in silico* study of the Dev two-component system where the green boxes are exclusively for truncated DevS201, blue boxes are exclusively for full-length DevS and gray boxes include both. The construction and reason

for preparation of DevS201 will be explained in Chapter 3. It is linked to the successful *in vitro* expression and purification of the components of the Dev TCS.

2.2.2.1 Identification of homologous proteins and multiple sequence alignment

The two major requirements for building a homology model are, firstly, a known protein sequence of the model to be built, and secondly at least one high-resolution structure of a related protein that can act as a template. The first step here was to find homologues to both the full length and truncated DevS (DevS201) by using BLAST algorithm as provided by PDB website (Bernstein *et al*., 1977; Altschul *et al*., 1990). For modelling purposes, a multiple alignment of the homologous proteins was constructed. This was done to aid the identification of the conserved catalytic core present in familial HKs. Alignment was carried out with DS Align123, a multiple sequence alignment method based on the CLUSTALW program using the GONNET scoring matrix in DS (Thompson *et al*., 1994). To optimize the quality of the alignment, the gap opening and gap extension penalties were set to 15 and 0.31 respectively. Furthermore, the alignment was manually adjusted to align the sequence motifs, as described by Grebe and Stock (1999). The delay divergent sequences property was set to 20% and the secondary structure predictions were used. Since very little structural information regarding the DevS protein is available, some time was spent to gain as much sequence information as possible to assist in the process of modelling the protein structure of DevS. The sequence was subjected to various secondary structure prediction algorithms, listed in Table 2.2. Secondary structure prediction in DS was done by using the Discrimination of protein Secondary structure Class (DSC) method (King and Sternberg, 1996).

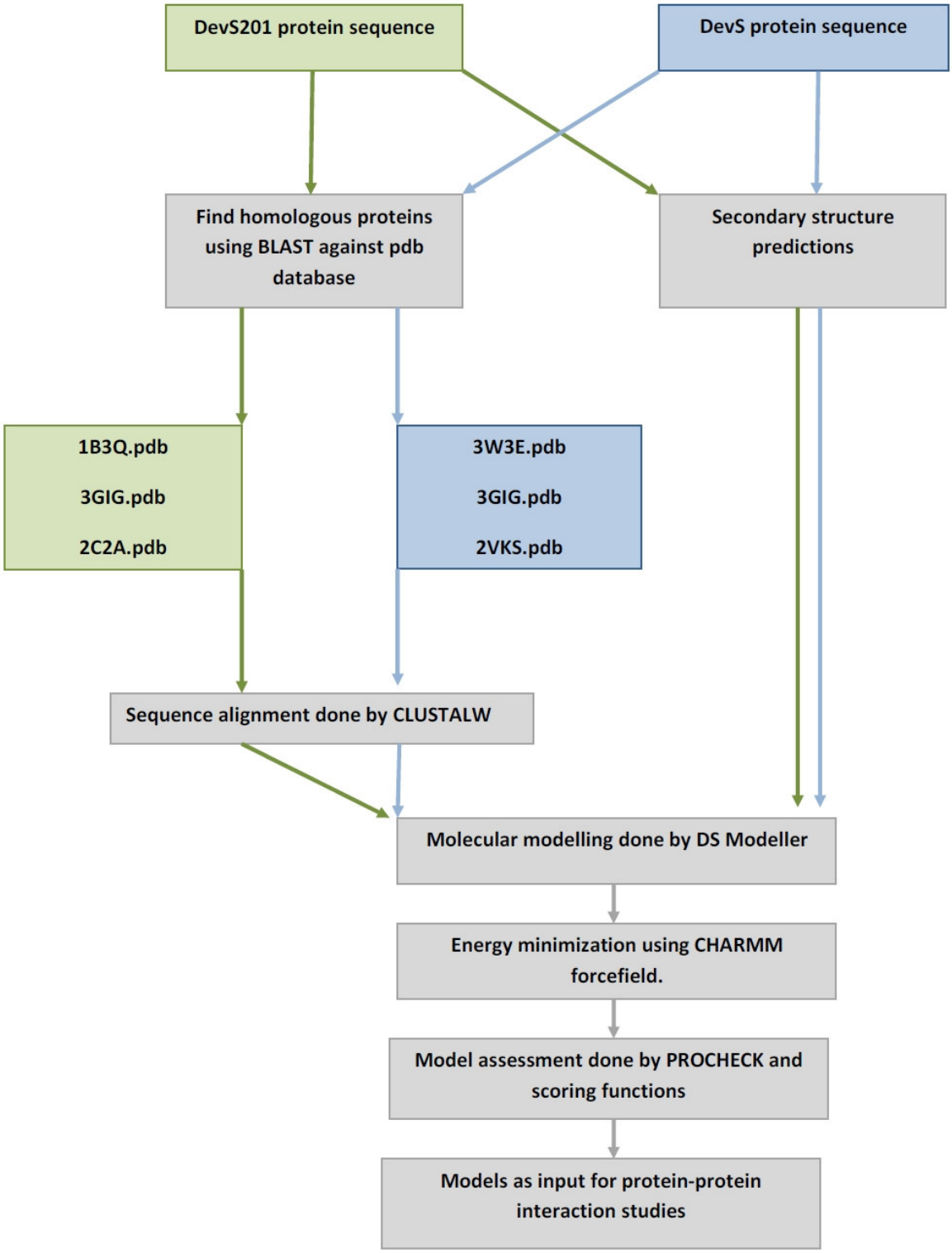


Figure 2.1: Summary of the methods used in the *in silico* study of the Dev two-component system. Green boxes are exclusive for DevS201, blue boxes are exclusive for DevS and gray boxes include both.

Table 2.2: Secondary structure prediction algorithms utilized in this study.

Program	Source
GARNIER	EMBOSS
JNEPRED JNETHMM JNETALIGN JNPSSM	http://www.compbio.dundee.ac.uk
JNETFREQ JPRED	http://www.compbio.dundee.ac.uk
PHD NNPREDICT HNNC SOPM SIMPA96 PREDATOR DSC Sec.Cons GOR4	http://npsa-pbil.ibcp.fr
HMMSTER	http://www.bioinfo.rpi.edu
PHD PSIPRED	http:// cubic.bioc.columbia.edu

2.2.2.2 Homology modelling

For DevS, three distinct structural templates identified in the previous section were used in succession to span the complete length of the DevS protein, to provide a larger set of template-target residue pairs. These structures were 2W3E, 3GIG and 2VKS. Three separate templates for the DevS201 homology model were also identified, namely 1B3Q, 3GIG and 2C2A. These templates were used and the models were built with DS MODELER using high refinement to generate 10 different conformer models (Sali and Blundell, 1993; Sali *et al*., 1995).

2.2.2.3 Energy minimization

Models generated for DevS and DevS201 were typed by the CHARMM forcefield before being energy-minimized. Energy minimization was used to find a local energy minimum for each molecule in the model; a total lower energy state indicates a more stable model (Brooks *et al*., 1983). The Steepest Descent algorithm for minimization was applied, followed by the Conjugate Gradient method for faster convergence towards a local minimum (Fletcher, 1969). The minimization was done for 2,000 steps with a root mean square (RMS) gradient of 0.1 Å using the Generalized Born implicit solvent model with a 14 Å non-bonded cut-off value, in addition to the DS's spherical cut-off parameters for electrostatic bond definition (Dominy and Brooks, 1999).

2.2.2.4 Model assessment after minimization

From these conformations, one representative model for each template was chosen that had a minimum number of residues in disallowed regions of a Ramachandran plot. The DOPE scores were also considered. To assess the quality of models after minimization, additional scoring functions were used and Ramachandran plots were obtained using DS MODELER. Protein structure and model assessment tools from the SWISSMODEL workspace were deployed, as well as the PROCHECK suite of programs (Laskowski *et al*., 1993; Arnold *et al*., 2006). PROCHECK assesses the stereochemical quality of a protein model and the output is displayed via a Ramachandran plot. For full length DevS, loop modelling refinement by DS MODELER was used to define one or more residues in the model that have no corresponding template residues, consequently found in the disallowed regions of the Ramachandran plots (Fiser *et al*., 2000).

2.2.3 Protein-protein interactions

Once suitable structures representing DevS and DevS201 were built, these structures were used to model and study protein-protein interactions between DevS and DevR. Two methods were used: homology docking and DS ZDOCK.

2.2.3.1 Protein homology docking

To produce a starting point for predicting protein-protein interactions, homology docking was used. This entailed modelling DevS and DevR onto the crystal structures of two other component histidine kinases. Two crystal structures were used where the response regulator and the HK are present in association. Since sequence homology between DevS and the template kinases was low, the structural similarities and topologies were used to superimpose the structures. The structural similarities of the template structures were quantified by secondary structure topologies present in the crystal structures. The topologies for DevS were

based on the homology models as well secondary structure predictions based on its sequence, using the DS DSC method (King and Sternberg, 1996).

Minimization with DS was used to arrive at a more accurate model of association. Three different minimization methods were used during this minimization step and the energy change was monitored closely. First, an Adopted Basis-set Newton-Raphson (ABNR) was used for 2,000 steps, (RMS gradient set to 0.1 kcal⁄mol/Å), followed by Steepest Decent for 4,000 steps, (RMS gradient 0.01 kcal/mol/Å), followed by Conjugated Gradient for 6,000 steps, (RMS gradient 0.001 kcal/mol/Å) and finally Steepest Decent 10,000 steps, (RMS gradient 0.001 kcal/mol/Å). The model with the lowest energy change over time was selected as the most accurate and the interaction sites were analyzed and compared to those found in the DS ZDOCK results.

2.2.3.2 Protein-protein docking using DS ZDOCK

DevS201 and DevR were typed by the CHARMM forcefield before docking commenced (Brooks et al., 1983). DS ZDOCK was run on default values with the distance interaction cut-off set to 10 Å and the angular step size at 6. The docked poses with a RMSD of less than 10 Å overall were clustered together into not more than 100 clusters (Comeau *et al*., 2004). DS ZRANK was used to rank the top 2,000 poses for further analysis (Pierce and Weng, 2007). The clusters were colour-coded based on their DS ZRANK score, red being indicative of a high score and blue indicative of a low score. All clusters with low scores were removed. Consequently, DS RDOCK was used for refinement of the top poses; using CHARMM with a dielectric constant of 4, the molecules were typed with CHARMM polar H to increase the speed of the scoring function.

2.3 Results

In the (PDB) there are currently 68,701 protein structures available, 89 of which are classified as HKs based on their EC number. That number drops to 60 when structures with more than 90% homology are removed, eliminating molecules that were crystallized more than once. SWISSPROT, a curated protein sequence database contains 47,734 HK sequences. When all redundancy was removed, a total of 5,573 unique HK sequences were available. These statistics emphasize the significant gap between sequence and structural data available for the HKs.

2.3.1 Identification of all structural data on two-component histidine kinases

For each PDB entry, the protein structures were compared with their respective protein sequences from the NCBI database and it was concluded that most of the structures were partial structures. Only 2 full length HK structures were available, CheA (*T. maritima*, 1UOS) and PhoR (*T. maritima*, 2C2A). The structural data of the kinases found in the PDB database was mined to see if there was a relationship between the structures and the functions of these sensor kinases. Table 2.3 shows the different classes that the histidine kinase sensors could be divided into, based on structure.

2.3.2 Classification of DevS

Kim and Forst (2001), took a phylogenetic approach to analyse the HK family in bacteria and archae. The HK groups were defined based on a phylogenetic analysis of the protein sequences. They concluded that all the HKs could be grouped in 5 different groups. These five common transmitter domains were used to BLAST against the DevS ORF to try and classify DevS into one of the five defined groups. It was found that DevS belongs to Class 3, along with only three other known (*E.coli*) HKs: NarX, NarQ and UhpE. When using the Grebe and Stock (1999), scheme it was found that DevS belongs to Group 7 (HPK7), which contains DegS and NarQ. The H-box in this group is distinguished by the presence of a negatively charged group two residues upstream from the conserved histidine and a positively

charged residue, usually an arginine, eight residues upstream. This places these residues adjacent to one another on one face of the presumed H-box α-helix. In addition, the phospho-accepting histidine is followed by an aspartate residue in all cases. The proline and phenylalanine that tend to be conserved in most kinase H-boxes are generally missing in HPK7 kinases. The N-box is also distinct: the second asparagine is conserved but in place of the first asparagine there is a glutamate. In addition, the F-box is absent and the distance between the D- and the G-boxes is reduced. This structural classification of DevS was needed to identify its closest relatives for homology modelling.

Table 2.3: The histidine kinase sensors can be divided into different classes based on structure

Class	**Sensor Kinase**	**Function**	**PDB accession number**
1	DctB	C_4 dicarboxylate sensor	3BY9
	DcuS	citrate sensor	3BY8
2	LuxQ	cell density sensor	2HJE
	CitA	citrate sensor	2V9A
3	AppS	function unknown	3CWF
	PhoQ	related to virulence	1YAX
4	CheA	flagellar motion	1UOS
5	EnvZ	osmoregulation	1BXD
	NtrB	nitrogen sensor	1R62
6	ArcB	aerobic sensor	1AOB
	YojN	involved with polysaccharide synthesis	1SR2
7	CheA	flagellar motion	2C2A
	PhoR	redox stated and Mg^{2+}	1B3Q
	PrrB	*M.tb* virulence	1YSR
8	FixL	oxygen sensor	1DO6, 1DRM
	NifL	cellular redox states	2GJ3
	KinA	sporulation	2VLG
	PprS	function unknown	1MZU

2.3.3 Structural modelling of DevS and DevS201

2.3.3.1 Identification of homologous proteins and multiple sequence alignment

The classification scheme provided by SCOP and ASTRAL may be used as a guideline for structure selection (Brenner *et al*., 2000; Lo Conte *et al*., 2000). This provided some useful insight where the homology modelling was concerned. For DevS201 (truncated version of DevS), three structurally different templates were found during the BLAST search, as seen in Figure 2.2(A). DevS201 showed 49.3, 39.0 and 39.4% sequence homology to 3GIG (*B. subtilis* DesK), 1B3Q (*T. maritima* CheA) and 2C2A (*T. maritime* PhoQ), respectively. Three separate models were built based on the three templates and the predicted protein structures were critically assessed to determine the more accurate model. No template could be found that spanned the complete length of the DevS sequence, so three different sequences had to be used as templates for the various parts of the full length sequence, as can be seen in Figure 2.2(B). From position 0 to 148, DevS shows 83.8% homology to 2W3E (*M. tb* DevS fragment), 15.7% homology to 2VKS (*M. smegmatis* DevS fragment) in position 173 to 322, and the C-terminal kinase core from position 305 to 525 has 19.4% homology to 3GIG. These three template structures were used simultaneously to build a model of full length DevS.

2.3.3.2 Homology modelling and model assessment

DS Modeller produced 10 different conformer models for each template used for DevS and DevS201. Energy minimization was used to find a local energy minimum for each molecule in the model; a total lower energy state indicates a more stable model. Homology model quality assessment is usually performed using two methods: statistical potentials or physics-based energy calculations. Both methods produce an estimate of the energy or an energy-like value that can be used to determine the structural accuracy of the model. Various scoring functions were used to identify the best model for DevS and DevS201. DS uses Probability

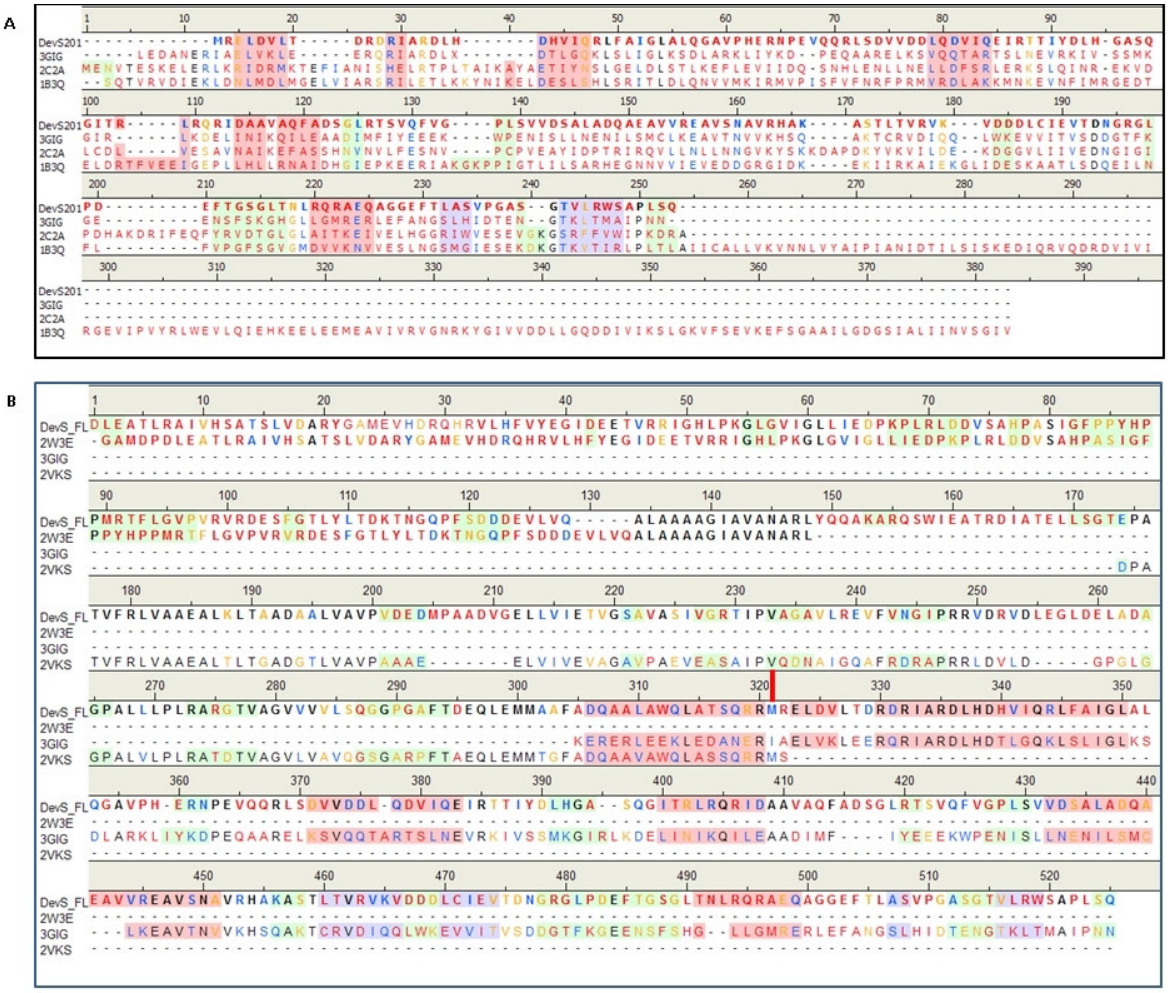


Figure 2.2: Sequence alignment between A) DevS201 and B) DevS and their closest homologues. Residue colours: Black is high sequence similarity, blue is strong, yellow is weak and red indicates no sequence similarity, respectively. Background colours indicate the predicted secondary structure, helices are red, sheets in blue and coils are green. The residues without a coloured background form parts of loops or there are mismatches in the prediction. The red line on the ruler in (B) indicated where the DevS201 protein will start, *i.e.* where DevS was truncated to form DevS201.

Density Functions (PDF) to calculate pseudo-energy terms and stereochemical pseudo-energy terms. DOPE is a knowledge-based scoring function that is often termed statistical potential scoring function. The DOPE scoring function is an atomic based statistical potential measures the relative stability of a conformation. The verification score for each amino acid residue is based on the secondary structure and indicates whether a residue is in the desired 3D environment. QMEAN is a composite scoring function consisting of 5 terms: torsion angle energy, distance-dependent pairwise residue-level potential used to assess long-range interactions, solvation energy, and agreement between predicted and observed secondary structure and agreement between predicted and observed solvent accessibility. Results of the

scoring functions were used to select the best conformers. The results for the best conformers for DevS201 are outlined in Table 2.4. Ramachandran plots were also used to select the best conformers (Figure 2.3), as well as monitoring the conformational changes due to energy minimization. Ramachandran plots were used to compare the residues in favourable, as well as disallowed, regions of each model.

Table 2.4: Results of the best conformers of the various scoring functions used during the model building of DevS201

Model	DOPE score	PDF total energy	PDF physical energy	Residues in favourable regions	Residues in disallowed regions	Verification score based on secondary structure	DFire score	QMean
1B3Q model 4	-12328.79	1433.53	578.95	50.3%	2.8%	7.04	-153.63	-153.63
2C2A model 8	-15715.77	1198.52	484.52	55.9%	1.1% %	43.31	-172.43	-33.585
3GIG model 9	-17141.98	907.1	444.44	70.4%	0%	39.5	-192.53	-47.8608
Full length DevS	-41063.00	1515.97	3528.71	75.9%	1.1%	64.61	-523.55	-43.0448

Based on the Ramachandran plots (Figure 2.3) and results from scoring functions (Table 2.4), the best DevS201 conformer from each template was chosen and can be seen in Figure 2.4 The 3GIG model was chosen as the best model for DevS201 based on the energy values predicted by the scoring functions mentioned in Table 2.4.

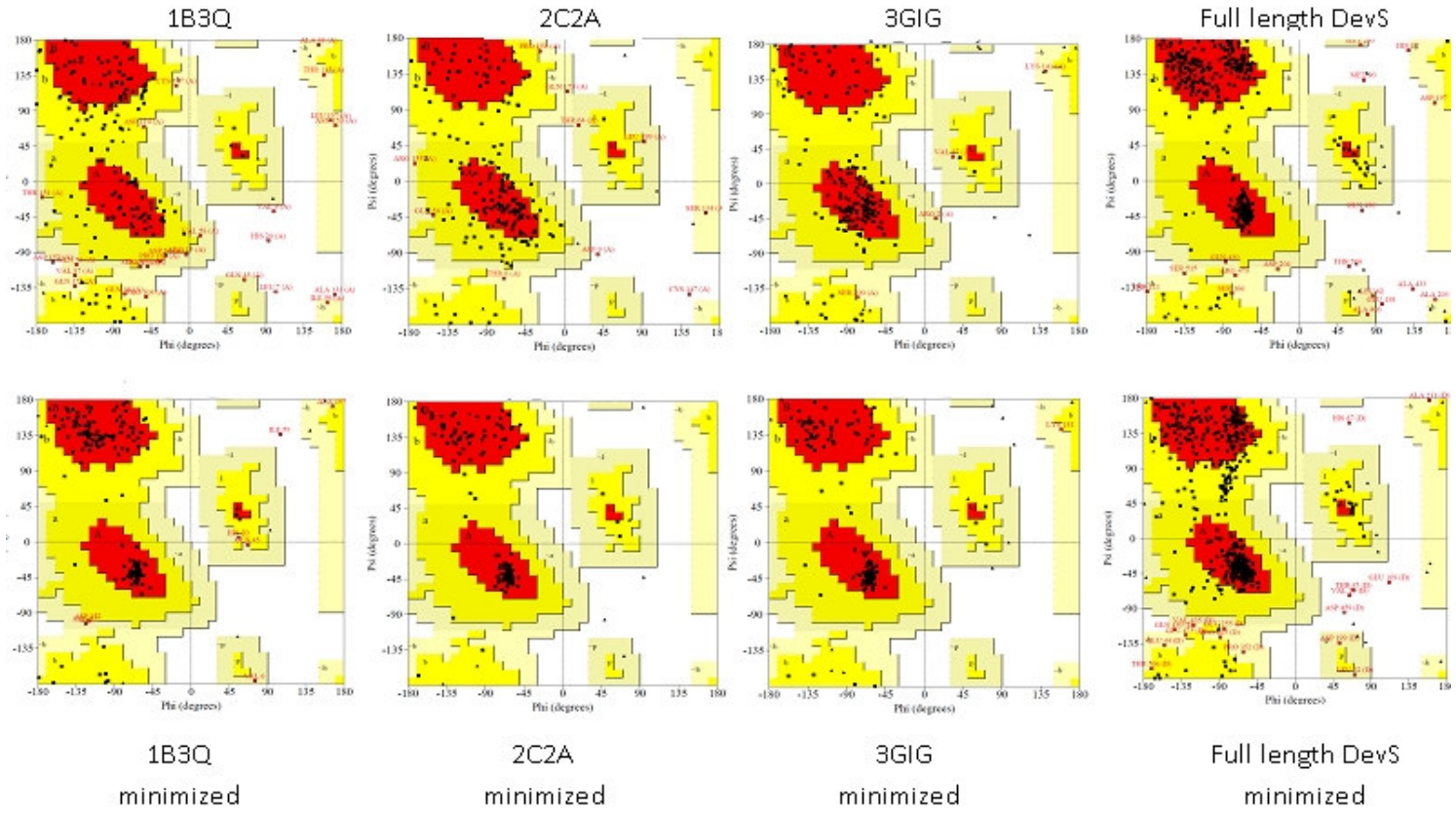


Figure 2.3: Ramachandran plots for the DevS201 models based on three templates and the full length DevS. The top row is the Ramachandran plots as built by DS MODELER using 1B3Q, 2C2A, 3GIG and the full length DevS. The bottom row indicates the Ramachandran plots after minimization runs. The Ramachandran plots were produced by PROCHECK. Most sterically favoured regions (red), additional allowed regions (dark yellow), generously allowed regions (light yellow) and disallowed regions (white). The x-axis shows the variance in the Phi angle and the y-axis the Psi angle. After energy minimization the number of amino acids in the disallowed regions decreased drastically for all DevS201 models. Comparison of the eight amino acids present in the disallowed regions of full length DevS before with the six amino acids present after energy minimization, indicated that the loops needs more optimization.

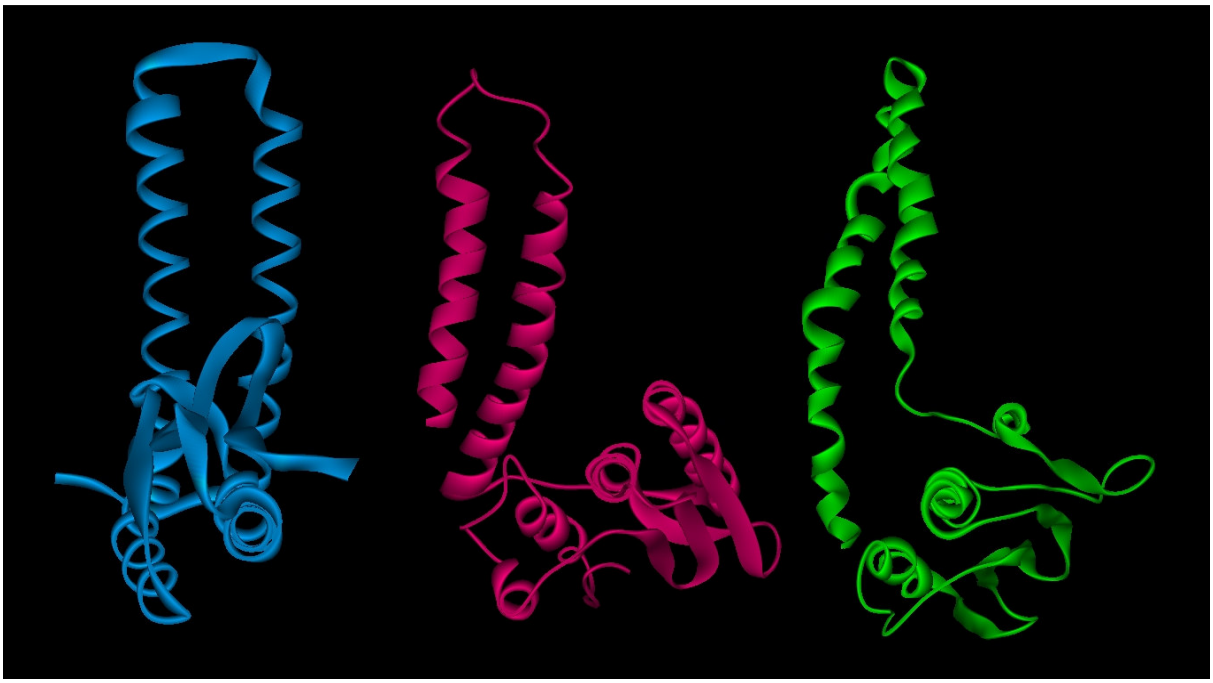


Figure 2.4: DevS201 models build by DS MODELER based on three different templates. The model on the left was built based on 1B3Q (blue), middle based on 2C2A (pink) and on the right based on 3GIG (green). The active residues are not shown in the models since the calculations do not impinge on them.

For DevS there was only a single model and the structure in Figure 2.5 (A) was produced. The comparison of the amino acids present in the disallowed regions of full length DevS before and after energy minimization (Figure 2.3), indicated that the loops needed more optimization. Loops can be defined as those pieces of protein where homology was low or no sequence was present. For DevS, there were a few stretches of protein that had no homology or very low homology, so loop refinement from DS MODELER was done. In Figure 2.5 (A), DevS is shown with the four loops identified as Asp121-Leu130, Arg142-Pro170, Tyr382-Ser388 and Phe405-Pro419. Figure 2.5(B) displays the loops after refinement in purple. Each LOOP REFINEMENT run produced ten different conformations of each loop. The best models were chosen based on the DOPE scores available for the loops. The only exception was the longest loop (Arg142-Pro170), where the DOPE score for the original conformation matched the predicted DOPE score of the refined conformation. No clear conclusion on the more correct orientation could therefore be reached. Figure 2.5(C-F) shows the original loops in green and the modelled loops in purple. When the DevS201 model and the DevS model,

with loop refinement, was superimposed a RMSD of 7.4 Å was observed. Figure 2.6 shows the similarities in the two structures.

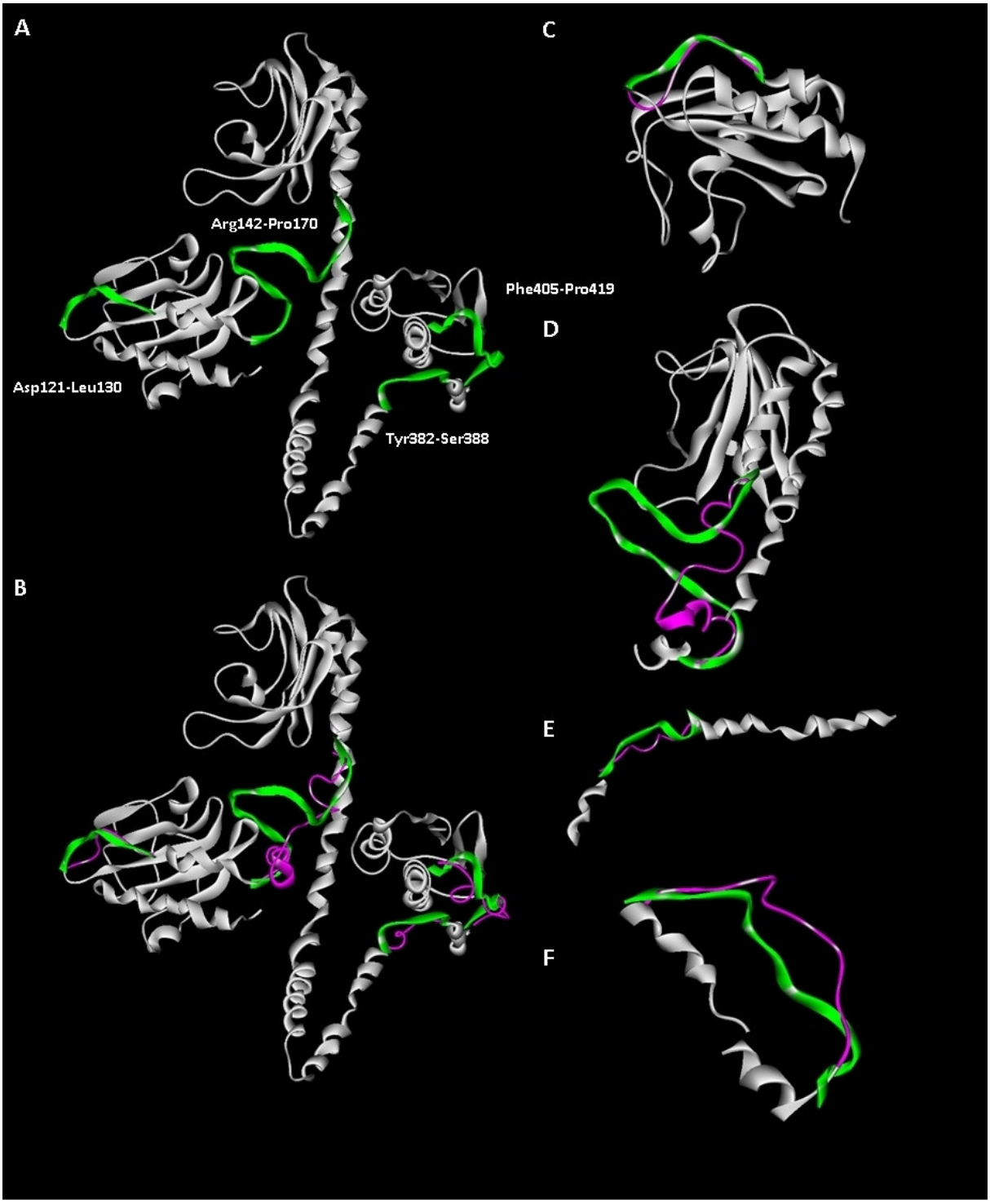


Figure 2.5: DevS full length model with loop modelling. (A) DevS model as produced by DS MODELER with four loops identified: Asp121-Leu130, Arg142-Pro170, Tyr382-Ser388 and Phe405-Pro419 (all showed in green). (B) The loops after loop refinement are shown in purple. (C) Closer look at loop Asp121-Leu130. (D) Loop Arg142-Pro170. (E) Loop Tyr382-Ser388. (F) Loop Phe405-Pro419.

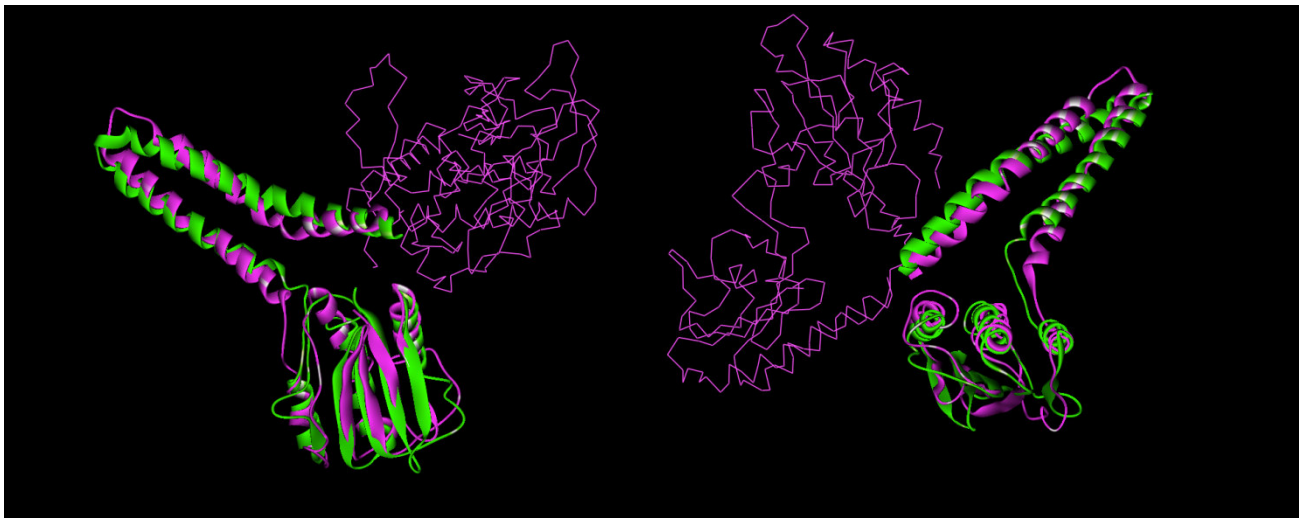


Figure 2.6: The DevS201 and DevS models superimposed. DevS201 model (in green) and the DevS (in dark pink) model superimposed to show similarities. The DevS transmembrane domain is not depicted as secondary structures.

2.3.4 Protein-protein interactions

Due to the limited nature of the structural data available for DevS and DevR, attempting *in silico* protein-protein docking without any prior knowledge would be impossible. Consequently, protein-protein docking was done based on structure similarity of known crystallized two-component kinases systems. Very little structural or biochemical information was available that could guide the docking. The ability of scoring functions and minimization techniques was used extensively to guide the study to reach a near-native structure.

2.3.4.1 Protein homology docking

During the similarity search, two reference structures were found in the PDB database of HKs that were crystallized while interacting with their corresponding response regulators. 3A0R contains *T. maritima* histidine kinase (HK) ThkA in complex with response regulator (RR) TrrA, while 3DGE contains *T. maritima* HK853 in complex with RR468. Table 2.5 contains the sequence similarity data between these crystal structures and the DevS and DevR

model structures. HK853 shows a high homology towards DevS, but it was only crystallized after the DevS201 models were built. A comparison between the DevS201 model and the HK853 showed a RMSD of 12.85 Å, which indicates the DevS201 is a good HK representative model.

Table 2.5 Information on sequence similarity between DevS and DevR and the respective templates

PDB accession number	**Histidine kinase (amino acid length)**	**Percentage similarity to DevS**	**Response regulator (amino acid length)**	**Percentage similarity to DevR**	**Organism**	**Literature Reference**
3AOR	ThkA (343)	21.7%	TrrA (115)	47.5%	*T. maritima*	Yamada *et al.*, 2006
1DGE	HK853 (237)	36.3%	RR468 (122)	46.0%	*T. maritima*	Casino *et al.*, 2009

The protein-protein interactions between DevS and DevR were modelled, based on these known structures of interacting two-component systems. The topology comparisons for the HKs are shown in Figure 2.7. A αβαββα motif was found in the DevS HK models, as in the ThkA and HK853 crystal structures.

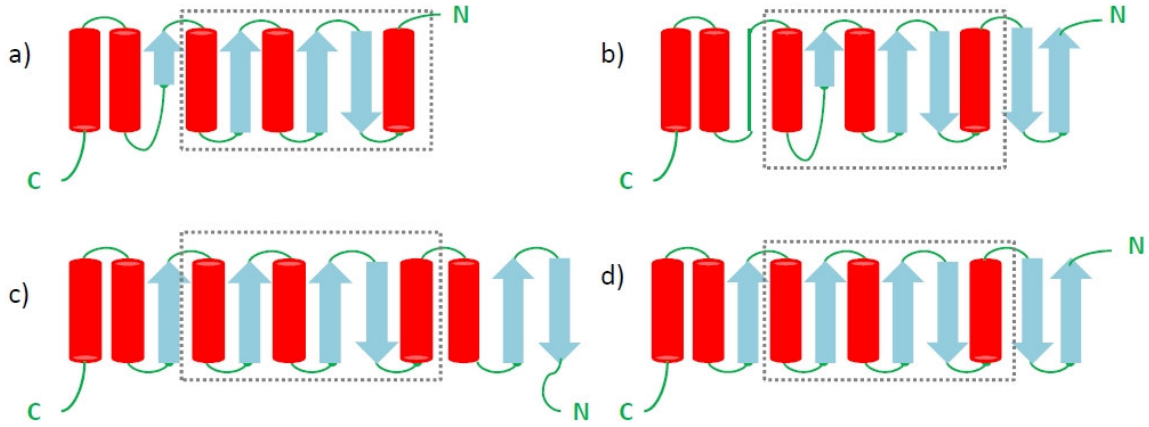


Figure 2.7: Topology comparisons between DevS201 and other known histidine kinases. Topologies for DevS201 based on template 2C2A (a), template 3GIG (b), ThkA from *T. maritima* (c) and HK853 from *T. maritima* (d). The red tube represents an σ-helix, the blue arrow represents a β-sheet and the loops are coloured green. The black blocks indicate the σβσββσ motif.

Optimal similarity between DevS and the template kinase was found when the topologies of the catalytic cores were compared. The catalytic core is more conserved than the sensor domain. Only the DevS201 models, containing the catalytic core, were therefore used to investigate protein-protein interactions. The assumption that the DevS201 can be used to model comprehensive protein-protein interactions was further investigated in Chapter 3.

Wisedchaisri *et al*. (2008) reported crystal structures of full-length unphosphorylated DevR at a 2.2 Å resolution. This structure was used in this investigation. When analyzing the topologies for the response regulators (Figure 2.8), an $(\alpha\beta)_5$ repetitive motif was found. The DevR structure does not have the β_9-sheet; however a loop was predicted for the corresponding amino acids. Benda *et al*. (2004) defined this distinctive fold of response regulators, where the 5β sheets are sandwiched between the 2α and 3α helices (Figure 2.8).

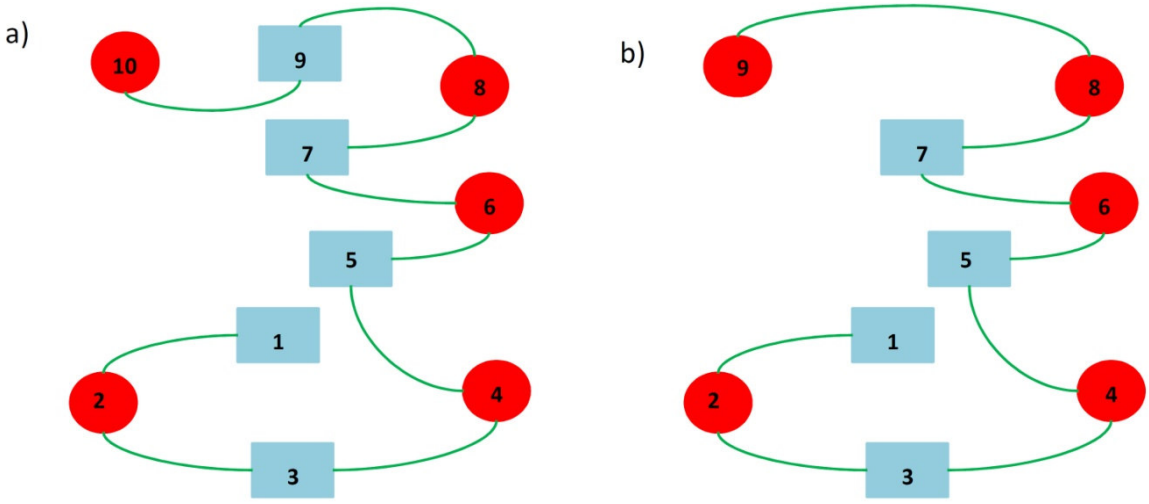


Figure 2.8: Topology comparisons between DevR and other known response regulators. The topologies for TrrA from *T. maritima* and RR468 from *T. maritima* were identical and depicted in (a). The topology for DevR is shown in (b). The red circle represents an σ helix, the blue square represents a β sheet and the loops are coloured green. The $(\sigma\beta)_5$ motif is observed and described for response regulators by Benda *et al*. (2004).

These topologies were used to superimpose the DevR and DevS structures on the 3A0R and 3GDE reference structures (Figure 2.9). The interactions between HK ThkA and its response regulator TrrA are outlined in Figure 2.9(A). Figure 2.9(B) shows the DevS201 superimposed on ThkA, Figure 2.9(C) shows DevR superimposed on TrrA and Figure 2.9(D) shows the DevS201 and DevR interaction. TCSs are known for their His-Asp phosphotransfer; the close proximity of His398 and Asp54 makes the model more plausible.

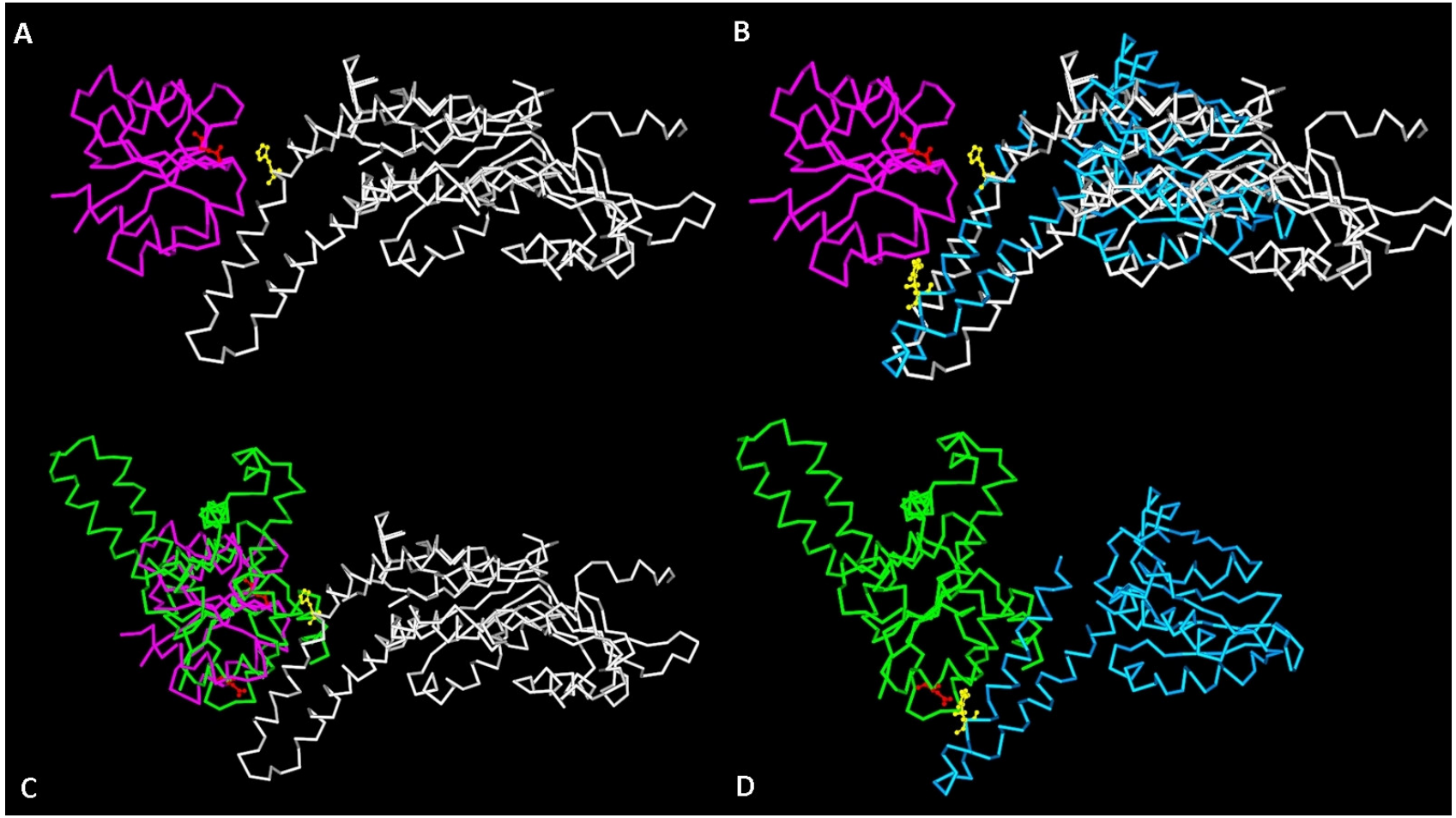


Figure 2.9: Homology docking based on the crystal structure of histidine kinase. (A) The histidine kinase ThkA (grey) in complex with the response regulator TrrA (purple) of *T. maritima*, the His547 is indicated in yellow and the Asp52 is indicated in red. (B) The histidine kinase DevS201 (blue) was superimposed onto ThkA, and His395 is indicated in yellow. (C) The response regulator DevR (green) was superimposed and the Asp54 is indicated in red. (D) The final binding conformation between DevR and DevS201, with the his-asp residues in close proximity.

In order to further refine the association between the DevS201 and the DevR molecule, a series of minimizations was done on the molecules individually and then as a dimer. The resulting energy values were: DevS201 -3972.17 kcal per mol, DevR -12153.77 kcal/mol and the dimer -11352.52 kcal/mol. The energy of the associated molecules was much lower than that of the separate molecules.

2.3.4.2 Protein-protein docking with DS ZDOCK

The aim of protein-protein docking is to assemble separate protein components into a protein-protein complex. The protein receptor (DevS) and a protein ligand (DevR) were docked to form an optimized conformation, such that the free energy of the overall system is minimized. A pose is defined as the relative orientation between the receptor protein and ligand protein. DS ZDOCK is a rigid body docking program that uses Fast Fourier Transformation which produced 54,000 docked poses meeting the criteria set by the distance cut-off filter. These molecules were clustered into 100 clusters, the largest of which contained 233 poses. The clustering analysis provided by the ZDOCK protocol is based on the RMSD matrix. Clustering is predominantly utilized to reduce the number of hits coming from ZDOCK, since the capacity to analyze each pose is far more limited than the capacity to generate hits. The purpose of scoring functions during protein-protein docking is to select the most favourable ranking solution that would give an insight into the affinity of the complex. For that reason, ZDOCK, RDOCK and ZRANK were used to distinguish near-native structures. Subsequently, all those molecules with positive values were removed, decreasing the dataset to 2,000 possible docking hits (also referred to as poses). Analysis based on the DS ZDOCK score, DS ZRANK score, overall rank and cluster size was then done and the results visualized using 3D plots (Figure 2.10). The criteria for a pose to be carried over to the next round of optimization were a high DS ZDOCK score, low DS ZRANK score, low overall rank and preferably part of a big cluster.

Figure 2.10(A) shows the 2,000 poses used for the analysis and Figure 2.10(B) shows the remaining 1,000 poses to be submitted for further analysis and optimization. DS RDOCK was used for the refinement of the top 30 docked complexes using a CHARMM-based procedure for refinement and scoring, identifying near-native structures (Brooks *et al*., 1983).

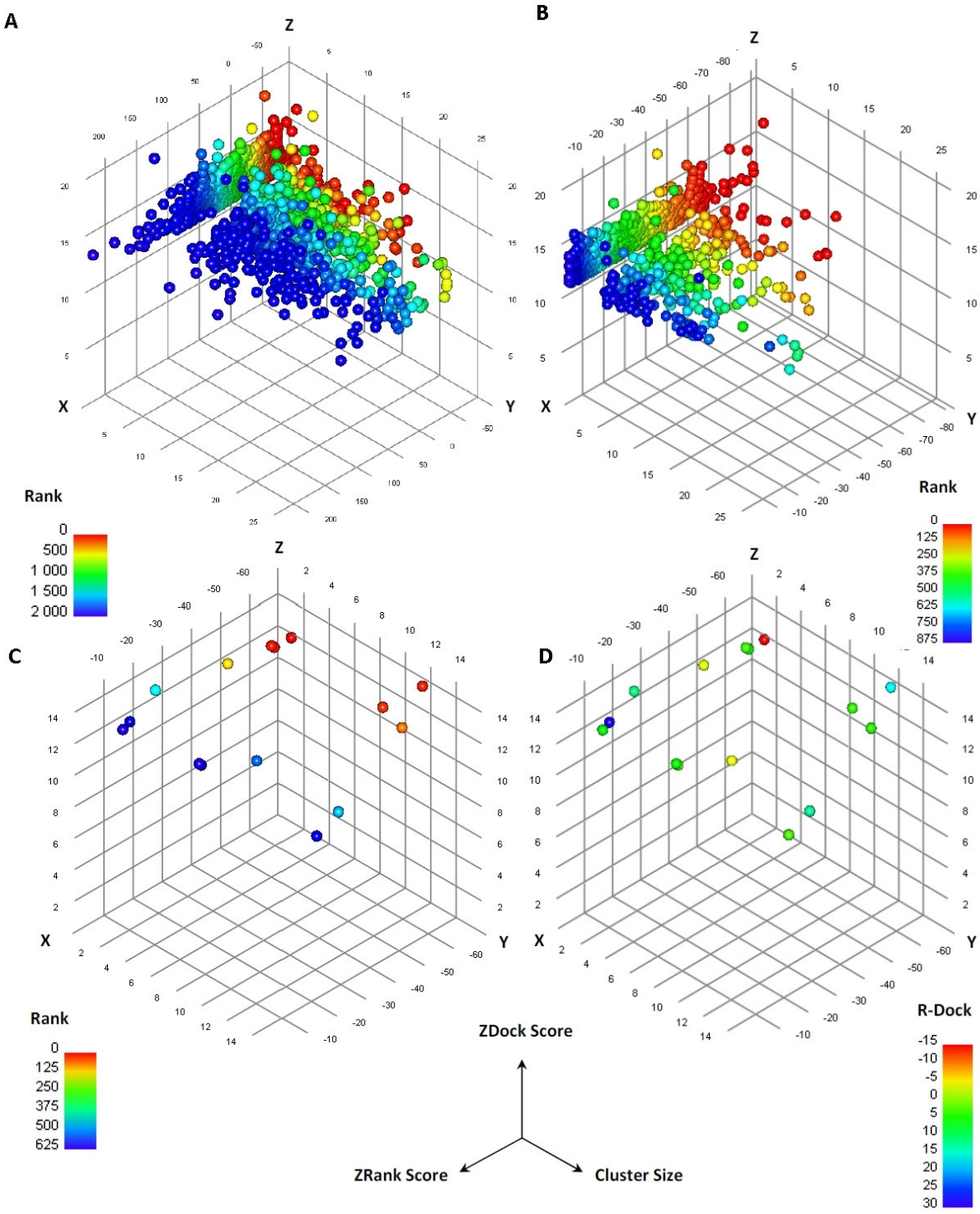


Figure 2.10: A 3D plot of ZRANK, ZDOCK and RDOCK scores. The dots represent cluster centres where DevR can potentially bind to DevS. The ZRANK and ZDOCK scores of the top 2,000 poses are shown in (A), with the high energy poses were removed in (B), and the final 16 poses in (C) and (D). For (A), (B) and (C), the colours of the dots represent the rank that pose has in relation to the others. In (D), the dots are coloured based on their RDOCK score.

Those poses with positive DS RDOCK values were removed from the dataset, as positive values can be indicative of clashes that could not be removed via minimizations. The

remaining 16 poses and their scores were visualized in Figure 2.10(C) and (D). The 16 best scoring docked poses were further analysed for their RMSD between different docked poses. The best poses from the docking results are shown in Figure 2.11. Only experimental data would be able to give a clear indication of which of the four models are correct.

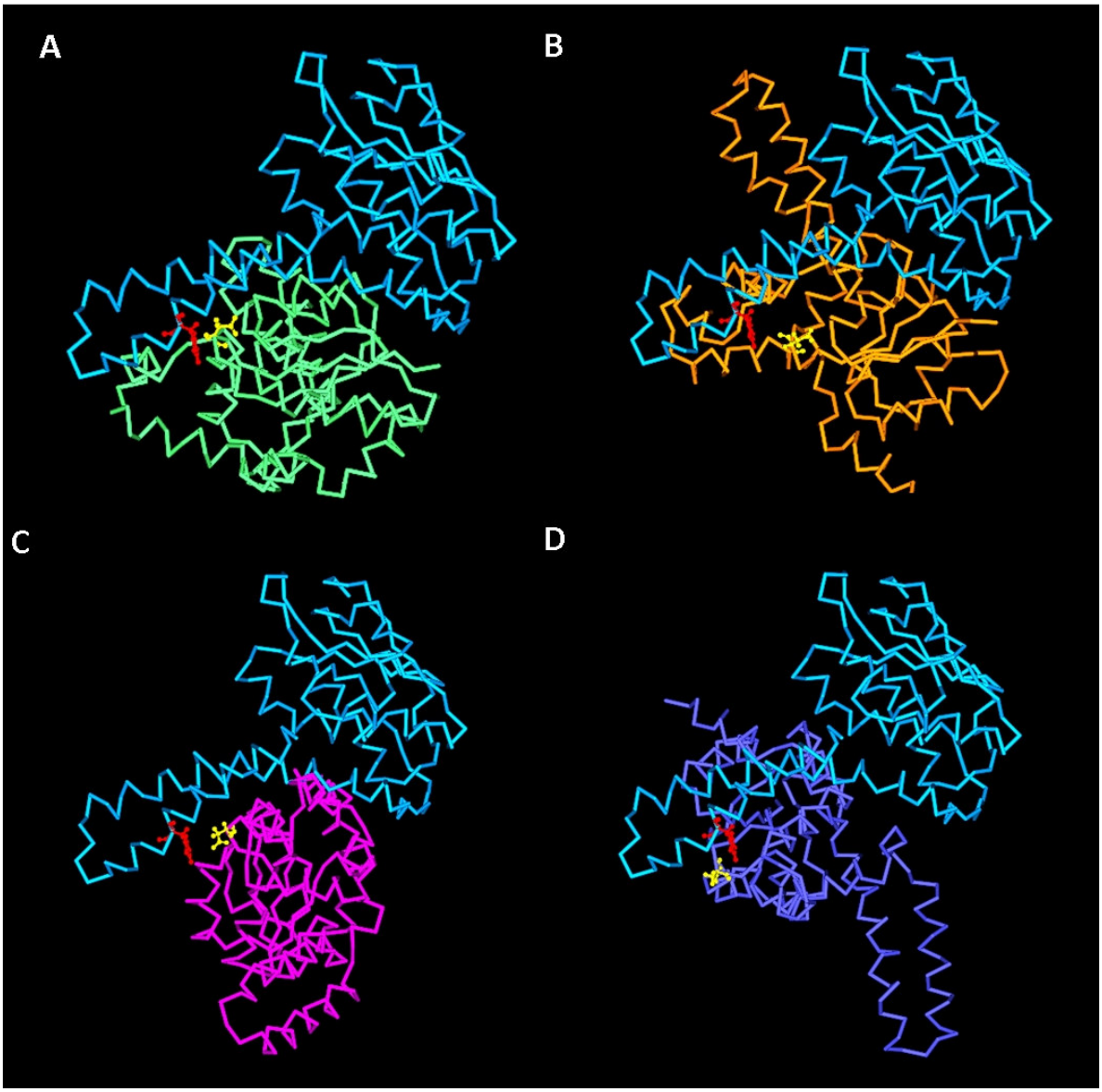


Figure 2.11: Final poses produced by ZDOCK for DevR binding to DevS201. The light blue molecule is DevS201. The different poses indicate the different ways in which DevR can bind to DevS. (A) Pose 1644 (lime green). (B) Pose 145 (orange). (C) Pose 324 (pink). (D) Pose 1508 (purple). In all of these poses, the His395 indicated in red is in close proximity to the Asp54 indicated in yellow.

The kinase core bears the phospho-accepting His residue within conserved sequence fingerprints called homology boxes. The H-box for DevS is RDL**H**DHVTQ, with His395

being the conserved histidine. This kinase core is a helical domain with structural similarity within the histidine kinase family, this suggested that the helical bundles facilitate histidine phosphorylation as well as dimerization. The histidine domain of DevS spans 67 residues starting at Asp386 to Gly452. The dimer interface is located in the helical-hairpin domains with the helices such that the cytosolic domains are adjacent, and in position to originate from the membrane. The helix α1 includes the His395 phosphorylation site and helix α2 makes helix–bundle contacts with helices of DevR.

2.4 Discussion:

Literature demonstrates that highly conserved regions are present in kinase proteins. HKs have these conserved domains in the regions where the phosphorylation takes place and at the ATP binding site. The kinase domain (catalytic core) consists of four conserved sequence motifs, namely the H, N, G1 and G2 boxes, and a more variable F-box. The HK superfamily has been classified according to various criteria. A protein family is defined as a group of proteins that share structure, sequence and functional similarity, suggesting they are derived from a common ancestral protein (Scheef and Bourne, 2005). The aim here was to place the HK catalytic core sequence motifs into a structural context, and create a model of the interactions between DevS and DevR. This was achieved in the manner described below.

BLAST, EC numbering, GO terms, CATH and SCOP databases were first used to find detailed structural information for the HK protein family. HKs are known to be insoluble when expressed *in vitro* as they are usually found as membrane-bound proteins (Saini *et al*., 2004a). Consequently, scientists tend to truncate the kinases and express, purify and crystallize only the domain they are interested in. As a result, it was found that there is a general lack of detailed structural information for full length HKs. Only 2 full length structures were available in the PDB database, CheA (*T.maritima*, 1UOS) and PhoR (*T.maritima*, 2C2A).

All the HK catalytic core structures clustered into 8 different classes. The CheA full length molecule was classified in its own class (class number 4). Structures with the same proven function, as well as structures from the same organism but different lengths were classified in

different classes, for example PhoQ classified into both class 3 and 5. The conclusion drawn from this data was that the partial structures do not contain enough information for comparative analysis between structure and function.

Homology modelling has five major weaknesses that should be kept in mind when structures are modelled: the correctness of the templates used, the alignment between the templates and the sequence to be modelled, regions without templates, backbone distortion and side chain packing. Each of the weaknesses was addressed during the molecular modelling of DevS and some shortfalls noted.

In order to ensure that the correct templates were chosen it was important to know where DevS belongs based on structural analysis. Template proteins require close homology in order for DevS to be modelled reliably. The DevS DNA sequence was compared to the known HK classes, and NarX, NarQ, UhpE and DegS were identified as being homologous to DevS. However, the aim was to produce a reasonable approximation of the DevS protein structure using the known structure of proteins related to DevS by homology. DevS was modelled using templates from DevS (*M. smegmatis*) and DesK (*B. subtilis*), while DevS201 was modelled using templates from DesK (*B. subtilis*), CheA (*T. maritima*) and PhoQ (*T. maritima*).

The accuracy of these models is dependent on the sequence identity between target and template, as well as the correct alignment. When the models were designed, special care was taken to realign DevS201 with its templates based on the homology boxes found in HKs. The secondary structures for those motifs were identical and this provided more information and therefore increased the quality of the model.

When choosing templates for the full-length DevS model, we were faced with a problem that there were portions of sequence that did not have a template or where the homology was so low that no secondary structure could be inferred. In these cases we relied on loop-modelling and loop refinement to arrive at the best conformation for those portions of sequences.

Energy minimization steps were used to improve the accuracy of the models. Energy minimization was used to find a local energy minimum for each molecule in the model: a total lower energy state indicating a more stable model (Brooks *et al*., 1983).

Homology model quality assessment is usually performed using two methods: statistical potentials or physics-based energy calculations. Both methods produce an estimate of the

energy or an energy-like value that may be used to determine the structural accuracy of the model. Various scoring functions were used to identify the best model for DevS.

DS uses Probability Density Functions (PDF) to calculate pseudo-energy terms and stereochemical pseudo-energy terms (Sali and Blundell, 1993). The PDF energy is the sum of energies of the stereochemical pseudo-energy terms which consist of valence bonds, valence angles and torsion angles, improper torsion angles, and soft-sphere repulsion (Sali and Blundell, 1993). Higher energy indicates a larger restraint violation in the model. These functions were used address backbone distortion and side chain packing in the DevS and DevS201 models to ensure that no torsion or spatial rules were violated during the building of the models.

The Discrete Optimized Protein Energy (DOPE) scores of each model were also assessed. A lower DOPE score statistically indicates a better model. DOPE is a knowledge-based scoring function that is often termed statistical potential scoring function. The energy values produced is based on the atomic distance-dependent statistical potential from known protein structures, based on probability theory (Shen and Sali, 2006).

The verification score determined for each amino acid residue is based on the secondary structure and indicates whether a residue is in the desired 3D environment. The higher the score, the greater the probability that the structure is more likely to be correct. DFire is a scoring function that calculates the residue-specific all-atom, distance-dependent potential of mean force by using a physical state of uniformly distributed points in finite spheres as an indication of the quality of the model (Zhou and Zhou, 2002). QMEAN is a composite scoring function consisting of 5 terms: torsion angle energy, distance-dependent pairwise residue-level potential used to assess long-range interactions, solvation energy, and agreement between predicted and observed secondary structure and agreement between predicted and observed solvent accessibility (Benkert *et al*., 2008).

Protein-protein docking is the computational modelling of the three dimensional structure of the interactions formed between two proteins as it would occur in a living organism. This produces probable candidate structures, which must be ranked using methods such as scoring functions to identify structures that are most likely to occur in nature. Probable candidate structures of DevS-DevR complex were modelled. Homology docking based on structure similarity of known crystallized two-component kinases systems was performed to reach a near-native structure for the binding of DevS201 to DevR. The *T. maritima* ThkA/TrrA and

HK853/RR468 TCSs were used as templates to model the 3D structure of the DevS-DevR complex. During the DevR and DevS dimerization analysis it was observed that the His395 and Asp54 residues were in different positions.

Marina *et al*., 2005 observed that there was a separation between ATP and the phosphoacceptor histidine in this structure of HK853-CD, suggest that the cytoplasmic portions of HK sensor must access multiple conformational states, some of which are critical for catalytic action. In a similar manner it would seem possible that the conformational state for phosphotransfer is different from the conformational state of the monomer. Since the DevR and DevS structures were first modelled using monomer structures as basis it could be possible that the conformation used in this homology docking is not the true form of both the proteins during dimerization. Support for this notion was found by Wisedchaisri *et al*., 2008 that reported that DevR undergoes conformational changes when phosphorylation occurs via a helix rearrangement mechanism. The full-length DevR structure indicated that in the inactive conformation, domain interaction buries Asp54 residues from accessibility to solvent (Wisedchaisri *et al*., 2008). Consequently the spatial orientation of asp residue observed in the DevR model used in the homology docking study might also be different during dimerization. This highlights one of the biggest short fall of homology modelling, the conformational changes cannot predicted accurately.

Molecular mechanics and molecular dynamics could be used to simulate the physical movement of atoms and molecules. Unfortunately both molecular mechanics and molecular dynamics for a dimer complex are very computationally expensive. In order to overcome the computational as well the size constraints a method like normal mode analysis could have been used to identify some of the potential conformational changes.

The ZDOCK docking algorithm was used to determine the quaternary structure of the complexes by allowing the relative orientations of the interacting proteins being modelled to vary, while holding the structure of each of the separate proteins fixed.

Scoring functions can be put to a number of good uses. Classically scoring functions are used to determine the lowest energy conformation of a protein, assuming the near-native structure (Anfinsen, 1973). Where protein-protein docking is concerned, the calculations of the energy minimization that goes together with these scoring functions become too extensive and too complex to explore exhaustively. Scoring functions like ZRANK and RDOCK that give good approximations of the minimum energy were used to ensure that undesirable (high energy)

conformations were removed from the dataset, rather than to assume the lowest energy is the native form (Lorenzen and Zhang, 2007; Pierce and Weng, 2007). DS ZRANK is an energy-based scoring function that predicts the strength of intermolecular interactions between two proteins. DS ZRANK uses three energy terms based on van der Waals interactions, electrostatic interactions and solvation energy to score all the poses. A positive value is an indication of very little or no interaction between the proteins or it could indicate residue clashes or other improper steric torsions.

After these steps, more than one conformation for the DevS-DevR complex with similar docked energies was observed. The decision on which conformation was the more correct form can only be estimated by comparison to known TCS systems. Alternatively, crystallization of this complex would be the ultimate proof for these models. A model of the ATP binding moiety is not available. None of the template models have a co-crystallized ATP that could be superimposed into the DevS models. Small ligand docking can be pursued in the future to try and co-ordinate the ATP to the amino acids it is known to bind.

Before any model can be deemed correct, a series of experimental data must be collected to ensure that the catalytic site, the binding site as well the other binding interactions can be explained by the model. Consequently, functional expression and activity testing of DevS and DevR were pursued, and are described in the following chapters.

The first priority was to determine if the binding affinity between DevR was the same for DevS as for DevS201; to validate our assumption to do the majority of our modelling based on DevS201. If this proves to be correct, the next step would be to do site-directed mutagenesis to determine which amino acids of DevS play a role in the binding between DevS and DevR.

Chapter 3 Expression, purification and activity studies of the Dev two-component system

3.1 Introduction

Due to the complexity of the Dev TCS, the techniques to assess this system were put in place using the *Bacillus subtilis* SpoII TCS system as a model. A functional kinase assay was needed to measure phosphate transfer, so the SpoII TCS was used as a model system to test the methods by which the radioactively labelled ^{32}P-ATP is used to track the phosphorylation. Mutation studies were done on SpoIIAB to assess the sensitivity of the assays for future usage in DevR mutation phosphorylation assays. Consequently, the SpoIIAA and SpoIIAB phosphorylation system was used to ensure the reliability and sensitivity of the kinase assay.

3.1.1 *B. subtilis* SpoIIAA and SpoIIAB system

B. subtilis uses the transcription factor σ^F to control transcription of sporulation-specific proteins (Partridge *et al*., 1991). σ^F is bound in an inactive form to the anti-σ factor SpoIIAB, which prevents transcription of relevant proteins. In the presence of ATP, SpoIIAB can either bind to and inhibit σ^F, or act as a specific protein kinase to phosphorylate SpoIIAA (Duncan and Losick, 1993; Min *et al*., 1993; Diederich *et al*., 1994). SpoIIE is a specific protein phosphatase whose substrate is the phosphorylated form of SpoIIAA (Duncan *et al*., 1995). Before sporulation, SpoIIAA is present largely or exclusively in the phosphorylated form, and consequently does not interact with SpoIIAB, leaving SpoIIAB unbound and free to inhibit σ^F (Diederich *et al*., 1994; Arigoni *et al*., 1996). During sporulation, SpoIIE becomes active and dephosphorylates SpoIIAA. Unphosphorylated SpoIIAA can now bind to SpoIIAB. Once the SpoIIAB becomes involved in the phosphorylation of the newly-released

SpoIIAA, it releases σ^F. The phosphorylation involves the transfer of the phosphate to Ser58 of SpoIIAA (Najafi *et al*., 1995). It was interesting to note a histidine close to the ATP binding site of SpoIIAB (Campbell *et al*., 2002). Although this histidine is not mentioned in the catalytic mechanism in literature, this histidine could be relevant (Diederich *et al*., 1994). The mutation of this His54 residue to an alanine may elucidate its potential role in the phosphorylation of SpoIIAB. Phe110 is within the ATP-binding site, and makes hydrophobic contacts with the face of the adenine ring (Campbell *et al*., 2002). Mutants at this position will decrease ATP binding activities, leading to the proposal that this residue functions in transition-state stabilization during the phosphorylation reaction.

3.1.2 Protein expression and purification

The quest for soluble functional protein has been an essential component of protein binding studies. One of the main complications during the expression of recombinant proteins in *E. coli* is the formation of inclusion bodies - aggregates of insoluble misfolded proteins (Baney and Mujacic, 2004). Inclusion body formation can be attributed to various factors, including the usage of strong promoters and high inducer concentrations (Tsumoto *et al*., 2003). The increase of foreign protein in the host's cellular environment can also escalate the rate of protein misfolding and aggregation. Unfortunately, bacteria lack the ability to support most post-translational modifications such as intra- or intermolecular disulphide bond formation and this can also lead to aggregation of proteins (Baney and Mujacic, 2004). There is a clear distinction between precipitated proteins and inclusion bodies. The former are proteins in their native form that randomly precipitate due to a reduction of solubility of the polypeptide chain, for example during salting out procedures. Inclusion bodies are believed to be the consequence of interactions between hydrophobic residues in unfolded or partially folded proteins that become exposed at the protein surface (Clark, 1998). The most frequently used scheme to recover biologically active proteins from inclusion bodies involves three steps: isolation of the inclusion bodies, solubilisation of the inclusion bodies and refolding of solubilised proteins. Although inclusion bodies can be a dead end to a soluble expression study, it does have the advantage that their isolation from the cell is very convenient as a first purification step (Rogl *et al*., 1998). Following either mechanical or chemical cell lysis, inclusion bodies can be collected by centrifugation. Washing the pellet with a detergent such

as Triton X100 or sodium deoxycholate should remove soluble proteins that might be stuck to the hydrophobic inclusion bodies (Clark, 1998). These proteins can hinder the refolding step. Otherwise, a sucrose gradient centrifugation step can be added to remove contaminating cell debris (Clark, 1998). Various solubilising conditions should be taken into consideration when proteins are refolded. Commonly used solubilisation agents are urea, guanidine-HCl and strong ionic detergents such as *N*-lauroylsarcosine and acids like formic acid (Tsumoto *et al*., 2003). Occasionally, chelating agents can be included in the solubilisation buffer to prevent oxidation reactions that may occur in the presence of metals. Temperature gradients between 30°C and 37°C are used to facilitate the process. The structure after solubilisation is dependent on the denaturant used: urea and guanidine-HCl produce flexible disordered structures, whereas detergents produce a much more ordered structure. However, with the controlled treatment of urea and guanidine-HCl, it is possible to get more efficient refolding because these chemicals undergo concentration-dependent binding to proteins (Clark, 1998; Tsumoto *et al*., 2003). The various approaches used to refold the proteins include dilution, dialysis and solid phase methods. The main aim of all these methods is to unfold the protein by using a detergent or alternative chaotropic chemical and then to stepwise remove the denaturing agent, allowing the protein to refold into its native conformation. In literature, it was shown that there is no universal refolding buffer or native structure recovery procedure. The refolding buffer usually refers to the buffer used to remove the denaturing agent and consequently is the buffer in which the protein is stable and soluble (Singh and Panda, 2005). The composition of the refolding buffer is protein-specific and depends on various factors such as ionic strength, pH, specific metal requirements and reducing agent (Vincentelli *et al*., 2004; Ventura, 2005). Nickel-chelating chromatography can be used as an example of solid phase refolding. This has the advantage of preventing further aggregation since the proteins are immobilized with their hydrophilic faces exposed, and are not free to associate with other proteins. This method combines the refolding step with a purification protocol (Rogl *et al*., 1998).

3.1.3 Factors influencing DevS solubility

It has been shown that the cytoplasmic domain of DevS is capable of autophosphorylation and phosphorylation of DevR (Saini *et al*., 2002). The N-terminal sensor domain is composed

of three transmembrane regions, and is also referred to as the membrane anchor domain. This hydrophobic part of the protein will influence protein folding and protein solubility. Erroneous hydrophobic interactions can cause misfolding and aggregation in inclusion bodies (Baney and Mujacic, 2004). For histidine kinase functionality, the only catalytic kinase core is needed, therefore by removing the hydrophobic part of the protein upstream from the His395 residue, the cytosolic region containing all the functional boxes will still be present. The location of the boxes as well as the phosphorylated His395 residue can be seen in Figure 3.1. Removing this hydrophobic domain should lead to better folding and increase solubility (Saini *et al*., 2002).

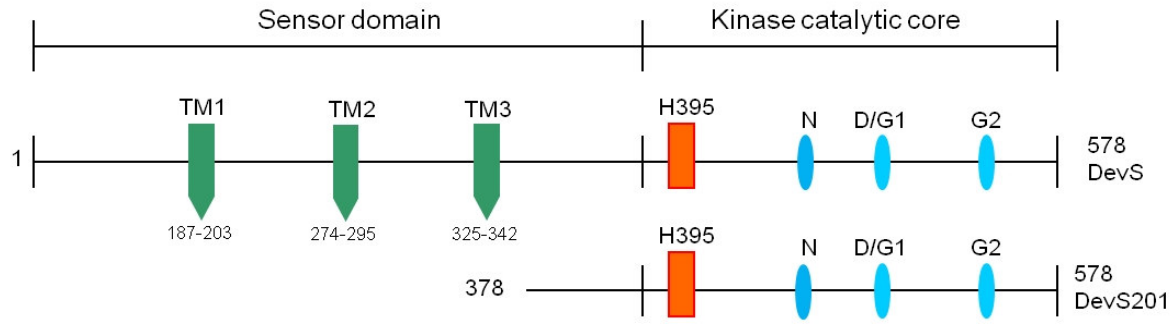


Figure 3.1: Significant features of DevS. The three predicted N-terminal transmembrane regions (TM1, TM2 and TM3) of DevS are indicated by the green arrows, with the corresponding residues indicated underneath. The His395 that forms part of the H box and which is autophosphorylated is indicated in red. The conserved N, D/G1 and G2 boxes are shown in light blue. The 201 amino acid cytosolic fragment of the C-terminal region of DevS is referred to as DevS201.

3.2 Materials and methods

3.2.1 SpoIIAA and SpoIIAB

Plasmids pEAA and pEAB, containing *B. subtilis* spoIIAA and spoII AB as *Nde*I and *Bam*HI inserts in vector pET3a (Novagen, Madison, Wisconsin, USA), were received from Professor MD Yudkin, Department of Biochemistry, Oxford University (Min *et al*., 1993).

3.2.1.1 Restriction enzyme digestion

In order to confirm the integrity of the received pEAA and pEAB plasmids, the following restriction enzyme digestions were set up: 50 ng plasmid DNA was digested with 1 U *Bam*HI (Fermentas) and 1 U *Nde*I (Fermentas), in the appropriate 10 x buffer (supplied by the manufacturer) at 37°C for 2 hours. The plasmid maps indicating the restriction sites can be seen in Figure 3.2.

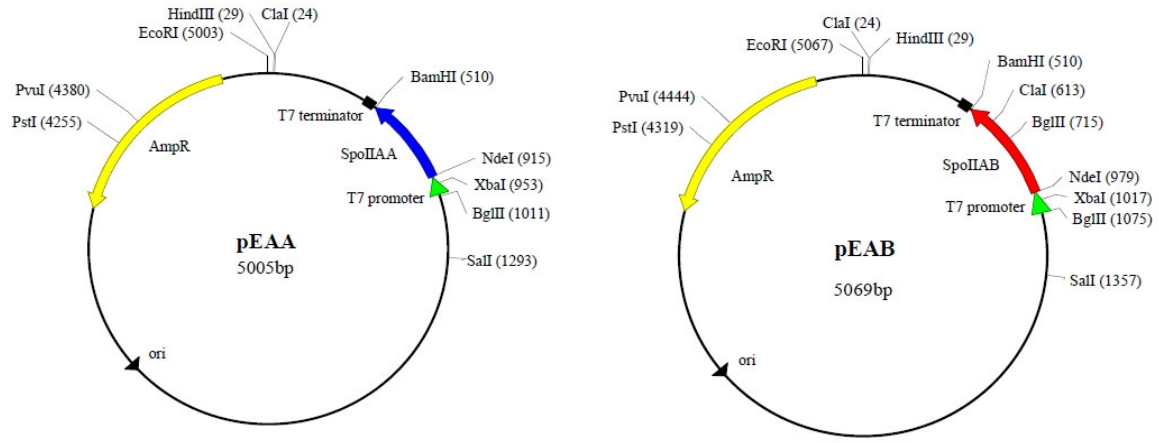


Figure 3.2: Plasmid maps of pEAA and pEAB, containing *B. subtilis* SpoIIAA and SpoIIAB, respectively. The ampicillin resistance gene (*bla*) is highlighted in yellow and the T7 promotor indicated in green.

3.2.1.2 Agarose gel electrophoresis

The digested products were separated by electrophoresis in a 1% agarose (Promega, Wisconsin, USA) gel. TAE buffer was used for electrophoresis. Ethidium bromide, a fluorescent DNA base interchelator (1.25 ng/ml) was included in the gel solution for DNA visualization. Samples were electrophoresed at 8 $V.cm^{-1}$ and the bands were visualized under UV illumination at 312 nm, on a Bio-Rad Chemidoc Gel Documentation System (Bio-Rad laboratories, USA), using the QuantityOne software.

3.2.1.3 Preparation of DH5α and BL21(DE3) electrocompetent cells.

Electrocompetent cells were prepared as described by Sambrook *et al*. (1989). A single colony was inoculated into 15 ml Luria Bertani (LB) media, and grown overnight at 37°C at 300 rpm. Fresh LB (250 ml) was inoculated with 1 ml of the overnight culture and grown at 37°C with shaking (220 rpm) until the cells reached early to mid-log phase (OD at 600nm of 0.3-0.6). To harvest, the broth was transferred to cold centrifuge tubes (250 ml) and centrifuged at 5,000*g* in a Sorvall RC-5C Plus, GSA rotor (Sorvall, USA) for 10 min at 4°C. All the subsequent steps were done at 4°C. After the supernatant was discarded, the cells were washed with 250 ml ice-cold sterile water. The suspension was centrifuged at 5,000*g* for 10 min. This washing step was repeated twice. After the final centrifugation step, the supernatant was immediately removed from the loose pellets. The pellets were resuspended into 10 ml ice-cold sterile 10% glycerol and incubated on ice for 30 minutes. Cells were subsequently harvested by centrifugation as described above, the supernatant removed by vacuum suction and the pellets resuspended in 800 µl ice-cold sterile 50% glycerol. This was divided into 100 µl aliquots and stored at -70°C.

3.2.1.4 Electro-transformation of pEEA and pEAB plasmids into E. coli DH5α

Electrocompetent DH5α cells were thawed on ice. To 2 ng of plasmid DNA (pEAA or pEAB), 100 µl of electrocompetent cells was added. This was then transferred to a pre-chilled electroporation cuvette (Hybaid, 2 mm gap) and a pulse of 2,500 V applied for 5 ms in a Bio-Rad MicroPulser electroporator (Bio-Rad Laboratories, Hercules, California) (Dower *et al*., 1988). LB media (1 ml) was added directly after electroporation and the cells were incubated for 1 hour at 37°C with shaking, and plated on LB-agar containing 100 µg/ml ampicillin for plasmid selection.

3.2.1.5 Plasmid isolation and insert preparation

In order to isolate the spoIIAA and spoIIAB genes from pEAA and pEAB, the Plasmid Miniprep Isolation Kit (PEQLab, Biotechnology GmbH) was used to isolate plasmid DNA from *E. coli* DH5α. This DNA was then subjected to a large-scale *Nde*I-*Bam*HI restriction enzyme digestion followed by agarose gel electrophoresis as described in Sections 3.2.1.1

and 3.2.1.2. For pEAA, this resulted in a plasmid backbone band of 4.6 kb, and a SpoIIAA insert band at 365 bp, while for pEAB there was also the 4.6 kb plasmid band and a SpoIIAB insert band at 422 bp. These small bands were excised from the agarose gel with a sterile scalpel blade, and the Zymoclean gel DNA Recovery Kit (Zymo Research Corporation, USA) was used to extract the DNA fragments from the agarose.

3.2.1.6 Sub-cloning SpoIIAA and SpoIIAB into pET16b.

pET16b (Novagen, Madison, Wisconsin, USA) was isolated, digested and agarose gel-purified as described above for pEAA and pEAB (Section 3.2.1.4). pET16b was chosen for expression of SpoIIAA and SpoIIAB as it contains a nucleotide sequence to express an N-terminal His_6-tag for easier protein purification. Ligations were set up at an insert: vector ratio of 20:1 using T4 DNA Ligase (Fermentas, Ontario, Canada) as per the manufacturer's instructions. The ligations were dried down completely using a speedy vac (Savant SC110) for 20 minutes, and the pellet resuspended in 1 ml 70% ethanol. This was centrifuged at 4°C at 10,000*g* for 30 minutes, the ethanol removed, the pellet air dried and resuspended in 1 µl dH_2O. The ligation mixture was then transformed into electrocompetent *E. coli* BL21(DE3) cells, using electroporation as described in the Section 3.2.1.4.

3.2.1.7 SpoIIAA and SpoIIAB protein expression

E. coli BL21(DE3), freshly transformed with the pET16b plasmids containing spoIIAA and spoIIAB genes, were plated onto LB plates containing 100 µg/ml ampicillin. One colony was used to inoculate 50 ml LB containing 100 µg/ml ampicillin and incubated overnight at 37°C with shaking. 250 ml LB liquid medium (containing 100 µg/ml ampicillin) was inoculated with 2.5 ml of the overnight culture, and incubated at 37°C with shaking until an optical density at 600 nm of approximately 0.6 (logarithmic growth phase) was reached. Subsequently, IPTG (pEQ Lab Biotechnology, GmbH) was added to a final concentration of 100 µM to induce protein expression. The cultures were then incubated at 37°C for 16 hours with shaking. The biomass was harvested by centrifugation at 10,000*g* for 10 minutes at 4°C

(Sorvall Plus RC-5C, GSA rotor, Sorvall, UK) and the pellet frozen for overnight storage (-20°C).

3.2.1.8 Extraction of the soluble protein fraction

Protein extraction and purification was done utilizing the methodology essentially described by Liebau *et al*. (2002). The frozen cells were allowed to thaw on ice and resuspended in Buffer A. A Complete Mini, EDTA-Free protease inhibitor cocktail tablet (Roche, Indianapolis, USA) was added and the cells sonicated for 10 cycles of 30-second pulsed sonication with 30 seconds incubation on ice water between cycles [(Vibracell sonicator, Sonics and Materials Inc. Danbury, Connecticut, USA) with the duty cycle set to 50%)]. The cell debris was removed by centrifugation at 1,000*g* for 10 minutes at 4°C (Sorvall Plus RC-5C, rotor SS34, Sorvall, UK), leaving the soluble protein fraction in the supernatant.

3.2.1.9 SpoIIAA and SpoIIAB protein purification

Affinity purification was carried out using the HIS™-select nickel affinity gel (Sigma-Aldrich, USA) essentially as per manufacturer's manual. A 5 ml HIS™-select nickel affinity gel column was used during the purification of His_6-SpoIIAA and His_6-SpoIIAB proteins and entailed the following steps: The column was washed with two column volumes of deionized water, followed by three volumes of equilibration buffer. The clarified crude extract from the previous section was loaded onto the column at a flow rate of 2 to 10 column volumes per hour. The resin and the crude extract were incubated at 4°C for half an hour before washing and elution. The column was washed with 5 column volumes of wash buffer. The protein bound to the column was eluted from the column using 3 to 10 column volumes of elution buffer. Fractions of 2 ml were collected. Following collection, the column was washed with 5 column volumes of binding buffer to remove the residual elution buffer and to restore the column.

3.2.1.10 SDS-PAGE analysis of SpoIIAA and SpoIIAB

The Laemmli method for SDS-PAGE was used to analyze the purified proteins, separating the protein fractions based on molecular mass (Laemmli, 1970). A 4% stacking gel and 14% running gel were prepared. A 15 µl purified protein sample (220 µg/ml) from each of the eluted fractions of His_6-SpoIIAA and His_6-SpoIIAB was transferred to a clean microcentrifuge tube, and an equal volume of Laemmli loading buffer was added. The protein samples were denatured at 90°C for 5 minutes and loaded onto the SDS-PAGE gel. Electrophoresis was performed in a Tris-Glycine buffer at 200 V in a Bio-Rad Mini Protean3 Electrophoresis system. Protein bands were visualized with PAGE staining solution, followed by gel destaining in a destaining solution.

3.2.1.11 Protein concentration determination

Protein concentration was determined by using the Quant-IT™ Protein Assay Kit (Invitrogen, USA) in conjunction with the QUBIT fluorimeter (Invitrogen, USA), essentially as described by the manufacturers. Dilutions were made when protein solution with concentrations higher than 1 mg/ml were expected. To 10 µl protein sample, 190 µl of the Quant-IT working solution (1 µl Quant-IT™ Reagent and 199 µl Quant-IT™ Buffer) was added. After a 15 minute incubation at room-temperature, the fluorescence was read by using the QUBIT fluorimeter and a concentration was provided by the reader.

3.2.1.12 Phosphorylation assays using radioactively labeled 32P-ATP

Eluted fractions of purified His_6-SpoIIAA and His_6-SpoIIAB with the highest protein concentrations (as determined in Sections 3.2.1.10 and 3.2.1.11), were pooled. The pooled samples were transferred into SnakeSkin Dialysis Tubing (Pierce Biotechnology, Rockford, Illinois, USA) The dialysis tubing was submerged in 1L PBS buffer and slowly stirred overnight at 4°C to remove the imidazole and other unwanted salts present in the elution buffer, to eliminating interference with the enzyme assay. EasyTides ATP [γ-^{32}P, 5 µCi/µl

packed in a 50 mM tricine buffer (pH 7.6) with green dye (Perkin Elmer Life and Analytical Science, Boston, USA)] was used to assess the reliability and sensitivity of kinase assays. SpoIIAB (kinase) and SpoIIAA (substrate) are stable for 24 hours in the assay reaction mix at 37°C. Hence, a single step reaction (both autophosphorylation of SpoIIAB and subsequent trans-phosphorylation of SpoIIAA) that ran for 16 hours was used. SpoIIAA protein (950 ng) was mixed with 300 ng SpoIIAB substrate in SpoII phosphorylation buffer, containing 5 µCi γ-^{32}P-labelled ATP and 1 mM ATP. All the reactions were performed in a final volume of 20 µl. The reactions were incubated at 37°C, an equal volume Leammli loading buffer added to stop the reactions, and they were then heated at 90°C for 5 minutes. These reactions were electrophoresed on a 14% SDS-PAGE gel (as per Section 3.2.1.10). The gels were stained with PAGE staining solution for 6 hours, with gentle shaking, and destained overnight in PAGE destaining solution containing 10% glycerol. The gels were transferred to Whatman 3MM paper, covered with cling-wrap, and vacuum dried for 6 hours at 65°C using a slab gel dryer (SAVANT, Forma Scientific, USA). The radioactive gel was exposed to an X-ray film (Hyperfilm MP, Amersham) for 48 hours at -70°C. The film was developed by soaking it for 2 minutes in Prefix fixer (Champion Photochemistry, RSA), washed for 2 minutes in water and left in PolyconA X-ray developer (May and Baker, Essex, UK) until the bands could be seen on the film.

3.2.1.13 Site-directed mutagenesis (SDM) of SpoIIAB

Two mutants of spoIIAB were created via site-directed mutagenesis. Histidine 45 was mutated to a glutamine (H54N) and the phenylalanine 110 was mutated to an isoleucine (F110I). The replacement of the His54 residue with a glutamine was performed to elucidate the potential role of the His54 in the phosphorylation of SpoIIAA, as other histidine kinases possess a conserved histidine that is autophosphorylated before transferring the phosphate to the substrate, while Phe110 is located within the ATP binding site of the kinase. The Phusion Site Directed Mutagenesis Kit (Finnzymes, Espoo, Finland) was used as described by the manufacturer, with some modifications on the standard protocol, as described below. For the H54N mutation, an *Nco*I restriction enzyme site was removed in a silent mutation to allow for selection of mutants, while for the F110I mutation, a silent *Pvu*I restriction enzyme site was created. Both of these changes are indicated in the primers in Table 3.1. Oligonucleotides

were obtained from IDT (Coraville, USA) and were dissolved in TE buffer at 37°C to give a final stock solution concentration of 100 µM. The stock solutions were stored at -20°C. Working stocks of 2.5 µM were prepared.

Table 3.1**:** Forward and Reverse oligonucleotides used for the site directed mutagenesis of SpoIIAB. These oligonucleotides were synthesized with 5'-phosphorylation

Primer Name	Sequence (5′ to 3′)	Size (bp)	°C
H54N-FW	GAATGCGATTATC**AATGG**ATATGAAGAGAACTG	33	58.4
H54N-RW	GTGACAGCCTCTGACACGACTGTTTTG	27	61.3
F110I-FW	TGAG**CGATCG**GAATGGGCATCACCATT	28	64.3
F110I-RW	AGCTCAGGCTTACTCGTAAATAGAGGCTGAC	31	64.3

Blue=*Nco*I site removed; original site was CATGG

Green=*Pvu*I site created; the original sequence was CGCTCT

PCR amplification was used in both the mutagenesis reactions of spoIIAB, using the mutagenic primer pairs above. Agarose gel electrophoresis of 5 µl of each PCR reaction was used to verify the success of the PCR amplification. Each PCR product (~25 ng) was circularized with 0.5 µl Quick T4 DNA Ligase in a 5 minute reaction at room-temperature. The ligated DNA was transformed into *E.coli* DH5α by electroporation (see Section 3.2.1.4) and transformants selected on LM media containing 100 µg/ml ampicillin. A single colony was cultured in 5 ml LB, supplemented with 100 µg/ml ampicillin, and plasmid isolated using the E.Z.N.A. plasmid miniprep kit (Omega Bio-Tek, USA). Restriction enzyme analysis confirmed either the absence of the *Nco*I site (H54N) or the presence of the *Pvu*I site (F110I)..

3.2.1.14 Cloning, expression and purification of mutated SpoIIAB genes

For expression and purification of the H54N and F110I His_6-SpoIIAB mutants, the pET16b vector (Novagen, EMD Biosciences, Germany) with a N-terminal His_6-tag was used for expression. A large scale restriction enzyme digestion was set up for H54N and F110I mutants of SpoIIAB, 800 ng plasmid DNA was digested with 1 U *Nde*I (Fermentas, USA) at 37°C for 6 hours and followed by the addition of 1 U *Bam*HI at 37°C for 15 hours in the

restriction buffer supplied by Fermentas. The digested products were separated by electrophoresis on a 1.5% agarose gel, as described before in Section 3.2.1.2, and the fragments purified from the agarose gel using the Zymoclean gel DNA Recovery Kit (Zymo Research, California,USA). pET16b was similarly digested and purified. Ligation reactions were set up at a picomole insert:vector ratio of 20:1 using T4 DNA Ligase (Fermentas, Ontario, Canada) as per the manufacturer's instructions. The ligations were dried and washed and then transformed into *E. coli* DH5α electrocompetent cells by electroporation, (see Section 3.2.1.4). Protein expression and purification of H45N-SpoIIAB and F110I-SpoIIAB were performed in exactly the same manner than the non-mutated His_6-SpoIIAB protein (Section 3.2.1.7 to Section 3.2.1.9), after transferring the plasmids to the expression host *E. coli* BL21(DE3).

3.2.2 Dev two-component system proteins

In various studies, attempts to obtain soluble and functional full-length HKs have consistently failed, consequently most studies have been performed with truncated versions of HKs that lack their transmembrane domains (Haydel *et al.*, 1999). This was clearly evident when structural data for HKs were investigated; only the soluble domains had been crystallized. Transmembrane regions of some HKs are lethal to *E.coli* during expression, so truncation was used (Haydel *et al*., 1999). Since the immediate goal here was to demonstrate DevS's biochemical function, the DevS used in this investigation was truncated to create DevS201, as done by Saini *et al*. (2004) in an attempt to improve solubility. It was also decided to append the His-tag at both its C-terminal and N-terminal, in the hope of improving solubility. However, none of these modified DevS proteins were expressed in the soluble fractions of *E. coli* BL21(DE3) in sufficient quantities for successful purification.

3.2.2.1 Oligonucleotides for devR, devS and devT amplification

Oligonucleotides (see Table 3.2) were obtained from IDT (Coraville, USA) and dissolved in TE buffer at 37°C to give a final stock solution concentration of 100 µM. The stock solutions were stored at -20°C. Working stocks of 2.5 µM were prepared. The forward primers (see

Table 3.2, the primers labelled FW with green base pairs) were designed to incorporate an *Nde*I restriction site on the 5' end. The reverse primers (see Table 3.2, the primers labelled RW with blue base pairs) were designed to incorporate *Bam*HI restriction site on the 3' end. For DevS, additional sets of primers were designed. To truncate the protein to DevS201, forward primer C-DevS-FW (Table 3.2,) was used with DevS-RW to leave only the cytoplasmic ATP binding and dimerisation fragments. Primer DevSc-RW (Table 3.2) was designed to allow sub-cloning of DevS or DevS201 into pET20b, which contains a C-terminal His_6-tag.

Table 3.2: Forward and reverse oligonucleotide primers used for the amplification of different Dev gene fragments.

Primer Name	Sequence (5′ to 3′)[∂]	Size (bp)	Tm (°C)
DevR-FW	GTCATATGGTAAAGGTCTTCTTGGTCG	27	57.0
DevR-RW	GAGGATCCTCATGGTCCAYCACCG	24	60.9
DevT-FW	GTCATATGAGCGATCCTCGGCCAG	24	61.1
DevT-RW	GTGGATCCCTACAAGTTGGCACTCG	26	63.4
DevS-FW	GTCATATGACAACAGGGGGCCTCG	24	61.4
DevS-RW	GTGGATCCTACTGCGACAACGGTG	24	63.1
CDevS-FW	GTCATATGCGCGAACTCGACGTAC	24	59.6
DevSc-RW	GTGGATCCTGCGACAACGGTG	21	60.7

Blue*: Bam*HI cleavage site.
Green: *NdeI* cleavage site.

3.2.2.2 Cloning of Dev TCS proteins

M. tb strain H37Rv (ATCC 25618) genomic DNA was used as the source of the *dev* genes, and was kindly provided by Prof I Wiid and Prof P van Helden, University of Stellenbosch. Each PCR reaction contained 10 ng of template, 250 nM each of forward and reverse primers, 200 µM dNTPs and a proofreading Taq polymerase, TaKaRa ExTaq (TaKaRa Biomedicals, Shuzo, Japan). The PCRs were carried out in a Mastercycler Personal 5332 (Eppendorf Version 2.12.32) PCR machine with the following settings: 95°C for 5 minutes; followed by 30 cycles of: 95°C for 1 minute, 65°C for 1 minute, 72°C for 1 minute; and a final step at 72°C for 10 minutes. The sizes of the PCR products were confirmed by agarose gel electrophoresis using a 1.5% agarose gel (Section 3.2.1.2). The PCR products were purified from the agarose gel using the Zymoclean gel DNA Recovery Kit. The gel-purified PCR products were ligated into the pGem T-Easy Vector System (Promega Corporation,

USA) as per the manufacturer's instructions. The ligated DNA products were transformed by electroporation into *E. coli* DH5α electrocompetent cells and transformants selected on LM agar containing 100 µg/ml ampicillin, 1 mM IPTG and 20 mg/ml X-GAL for blue-white selection (see Section 3.2.1.3 and Section 3.2.1.4).

3.2.2.3 Sub-cloning the dev genes into pET vectors for expression and affinity purification.

Large scale restriction enzyme digestions were set up for all the pGEM-T Easy constructs: 800 ng of plasmid DNA was digested with 1 U *Nde*I (Fermentas, USA) at 37°C for 6 hours, followed by the addition of 1 U *Bam*HI at 37°C for 15 hours. The digested products were separated by electrophoresis on a 1.5% agarose gel and the desired fragments purified from the agarose gel using the Zymoclean gel DNA Recovery Kit. The pET16b vector (Novagen, EMD Biosciences, Germany) was chosen for the N-terminal His_6-tagged expression, while the pET20a vector (Novagen, EMD Biosciences, Germany) was used to obtain C-terminal His_6-tagged proteins. Ligation reactions of *devR*, *devS*, *devT* and *devS*201 with pET16b, as well as *devS*c and *devS*201c into pET20a, were set up at a picomole insert:vector ratio of 20:1 using T4 DNA Ligase (Promega Corporation, USA) as per the manufacturer's instructions. The ligated DNA was dried, washed and transformed into *E. coli* DH5α and subsequently BL21(DE3) by electroporation, as described in Section 3.2.1.3. The plasmid maps indicating the different constructs produced for DevS and DevS201 can be seen in Figure 3.3, while Figure 3.4 shows the plasmid maps for and DevR.

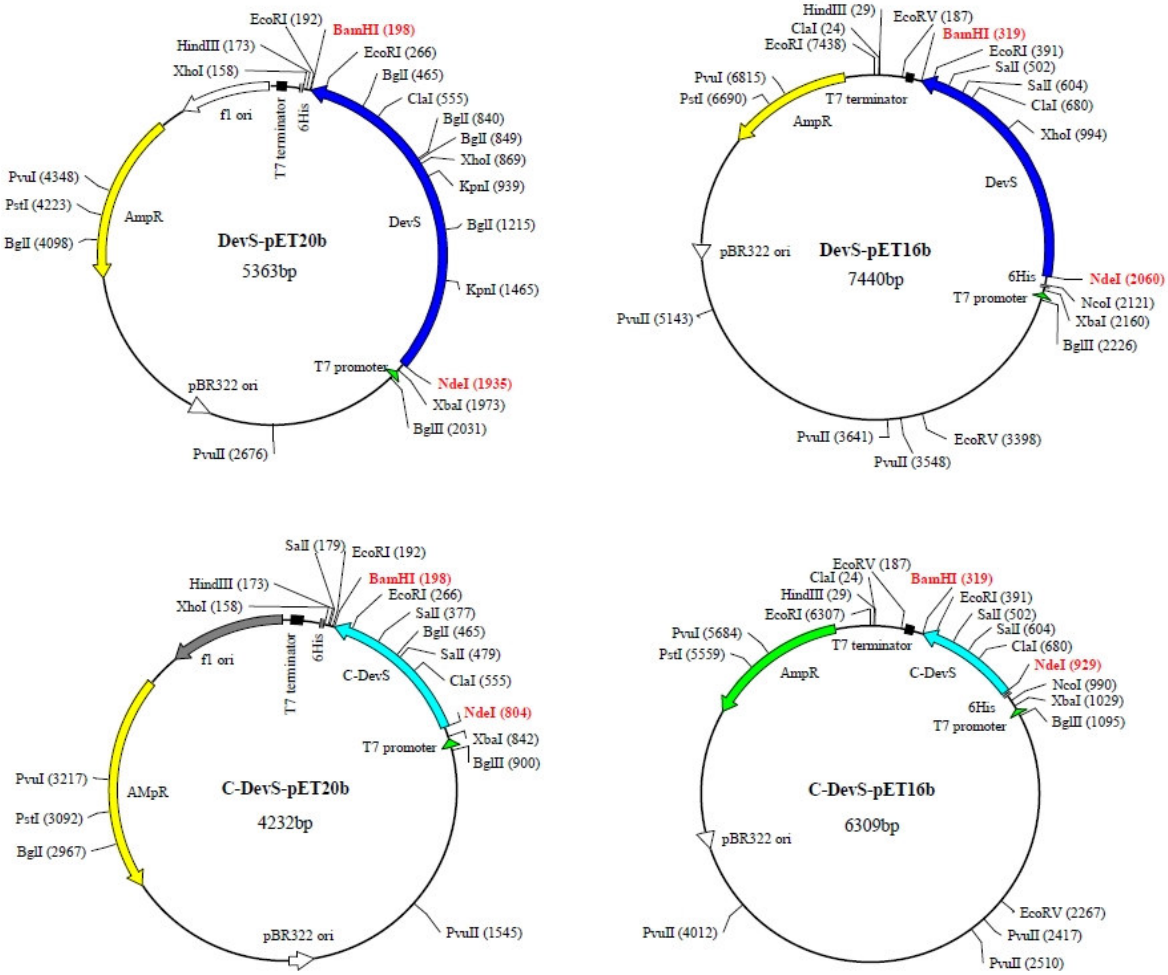


Figure 3.3: Plasmid maps for the different DevS gene constructs. The DevS indicates the full length DevS and the C-DevS is the truncated DevS201. As indicated, the pET20b vector contributes a C-terminal His_6-tag to the protein and the pET16b adds an N-terminal His_6-tag. The *Nde*I and *Bam*HI sites used for restriction analysis are colored red. The ampicillin resistance gene (bla) is highlighted in yellow/green.

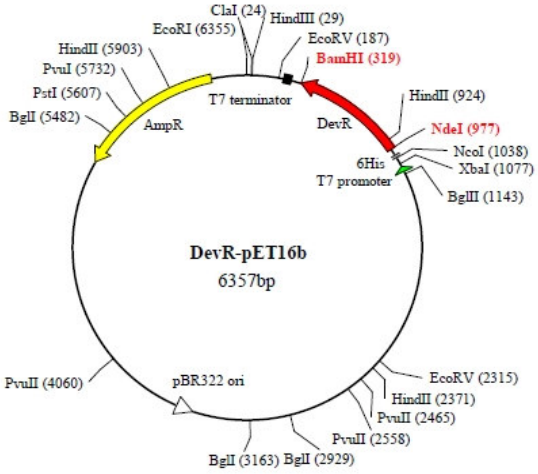


Figure 3.4: The plasmid map of the response regulator DevR in pET16b. The *Nde*I and *Bam*HI sites used for restriction analysis are colored red. The ampicillin resistance gene is highlighted in yellow.

3.2.2.4 Protein expression and purification

Protein expression, soluble protein extraction and purification was carried out as described in Sections 3.2.1.7 - 3.2.1.9. For DevR and DevT, sufficient soluble protein in the cell-free extracts was noted to continue with non-denaturing purification of His_6-tagged proteins as explained in Section 3.2.1.9. Neither truncation, nor the C-terminal nor N-terminal His_6-tags improved the solubility of the expressed DevS, and another avenue needed to be explored.

3.2.2.5 Matrix assisted refolding of DevS and DevS201

Protein expression of C- and N-terminal his-tagged DevS and DevS201 was carried out as described in Section 3.2.1.7. The biomass was sonicated and then centrifuged as described in 3.2.1.8. The pellet included the insoluble proteins sequestered in inclusion bodies, and was dissolved in denaturation buffer (using 8M urea as solubilisation agent) and incubated at 37°C for 2 hours, shaking at 220 rpm. The remaining insoluble proteins were removed by centrifugation for 5 minutes at 3,500*g*. The solubilised denatured proteins in the supernatant were loaded onto a 5 ml column containing pre-equilibrated HIS™-select nickel affinity gel (Sigma-Aldrich, USA) and incubated for 30 minutes at room temperature. The unbound protein was collected as flow-through. The column was washed with 3 column volumes of denaturation buffer containing 20 mM imidazole and then washed stepwise with renaturation buffer containing decreasing concentrations of urea (8 M to 1 M). The column was subsequently washed with 2 bed volumes of renaturation buffer without urea to further remove any contaminating *E. coli* proteins and remaining urea. The bound His_6-tagged protein was eluted with 20 ml renaturation buffer containing 250 mM imidazole, in fractions of 2 ml. The fractions were analyzed by SDS-PAGE (Section 3.2.1.10) and the peak fractions were pooled, dialyzed against PBS buffer, and stored at 4°C.

3.2.2.6 Radioactive phosphorylation assays

Phosphorylation assays using γ-^{32}P-labeled ATP (EasyTides), 5 µCi/µl packed in a 50 mM tricine buffer (pH 7.6) with green dye (Perkin Elmer Life and Analytical Science, Boston, USA) was used for the kinase assays. The phosphorylation of the Dev two-component system occurs in a two step reaction that is different from the SpoII system that was described in Section 3.1.12. According to Saini *et al*. (2002), the maximum signal for phosphorylated DevR (DevR~P) is obtained within 2 minutes and total dephosphorylation of DevR occurs within 10 minutes. To detect autophosphorylation of DevS, purified and refolded DevS (30 µM) was incubated with 5 µCi of [γ-^{32}P]ATP in 20 µl Dev phosphorylation buffer at 25°C. This autophosphorylation reaction was terminated after 1 hour with Laemmli loading buffer, followed by 5 minutes of heating at 90°C. For the phosphotransfer assays, DevR (15 µM) were added to the autophosphorylated kinase (DevS~P). The reactions were terminated after 2 minutes. The same phosphorylation reaction was repeated by replacing DevS with DevT, since it is hypothesized that autophosphorylation of DevT might also initiate phosphotransfer to DevR. These reactions were analyzed on 12% SDS-PAGE gel (as per Section3.2.1.10). After destaining, the gel was rinsed in gel storage buffer for about 15 minutes in order to try and prevent unnecessary cracking of the gels during the drying procedure. The gels were transferred to Whatman paper, covered with cling-wrap, and dried for 6 hours at 65°C using a slab gel dryer (SAVANT, Forma Scientific, USA). The radioactive gel was exposed to an X-ray film (Hyperfilm MP, Amersham) for 48 hours at -70°C. Subsequently the film was soaked for 2 minutes in Prefix fixer (Champion Photochemistry, RSA), washed for 2 minutes in water and left in PolyconA X-ray developer (May and Baker, Essex, UK) until the bands could be seen on the film.

3.3 Results

3.3.1 *Bacillus subtilis* SpoIIAA and SpoIIAB

A representative HK-like two-component signalling system, *B. subtilis* SpoIIAA and SpoIIAB, was chosen as a model to set up an investigation into HK active sites. After sub-cloning *spoIIAA* and *spoIIAB* from pEAA and pEAB into pET16b, the plasmids were transformed into *E. coli* DH5α cells via electroporation. Plasmid-containing cells were selected with 100 µg/ml ampicillin. After plasmid isolation, the presence of the *spoIIAA* and *spoIIAB* genes in the plasmids was confirmed with restriction enzyme digestion using *Nde*I and *Bam*HI. SpoIIAA and SpoIIAB were successfully expressed in the soluble fractions of *E. coli* expression host BL21(DE3), in sufficient quantities for affinity purification, as can be seen in the 14% SDS-PAGE gel shown in Figure 3.5. After the presence of the proteins in the soluble fractions was confirmed, the fractions were used in nickel affinity chromatography. From the 14% SDS-PAGE gel in Figure 3.6, it was evident that the protein affinity purification was successful, since there were bands of the expected size, at approximately 14 and 17 kDa. Contaminating bands were observed above 30 kDa molecular weight. The purified proteins were then dialyzed overnight in PBS

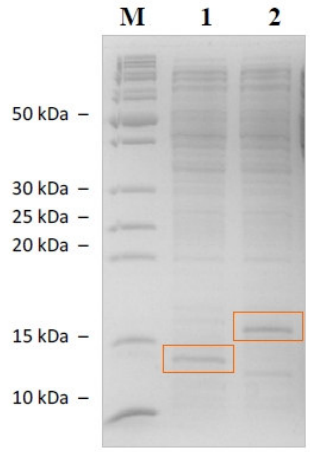


Figure 3.5: 14% SDS-PAGE analysis of the soluble protein expression of SpoIIAA and SpoIIAB. Lane M: Protein PAGERuler (Fermentas). Lane 1: SpoIIAA soluble fraction. Lane 2: SpoIIAB soluble fraction. The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, provided by Fermentas. The boxes indicate SpoIIAA at 14.2 kDa (lane 1) and SpoIIAB at 14.7 kDa (lane 2).

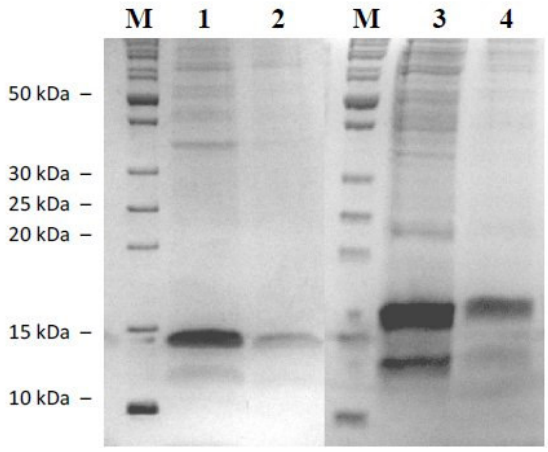


Figure 3.6: 14% SDS-PAGE analysis after affinity protein purification with HIS™-select nickel affinity gel . Lane M: Protein PAGERuler (Fermentas). Lane 1: Sample of SpoIIAA that was loaded on the HIS™-select nickel affinity gel. Lane 2: Purified SpoIIAA (20.7 µg/ml). Lane 3: SpoIIAA protein sample loaded on the HIS™-select nickel affinity gel. Lane 4: Purified SpoIIAB (40.8 µg/ml). The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, provided by Fermentas.

The phosphorylation of SpoIIAA by SpoIIAB was determined by means of radio-active phosphorylation assays. The end result of the SpoII phosphorylation assay is a stable phosphorylated protein, (SpoIIAB~P during autophosphorylation and SpoIIAA~P during trans-phosphorylation reactions), as shown in Figure 3.7. In Figure 3.7, lane 1 shows the autophosphorylated SpoIIAB, lane 2 indicates no autophosphorylation for SpoIIAA, while the trans-phosphorylation of SpoIIAA by SpoIIAB is shown in lanes 3 and lane 4.

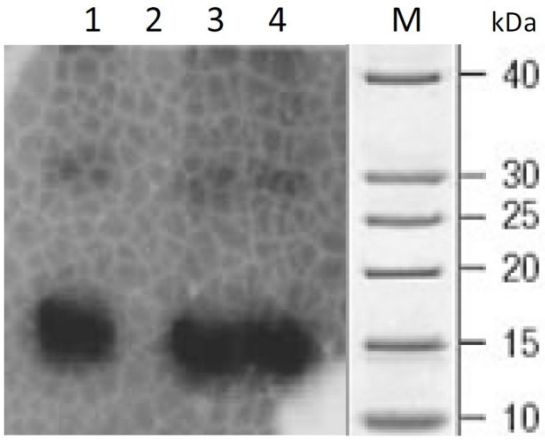


Figure 3.7: Radio-active phospho-transfer analysis of SpoIIAA and SpoIIAB protein activity. Lane 1: Autophosphorylation of SpoIIAB (kinase) Lane 2: Negative control SpoIIAA (substrate). Lane 3 and 4: SpoIIAA and SpoIIAB in a phosphorylation assay. Lane M indicates the molecular mass of the relevant bands on the PAGERuler, as provided by Fermentas.

In order to show that the site-directed mutagenesis of SpoIIAB was successful, *Nco*I (removed in H54N) and *Pvu*I (added in F110I) restriction enzyme digestions were done with both mutated and wild-type spoIIAB plasmid constructs, which can be seen in Figure 3.8. The H54N mutation removes one of the two *Nco*I restriction site present in the construct, therefore the mutant will present as a single band at 2.8 kb (Figure 3.8, lane 2) but the wild-type will have two bands (2.6 kb and 0.2 kb) seen in lane 3. The 0.2 kb band cannot be seen on the gel because it is too small. The F110I mutant adds a second *Pvu*I site; consequently the wild type has a single band at 2.8 kb (lane 6) and the mutant have two bands (1.9 kb and 0.9 kb) in lane 7. After sub-cloning the H54N- and F110I-mutated spoIIAB genes into pET16b for His-tagged expression, the mutated proteins were expressed as described for the wild-type, and expression analyzed on SDS-PAGE gels.

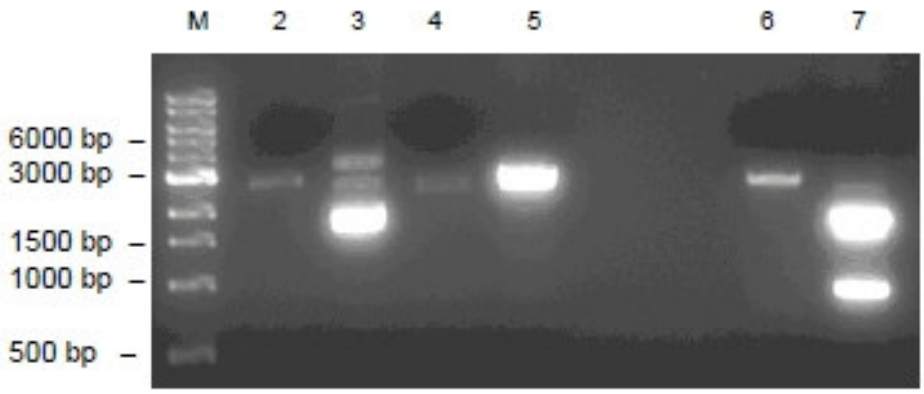


Figure 3.8: Restriction enzyme digestion of the DNA from mutated and non-mutated SpoIIAB. Lane M: 1 kb DNA ladder (New England Biolabs). Lane 2: H54N-SpoIIAB: *Nco*I. Lane 3: Wild type SpoIIAB: *Nco*I (2.6 kb and 0.2 kb not visible). Lane 4: Wild type SpoIIAB: *Xba*I. Lane 5: H54N-SpoIIAB: *Xba*I. Lane 6: Wild type SpoIIAB: *Pvu*I (single band 2.8 kb). Lane 7: F110I-SpoIIAB: *Pvu*I (1.9 kb and 0.9 kb).

The lack of phosphotransfer by the H54N and F110I to SpoIIAA, compared to the control phosphorylation SpoIIAB, was shown by means of radio-active phosphorylation assays (Figure 3.9). Lane 2 shows the autophosphorylations of SpoIIAB and the consequent transphosphorylation of SpoIIAA in Lane 5. Autophosphorylation was absent for the H54N mutant in lane 3, with no transphosphorylation in lane 6. Similarly, no autophosphorylation is seen in lane 4 for the F110I mutant and its phosphotransfer reaction between F110I and SpoIIAA (lane 7).

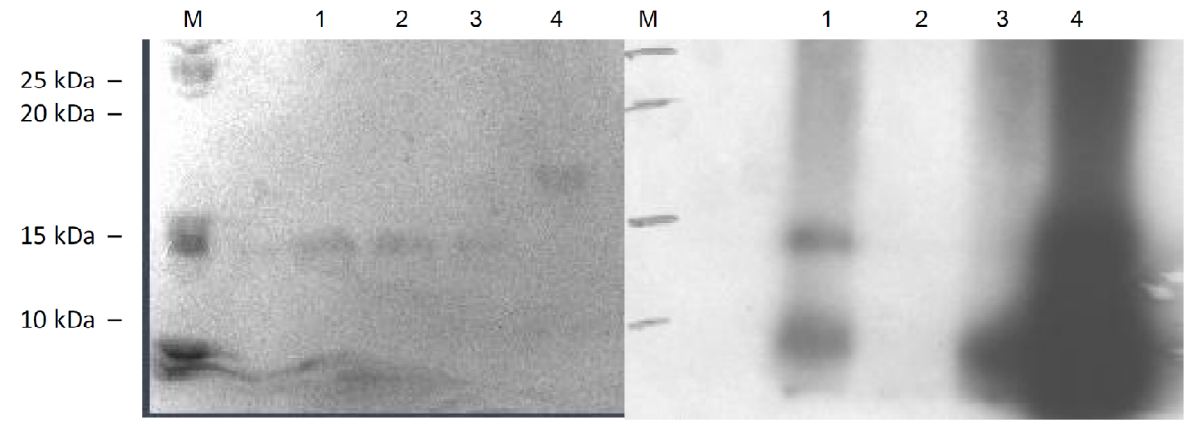


Figure 3.9: Phosphorylation analysis: The purification of the mutants on a 14% SDS-PAGE (right) and the radiogram of the experiment (left). One the 14% SDS-PAGE (right): Lane M: PageRuler Protein Ladder. Lane 1: Purified SpoIIAB. Lane 2: Purified H54N. Lane 3: Purified F110I. Lane 4: Purified SpoIIAA. The radiogram of the experiment: Lanes M: PageRuler Protein Ladder. 1: SpoIIAB (autophosphorylation reaction). Lane 2: H54N (autophosphorylation reaction). Lane 3: F110I (autophosphorylation reaction). Lane 4: Phosphotransfer reaction between SpoIIAB and SpoIIAA. The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas. On the radiogram the 15 kDa 20 kDa and 50k Da bands of the PAGE ruler were indicated during development with a permanent marker.

3.3.2 *M. tb* two-component signal transduction systems

The full-length *devR*, *devS* and *devT* genes, as well as the truncated *devS*201, were amplified by PCR, with the primers designed to flank the genes in the genome of *M. tb* H37Rv strain. The final PCR products are shown in Figure 3.10. After cloning the PCR products into pET16b (*devR*, *devS*, *devS*201 and *devT*) and pET20a (*devS*c and *devS*201c), respectively, the presence of the *dev* genes in the plasmid was confirmed with restriction enzyme digestion (data not shown).

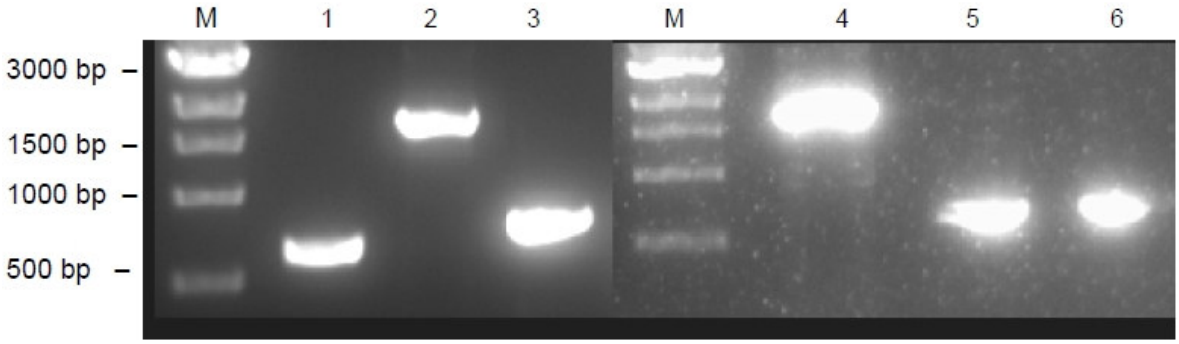


Figure 3.10: PCR products from gene amplification from *M. tb* H37Rv strain. Lanes M: 1 kb Base pair ladder (New England Biolabs). Lane 1: *devR* (654 bp). Lane 2: *devS* (1737 bp). Lane 3: *devT* (807 bp). Lane 4 *devS* (1737 bp) Lane 5: *DevS201* (650 bp). Lane 6: *devS*201c (650 bp).

The full length N-terminal His_6-tag DevR and DevT proteins were both expressed in the soluble fractions of *E. coli* BL21(DE3), as in Figure 3.11, lanes 1 and 3, respectively. DevS was only present in the insoluble fraction as can be seen in Figure 3.11, lane 5. Soluble DevR and DevT were successfully purified by affinity purification, using HIS™-select nickel affinity gel chromatography. From the SDS-PAGE gel in Figure 3.12, it was evident that the purification of DevR (lanes 2 and 3) as well as DevT (lanes 4 and 5) was successful. Initial attempts at soluble expression of sufficient quantities of both full length DevS and truncated DevS210 did not succeed, whether containing an N- or C-terminal His-tag, although there were high levels of expression of insoluble protein, as can be seen in Figure 3.13 (lanes 4-6).

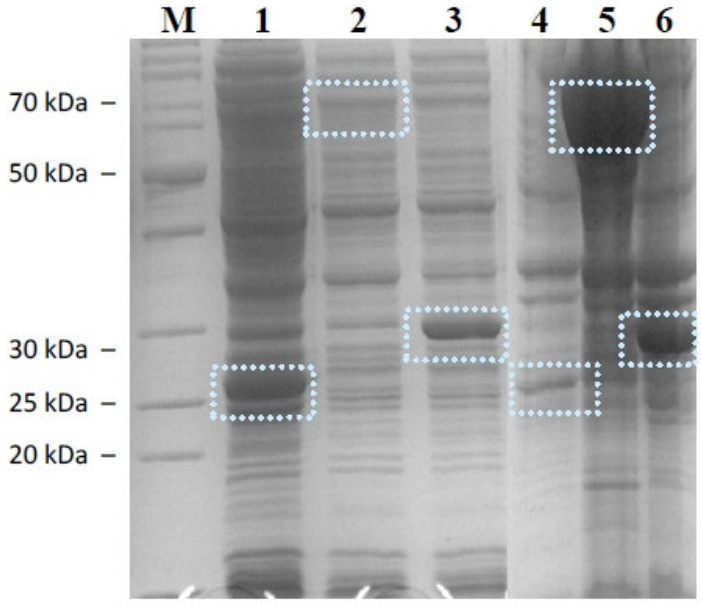


Figure 3.11: 12% SDS-PAGE analysis of protein expression of DevR, DevS and DevT. Lane M: Protein PAGERuler (Fermentas). Lane 1: DevR (25 kDa without His_6-tag) soluble fraction. Lane 2: DevS soluble fraction (62.3 kDa without His_6-tag). Lane 3: DevT soluble fraction (28 kDa without His_6-tag). Lane 4: DevR insoluble fraction (pellet). Lane 5: DevS insoluble fraction (pellet). Lane 6: DevT insoluble fraction (pellet). The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas.

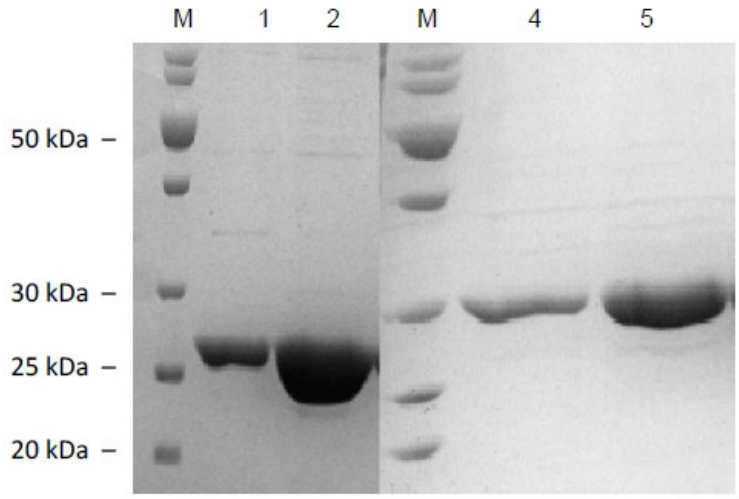


Figure 3.12: 12% SDS-PAGE analysis of DevR and DevT proteins purified by affinity purification with HIS™-select nickel affinity gel. Lane M: PageRuler Protein Ladder. Lane 1: Purified DevR. Lane 2: Purified DevR Lane M: PageRuler Protein Ladder. Lane 4: Purified DevT Lane 5: Purified DevT. The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas.

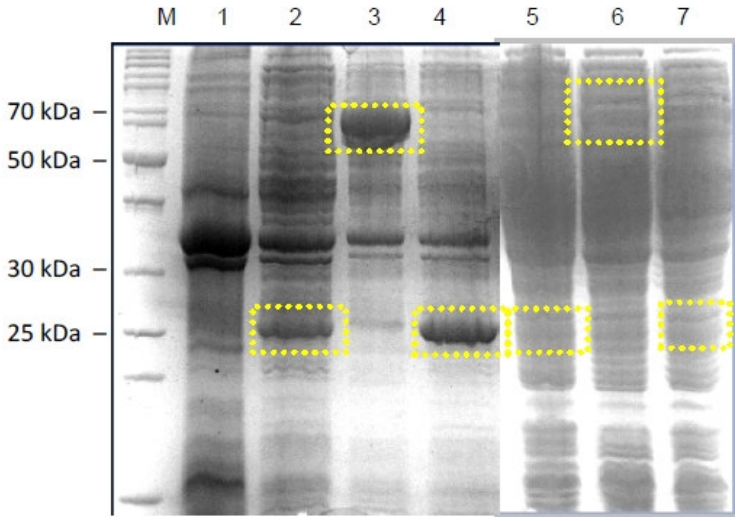


Figure 3.13: 12% SDS-PAGE analysis of protein expression of the C-terminal His_6-tagged DevSc, the truncated N-terminal His_6-tagged DevS210 and the truncated C-terminal His_6-tagged DevS210c. Lane M: PageRuler Protein PAGERuler (Fermentas). Lane 1: BL21(DE3) pellet fraction without any construct. Lane 2: DevS210c insoluble fraction (26.5 kDa). Lane 3: DevSc insoluble fraction (62.3 kDa). Lane 4: DevS210 insoluble fraction (26.5 kDa). Lane 5: DevS210c soluble fraction. Lane 6: DevSc soluble fraction. Lane 7: DevS210 soluble fraction. The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas.

Another method to resolve the insolubility of DevS was to use matrix assisted refolding, which was carried out according to Saini *et al*. (2002). As indicated by Figure 3.14, this method worked very efficiently for all the constructs DevSc, DevS201 and DevS201c.

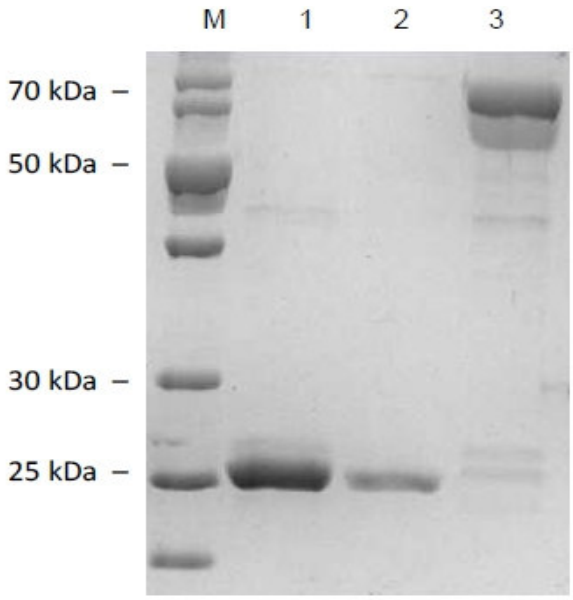


Figure 3.14: 12% SDS-PAGE analysis of refolded truncated N-terminal His_6-tagged DevS201, truncated C-terminal His_6-tagged DevS201c and full-length C-terminal His_6-tagged DevSc proteins, purified by matrix assisted refolding on HIS™-select nickel affinity gel. Lane M: PageRuler Protein Ladder. Lane 1: Purified DevS201 (180 µg/ml). Lane 2: Purified DevS201c (82 µg/ml). Lane 3: Purified DevSc (520 µg/ml). The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas.

After the successful expression of the kinases and substrate proteins had been confirmed on SDS-PAGE gels, the protein activity was identified by means of radio-active phosphorylation assays. Firstly, the autophosphorylation capacity of the different DevS constructs needed to be determined, in order to ascertain whether the truncation caused any loss of autophosphorylation ability. This reaction is problematic because the phosphorylated form of DevS construct is transient. Figure 3.15 shows no autophosphorylation, which may be due to the unstable nature of the phosphorylated kinase. However, trans-phosphorylation of DevR was successfully shown. Figure 3.16, lane 3 shows DevR phosphorylation when combined with DevS201, showing that truncation of DevS does not interfere with its functionality. No phosphorylation of DevR was shown with either the full length DevS or the DevT kinase.

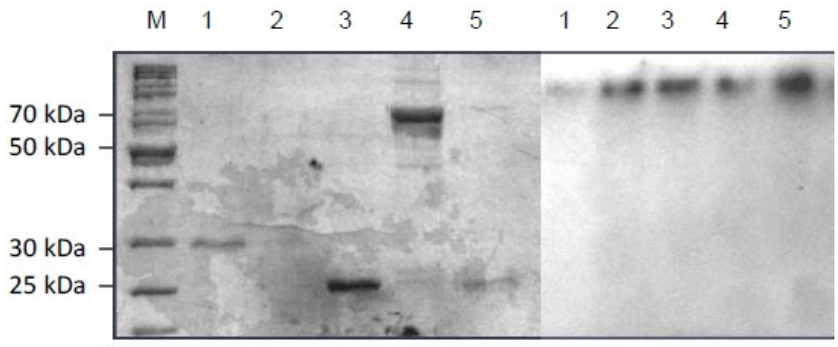


Figure 3.15: Autophosphorylation assay: 12% SDS-PAGE (left) and the radiogram of the assay (right). Lane M: PageRuler Protein Ladder (Fermentas). Lane 1: DevT (sample 1). Lane 2: DevT (sample 2). Lane 3: DevS201. Lane 4: DevSc. Lane 5: DevS201c.

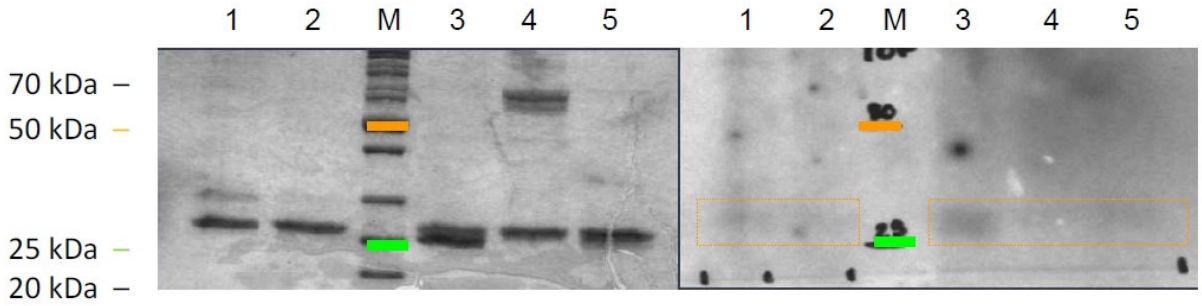


Figure 3.16: Trans-phosphorylation assay: 12% SDS-PAGE (left) and the radiogram of the assay (right). Lane M: PageRuler Protein Ladder. Lane 1: DevT (sample 1). Lane 2: DevT (sample 2) Lane 3: DevS201. Lane 4: DevSc. Lane 5: DevS201c. The molecular mass of the relevant bands on the PAGERuler is indicated on the left side, as provided by Fermentas. On the radiogram the 25 kDa and 50k Da bands of the PAGE ruler were indicated during development.

3.4 Discussion

At the outset, the aim was to use the SpoII TCS as a model system to demonstrate functional kinase expression and allow for the optimization of the radio-active phosphorylation kinase assay. The following was successfully carried out:

- The histidine kinase SpoIIAB and the substrate SpoIIAA was successfully cloned, expressed and purified using HIS™-select nickel affinity gel.
- The autophosphorylation of SpoIIAB and the subsequent phosphotransfer to SpoIIAA was visualized using radio-labelled ATP.
- The protocols for performing site-directed mutagenesis were developed and tested by making two SpoIIAB mutations.

- The sensitivity of the phosphorylation assay was measured by comparing the wild type SpoIIAB activity to the mutant SpoIIAB activity.

Functional expression and purification of the Dev TCS components was undertaken: N-terminal His_6-tagged DevR and DevT were expressed in moderate levels of solubility and the proteins purified on a native HIS™-select nickel affinity gel. Both N-terminal and C-terminal His_6-tagged DevS were expressed insolubly and aggregated to form inclusion bodies.

It has been demonstrated that removal of the N-terminal membrane-associated region of DevS should also increase solubility (Saini *et al*., 2004a). Two different truncated forms of DevS were therefore produced: N-terminal His_6-tagged truncated DevS (DevS201) and C-terminal His_6-tagged truncated DevS (DevS201c). Neither of the truncations improved the solubility of the DevS, but denaturing HIS™-select nickel affinity purification in the form of matrix assisted refolding led to the production of soluble DevS protein (Saini *et al*., 2002). The proteins were denatured with high concentrations of urea (6M-8M) and then loaded onto the HIS™-select nickel affinity gel. By systematically decreasing the concentration of urea (from 8M tot 0.5M) in the buffer, the proteins were allowed to slowly refold to their native conformation while still bound to the column. When the concentration of urea reached zero, the refolded proteins were eluted from the column with an increased concentration of imidazole.

The question remained as to whether DevS201 has the same functionality as DevS with regards to autophosphorylation and phosphotransfer. The autophosphorylation data was inconclusive. As shown in Saini *et al*. (2004a), the phosphorylated DevS is intrinsically stable; even though various attempts was made to get the same result it was unsuccessful. Saini *et al*. (2002) showed that phosphorylated DevS in the presence of DevR was labile and rapidly transferred its phosphoryl group to generate the phosphorylated DevR. DevS autophosphorylation appears to be transient. Consequently no autophosphorylation of DevS could be seen in the phosphorylation reactions as phosphorylated DevS was not sufficiently stable or the phosphoryl group was transferred to DevR before the reaction could be stopped. The positive results from the phosphotransfer assays between DevS201 and DevR proved functionality of the truncated form of the kinase. This allowed for further investigations to be carried out on the DevR and DevS interactions

Chapter 4 *In silico* and *in vitro* DNA binding analysis of DevR

4.1 Introduction

4.1.1 Known information on how DevR binds DNA

4.1.1.1 The 20 bp palindromic consensus sequence

The DevR response regulator functions as a transcription factor of the two-component response regulator class (Dasgupta *et al*., 2000). Park *et al*. (2003) searched for shared sequence motifs upstream of hypoxic response genes by doing a computer analysis with the motif discovery program YMF (Sinha and Tompa, 2002). The analysis uncovered a 20 bp palindromic consensus sequence 5′-TTSGGGACTWWAGTCCCSAA-3′ (where S = C/G; W = A/T), a variation of which could be found upstream of 84% of the *M. tb* genes which are rapidly induced at least twofold by hypoxia (Park *et al*., 2003). Previous research suggested that bases G_4, G_5, G_6, and C_8 in the 20 bp palindromic consensus sequence are crucial for DNA interaction with DevR (Park *et al*., 2003; Wisedchaisri *et al*., 2008). Mutating G_4 to A could be tolerated for DevR and DNA interaction, but no activity was present as it is the site of phosphorylation by DevS. In contrast, G_5, G_6, and C_8 nucleotides were conserved in most of the motifs binding to DevR, forming the basis of recognition (Chauhan and Tyagi, 2008a). NarL, a well-studied response regulator and transcription factor, interacts specifically with its cognate DNA motif in its phosphorylated form. Similarly, phosphorylated and not unphosphorylated DevR forms a DNA interaction with sequences containing the palindromic consensus sequence (Chauhan and Tyagi, 2008a). The Asp54 site of phosphorylation and other amino acids involved in DNA binding can be seen in Figure 4.1. The goal of this investigation was to resolve the regulatory function of DevR, by determining the binding of DevR to recognition sequences in *narK* and *hspX*, both hypoxia-regulated genes. NarK is a

nitrate/nitrite transporter protein. The increase in nitrate reduction under hypoxic conditions is attributed to the increase in nitrate/nitrite transport (Chauhan and Tyagi, 2008a). The intergenic regions of *narK* contain three putative Dev boxes (recognition motifs), to which DevR can bind (Chauhan and Tyagi, 2008a). During latency, expression of the hspX protein is up-regulated. The *hspX* promoter sequence includes multiple DevR binding sites which overlap with the -35 promoter element (Chauhan and Tyagi, 2008b). Since both of these elements contains DevR binding sites and were previously proven to bind DevR, they were used to study the influence that mutations will have on DevR binding.

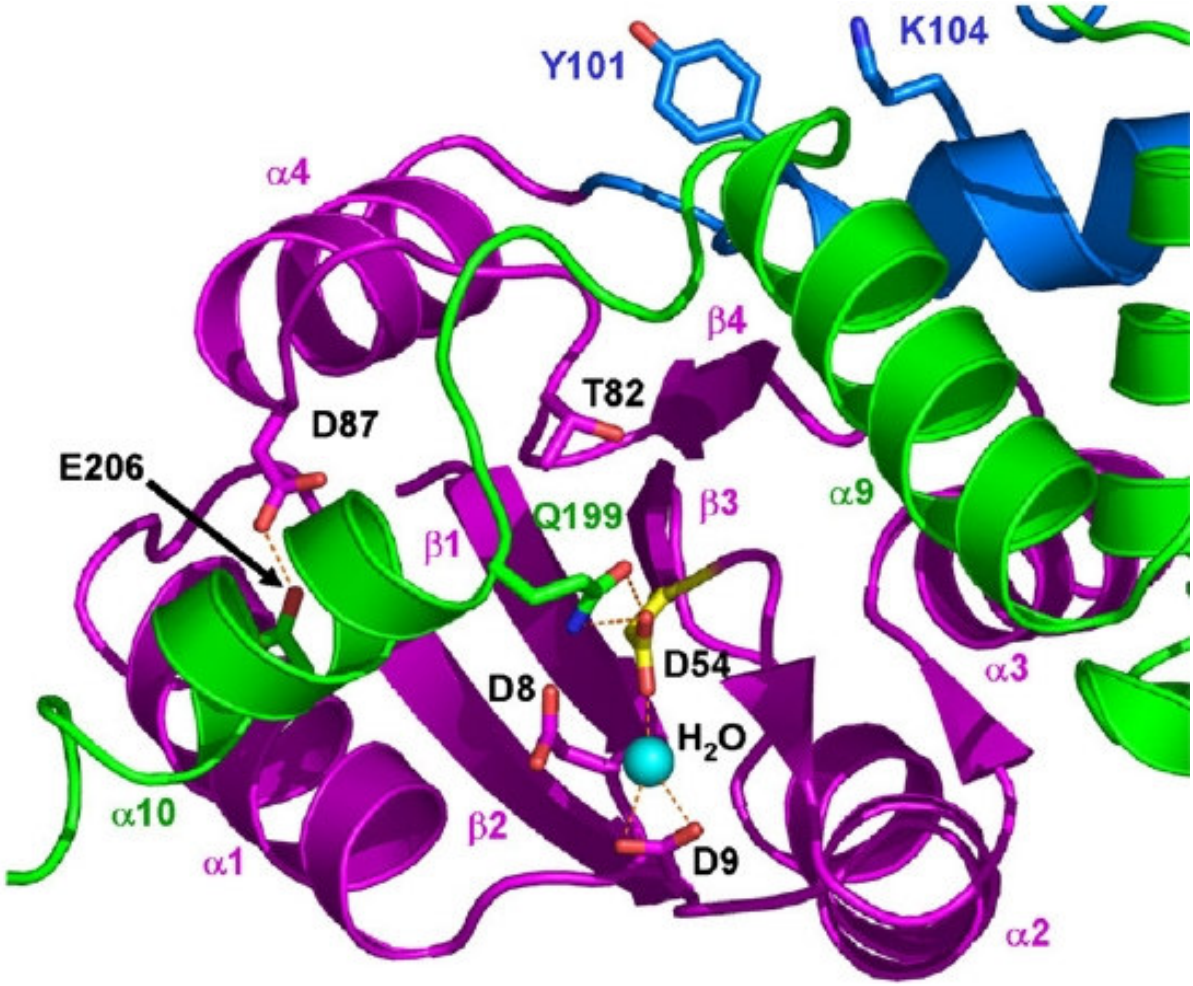


Figure 4.1: The DevR phosphorylation site. The side chains of Asp54 (highlighted in yellow), Thr82 and Asp87 in the N-terminal domain (highlighted in purple), Tyr101 and Lys104 in the linker (highlighted in blue), Gln199 and Glu206 in the C-terminal domain (highlighted in green) are shown as stick models (Wisedchaisri *et al*., 2008). These amino acids were hypothesized by Wisedchaisri *et al*. (2008) to play a role in the trans-phosphorylation of DevR by DevS.

4.1.1.2 Structural information on DevR

Wisedchaisri *et al*. (2008) reported crystal structures of full-length unphosphorylated DevR at 2.2 Å resolution and its C-terminal DNA binding domain at 1.7 Å resolution. DevR forms a dimer related by a non-crystallographic 2-fold axis (Figure 4.2). The two-helix linker (helices α5 and α6) from both subunits form a four-helix, as indicated in blue and gray in Figure 4.2. The second part of the dimerization interface comes from interactions between the N-terminal domain helices α2 and α3 of one subunit and the C-terminal domain helices α7 and α8 of the other subunit (Wisedchaisri *et al*., 2008). Wisedchaisri *et al*. (2005) hypothesized that the following amino acids play a role in the DevR–DNA complex: Lys179, Lys182, and Asn183 could interact with nucleotide bases, while Arg196, Thr198, Val202, and Thr205 could contribute to the dimerization interface. Further experimental analysis is necessary in order to understand the binding of DevR to these consensus sequences and to determine the nature of these interactions.

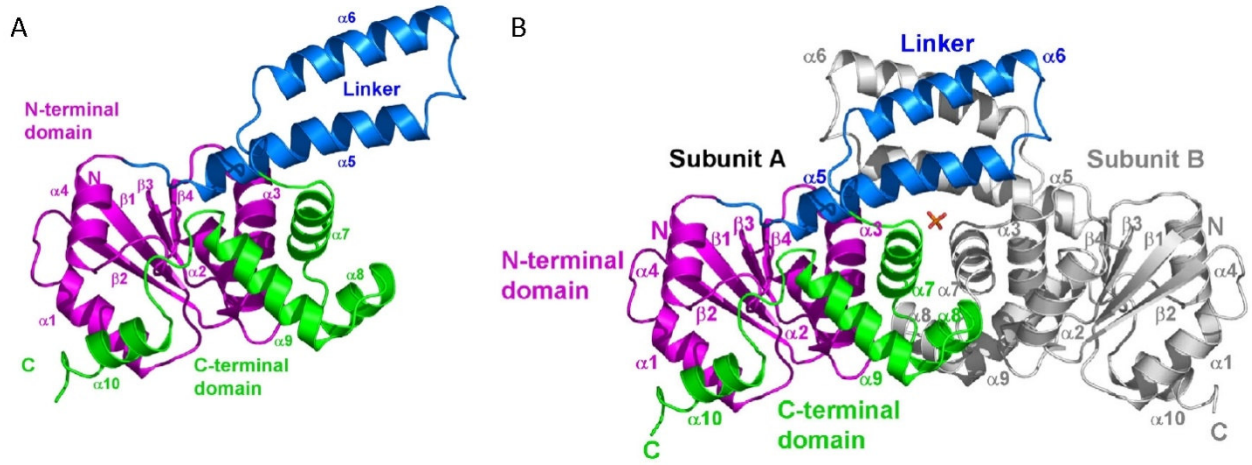


Figure 4.2: Crystal structure of full-length DevR. (A) The DevR monomer. The structure is shown in purple for the N-terminal domain, blue for the linker, and green for the C-terminal DNA binding domain. (B) The DevR dimer, subunit A is shown in (A). Subunit B is shown in gray. A sulfate molecule (shown as a stick model) is found in the dimer interface between the two DevR molecules (Wisedchaisri *et al.,* 2008).

4.1.2 Electrophoretic Mobility Shift Assay (EMSA)

The interaction of proteins with DNA is central to many cellular processes. EMSA is a technique used to study DNA-protein interactions. This technique is based in the observation

that DNA-protein complexes migrate more slowly than unbound DNA fragments when electrophoresed on a native polyacrylamide gel, and is also referred to as mobility shift or gel retardation assays (Fried and Crothers, 1981; Garner and Revzin, 1981).

4.2 Methods

4.2.1 DevR binding DNA modelling

Structural data of DevR binding to a 20 bp palindromic sequence containing inverted repeats of the $G_4G_5G_6A_7C_8T_9$ recognition motif were available in two different PDB database structures: 1ZLK and 3C3W (Wisedchaisri *et al*., 2008). Discovery Studio (DS) was used to visualize all the hydrogen bonds that could potentially form between the protein and DNA. DS was also used to model mutations in the identified amino acid residues. The structural modelling of the mutations was followed up by biochemical investigations in order to test predictions made using the model regarding residues that may be important to the binding of DNA.

4.2.2 Generation of PCR fragments for DNA-protein binding

M. tb strain H37Rv (ATCC 25618) genomic DNA was used as the source of the *narK* and *hspX* target DNA fragments. The primers used for PCR amplification of these DNA fragments are shown in Table 4.1. A temperature gradient PCR was set up for all the different constructs to determine the optimal annealing temperatures. The optimum annealing temperature was determined to be 53.3°C. Each PCR reaction contained 50 ng template DNA (received from Prof P v Helden, University of Stellenbosch, South Africa), 250 nM of each forward and reverse primers (from IDT Coraville, USA) as shown in Table 4.1, 200 µM dNTPs and 1 U TaKaRa ExTaq (TaKaRa Biomedicals, Shuzo, Japan). The PCR was carried out in a Mastercycler (Personal 5332 Eppendorf Version 2.12.32) with the following settings: 95°C for 5 minutes was followed by 30 cycles of: 95°C for 1 minute, 53.5°C for 1 minute, 72°C for 1 minute, and a final extension at 72°C for 10 minutes. The size of the PCR products was confirmed by agarose gel electrophoresis using a 1.5% agarose gel (as per

Section 3.2.1.2). The PCR products of the respective target DNA fragments were purified from the agarose gel with the Zymoclean gel DNA Recovery Kit (see Section 3.2.1.5).

Table 4.1 HspX and NarK primers designed to generate PCR fragments for DNA-protein binding

narKF	5'- GACACGATCCGGGGTCTCG -3'
narKR	5'- TCCCCTTTCCAGTGGCGACC -3'
hspXF	5'- TCTGAACGGCGGTTGGCAGACA -3'
hspXR	5'- CGGGAAGGGTGGTGGCCATTTG -3'

4.2.3 Site directed mutagenesis (SDM) of DevR

Point mutations in DevR were created to elucidate the identified amino acids role in DNA-protein binding: Lys179 was changed to isoleucine (K179I), Lys182 was changed to isoleucine (K182I), Asp183 to leucine (N183L), and Tyr205 to valine (T205V). The Phusion Site Directed Mutagenesis Kit (Finnzymes, Espoo, Finland) was used as described by the manufacturer. Mutations of the polar amino acids K179I, K182I and N183L to non-polar amino acids could elucidate their potential role in the binding of DevR to the DNA regions. The T205V mutation is located on the DevR dimerization interface and it might interfere with DevR dimerization or the binding to DevS during phosphorylation. For the K179I mutation, a *Pvu*I restriction enzyme site was created by silent mutation to allow for selection of mutants, while for the K182I and N183L mutation, a *Sna*BI restriction enzyme site was created. For the T205 mutation, an *Afl*II restriction enzyme site was created to allow for selection of mutants. All of these changes are indicated in Table 4.2. The primers were synthesized by IDT (Coraville, USA) and were dissolved in TE buffer at 37°C to give a final stock solution concentration of 100 µM. Working stocks of 2.5 µM were prepared. PCR amplification was used in all the mutagenesis reactions, using the mutagenic primers in Table 4.2 and *M. tb* strain H37Rv (ATCC 25618) genomic DNA was used as the template. The PCR parameters were adjusted to suit the Phusion Hot Start DNA polymerase: the enzyme activation took place at 98°C for 30 seconds, followed by 25 cycles of: denaturation at 98°C for 30 seconds, annealing for 3 minutes at 71.9°C (for K179I, K182I and N183L) or 67.6°C (T205V), and extension at 72°C for 30 seconds, followed by a final extension was 72°C for 5 minutes. During ligation, 25 ng of each purified PCR product was circularized with 0.5 µl Quick T4

DNA Ligase in a 5 minute reaction, with 2x Quick ligation buffer at room-temperature. These ligation reactions were transformed into *E. coli* DH5α electrocompetent cells by electroporation and transformants selected on LB agar containing 100 µg/ml ampicillin (see Section 3.2.1.4) A resultant colony was cultured in 5 ml LB media containing 100 µg/ml ampicillin, and the E.Z.N.A. plasmid miniprep kit (Omega Bio-Tek, USA) was used to obtain plasmid DNA for restriction enzyme verification. The plasmids from the respective mutations were digested with the selected restriction enzymes to confirm the presence of the mutations. Following the confirmation of the silent mutations, the mutations were confirmed by sequencing (Inqaba Biotec, Pretoria, South Africa), and the plasmids transformed into the *E. coli* expression host Bl21(DE3) cells by electroporation.

Table 4.2: PCR primers used to insert point and silent mutations in DevR.

Mutation	Primers	Wild type codon	Restriction enzyme site for screening
K179I	F: 5′-TAGCCGAAATCACGGTGAAGAACTACGTGT-3′	AAG	
	R: 5′-GGAACATTCGATCGGCGATCTGCTT-3′		PvuI
K182I	F: 5′-AGACGGTGATCAACTACGTATCGCGGTTG-3′	AAG	SnaBI
	R: 5′-TTTCGGCTAGGAACATTCGGTCG-3′		
N183L	F: 5′- AGACGGTGAAGCTCTACGTATCGCGGTTG-3′	AAC	SnaBI
	R: 5′-TTTCGGCTAGGAACATTCGGTCG-3′		
T205V	F: 5′-GTATTCGCGGTAGAGCTTAAGCGCTCG-3′	ACG	AflII
	R: 5′-CGCGGCTTGCGTCCGA-3′		

The wild type codons correspond to the codons on the primer where the point mutations were made to change to amino acid (highlighted turquoise). The codons highlighted in grey were changed to form silent mutations, inserting a restriction enzyme site that can be used for screening.

4.2.4 DevR-DNA Binding assay

Wild-type DevR, DevS and DevT, as well as the DevR mutants prepared in Section 4.2.3, were expressed and purified as previously described in Section 3.2.2.4 and 3.2.2.5. Phosphorylation of DevR was performed in two ways: chemical phosphorylation with acetyl phosphate, or enzymatic phosphorylation with DevS. Acetyl phosphate has been shown to activate response regulators as efficiently as their corresponding kinases (McCleary and Stock, 1994). The DevR-DNA binding reaction took place in standard EMSA binding buffer, with 15-20 ng *hspX* or *narK* DNA, 10 - 40 µg DevR protein and 50 mM acetyl phosphate. For enzymatic phosphorylation, 20 µg DevS or DevT was used and 9.6 mM ATP was added. The reactions were incubated at 37°C for 30 minutes, then 2 µl Fermentas Orange DNA Loading Dye was added to stop the reaction.

4.2.5 Native PAGE

Non-denaturing TBE-polyacrylamide gels were used to resolve the DNA-DevR interactions. A 4% native stacking gel and a 5% running gel were used, with gels pre-run at a constant voltage of 80 V until the current has stabilised, using 0.5% TBE. This was done to remove all traces of ammonium persulphate, to evenly distribute the ions throughout the gel and to maintain a constant gel temperature during electrophoresis. The 20 µl DevR-DNA binding reactions from Section 4.2.4 were separated at 120 V in a Bio-Rad Mini Protean3 electrophoresis system, in 0.5% TBE running buffer. The gel was stained with ethidium bromide in 0.5% TBE solution for one hour and the DNA bands were visualized and photographed under UV light using the Bio-Rad ChemiDoc system (Bio-Rad Laboratories, USA), with QuantityOne software.

4.2.6 Phosphorylation of DevR via Acetyl phosphate and Mass Spectrometry (LC-MS using C4 Jupiter column)

Mass spectrometry was done by Dr. S Stoychev (CSIR, Pretoria, South Africa). Two samples were provided: unphosphorylated DevR in PBS buffer pH 7.4 (concentration of 300 µg/ml therefore 11.5 pmol/µl) and DevR phosphorylated with acetyl phosphate in MS phosphorylation buffer pH 7.4 (concentration of 144 µg/ml equivalent to 5.5 pmol/µl). The sample analysis was done as follows: 2 µl of unphosphorylated DevR (approximately 20 pmol per experiment) and 10-30 µl of phosphorylated DevR (approximately 55-65 pmol per experiment) were loaded. Protein desalting was done on a C4 Jupiter Reverse phase column (1x150 mm) at 150 µl/min using a gradient wash step starting at 20% MS buffer B, increasing to 80% MS buffer B over a period of 10 minutes, then a continuous wash at 80% MS buffer B for 5 minutes and then decreasing the concentration again to 20% MS buffer B (where buffer A: 0.5% acetonitrile and 0.1% formic acid and buffer B: 100% acetonitrile and 0.1% formic acid). The C4 column is connected via a 6 port switch valve to QSTAR-ELITE with Turbo-Ion source installed. Data acquisition performed in TOF-MS mode using a scan

range of 600-1700 mz. During data processing the protein isotopic profiles were deconvoluted by the "Bayesian Protein Reconstruct" tool found in Analyst QS v2.0.

4.3 Results

4.3.1 *In silico* modelling of DevR binding DNA

In silco modelling of DevR binding to a DNA target identified three amino acids that potentially play a role in the binding of DevR to the promoter DNA of genes it regulates, namely Lys179, Lys182 and Asp183 (Figure 4.3). Thr205 was thought to play an important role in the dimerization of the DevR molecules upon binding DNA. The changes in DevR protein structure caused by the mutations K179I, K182I and N183L were modelled *in silico*, and the collective effect is shown in Figure 4.4. The wild type DevR is shown in Figure 4.4 (A). A positively charged region comprising the two lysines and the asparagine clustered together, is shown in blue. The red indicates a more negatively charged region and white indicates a neutral charge. When the K179I, K182I and N183L mutations are made (B), there is a net loss of charge in that region of the protein indicated by the white coloured region. Spatially, the bulky side chains of the wild type K179 and K182 (C) are replaced by a smaller side chain I179 and I182 (D). Both the charge and the size differences could possibly affect the binding of DevR to its target DNA.

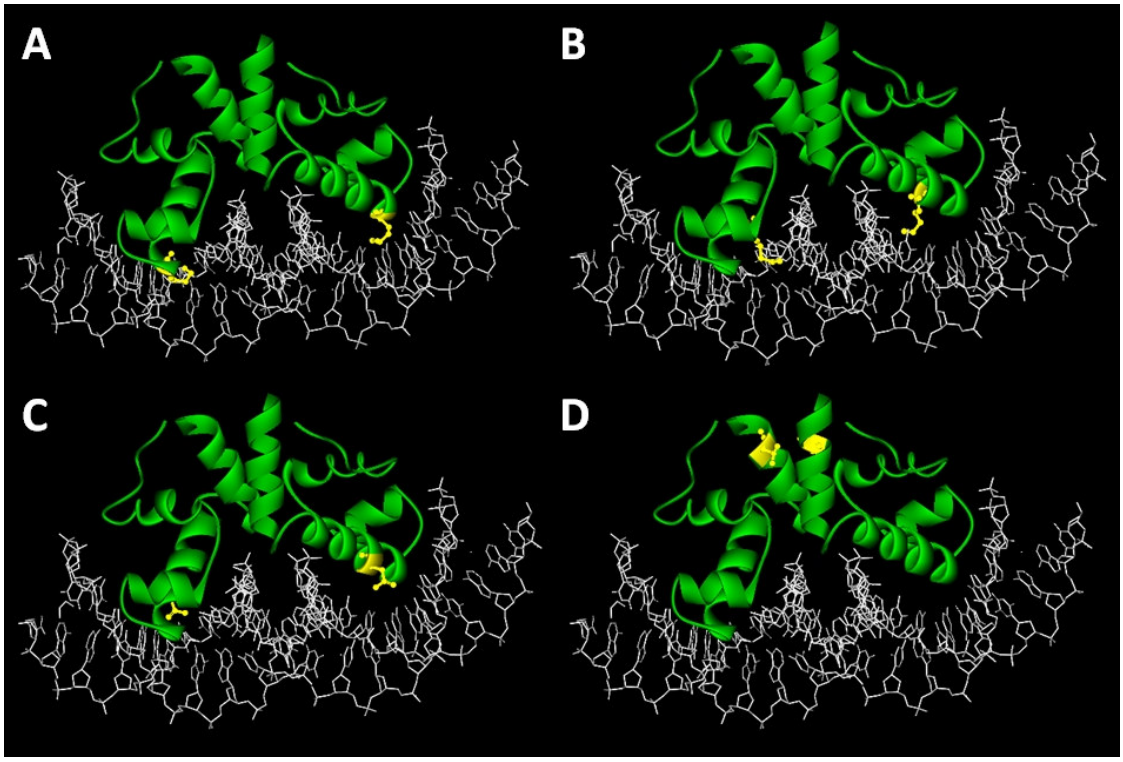


Figure 4.3: *In silico* modelling of DevR interacting with a target DNA fragment identified amino acids to be mutated: K179I, K182I, N183L and T205V. The amino acids that were mutated are highlighted in yellow A) K179, B) K182, C) N183. In D) the T205 that sits on the interface of the dimer can be seen.

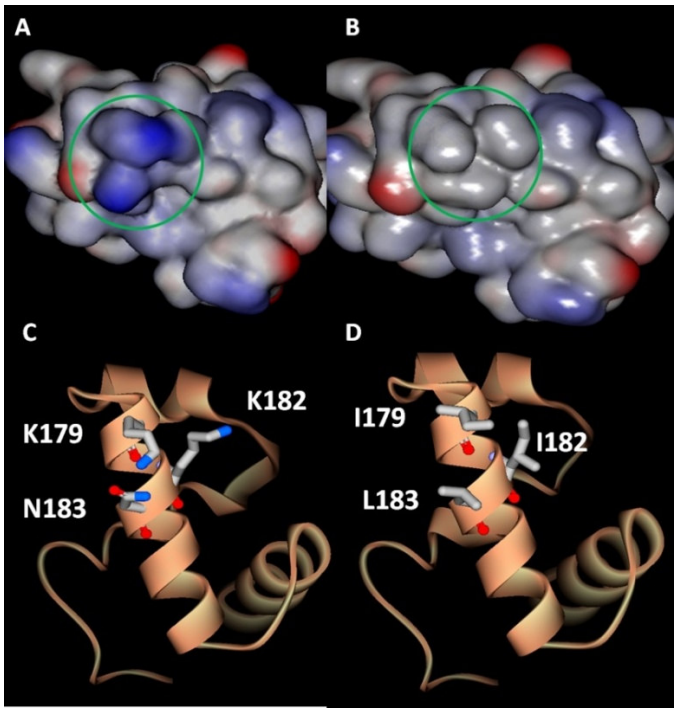


Figure 4.4: Implications of mutations on the charge and size of the DevR. (A) shows the wild type DevR and (B) shows the mutated DevR with emphasis on the charge change that occurs during mutations. The more positive charged regions are shown in blue, while red indicates more negatively charged regions and white regions are uncharged. (C) shows the wild type having more bulky side chains than the mutant DevR (D). K179 and K182 have larger side chains than I179 and I182, similarly N183 is bulkier than L183.

4.3.2 Electrophoretic Mobility Shift Assay

The *narK* and *hspX* target DNA fragments were amplified by PCR, with the primers designed to flank the DNA fragments in the genome of *M. tb* H37Rv strain. The final PCR products are shown in Figure 4.5. The *hspX* and *narK* fragments had expected sizes of 186 bp and 277 bp, respectively.

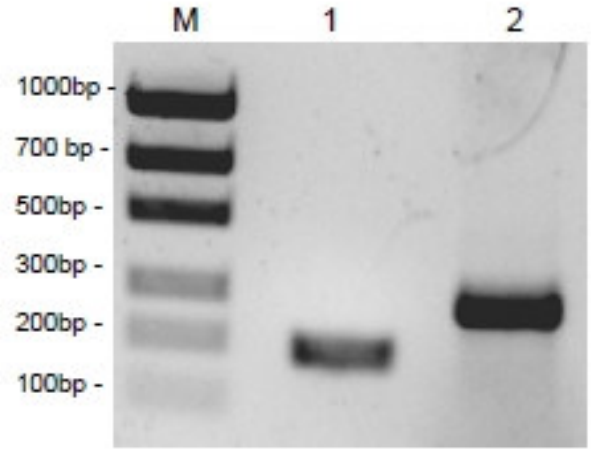


Figure 4.5: PCR products from NarK and HspX amplification from *M. tuberculosis* H37Rv strain. Lane M: MassRuler Express, Low range (forward) (Fermentas). Lane 1: HspX 186 bp. Lane 2: NarK 277 bp.

Dev proteins were purified using matrix assisted refolding, and results are shown in Figure 4.6. Phosphorylation of DevR can be done in two ways, either chemically by using acetyl phosphate or enzymatically by using DevS or DevT and ATP (McCleary and Stock, 1994; Saini *et al*., 2004b). Results shown in Figure 4.7 indicate that the acetyl phosphate and DevS methods were equally effective, with the phosphorylated DevR resulting in the same binding (shift) effect for both ways of phosphorylation. In Figure 4.7, the shift seen in lanes 1-3 is as result of DevR being phosphorylated by acetyl phosphate. Lanes 5-7 shows the resulting shift caused by DevT phosphorylation, which was much less effective. Different concentrations of DevS and truncated DevS201 were used to phosphorylate DevR and the resulting shifts can be seen in Figure 4.7 (lanes 8-16). Since using acetyl phosphate is less variable, it was decided to continue using that protocol.

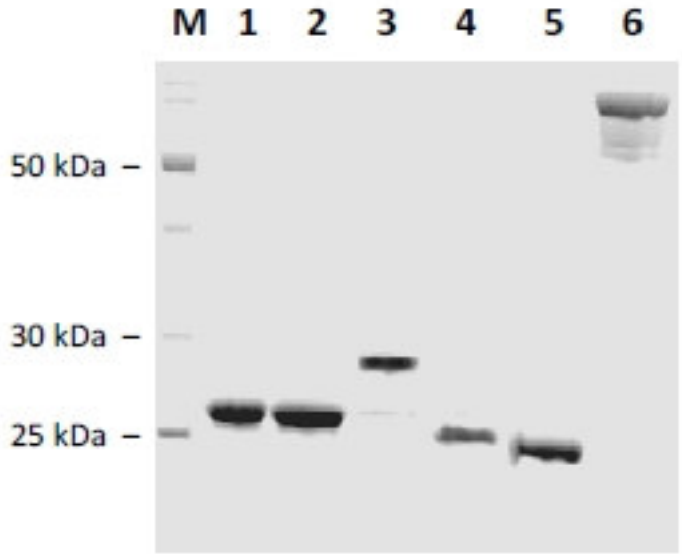


Figure 4.6: A 12% SDS-PAGE analysis of refolded DevR, DevT, truncated N-terminal His_6-tagged DevS201, truncated C-terminal His_6-tagged DevS201c and full-length C-terminal His_6-tagged DevSc proteins, purified by matrix assisted refolding on HIS™-select nickel affinity gel. Lane M: PageRuler Protein Ladder (Fermentas). Lane 1-2: Purified DevR. Lane 3: Purified DevT. Lane 4: Purified DevS201. Lane 5: Purified DevS201c. Lane 6: Purified DevSc.

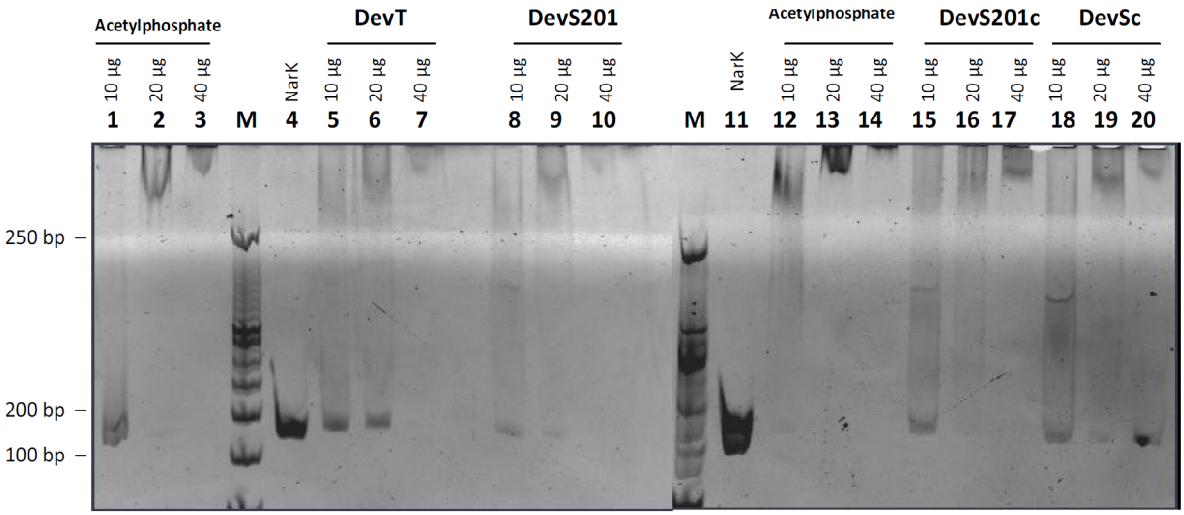


Figure 4.7: Comparison of the phosphorylation methods used to phosphorylate DevR. Lane M: O'GeneRuler™ 50 bp DNA Ladder (Fermentas). Lane 1- 3: NarK with DevR (phosphorylated by 50 mM acetyl phosphate). Lane 4: NarK without DevR. Lane 5-7: NarK with DevR phosphorylated by DevT. Lane 8-10: NarK with DevR phosphorylated by DevS201. Lane 11: NarK without DevR. Lane 12-14: NarK with DevR phosphorylated by 50 mM acetyl phosphate. Lane 15-17: NarK with DevR phosphorylated by DevS201c. Lane 18-20: NarK with DevR phosphorylated by DevSc. The DevR concentration is indicated. The NarK concentration was kept constant.

4.3.3 Site Directed Mutagenesis of DevR

The first step in producing mutants of DevR was to amplify the gene and incorporate the mutation by using 5'-phosphorylated mutagenic primers. The PCR reaction was subjected to agarose gel electrophoresis to verify the success of the PCR amplification (as seen in Figure 4.8 These PCR reactions produced linear plasmids with the mutated genes. Before transformation, the linear plasmids had to be circularised.

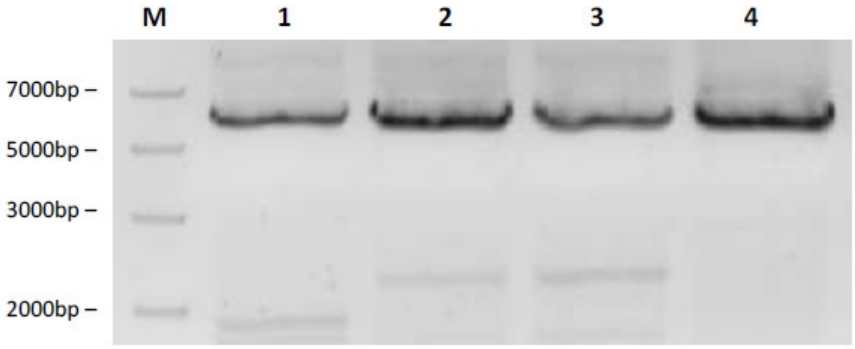


Figure 4.8: PCR products from mutant DevR amplification from *M. tuberculosis* H37Rv strain. The expected size from all four amplifications is 6354 bp. Lane M: MassRuler Express, High range (forward) (Fermentas). Lane 1: K179I. Lane 2: K182I. Lane 3: N183L. Lane 4: T205V.

The circularised plasmids were transformed into BL21(DE3) and expression was monitored via SDS-PAGE, as seen in Figure 4.9. K179I, K182I, N183L and T205V were used in the binding assays with both HspX and NarK DNA fragments. During electrophoresis, the *hspX* and *narK* DNA fragments were used as negative controls and the shift that takes place when phosphorylated wild-type DevR binds to the *hspX* or *narK* fragments was used as a positive control. In Figure 4.10, the presence of a shift can be seen when K182I and T205V binds to the *narK* fragment. No shift takes place when K179I and N183L were used, indicating that binding to the DNA was prevented. The same results can be seen in Figure 4.11 when the *hspX* DNA fragment is used.

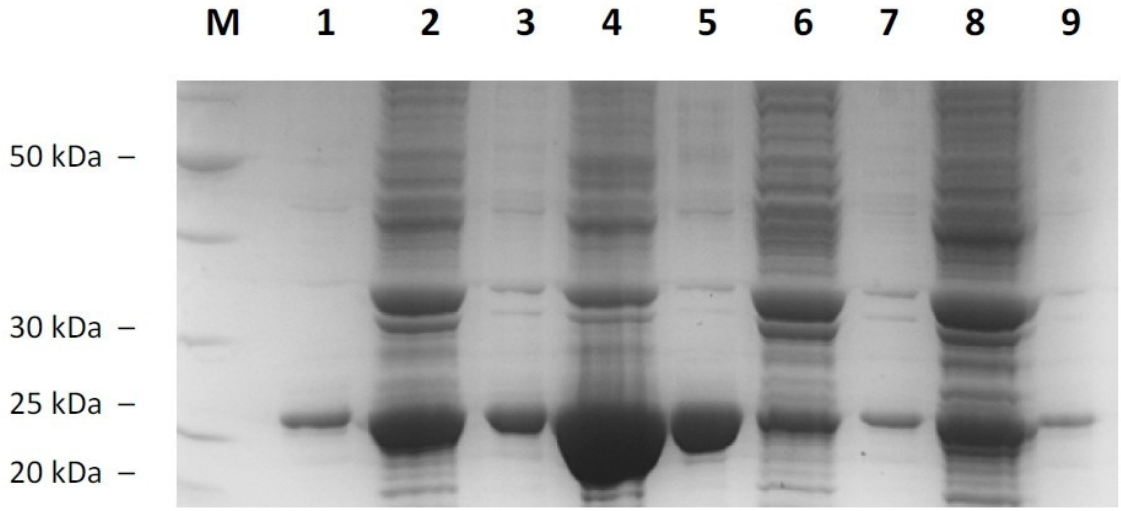


Figure 4.9: Expression and purification of DevR mutants K179I, K182I, N183L and T205V. Lane M: PageRuler Protein Ladder (Fermentas). Lane 1: Purified DevR. Lane 2: K179I (cell free extract). Lane 3: Purified K179I. Lane 4:K182I (cell free extract). Lane 5: Purified K182I. Lane 6: N183L (cell free extract). Lane 7: Purified N183L. Lane 8: T205V (cell free extract). Lane 9: Purified T205V.

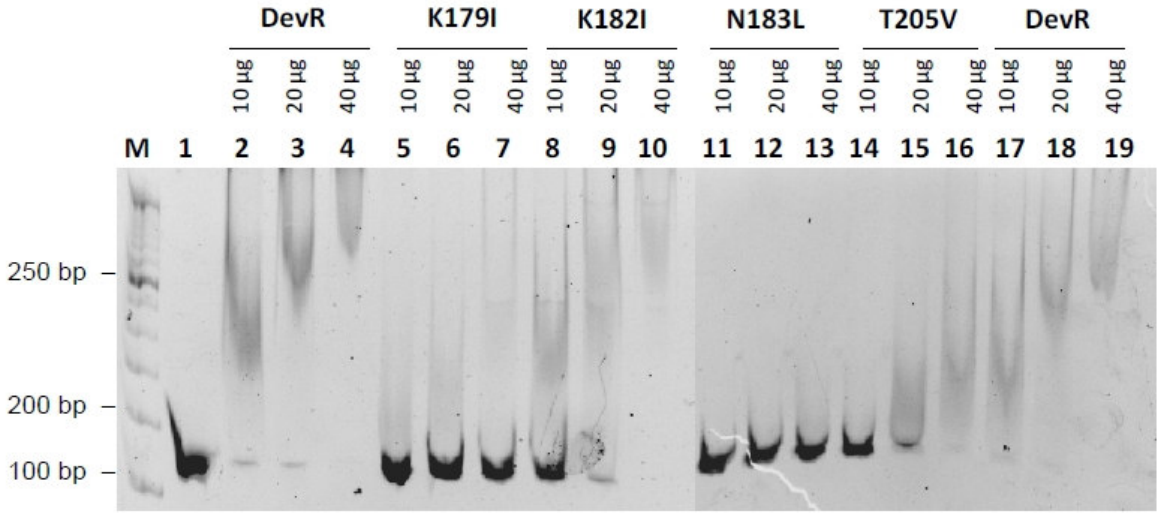


Figure 4.10: Analysis of DevR and mutants binding with NarK DNA fragment. Lane M: O'GeneRuler™ 50 bp DNA Ladder (Fermentas). Lane 1: NarK without DevR. Lane 2-4: NarK with DevR. Lane 5-7: NarK with K179I. Lane 8-10: NarK with K182I. Lane 11-13: NarK with N183L. Lane 14-16: NarK with T205V. Lane 17-19: NarK with DevR. Concentrations of protein are indicated and the DNA concentrations were kept constant.

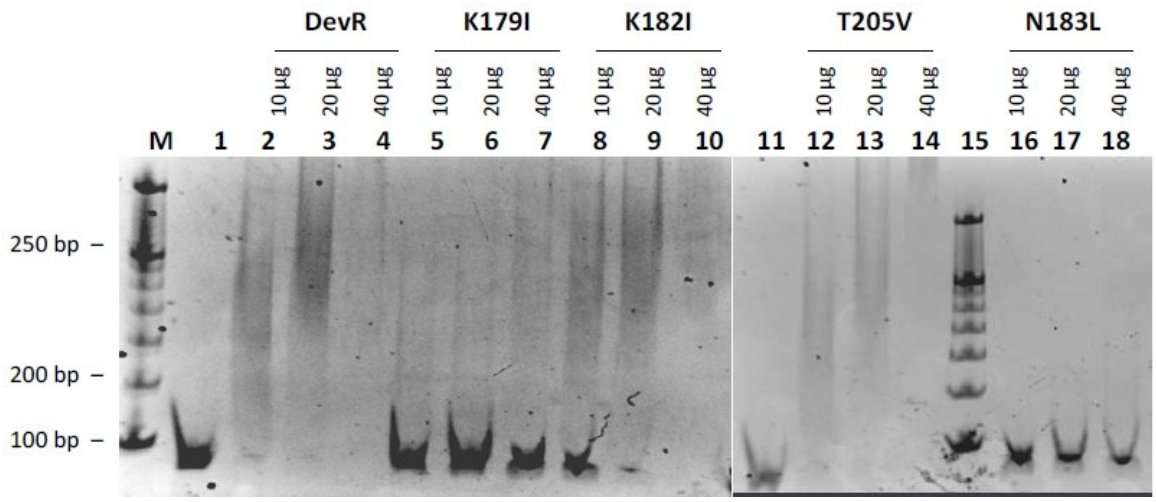


Figure 4.11: Analysis of DevR and mutants binding with HspX DNA fragment. Lane M and 15: O'GeneRuler™ 50 bp DNA Ladder (Fermentas). Lane 1 and 11: HspX without DevR. Lane 2-4: HspX with DevR . Lane 5- 7: HspX with K179I. Lane 8-10: HspX with K182I. Lane 12- 14: HspX with T205V. Lane 16-18: HspX with N183L. Concentrations of protein are indicated and the DNA concentrations were kept constant.

4.3.4 Mass spectrometry

The aim of the MS experiments was to prove the acetyl phosphate does phosphorylate the DevR. Radioactively-labelled ATP was used in Chapter 3 to show that DevS201 can phosphorylate DevR. No radioactively-labelled acetyl phosphate could be obtained commercially, so MS was used to monitor phosphorylation. Phosphorylation causes a molecular mass shift of 80 Da. The MS results of the acetyl phosphate-treated DevR showed that the protein appears to be phosphorylated at multiple sites, as there is an increase in molecular mass in steps of 80 Da, as seen in Figure 4.12. The inspection of DevR revealed 18 aspartic acids, as indicated in Figure 4.13, these could explain the multiple phosphorylations on DevR, as acetyl phosphate phosphorylates at aspartic acids. On the isotopic profiles for phosphorylated DevR there are only 14 adducts, which could be explained if solvent accessibility, as shown in Figure 4.13 (B), is considered.

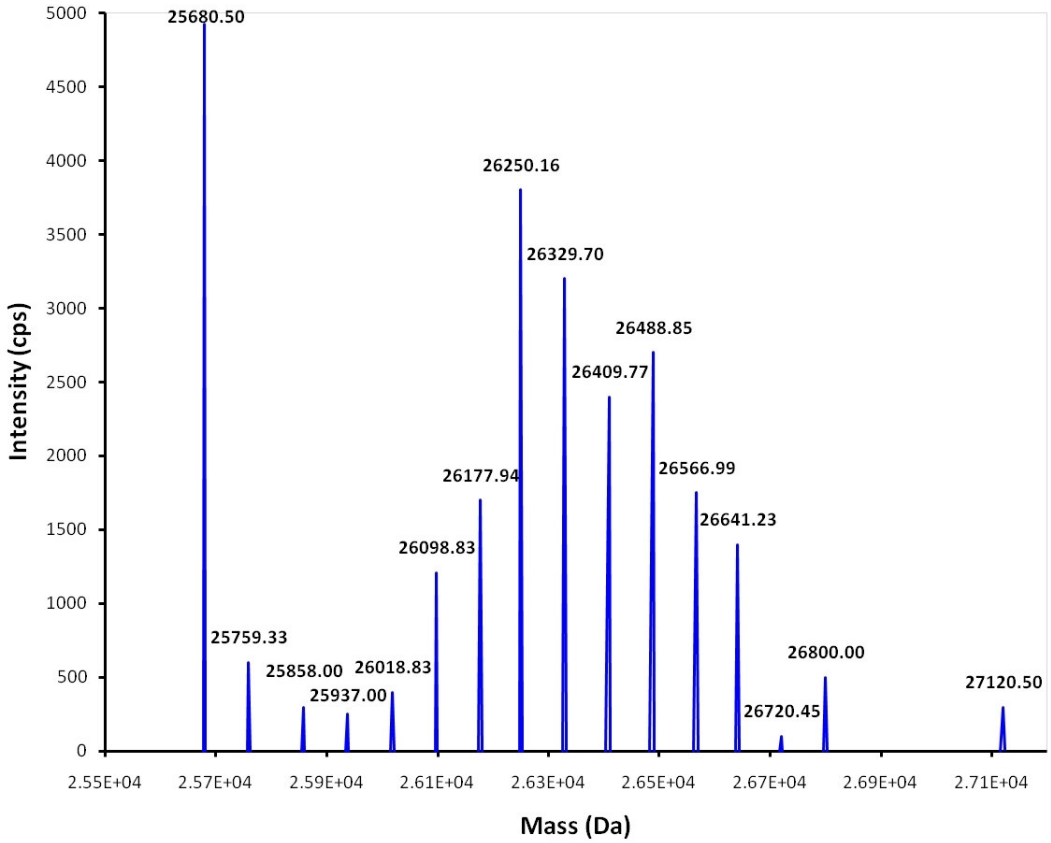


Figure 4.12: The protein isotopic profiles for phosphorylated DevR (with acetyl phosphate). The first peak is unmodified DevR, with an average molecular weight of 25.6805 kDa. A number of modifications can be seen 80 Da apart, which is assumed to be various phosphorylation products.

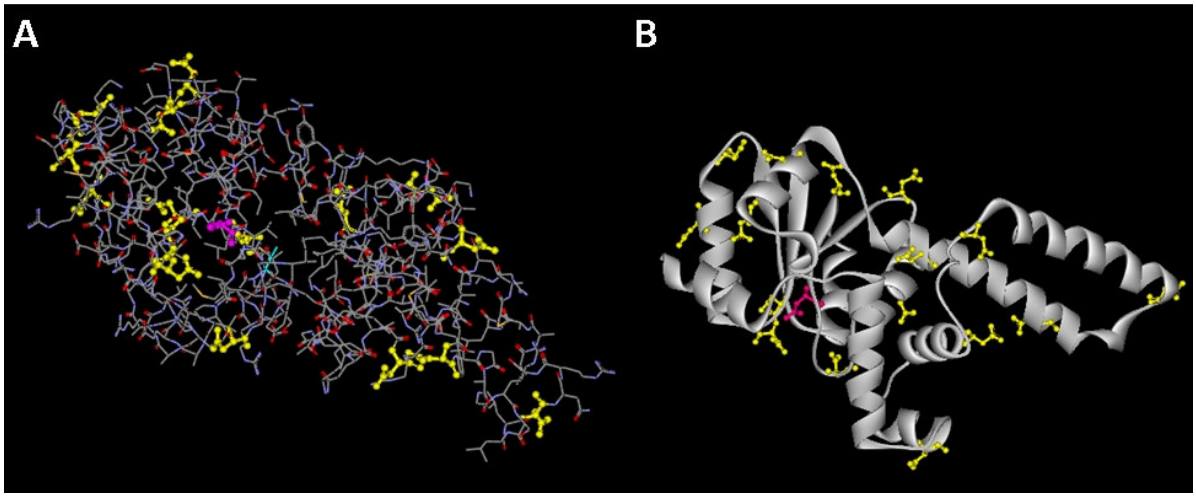


Figure 4.13: All aspartic acids in DevR that could potentially be phosphorylated by acetyl phosphate. (A) Highlights 18 aspartic acids in DevR in yellow, Asp54 is highlighted in pink. (B) Shows that some of the aspartic acids are not solvent accessible, this can explain the peaks on the isotopic profiles for phosphorylated DevR.

4.4 Discussion

Wisedchaisri *et al*. (2005) suggested a set of amino acids that could play a role in the binding of DevR to the consensus sequence in the promoter DNA of the genes it regulates. The only structure available is the truncated C-terminal domain of DevR binding to a 20 bp palindromic sequence containing inverted repeats of the GGGACT recognition motif. To understand the functionality of binding of DevR to DNA consensus sequences as well as the nature of these interactions the following was undertaken:

EMSA was used to identify sequence specific DNA-protein interactions. Mutation studies were used to identify which amino acid residues are involved in the sequence specific DNA-protein interactions. Mutations were modelled *in silico* and mutant DevR proteins were produced to confirm the effect.

The structure of the truncated DevR has provided useful insight into the DNA binding in this domain, however a structure of the full length DevR interactions is lacking. The full length DevR dimerization and DNA binding needs to be investigated. Therefore, a model was built of the full length DevR dimer binding to the DNA consensus sequence.

From the model, it was also postulated that the following amino acids were necessary for DevR binding to DNA: K179, K182 and N183. In addition, it was proposed that T205 was needed for dimerization. When the mutations K179I, K182I, N183L and T205V were modelled, hydrogen bonding side chains were no longer available, the local charge changed from partial positive to neutral, and the smaller side chains had a stereochemical influence. Subsequently these single mutations were made to DevR and assessed *in vitro* to validate the model.

The mutated DevR proteins were analyzed *in vitro* to determine their influence on the binding of DevR to DNA. During EMSA analysis of the K179I and N183L mutants, no shift was observed, indicating the role of K179 and N183 in DNA binding. Therefore, inhibitors interacting with these two amino acids may prevent DevR binding to DNA. The DevR N182 and T205 mutations still bound to the DNA as the EMSA shift still occurred in assays performed using these mutants. Before EMSA can be used as a routine screen there are certain shortfalls must be addressed and solved:

The loss of binding interactions can be attributed to more than just the loss of electrostatic contacts, as hydrogen bonding as well as the inappropriate steric volumes can also play a role.

The DNA did not seem to form a consistent band but rather a smear of DNA fragments. This happens when the DNA fragments could not withstand the gel running conditions and therefore dissociated from the complexes during electrophoresis. The gel should be run at low temperature (+4°C) or at low voltage, to stabilize protein-DNA interactions to decrease the smears of fragile complexes (Hellman and Fried, 2007). During these EMSA studies the gels were ran at low temperature (+4°C) and the buffer was sufficiently cold as well in order to try and prevent the smearing. However, neither lowering the temperature nor decreasing the voltage helped to stabilize the complex.

The shift is not linear. A higher degree of retardation is seen with an increase of the protein concentration when the DNA concentration is kept constant. Cicero *et al*. (1998) experienced a similar problem and the conclusion was reached that by using shorter DNA sequences as substrates they blocked the slower migrating DNA and linear shifts could be observed. At constant protein concentrations, a linear shift would then be seen but with an increase in protein concentration, additional DNA or protein interactions may take place that will cause the different migration patterns.

Another explanation for both the smearing and the non linear migration could be that at high protein concentrations the equilibrium to binding is for all the DNA to be completely bound therefore you see a band high up on the gel. At lower concentrations the equilibrium is shifted and more protein is able to dissociate while you are running the gel that is why you smearing is observed. Poor resolution may also be due to the detection system, radioactive labelling may have improved the results seen.

Mass spectrometry experiments were carried out to demonstrate the acetyl phosphate phosphorylation of DevR. Literature reports that DevR has a molecular mass of 25.5 kDa (Saini *et al*., 2004b). Based on the amino acid sequence NP_217649 (two-component transcriptional regulatory protein DevR [*Mycobacterium tuberculosis* H37Rv]) found in the NCBI database, the molecular mass was calculated to be 23.294 kDa. The differences in mass between the gene sequence and the expressed protein was as a result of the cloning vector that adds amino acids to both ends of the protein. This resulted in an expected molecular weight of 25.683 kDa. The MS determined mass was 25.680 kDa. Phosphorylation causes a

molecular mass shift of 80 Da. The MS results of the acetyl phosphate-treated DevR however shows that the protein is phosphorylated at multiple sites, increasing the molecular mass in steps of 80 Da. The phosphorylation of DevR with acetyl phosphate is less specific than enzymatic phosphorylation. The inspection of DevR revealed 18 aspartic acids, which could explain the multiple phosphorylation of DevR. On the isotopic profiles for phosphorylated DevR, there are only 14 adducts, which could be explained if solvent accessibility is considered.

During DNA-protein complex formation, the protein undergoes several conformational changes relative to the crystal structure of the native protein. These conformational changes were unaccounted for in the models and this could lead to the discrepancies between the amino acids identified to bind to DNA in the model, and those that are actually utilized to bind the DNA *in vitro*. Hence, the low level of binding of the DevR mutant enzymes, K182I and T205V, is challenging the assumptions of their role in DNA binding, and consequently challenging the models built. The role of these amino acid side chains in DNA binding may be better explained in higher resolution models and crystal structures.

Chapter 5 *In vitro* protein-protein interaction analysis between DevS and DevR

5.1 Introduction

Specific interactions are involved in almost any physiological process. Protein-protein interactions are crucial to the operations of nearly every cellular process, but our understanding of interaction specificity is limited. The specificity of two-component signalling pathways relies on the intrinsic ability of a histidine kinase to discriminate its cognitive response regulator from all other substrates (Puttick *et al.,* 2008). However, identifying the amino acids responsible for such molecular level discrimination is difficult (Wisedchaisri *et al*., 2008). Site directed mutagenesis can identify important residues but it is often insufficient to discriminate between residues necessary for catalysis and those determining substrate selectivity.

5.1.1 Surface Plasmon Resonance (SPR)

SPR is an established and current technique used for biomolecular interaction analysis (BIA) (Phillips and Cheng, 2007). Using a SPR biosensor like the BIAcore system, enables the measurement of analytes binding to immobilized biomolecules (also called the ligand) without using labels (Karlsson, 2004). The BIAcore system (www.BIAcore.com) comprises a flow system for continuous buffer and efficient sample delivery along with the real time monitoring of complex formation and dissociation of the analyte from the ligand. Ligands can be immobilized onto a variety of surfaces using either non-covalent or covalent bonds (Willard and Siderovski, 2006). The response units can be quantitatively used to determine the reaction rate constants of the binding interaction (Karlsson, 2004). Surface regeneration is

done to ensure that the same number of ligand binding sites is available each time a new analyte sample is used. Using a range of analyte concentrations should ensure sufficient data for curve fitting and rate determination (Karlsson, 2004). The Langmuir equation is used to relate binding of the analyte to a ligand as a 1:1 binding model. The equation was developed by Irving Langmuir in 1916.

5.1.2 Biosensors

Biomolecular interaction analysis (BIA) is used to resolve the multitude of interconnected pathways and systems whose balanced function is essential. Biosensors rely on the specific interactions between biomolecules. During binding, a change is observed in different physical parameters, such as mass or enthalpy and this can be quantified as interaction kinetics, specificity and strength, as well as any conformational changes. Based on literature, the BIA market is dominated by BIAcore, whose SPR systems use optical sensors to measure the rate of mass addition and removal from which kinetic and affinity parameters can be calculated. One the biggest advantages of using the BIAcore is the small amount of ligand and analyte needed to obtain valuable information, speed of measurement, real time kinetic analysis as well as stoichiometry (Philips and Cheng, 2007).

Figure 5.1 shows a sensorgram produced by the BIAcore system correlated with sensorgram events as adapted from Karlsson (2004). When using the biosensor to study biomolecular interactions one molecule called the ligand will be mobilized to the sensor chip. The ligand is shown in orange in Figure 5.1, with the analyte represented by the green spheres. The analyte is washed over the ligand and the interaction measured.

The degree of binding of the ligand to the sensor chip is measured in Response Units (RU) and after immobilization a baseline is established; when the analyte binds to the ligand this interaction is also quantified in RUs (Philips and Cheng, 2007). The kinetics of an interaction is based on the rates of association (k_a) and dissociation (k_d). The values of k_a and k_d are determined by fitting the RU data to the 1:1 Langmuir model. The equilibrium dissociation constant KD is calculated by the ratio of the kinetic rate constants (k_d/k_a), measured in M.

High affinity interactions is defined by KD<10 nM and weak affinity interactions are defined by KD>100 µM.

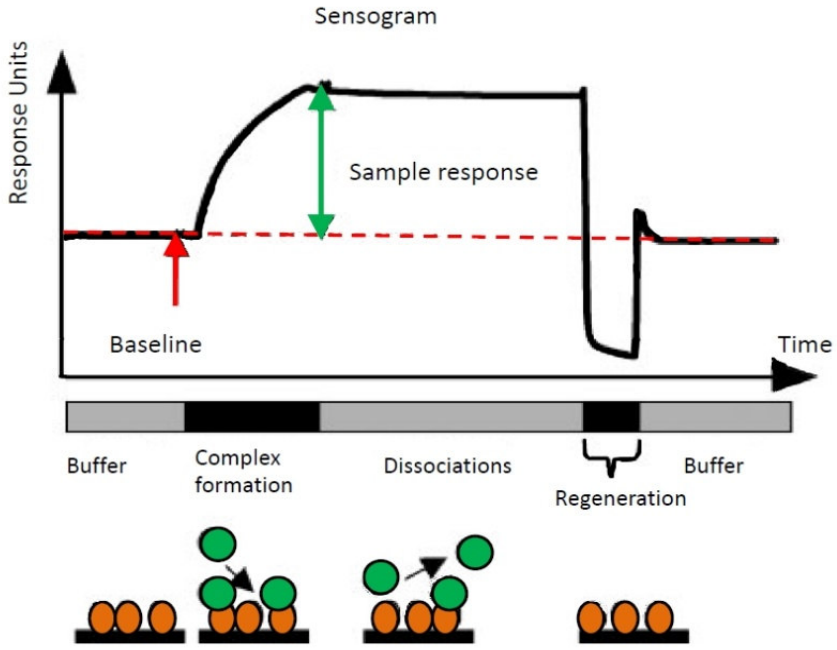


Figure 5.1: The sensorgram and sensorgram events. Initially the ligand (orange spheres) is mobilized to the sensor chip, and a baseline level indicated by the red dotted line is established. The analyte sample (green spheres) is injected and complex conformation occurs during the sample response. After the sample injection stops, the analyte starts to dissociate, a regeneration solution can be injected to speed up the dissociation process and restore the surface for further experiments.

5.1.3 Rate equations for 1:1 kinetics

When studying biomolecular interactions using a biosensor, the phase during which analyte is being injected and a complex is formed with the bound ligand, is labelled the 'association phase', while the period following the end of the injection is termed the 'dissociation phase'. Equilibrium is reached when the association rate equals the dissociation rate. The kinetics of an interaction based on the rates of association (k_a) and dissociation (k_d) can be determined, as seen from Figure 5.2. A is the analyte in solution, whose concentration is maintained by a constant flow system. B is the ligand on the surface. Total concentration can be expressed in RU. AB is the complex where the concentration of complex is also measured directly in RU.

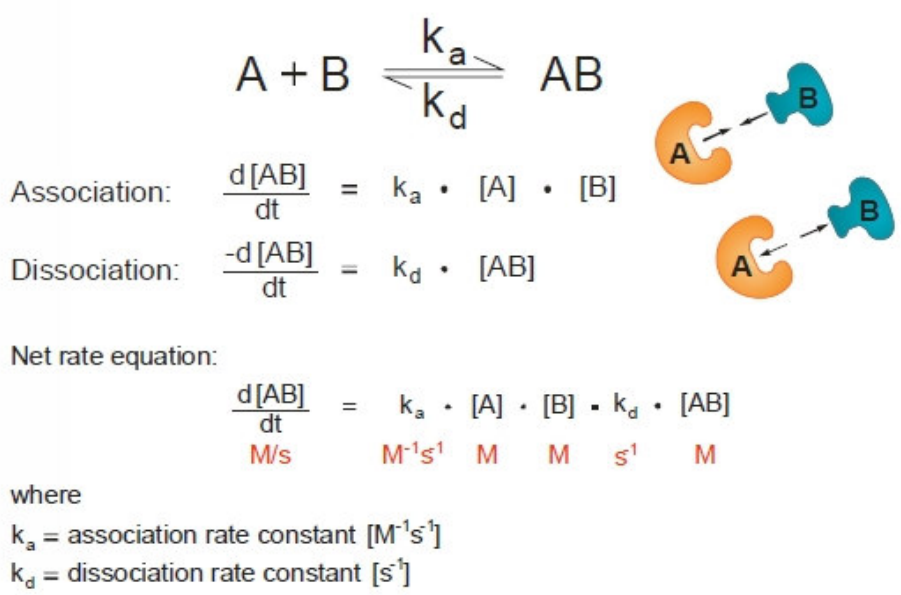


Figure 5.2: Rate equations for 1:1 kinetics. A is the analyte in solution, a constant concentration is maintained via the flow system. B is the ligand on the surface. Total concentration can be expressed in RU. AB is the complex where the concentration of complex measured directly in RU.

5.1.4 Protein-protein interactions between DevS and DevR

Two-component histidine kinases play a significant role in the regulation of bacterial homeostasis. This is especially true in disease causing organisms such as *Mycobacteria tuberculosis.* Two-component signalling represents an excellent system for probing the molecular basis of specificity in a large paralogous signalling family. This study is particularly valuable because no high-resolution structures are available for full length or truncated DevS in complex with DevR. The overall aim of the project is to identify a novel mechanism to target tuberculosis in a rational drug design program. If DevR and DevS binding are to be used in a drug development program it is essential to have the protocols to accurately measure their interaction in addition to develop a fundamental understanding of how their interactions occur. The question remained as to whether DevS201 has the same binding affinity to DevR in comparison to DevS. These experiments will explore the difference in binding affinity of DevR to both DevS and DevS201.

5.2 Methods

5.2.1 Ligand immobilization

All the experiments were conducted on a BIAcore 3000 instrument (Stevenage, United Kingdom), using the HBS-N buffer for the CM5 biosensor chip (research grade; BIAcore) with a carboxylmethyl dextran surface. The temperature was maintained at 25°C during the whole experiment. The DevS or DevS201 proteins were immobilized to the surface using amine coupling through lysine residues following the standard protocol, as explained. The carboxymethyl groups were activated with 0.2 M EDC and 0.5 M NHS, thus creating highly reactive succinimide esters which react with amine and other nucleophilic groups on proteins. The second step was the coupling step during which ligand is injected into the flow cell at a buffer pH which causes the proteins to be positively charged, thereby driving the electrostatic interactions with the negatively charged carboxylated dextran matrix. To optimize binding, the reaction takes place in a buffer with a pH below the protein's isoelectric point, yielding a net positive charge. Both DevS and DevS201 bound optimally in 10 mM sodium acetate, pH 3.5. During the final blocking step, the activated carboxymethyl groups that are not involved in ligand binding were neutralized by very high concentrations of ethanolamine (1 M ethanolamine, pH 8.0), which would elute any non-covalently bound material.

Following analysis, a regeneration step was carried out to remove the analyte from the ligand so the surface may be re-used. For regeneration, 10 mM glycine at pH 2.5 was used.

5.2.2 DevR binding analysis

The affinity of the interaction between response regulator (DevR) and the sensor kinase (DevS or DevS201) was measured on a BIAcore 3000 instrument. The bound ligand and analyte were measured as a change of the refractive index, recorded in RU (Joubert *et al*., 2010). The temperature was maintained at 25°C during the whole experiment and the flow rate at 1 µl/min (Dey *et al*., 2005). The proteins were immobilized to the surface using amine coupling through lysine residues following the standard protocol (Karlsson *et al*., 2004,

Joubert *et al*., 2010). In order to perform kinetic analysis, the ligand (either DevS or DevS201) was immobilized at three different concentrations in three different flow cells. Measured in RU, for DevS, the concentrations were 11,489.37 RU, 12,598.42 RU and 13,731.07 RU. For DevS201, they were 7,144.61 RU, 10,735.60 RU and 12,665.33 RU. DevR (analyte) was prepared at a range of concentrations (100 to 2,000 nM). A 75 µl sample was injected (KINJECT procedure) at 5 µl/min using an association time of 300 seconds, followed by 10 µl of 10 mM glycine (pH 2.5) with a 300 second dissociation. Injection of the analyte was initially done starting from lower concentrations to the highest concentrations, and then repeated by starting at the highest concentration going down to the lowest. Analyses of association and dissociation curves were performed using BIAevaluation 4.1 software (BIAcore) by the so-called global method. The data were fitted to the 1:1 Langmuir model. The dissociation (k_d) data were fitted first, followed by the complex formation (k_a) data, both done globally. For the DevS201 analysis, 9 curves were fitted for the following analyte concentrations: 100 nM, 200 nM, 250 nM, 375 nM, 500 nM, 750 nM, 1000 nM, 1,500 nM and 2,000 nM. For DevS, 10 curves were fitted for the following analyte concentrations: 100 nM, 125 nM, 200 nM, 250 nM, 328 nM, 656 nM, 750 nM, 1,312 nM, 1,500 nM and 2,000 nM). The affinity of binding (KD) was determined by the ratio of the kinetic rate constants, k_d/k_a.

5.3 Results

During the BIAcore analysis two methods of ligand immobilization were considered: NTA immobilization for His_6-tagged proteins and amine coupling. Initially, the NTA coupling was considered since the Dev proteins were already His_6-tagged for purification. Unfortunately a common deficiency of NTA-immobilized proteins is the slow and continuous dissociation of the proteins from the surface (Willard and Siderovski, 2006). A secondary problem was that both the ligand and the analyte were His_6-tagged, increasing non-specific binding because the analyte will bind to the chip surface and not specifically to the ligand. The BIAcore system is unable to measure the difference. Consequently, amine coupling was successfully used to bind DevS and DevS201 onto the chip.

The success of the amine coupling was measured in RU units, and if the coupling was unsuccessful the binding would be very low, approximately 7,000 RU. Values above this are regarded as successful binding. The kinetics of an interaction, and the rates of association (k_a) and dissociation (k_d), can be determined from the information in a sensorgram such as that shown in Figure 5.3, obtained for one of the flow cells with DevS as ligand. Using the data obtained, the KD was calculated to be 255 nM for DevS201 and 184 nM for DevS. This was an average from the three flow cells with different concentrations of bound ligand. The data were reproducible from one flow cell to another with a standard deviation of 0.06 (DevS) and 0.31 (DevS201). These KD values show that the truncated DevS201 has a similar affinity to the full length DevS for DevR. A Student's t-test was done on the data and found to be 0.019.

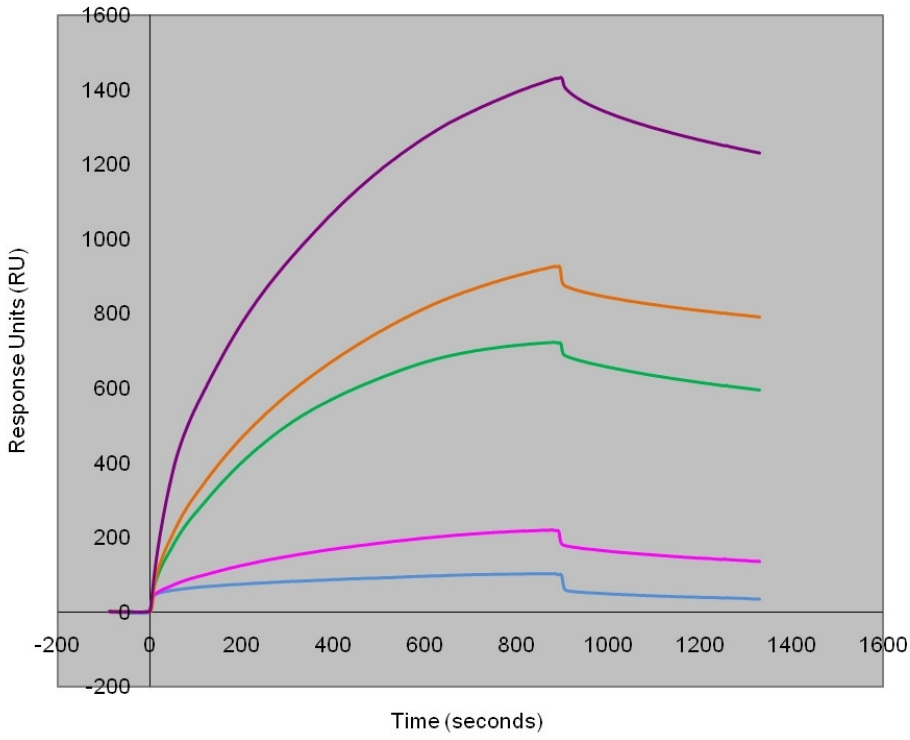


Figure 5.3: Overlay plots of sensorgram of DevS interactions of various concentrations of DevR. The following concentrations are shown on the graph: 100 nM (blue), 250 nM (pink), 656 nM (green), 750 nM (orange) and 1,500 nM (purple). The increase in signal is indicative of the association that is taking place, the quick drop is the due to the change in buffer and the slow decrease in signal is the dissociation taking place.

5.4 Discussion

Sensor kinases are membrane-bound with abundant transmembrane regions, which influences solubility during *in vitro* expression and purification. It is therefore possible that the absence of these regions in DevS201 could induce conformational changes during *in vitro* refolding. The consequential aim of this research was to determine the effect of the truncation of DevS on its binding to DevR, to establish whether DevS201 can be used in future binding and *in silico* modelling studies.

When binding takes place, the response in the sensorgram increases. The binding orientation of DevS to the chip can compromise the DevR binding site, however the RU values indicated that protein-protein binding did take place. Once equilibrium is reached, a constant signal is seen. Dissociation of the bound molecules causes the response to decrease. BIAcore evaluation software generated the values of k_a and k_d by fitting the data to the 1:1 Langmuir model. The affinity of binding is determined by the equilibrium constant, KD. KD is calculated by the ratio of the kinetic rate constants, k_d/k_a. Using the data obtained, the KD was calculated to be 255 nM for DevS201 and 184 nM for DevS. It is evident that the truncated DevS201 binds moderately better to DevR than DevS. The lower the KD, the higher affinity the ligand has for the analyte. Changes in protein conformation could influence the accessibility of the DevR binding regions, thereby interfering with binding. DevS201 comprises the soluble cytosolic catalytic core; the loss of binding activity could be linked to changes in conformation due to truncation.

The kinetic constants obtained were compared to other reported biosensor-derived constants for TCSs. During biosensor experiments done by Perraud *et al*. (2000), the authors found that EvgA showed a very weak interaction with the kinase domain of EvgS, with an affinity constant of 806 nM. A stronger affinity was found between HupT and HupUV, with a KD of 430 nM (Elsen *et al*., 2003). Two other studies also had KD values in the nanomolar range: the VanSR system had a KD≈ 30 nM, (Fisher *et al*., 1996) and CheAY had a KD of 30 nM (Schuster *et al*., 1993). The KDs determined for DevS and DevS201 were also in the nanomolar range, at 184 nM and 255 nM, respectively. For sufficiently strong binding between a histidine kinase and a response regulator, the KD needs to be in the nM range. Therefore it can be concluded that DevS201 binds DevR strongly enough to be used in future studies. The low consumption of target proteins as well as the sensitivity for detection of low-

affinity binding confirms that the BIAcore system is an excellent way of studying complex drug interactions. Hence the BIAcore could be used to screen small-molecule inhibitors of DevR-DevS interactions and potentially DevR-DNA binding as well.

Chapter 6 Concluding discussion

The focus of this study was to investigate the binding interactions of *Mycobacterium tuberculosis* Dev two-component system with the aim of identifying potential loci for drug design. The following facets were investigated:

The basic biochemistry of the TCS starts with DevS's capability to recognize the change in the environment, autophosphorylation and subsequent phosphorylation of DevR. Accordingly, if phosphorylation is blocked by any mechanism it will have a detrimental effect on the Dev TCS system's control of the dormancy regulon. The protocols that were developed in testing the efficiency of DevS and DevS201 phosphorylation can be used to screen inhibitors designed to block phosphorylation. The identification of lead molecules that inhibit phosphotransfer is an avenue that is relatively unexplored for inhibitor screening.

To facilitate phosphorylation, the DevS-DevR complex formation is vitally important. Protein-protein docking provides us with a series of candidate structures for the DevS201-DevR complex. The native structure can only be corroborated if the complex is crystallized and the structure determined. The expression of high concentrations of all the components of the Dev TCS was achieved. This paves the way for future crystallization or NMR studies to elucidate the specific amino acids that play a role in the complex formation. High throughput screening can be used to find a drug that will target protein complex formation. Consequently, a protocol using the BIAcore SRP technology was developed to quantify DevR and DevS binding. It was established that this method is sensitive enough to be used in high throughput screening. It also proved that the truncation of DevS did not compromise the binding constant significantly. Since the protocol for DevR binding to both DevS and DevS201 was effectively achieved, one can modify this screen to incorporate DevT as well. Based on the BIAcore results, it was also concluded that using DevS201 in the structural design strategies will give us a near native presentation of the DevS-DevR complex. This assumption is supported by the results obtained from the phosphotransfer assays, where DevS201 is able to phosphorylate DevR as efficiently as the full length DevS.

Financial constraints were the biggest issue in repeating these experiments with the DevR mutants. Because the experiments needed to run in triplicate to make the results statistically significant three chips was used to do the DevR-DevS interaction analysis. This means that an additional 3 chips were needed to test each mutant. The stability of the immobilized protein becomes questionable after a specified amount of runs and consequently the DevS mobilized for the DevS-DevR interactions could not be re-used efficiently. In conclusion, it was not financially possible for us to obtain another 12 chips to test the mutant DevR. It would have been a great contribution to the thesis if it was possible to study the DevR mutant interactions with DevS using BIAcore. Future work should include testing the DevR mutants especially the T205 mutant.

After DevR is phosphorylated, the activated response regulator induces the dormancy regulon by interacting with specific palindromic DNA sequences that precedes the genes of the regulon. Compounds that inhibit DevR binding to target DNA can down-regulate dormancy gene transcription, and drastically reduce survival of hypoxic bacteria (Gupta *et al*., 2009). The model of the DevR binding to DNA gave us an insight into the amino acids used for binding. K179I and N183L were identified as amino acids that can be used as targets in a structure-based drug design strategy. SDMs of these amino acids confirmed this and can be used in future to map the exact DNA binding site. The EMSA protocol developed here can be used routinely to screen for inhibitors against DevR-DNA binding.

The intention was to investigate the binding interactions of DevS-DevR, with the aim of improving our understanding of the detailed mechanism associated with *M. tb* latency and conceivably provide good drug targets for persistent tuberculosis. In conclusion four different aspects were identified as drug target against the Dev TCS: The autophosphorylation of DevS, the phosphorylation of DevR by DevS, binding interaction between DevR and DNA, and DevS and DevR complex formation.

Many response regulators can be phosphorylated albeit less efficiently by low molecular weight phosphor donor such as acetyl phosphate (McCleary and Stock, 1994). Since phosphorylated DevR rapidly loses its phosphosignal in the presence of both DevS and DevT,

the DevR was phosphorylated using acetyl phosphate. This was done to ensure reproducible results that were only dependent on one protein species rather than adding the additional variability of a second protein species to the assay.

Preventing the signal transduction system between the histidine kinase and the reaction regulator could thus prevent the establishment of dormancy of people infected by M.tb. Consequently by establishing assays to indentify inhibitor that will either block the phosphotransfer or dimerization will be more relevant to the drug discovery process. The his-asp signal transduction pathways are distinctly different from the pathways seen in humans hence drugs that will inhibit the Dev TCS are expected to specifically target bacteria. The handling of radioactive isotopes is a perceive disadvantage to studying the phosphotransfer interactions. By using the BIAcore system this will be circumvented but ultimately enabling the rapid screening of molecules that will prevent the signal transduction by intervening with the DevR and DevS binding. The inherent difficulty in targeting the Dev TCS is that both DevS and DevT can phosphorylate DevR

Although the signal transduction seems to be the superior target due to its specificity, DevR has merits as a drug target as well because of its key role it plays in the signalling cascade. Consequently multiple signalling systems could be inhibited by a single molecule if the DevR binding interactions with DNA could be prevented. Developing high throughput assays based on EMSA might be more difficult.

It would therefore be an optimal situation if one could combine the findings of this study into one solution. The BIAcore was the instrument that worked the best in terms of quantitative measuring of binding interaction and by proxy also inhibition. However, targeting the response regulator will interfere with pathways that are vital for *M. tb* survival, in the host. For future studies it would be worthwhile investigating the ability to study the DevR DNA binding interactions using the BIAcore system.

In conclusion, this investigation provided significant insights into the molecular basis of dormancy as regulated by the Dev TCS. The investigation into both the protein-protein interactions as well as the DNA-protein interactions provides a set of drug targets that can be used against persistent tuberculosis. A more detailed rational drug design program can now be put into place.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215 (3), 403-410.

Altendorf, K., Voelker, P. and Puppe, W. (1994). The sensor kinase KdpD and the response regulator KdpE control the expression of the kdpFABC operon in *E. coli*. *Research in Microbiology* 145 (5-6), 374-381.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* 181 (4096), 223-230.

Arigoni, F., Duncan, L., Alper, S., Losick, R. and Stragier, P. (1996). SpoIIE governs the phosphorylation state of a protein regulating transcription factor sigma F during sporulation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* 93 (8), 3238-3242.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22 (2), 195-201.

Baney, F. and Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli. Nature Biotechnology* 22 (11), 1399-1408.

Beier, D. and Gross, R. (2006). Regulation of bacterial virulence by two-component systems. *Current Opinion in Microbiology* 9 (2),143-152.

Benda, C., Scheufler, C., de Marsac, N. and Gärtner, W. (2004). Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. *Biophysical Journal* 87 (1), 476-487.

Benkert, P., Tosatto, S.C.E. and Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Protein: Structure, Function and Bioinformatics* 71, 261–277.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology* 112 (3), 535–542.

Bijlsma, J.J.E. and Groisman, E.A. (2003). Making informed decisions: regulatory interactions between two-component systems. *Trends in Microbiology* 11, 359–366.

Bilwes, A.M., Quezada, C.M., Croal, L.R., Crane, B.R. and Simon, M.I. (2001). Nucleotide binding by the histidine kinase CheA. *Nature Structural Biology* 8 (4), 353–360.

Boon, C. and Dick, T. (2002). *Mycobacterium bovis* BCG response regulator essential for hypoxic dormancy. *Journal of Bacteriology* 184 (24), 6760-6767.

Brenner, S.E., Koehl, P. and Levitt, M. (2000). The ASTRAL compendium for protein structure and sequence analysis. *Nucleic Acids Research* 28 (1), 254-256.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 4, 187–217.

Buckler, D.R., Anand, G.S. and Stock, A.M. (2000). Response-regulator phosphorylation and activation: a two-way street? *Trends in Microbiology* 8 (4), 153-155.

Campbell, E.A., Masuda, S., Sun, J.L., Muzzin, O., Olson, C.A., Wang, S. and Darst, S.A. (2002). Crystal structure of the *Bacillus stearothermophilus* anti-σ factor SpoIIAB with the sporulation sigma factor (σ^F). *Cell* 108, 795–807.

Casino, P., Rubio, V. and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* 139 (2), 325-336.

Chauhan, S. and Tyagi, J.S. (2008a). Cooperative binding of phosphorylated DevR to upstream sites is necessary and sufficient for activation of the Rv3134°C -devRS operon in *Mycobacterium tuberculosis*: Implication in the induction of DevR target genes. *Journal of Bacteriology* 190 (12), 4301-4312.

Chauhan, S. and Tyagi, J.S. (2008b). Interaction of DevR with multiple binding sites synergistically activates divergent transcription of narK2-Rv1738 genes in *Mycobacterium tuberculosis*. *Journal of Bacteriology* 190 (15), 5394-5403.

Cicero, M.P., Alexander, K.A. and Kreuzer K.N. (1998). The MotA transcriptional activator of bacteriophage T4 binds to its specific DNA site as a monomer. *Biochemistry* 37, 4977-4984.

Clark, E.D. (1998). Refolding of recombinant proteins. *Current Opinion in Biotechnology* 9, 157-163.

Clemens, D.L. and Horwitz, M.A. (1995). Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited. *Journal of Experimental Medicine* 181, 257–270.

Clover, R.T., Kriakov, J., Garforth, S.J., Baughn, A.D. and Jacobs, W.R. (2007). The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in *Mycobacterium smegmatis*. *Journal of Bacteriology* 189 (15), 5495-5503.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.

Comeau, S.R., Gatchell, D.W., Vajda, S. and Camacho, C.J. (2004). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 20, 45-50.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C. and Dye, C. (2003). The growing burden of Tuberculosis. *Archives of Internal Medicine* 163, 1009-1021.

Cosma, C.L., Sherman, D.R. and Ramakrishna, L. (2003). The secret lives of the pathogenic Mycobacteria. *Annual Review of Microbiology* 57, 641-676.

Dasgupta, N., Kapur, V., Singh, K.K., Das, T.K., Sachdev, S., Jyothisri, K. and Tyagi, J.S. (2000). Characterization of a two-component system, DevR-DevS, of *Mycobacterium tuberculosis*. *Tubercle and Lung Disease* 80 (3), 141-159.

Dey, A.K., Griffiths, C., Lea, S.M. and James, W. (2005). Structural characterization of an anti-gp120 RNA aptamer that neutralizes the R5 strain of HIV-1. *RNA* 11, 873-884.

Diederich, B., Wilkinson, J.F., Magnin, T., Najafi, S.M.A., Errington, J. and Yudkin, M.D. (1994). Role of interactions between SpoIIAA and SpoIIAB in regulating cell-specific transcription factor σ^F of *Bacillus subtilis*. *Genes and Development* 8, 2653–2663

Dominy, B.N and Brooks, C.L. (1990). Development of a generalized Born model parametrization for proteins and nucleic acids. *Journal of Physical Chemistry* B 103 (18), 3765–3773.

Dons, L., Eriksson, E., Jin, Y., Rottenberg, M.E., Kristensson, K., Larsen, C.N., Bresciani, J. and Olsen, J.E. (2004). Role of flagellin and the two-component CheA/CheY system of *Listeria monocytogenes* in host cell invasion and virulence. *Infection and Immunity* 72, 3237-3244.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988). High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Research* 16 (13), 6127-6145.

Duncan, L. and Losick, R. (1993). SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein sigma F from *Bacillus subtilis. Proceedings of the National Academy of Sciences* 90, 2325–2329.

Duncan, L., Alper, S., Arigoni, F., Losick, R. and Stragier, P. (1995). Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science* 270, 641–644.

Elsen, S., Duché, O. and Colbeau, A. (2003). Interactions between the H2 sensor HupUV and the histidine kinase HupT controls the HupSL hydrogenase synthesis in *Rhodobacter capsulatus. Journal of Bacteriology* 185 (24), 7111-7119.

Ewann, F., Locht, C. and Supply, P. (2004). Intracellular autoregulation of the *Mycobacterium tuberculosis* PrrA response regulator. *Microbiology* 150, 241-246.

Fätkenheuer, G., Taelman, H., Lepage, P., Schwenk, A. and Wenzel, R. (1999). The return of tuberculosis. *Diagnostic Microbiology and Infectious Disease* 34 (2), 139-146.

Fiser, A., Kinh Gian Do, R. and Sali, A. (2000). Modelling of loops in protein structures. *Protein Science* 9, 1753-1773

Fisher, S.L., Kim, S.K., Wanner, B.L. and Walsh, C.T. (1996). Kinetic comparison of the specificity of the vancomycin resistance VanS for two response regulators, VanR and PhoB. *Biochemistry* 35, 4732-4740.

Fisher, M.A., Plikaytis, B.B. and Shinnick, T.M. (2002). Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *Journal of Bacteriology* 184 (14), 4025-4032.

Fletcher, R. (1969). Optimization. Academic Press: New York and London.

Flynn, J.L. and Chan, J. (2001). Tuberculosis: Latency and Reactivation. *Infection and Immunity* 69 (7), 4195-4201.

Fontan, P.A., Walters, S. and Smith, I. (2004). Cellular signalling pathways and transcriptional regulation in *Mycobacterium tuberculosis*: stress control and virulence. *Current Science* 86, 122-134.

Foussard, M., Cabantous, S., Pédelacq, J., Guillet, V., Tranier, S., Mourey, L., Birck, C. and Samama, J. (2002). The molecular puzzle of two-component signaling cascades. *Microbes and Infection* 3 (5), 417-424.

Fried, M. and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* 9 (23), 6505-6525.

Galperin, M.Y. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *Journal of Bacteriology* 188 (12), 4169-4182.

Garner, M.M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Research* 9 (13), 3047-3060.

Gelber, R.H., Andries, K., Paredes, R.M.D., Evelyn, C., Andaya, S. and Burgos, J. (2009). The Diarylquinoline R207910 is bactericidal against *Mycobacterium leprae* in mice at low dose and administered intermittently. *Antimicrobial Agents and Chemotherapy* 53 (9), 3989-3991.

Grebe, T.W. and Stock, J.B. (1999). The histidine protein kinase super family. *Advances in Microbial Physiology* 41, 139-227.

Grigoroudis, A.I., Panagiotidis, C.A., Lioliou, E.E., Vlassi, M. and Kyriakidis, D.A. (2007). Molecular modelling and functional analysis of the AtoS-AtoC two-component signal transduction system of *Escherichia coli. Biochimica et Biophysica Acta* 1770, 1248-1258.

Groisman, E.A. (2001). The pleiotropic two component regulatory system PhoP-PhoQ. *Journal of Bacteriology* 6, 1835-1842.

Gupta, R.K., Thakur, T.S., Desiraju, G.R. and Tyagi, J.S. (2009). Structure-based design of DevR inhibitor active against non replicating *Mycobacterium tuberculosis*. *Journal of Medicinal Chemistry* 52 (20), 6324–6334.

Havlir, D. and Ellner, J. (2000). *Mycobacterium avium* complex. Principles and practices of infectious diseases. Churchill Livingstone. Philadelphia, PA.

Haydel, S.E. and Clark-Curtiss, J.E. (2004). Global expression analysis of two-component system regulator genes during *Mycobacterium tuberculosis* growth in human macrophages. *FEMS Microbiology Letters* 236 (2), 341-347.

Haydel, S.E., Dunlap, N,E. and Benjamin Jr., W.H. (1999). *In vitro* evidence of two-component system phosphorylation between the *Mycobacterium tuberculosis* TrcR/TrcS proteins. *Microbial Pathogenesis* 26 (4), 195-206.

Hellman, L.M. and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols* 2, 1849–1861.

Illergard, K., Ardell, D.H. and Elofsson. A. (2009). Structure is three to ten times more conserved than sequence - A study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics* 77 (3), 499 - 508.

Ioanoviciu, A., Yuk, E.T., Moenne-Loccoz, P. and Ortiz de Montellano, P. (2007). DevS, a heme-containing two component oxygen sensor of *Mycobacterium tuberculosis*. *Biochemistry* 46 (14), 4250-4260.

Joubert, M.K., Kinsley, N., Capovilla, A., Sewell, B.T., Jaffer, M.A. and Khati, M. (2010). A modelled structure of an aptamer-gp120 complex provides insight into the mechanism of HIV-1 neutralization. *Biochemistry* 49, 5880-5890.

Karlsson, R. (2004). SPR for molecular interaction analysis: a review of emerging application areas. *Journal of Molecular Recognition* 17 (3), 151-161.

Kendall, S.L., Rison, S.C.G., Movahedzadeh, F., Frita, R. and Stoke, N.G. (2004). What do microarrays really tell us about *M. tuberculosis*? *Trends in Microbiology* 12 (12), 537-544.

Kim, D. and Forst, S. (2001). Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* 147, 1197-1212.

King-Scott, J., Nowak, E., Mylonas, E., Panjikar, S., Roessle, M., Svergun, D.I. and Tucker, P.A. (2007). The structure of a full-length response regulator from *Mycobacterium tuberculosis* in a stabilized three-dimensional domain-swapped, activated state. *The Journal of Biological Chemistry* 282, 37717-37729.

King, R.D. and Sternberg, M.J.E. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science* 5, 2298-2310.

Klumpp, S and Krieglstein, J. (2002). Serine/threonine protein phosphatases in apoptosis. *Current Opinion in Pharmacology* 2 (4), 458-462.

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

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283–291.

Laub, M.T. and Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annual Review of Genetics* 41, 121–145.

Liebau, E., Bergmann, B., Campbell, A.M., Teesdale-Spittle, P., Brophy, P.M., Luersen, K. and Walter, R.D. (2002). The glutathione *S*-transferase from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 124, 85-90.

Lo Conte, L., Ailey, B., Hubbard, T.J.P., Brenner, S.E., Murzin, A.G. and Chothia, C. (2000). SCOP: a Structural Classification of Proteins database. *Nucleic Acids Research* 28 (1), 257-259.

Lorenzen, S. and Zhang, Y. (2007). Identification of near-native structures by clustering protein docking conformations. *Proteins* 68, 187-194.

Maartens, G. and Wilkinson, R.J. (2007). Tuberculosis. *The Lancet* 370, 2030-2043.

Malhotra, V., Sharma, D., Ramanathan, V.D., Shakila, H., Saini, D.K., Chakravorty, S., Das, T.K., Li, Q., Silver, R.F., Narayanan, P.R. and Tyagi, J.S. (2004). Disruption of response regulator gene, devR, leads to attenuation in virulence of *Mycobacterium tuberculosis*. *FEMS Microbiology Letters* 231 (2), 237-245.

Marina, A., Motts, C., Auzenberg, A., Hendrickson, W.A. and Waldburger, C.D. (2001). Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. *The Journal of Biological Chemistry* 276 (44), 41182-41190.

Marina, A., Waldburger, C. D. and Hendrickson, W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *The EMBO Journal* 24, 4247 – 4259.

Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B., and Sali, A. (2002). Reliability of assessment of protein structure prediction methods. *Structure* 10, 435–440.

McCleary, W.R. and Stock, J.B. (1994). Acetyl phosphate and the activation of two component response regulators. *The Journal of Biological Chemistry* 16, 31567-31572.

Mika, F. and Hengge, R. (2005). A two-component phosphotransfer network involving ArcB, ArcA and RssB coordinates synthesis and proteolysis of sigmaS (RpoS) in *Escherichia coli. Genes and Development* 19, 2770-2781.

Min K.T., Hilditch, M.C., Diederich, B., Erington, J. and Yudkin, M.D. (1993). σ^F, the first compartment-specific transcription factor of *B. subtilis*, is regulated by an anti-σ factor that is also a protein kinase. *Cell* 74, 735-742.

Najafi, S.M.A., Willis, A.C. and Yudkin, M.D. (1995). Site of phosphorylation of SpoIIAA, the anti-anti-sigma factor for sporulation-specific σF of *Bacillus subtilis*. *Journal of Bacteriology* 177, 2912–2913.

Ninfa, E.G., Atkinson, M.R., Kamberov, E.S. and Ninfa, A.J. (1993). Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. *Journal of Bacteriology* 175, 7024–7032.

O'Toole, R., Smeulders, M.J., Blokpoel, M.C., Kay, E.J., Lougheed, K. and Williams, H.D. (2003). A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in *Mycobacterium smegmatis. Journal of Bacteriology* 185 (5), 1543-1554.

Park, H.D., Giunn, K.M., Harrel, M.I., Liao, R., Voskuil, M.I., Tompa, M., Schoolnik, G.K. and Sherman, D.R. (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Molecular Microbiology* 48, 833-843.

Parish, T., Gordhan, B.G., McAdam, R.A., Duncan, K., Mizrahi, V. and Stoker, N.G. (1999). Production of mutants in amino acid biosynthesis genes of *Mycobacterium tuberculosis* by homologous recombination. *Microbiology* 145, 3497-3503.

Parish, T., Smith, D.P., Roberts, G., Betts, J. and Stoker, N.G. (2003). The senX3–regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. *Microbiology* 149, 1423-1435.

Partridge, S.R., Foulger, D., Errington, J. (1991). The role of sigma F in prespore-specific transcription in *Bacillus subtilis*. *Molecular Microbiology*. 5(3), 757–767.

Perraud, A., Rippe, K., Bantscheff, M., Glocker, M., Lucassen, M., Jung, K., Sebald, W., Weiss, V. And Gross, R. (2000). Dimerization of signalling modules of the EvgAS and BvgAS phosphorelay system. *Biochemica et Biophysica Acta* 1478, 341-354.

Phillips, K.S. and Cheng, Q. (2007). Recent advances in surface plasmon resonance based techniques for bioanalysis. *Analytical and Bioanalytical Chemistry* 387 (5), 1831-1840.

Pierce B. and Weng Z. (2007). ZRANK: Reranking protein docking predictions with an optimized energy function. *Proteins* 7 (4), 1078-1086.

Puttick, J., Backer, E.N. and Delbaere, L.T.J. (2008). Histidine phosphorylation in biological systems. *Biochemica et Biophysica Acta* 1781, 100-105.

Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B. and Martin, C. (2001). An essential role for PhoP in *Mycobacterium tuberculosis* virulence. *Molecular Microbiology* 41 (1), 179–187.

Raghavan, V. and Groisman, E.A. (2010). Orphan and hybrid two-component system proteins in health and disease. *Current Opinion in Microbiology* 13 (2), 226-231.

Rhen, M., Eriksson, S., Clements, M., Bergstrom, A. and Normark, S.J. (2003). The basis of persistent bacterial infections. *TRENDS in Microbiology* 11 (2), 80-86.

Roberts, D.M., Liao, R.P., Wisedchaisri, G., Hol, W.G.J. and Sherman, D.R. (2004). Two sensor kinases contribute to the hypoxic response of *Mycobacterium tuberculosis*. *Biological Chemistry* 279 (22), 23082-23087.

Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. and Smith, I. (2002) IdeR, an essential gene in *Mycobacterium tuberculosis*: Role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infection and Immunity* 70, 3371–3381.

Rogl, H., Kosemund, K., Kuhlbrandt, W. and Collinson, I. (1998). Refolding *Escherichia coli* produced membrane protein inclusion bodies immobilized by nickel chelating chromatography. *FEBS Letters* 432, 21-26.

Saini, D.K., Pant, N., Das, T.K. and Tyagi, J.S. (2002). Cloning, overexpression, purification, and matrix-assisted refolding of DevS (Rv 3132c) histidine protein kinase of *Mycobacterium tuberculosis*. *Protein Expression and Purification* 25 (1), 203-208.

Saini, D.K., Malhotra, V., Dey, D., Pant, N., Das, T.K. and Tyagi, J.S. (2004a). DevR-DevS is a bona fide two component system of *Mycobacterium tuberculosis* that is hypoxia responsive in the absence of the DNA-binding domain of DevR. *Microbiology* 150, 865-875.

Saini, D.K., Malhotra, V. and Tyagi, J.S. (2004b). Cross talk between DevS sensor kinase homologue Rv2027c, and DevR response regulator of *Mycobacterium tuberculosis*. *FEBS letters* 565, 75-80.

Sali, A. and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234, 779–815.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. and Karplus, M. (1995). Evaluation of comparative protein modelling by MODELLER. *Proteins*. 23, 318-326.

Sambrook, J.; Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, vol. I. 2nd edition. Cold Spring Harbor Laboratory Press, 1989. ISBN 0-87969-309-6.

Sander, C. and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, 9, 56-68.

Saviola, B. and Bishai, W. (2006). The genus Mycobacterium – Medical. Prokaryotes 3, 919-933.

Scanu, A.M., Bull, T.J., Cannas, S., Sanderson, J.D., Sechi, L.A., Dettori, G, Zanetti, S. and Hermon-Taylor, J. (2007). *Mycobacterium avium* subspecies paratuberculosis infection in cases of irritable bowel syndrome and comparison with Crohn's disease and Johne's disease: Common neural and immune pathogenicities. *Journal of Clinical Microbiology* 45 (12), 3883-3890.

Scheef, E. D. and Bourne, P. E. (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Computational. Biology* 1(49), 359-381.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C. and Schoolnik, G.K. (2003). Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. *Journal of Experimental Medicine* 198 (5), 693-704.

Schnell, R., Agren, D. and Schneider, G. (2008). 1.9 Å structure of the signal receiver domain of the putative response regulator NarL from *Mycobacterium tuberculosis*. *Acta Crystallographica Section F* 64 (12), 1096-1100.

Schuster, S.C., Swanson, R.V., Alex, L.A., Bourret, R.B. and Simon, M.I. (1993). Assembly and function of a quaternary signal transduction complex monitored by surface plasmon resonance. *Nature* 365, 343–347.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31: 3381- 3385.

Shen, M.Y. and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science* 15, 2507-2524.

Sherman, D. R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M. I. and Schoolnik, G. K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding α-crystallin. *Proceedings of the National Academy of Sciences* USA 98, 7534-7539.

Shinnick, T.M. and Good, R.C. (1994). Mycobacterium taxonomy. *European Journal of Clinical Microbial Infectious Diseases* 13 (11), 884-901.

Singh, S.M. and Panda, A.K. (2005). Solubilisation and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering* 99 (4), 303-310.

Sinha, S. and Tompa, M. (2002). YMF: a program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Research* 31, 3586–3588.

Sousa, E.H.S., Tuckerman, J.R., Gonzalez, G. and Gilles-Gonzalez, M. (2007). DosT and DevS are oxygen-switched kinases in *Mycobacterium tuberculosis*. *Protein Science* 16 (8), 1708–1719.

Stock, A., Chen, T., Welsh, D. and Stock, J. (1988). CheA Protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions. *Proceedings of the National Academy of Sciences* 85 (5), 1403-1407.

Stock, A.M., Robinson,V.L. and Goudreau,P.N. (2000). Two-component signal transduction. Annual Review of. *Biochemistry* 69*,* 183–*215.*

Sun, J., Daniel, R., Wagner-Döbler, I. and Zeng, A.P. (2004). Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. BMC *Evolutionary Biology* 4 (36) doi:10.1186/1471-2148-4-36.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22 (22), 4673-4680.

Toosi, Z. and Ellner, J. (1998). *Mycobacterium tuberculosis* and other mycobacteria. Infectious Diseases. Saunders and company. Philidelphia, PA.

Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification* 28, 1-8.

Ulrich, L.E., Koonin, E.V. and Zhulin, I.B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology* 13 (2), 52-56.

Ventura, S. (2005). Sequence determinants of protein aggregation: tools to increase protein solubility. *Microbial Cell Factories* 4, 11-18.

Vescovi, E.G., Ayala, Y.M., Di Cera, E. and Croisman, E.A. (1997). Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg^2 and Ca^2. *Journal of Biological Chemistry* 3, 1440-1443.

Vincentelli, T., Canaan, S., Campanacci, C., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. and Bignon, C. (2004). High-throughput automated refolding screening of inclusion bodies. *Protein Science* 13, 2782-2792.

Voskuil, M.I., Schnappinger, D., Visconti, K.C., Harrell, M.I., Dolganov, G.M., Sherman, D.R. and Schoolnik, G.K. (2003). Inhibition of respiration by nitric oxide induces a *Mycobacterium* dormancy program. *Journal of Experimental Medicine* 198, 705-713.

Voskuil, M.I., Visconti, K. and Schoolnik, G. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis* 84 (3), 218-227.

West, A.H. and Stock. A.M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences* 26 (6), 369-376.

Willard, F.S. and Siderovski, D.P. (2006). Covalent immobilization of histidine-tagged proteins for surface plasmon resonance. *Analytical Biochemistry* 353 (1), 147-149.

Wisedchaisri, G., Wu, M., Rice, A.E., Roberts, D.M., Sherman, D.R. and Hol, W.G. (2005). Structures of *Mycobacterium tuberculosis* DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. *Journal of Molecular Biology* 354, 630–641.

Wisedchaisri, G., Wu, M., Sherman, D.R. and Hol, W.G. (2008). Crystal structures of the response regulator DosR from *Mycobacterium tuberculosis* suggest a helix rearrangement mechanism for phosphorylation activation. *Journal of Molecular Biology* 378 (1), 227-242.

Wolanin, P.M., Thomason, P.A. and Stock, J.B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biology* 3 (10), 3011-3013.

Yamada, S., Akiyama, S., Sugimoto, H., Kumita, H., Ito, K., Fujisawa, T., Nakamura, H. and Shiro, Y. (2006). The signaling pathway in histidine kinase and the response regulator complex revealed by X-ray crystallography and solution scattering. *Journal Molecular Biology* 362, 123-139

Zahrt, T.C. and Deretic, V. (2000). An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *Journal of Bacteriology* 182 (13), 3832-3838.

Zhou, H. and Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science* 11, 2714-2726.