

https://dspace.org/

Structural and functional characterisation of escherichia coli stringent starvation protein A (EcSspA)

Matenchi, Kamogelo Given

2018

https://hdl.handle.net/10539/26694

Downloaded from a DSpace repository. https://dspace.org/

School of Molecular and Cell Biology

University of the Witwatersrand

Johannesburg

Structural and functional characterisation of *Escherichia coli* stringent starvation protein A (EcSspA)

Kamogelo Matenchi

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the degree of Master of Science

Johannesburg, September 2018

Declaration

I, **Kamogelo Matenchi** (669478) am a student registered for the degree of Master of Science (MSc) in the academic year 2016.
I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.
- I have not submitted this work before for any other degree or examination at this or any other University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.



Signature 14th day of September 2018

Abstract

Escherichia coli Stringent starvation protein A (EcSspA) is an RNA polymerase associated transcription factor required for the activation of phage P1 late genes and plays an important role in stationary phase induced acid tolerance of *Escherichia coli*. The crystal structure of the *Yersinia pestis* orthologue of SspA (YpSspA) shows that the protein assumes a glutathione *S*-transferase (GST) fold and contains both the N and C-terminus GST-like domains. However, little is known about the structure of EcSspA and the functional roles of the GST domains within the protein. This study focuses on the characterisation of EcSspA in terms of its secondary, tertiary and quaternary structure, and functional characterisation through an enzyme assay and ligand binding studies. Secondary structure analysis by far-UV circular dichroism spectroscopy indicates that EcSspA is a predominantly alpha helical protein. Intrinsic tryptophan fluorescence has shown that EcSspA tryptophan residues are less solvent accessible and predominantly hidden within the interior of the protein. Like the YpSspA orthologue and other GST-like proteins, EcSspA is a homodimeric protein with a structure similar to that of YpSspA as shown by homology modelling. Using both the standard GSH/CDNB conjugation assay and a glutathione binding test it was confirmed that EcSspA does not have GST activity and does not bind GSH. Binding studies with the amphipathic ligand 8-anilino-1-naphthalene sulfonate (ANS) using fluorescence spectroscopy and isothermal titration calorimetry (ITC) indicated for the first time that EcSspA possesses ligandin function and that the interaction is exothermic. Molecular docking of ANS to the homology model of EcSspA indicates the binding of a single molecule of ANS to occur at the dimer interface of EcSspA involving weak interactions from hydrogen bonding and electrostatic interactions. The results of this study show that some functional and structural properties of eukaryotic GSTs are conserved in the bacterial EcSspA homologue.

"The paradox of education is precisely this - that as one begins to become conscious one begins to examine the society in which he is being educated" James Baldwin

"O my body, make of me always a man who questions!"

Frantz Fanon, Black Skin, White Masks

Table of contents

Declaration ..... **i**
Abstract ..... **ii**
Table of contents ..... **iv**
Acknowledgments ..... **vi**
List of figures ..... **vii**
List of tables ..... **ix**
List of abbreviations ..... **x**
1. CHAPTER ONE: INTRODUCTION **1**
1.1.GST protein superfamily 1
1.2.Catalytic and ligandin functions of GSTs 1
1.3. The GST-fold protein structure 2
1.4. GST-like proteins and their functions 3
1.5. SspA family of proteins 5
1.6. SspA as a member of the GST superfamily: protein structure 5
1.7. *Escherichia coli*: Health and disease implications 8
1.8. *Escherichia coli* stress response mechanisms and survival 10
1.9. Role of EcSspA in stress response and pathogenesis 11
1.10. Study rationale and motivation 11
1.11. Aim and objectives 12
1.11.1. Aim 12
1.11.2. Objectives 12
2. CHAPTER TWO: EXPERIMENTAL PROCEDURES **14**
2.1. Materials 14
2.2. Methods 14
2.2.1. Bacterial transformation and cDNA sequence verification 14
2.2.2. Heterologous overexpression of EcSspA and protein purification 15
2.2.3. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) 17
2.2.4. Protein concentration determination 18
2.2.5. Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) 18
2.2.6. Structural characterisation of EcSspA 19
2.2.7. Homology modelling 21

2.2.8. Molecular docking 21
2.2.8. GSH/CDNB conjugation assay 22
2.2.10. GSH agarose affinity chromatography 23
2.2.11. ANS binding studies 24
2.2.12. Crystallisation of EcSspA 25
3. CHAPTER THREE: RESULTS 27
3.1. cDNA sequence verification 27
3.2. Expression trials of SspA in *Escherichia coli* BL21 (DE3) pLysS cells 27
3.3. Protein overexpression and purification 31
3.4. Protein size estimation 33
3.5. Protein concentration determination and purity assessment 33
3.6. Confirmation of protein identity 35
3.7. Structural characterisation of EcSspA 38
3.7.1. Structure modelling and analysis 38
3.7.2. Far-UV circular dichroism: secondary structural characterisation 39
3.7.3. Intrinsic fluorescence: tertiary structural characterisation 40
3.7.4. SE-HPLC: quaternary structural characterisation 41
3.8. Thermal unfolding properties of EcSspA 43
3.9. Measurement of glutathione *S*-transferase activity and GSH binding test 44
3.10. ANS binding studies 47
3.11. Thermodynamics of ANS binding to EcSspA 49
3.12. Molecular docking studies of ANS binding to EcSspA 51
3.13. Crystallographic studies 53
4. CHAPTER FOUR: DISCUSSION 54
4.1. EcSspA is a predominantly alpha helical protein 54
4.2. EcSspA tryptophan residues are less solvent accessible 55
4.3. Oligomeric state of EcSspA 57
4.4. EcSspA does not bind GSH and shows no catalytic activity for GSH/CDNB conjugation 57
4.5. EcSspA possesses ligandin function towards ANS 60
5. REFERENCES 64

Acknowledgments

I am greatly indebted to my supervisor Prof Heini Dirr for his guidance and encouragement throughout my MSc as well as for affording me the opportunity to work in his lab

I am grateful to Dr Ikechukwu Achilonu for his useful discussions and advice with regard to the project

I would also like to thank members of the PSFRU for their help and for providing a conducive environment to learn and cultivate growth.

I would also like to thank Dr Stoyan Stoychev CSRIR (Centre for scientific and industrial research) for carrying out the mass spectrometry mentioned in this dissertation and helping me interpret the results

I would also like to thank the National research fund for funding this project

I am grateful to my family and friends for their love and support

List of figures

Caption	Page number
Figure 1.1: The GST-fold.	2
Figure 1.2: YpSspA sequence and secondary structure topology	7
Figure 1.3: The ribbon diagram representation of dimeric YpSspA	7
Figure 1.4: Sequence alignment of EcSspA to that of YpSspA	8
Figure 1.5: A schematic illustration of the domain structure of EcSspA as predicted by the ScanProsite server	8
Figure 3.1: Confirmation of EcSspA cDNA	28
Figure 3.2: Expression trials of EcSspA in *Escherichia coli* BL21 (DE3) pLysS cells in 2x YT media at 37 °C.	29
Figure 3.3: Expression trials of SspA in *Escherichia coli* BL21 (DE3) pLysS cells in 2x YT media at 30 °C.	30
Figure 3.4: IMAC Elution profile for EcSspA purification	32
Figure 3.5: SDS-PAGE analysis of purified EcSspA	33
Figure 3.6: The absorbance spectrum of purified EcSspA and protein concentration determination	34
Figure 3.7:LC-MS/MS spectrum for a selected peptide match (IEKDWYTLMN) derived from in-gel trypsin digest represented by a series of b and y ions	36
Figure 3.8: A representation of b and y ions that correspond to identified fragment masses corresponding to the amino acid sequence of a single peptide match to the EcSspA protein band	36
Figure 3.9: SWISS Model homology structure prediction of EcSspA	38
Figure 3.10: Far-UV circular dichroism spectrum of EcSspA	39
Figure 3.11: Intrinsic fluorescence emission spectrum of EcSspA	41
Figure 3.12:SE-HPLC elution profiles of EcSspA and hGSTP1-1	42
Figure 3.13: Melting curve of EcSspA using circular dichroism as a probe	43
Figure 3.14: Progress curves of the CDNB/GSH catalysis	45
Figure 3.15: An SDS-PAGE showing the binding capability of pure EcSspA and hGSTP1-1 on a GSH-agarose column	46
Figure 3.16: Fluorescence emission spectra of ANS bound to EcSspA	47

Figure 3.17: An isotherm showing the binding of ANS to EcSspA obtained using fluorescence spectroscopy at pH 6.5 at 20 °C 48

Figure 3.18: Representative calorimetric titration profile of the binding of the anionic dye ANS to EcSspA protein 50

Figure 3.19: Homology model of EcSspA in complex with ANS 51

Figure 3.20: Simulated binding of ANS to EcSspA homology model 52

Figure 3.21: Three-dimensional protein crystals of size ~ 100 µm of EcSspA 53

Figure 4.1: Tryptophan residue distribution in dimeric EcSspA 56

Figure 4.2: Structural comparison of EcSspA and hGSTP1-1 59

Figure 4.3: Simulated binding isotherms for various c-value parameters and extraction of the binding parameters 63

List of tables

Title	Page number
Table 1.1: Proteins that contain a domain with a GST-like fold and their biological functions	4
Table 3.1: Data extracted from Protein Pilot Software analysis indicating protein characterisation and sequence coverage for SDS-PAGE protein band	37
Table 3.2: The secondary structure content of SspA proteins	40
Table 3.3: Thermodynamics parameters of EcSspA-ANS interaction titration outlined	50

List of abbreviations

ANS	8-anilino-1-naphthalene sulfonate
BLAST	Basic Local Search Alignment Tool
cDNA	coding deoxyribonucleic acid
CDNB	1-chloro-2,4-dinitrobenzene
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
EcSspA	*Escherichia coli* stringent starvation protein A
EAEC	Enteroaggregative *Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EIEC	Enter invasive *Escherichia coli*
EHEC	Enterohemorrhagic *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
Far-UV CD	far-ultraviolet circular dichroism
FPLC	fast protein liquid chromatography
GMQE	global model quality estimation
GSH	reduced glutathione
GST	glutathione *S*-transferase
H-NS	Histone-like nucleoid structuring
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
hGSTP1-1	human class pi GST with two type 1 subunits
kDa	kilodaltons
K_d	dissociation constant

LC-MS/MS	liquid chromatography mass spectrometry/mass spectrometry
mAu	milli absorbance units
MglA	macrophage growth locus protein A
MRE	mean residual ellipticity
NRMSD	normalised root mean square deviation
OD	optical density
ORF	open reading frame
PDB	protein data bank
PMSF	phenylmethylsulfonyl fluoride
RNAP	ribonucleic acid polymerase
RPM	revolutions per minute
RMSD	root mean square deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE-HPLC	size exclusion high pressure liquid chromatography
SjGST	*Schistosoma Japonica* glutathione *S*-transferase
UV	ultraviolet
V/V	volume by volume
YpSspA	*Yersinia pestis* stringent starvation protein A
YT	yeast tryptone

Amino acids are named according to the IUPAC one and three letter code

1. CHAPTER ONE

INTRODUCTION

1.1. GST protein superfamily

The GST (Glutathione transferase) (EC 2.5.1.18) superfamily is a vast and diverse group of complex proteins with multiple functions. It is an ancient superfamily of proteins that is believed to have arose from a thioredoxin-like ancestor in response to an oxidative environmental stress. The superfamily has further been subdivided into four subfamilies that includes; the soluble cytosolic GSTs, the microsomal GSTs, the mitochondrial GSTs and the bacterial fosfomycin resistance GSTs (Sheehan *et al*., 2001). The sequence resemblance between the different classes has been shown to be as low as 20 %, despite the low sequence resemblance between the classes; members of the GST superfamily are related by a common fold and not by functions (Sheehan *et al*., 2001). Cytosolic GSTs are the most well studied class of GSTs from several species and are classified into a group of sixteen. Members of the same class possess a very high sequence resemblance (Sheehan *et al*., 2001; Hayes and Pulford, 1995). Half of the eighteen class of cytosolic GSTs are of a mammalian descent, which includes; Alpha, Mu, Pi, Sigma, Theta, Omega, Kappa and Zeta (Mannervik *et al*., 1985).

1.2. Catalytic and ligandin functions of GSTs

GSTs are phase II metabolic enzymes that have been shown to play a role in the protection of cellular organs from a number of foreign xenobiotics (Armstrong, 1991; Sheehan *et al*., 2001). GSTs are responsible for catalysing GSH (γ -glutamyl-cysteinyl-glycine) to a broad range of electrophilic compounds. The proteins catalyse the conjugation of GSH to an electrophilic substrate resulting in soluble and non-toxic compounds. The crystal structure of GSTs displays the presence of two characteristic ligand binding sites. These sites are called the G and H-sites and for the binding of GSH and electrophilic substrates, respectively (Dirr *et al*., 1994; Oakley *et al*., 1999). Both the sites differ in their level of specificity with the latter being less specific (Oakley *et al*., 1999). McTigue *et al*. has shown that a ligand binding site is also present at the interface of dimeric GSTs specifically the *Schistosoma japonica* GST which was shown to bind a drug molecule at its interface (McTigue *et al*., 2005).

The non-catalytic ligand binding property of GSTs was initially observed for an abundant protein from the rat liver which was later established to be a GST (Litwack *et al*., 1971). These ligands included molecules such as haemin, biological hormones and drugs of therapeutic nature (Ketterer *et al*., 1978). The ligand binding property was later discovered to be a characteristic of many other forms of GSTs (Caccuri *et al*., 1990). However, the role of GSTs is not limited to catalytic and ligandin functions but also includes the regulation of cell signalling pathways through GST-protein kinase interactions (Tew *et al*., 2011). Furthermore, there are several proteins with a GST-like fold that have been identified and attest to the multifunctional nature of the superfamily (see section 1.4).

1.3. The GST-fold protein structure

The GST superfamily is characterised as a member of the thioredoxin-like superfamily of proteins. Members of this superfamily are made up of four β-strands which are shielded by two alpha helices on each side resulting in an overall secondary structural topology of βαβαββα (Figure 1.1). Thioredoxin proteins were the initial protein with this type of arrangement to be characterised (Holmgren *et al*., 1975). The Trx-like fold characterises the N-terminal domain of GSTs. All members of the GST family possess this type of arrangement and fold in domain 1. The GST fold is characterised by the presence of both a thioredoxin-fold and alpha helical domain at the N and C terminal, respectively.

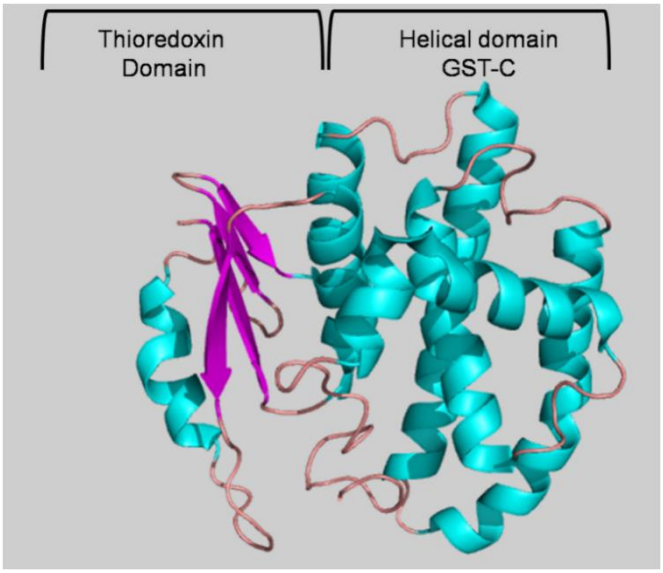


Figure 1.1: The GST-fold. A Ribbon diagram representation ofhGSTM2-2 which is characterised by the GST fold. The N-terminal thioredoxin fold and C-terminal alpha helices are annotated on the image. The image was generated using PyMOL (DeLano, 2002) PDB ID: 1XW5.

1.4. GST-like proteins and their functions

GST-like proteins within the context of this study will be defined as proteins that belong to the glutathione structural family and have been shown to be structurally related to canonical GSTs by sequence and structural alignment. These proteins have been shown to act as regulatory proteins or perform other functions besides the catalytic conjugation of glutathione. Some of the GST related proteins and their biological functions are tabulated in Table 1.1. The *Drosophila* GST delta was shown to have a sequence resemblance to the stringent starvation protein of *Escherichia coli*. Both proteins were confirmed to be made of two identical subunits of 211 amino acids and the proteins have an overall similar amino acid composition (Toung and Tu, 1992). The sequence homology between the proteins gave rise to the initial functional speculation, therefore suggesting that some functions may be conserved in different organisms. Similarly, the eukaryotic elongation factor 1 gamma contains an N-terminal GST-like domain which is related to the GST Theta-like proteins and contains a glutathione binding domain (Koonin *et al*., 1994). The GST-like domain has been shown to be present in proteins that play a functional role in nitrogen metabolism, oxidoreductases and the formation of chloride channels (Harrop *et al.*, 2001; Aslund *et al*., 1994; Coschigano and Magasanik, 1991). The role of the GST-like domain in other proteins and its functional implications remain of interest. Of recently, Achilonu *et al*. (2014) has shown that the N-terminal GST-like domain of the human eukaryotic elongation factor 1 gamma (eEF1γ) binds GSH and has some enzyme activity (Achilonu *et al*., 2014).

Table 1.1: Proteins that contain a domain with a GST-like fold and their biological functions

Protein name	Biological function	Reference
Escherichia coli glutaredoxins	Functions as oxidoreductases in *Escherichia coli*	(Fernandes *et al*., 2005 ; Aslund *et al*., 1994)
β-etherases	Protein acts as an enzyme that catalyses the unique cleavage and reductive cleavage of arylglycero-beta-aryl ether	(Masai *et al*., 1993)
Yeast nitrogen metabolism regulator URE2	The gene plays an important role in the cellular response to the nitrogen source	(Coschigano and Magasanik, 1991)
S-crystallins	Catalyse a number of biological molecules	(Tomarev *et al*., 1995; Tomarev and Zinovieva, 1988)
Chloride intracellular channel (CLIC)	Function as chloride ion channels in biological membranes	(Harrop *et al*., 2001)
Gamma subunit of eukaryotic translation factor 1 (EF1γ)	Functions as an elongation factor	(Koonin *et al*., 1994)

1.5. SspA family of proteins

Members of the stringent starvation protein A (SspA) family are a family of bacterial transcription factors that do not contain a typical DNA binding domain, but they regulate gene expression through protein-protein interactions with the (RNA polymerase) RNAP (Isihama and Saitoh, 1979). EcSspA is a transcriptional activator for the lytic development of phage P1 (Hansen *et al*., 2003). SspA orthologues found in a number of bacteria are implicated in the expression of genes involved in pathogenesis. MglA (Macrophage growth locus protein A) shows a strong sequence homology to SspA proteins (Cuthbert *et al*., 2015). However, the sequence identity between MglA protein and SspA proteins is approximately 25 %. Despite the low sequence identity, the proteins appear to be structurally similar (Cuthbert *et al*., 2005). MglA forms a heterodimer with SspA in *Francisella tularensis* to regulate virulence gene expression (Wrench *et al*., 2013). RegF is an SspA homologue in *Neisseria gonorrhoeae* which regulates the expression of virulence factors during *gonococcal* infection. RegF shows a 42 % identity to EcSspA (De Reuse and Taha, 1997). The widespread of SspA homologues in Gram-negative bacteria plays a significant role in an understanding of the mechanism of the control of gene expression by SspA. Therefore, orthologous proteins can contribute to the gain of insight into the structural basis for the function of EcSspA.

1.6. SspA as a member of the GST superfamily: protein structure

The structure of SspA proteins from *Haemophilus influenza*, *Pseudomonas fluorescens*, *Pseudomonas putida* and *Yersinia pestis* have been determined and show that SspA is a member of the cytosolic GST superfamily. SspA protein is a highly conserved protein amongst the Gram-negative bacteria. SspA homologues are found in more than fifty Gram-negative bacteria, which includes human, animal and plant pathogens, as well as in non-pathogenic bacteria. Multiple sequence alignment of SspA in Gram-negative bacteria has shown that SspA protein is highly conserved. A 40-100 % sequence identity between *Escherichia coli* strain K-12 SspA and fifty SspA orthologues is observed, this high level of sequence similarity suggests that SspA orthologues share the same structure and will therefore have the same function (Hansen *et al*., 2005a). Orthologous proteins are defined as proteins with similar or identical amino acid sequences and are therefore assumed to perform the same function (Makarova *et al*., 1999). There is a correlation between amino acid sequences of orthologous proteins and the evolutionary relationship between them

(Andrykovitch *et al*., 2003). The structure of an orthologue can be used to predict the structure of a protein. There is no structure available for EcSspA. There is however, a structure available of *Yersinia pestis* SspA (YpSspA).

The crystal structure of YpSspA was determined at a resolution of 2.0 Å (Hansen *et al*., 2005a). The primary sequence and secondary structure topology are illustrated in Figure 1.2. The protein is characterised by a predominantly alpha helical content with a few beta sheets present. Based on the solved crystal structure of YpSspA, it was observed that the protein assumes the characteristic fold of a GST similar to the GSTB1-1 (Nishida *et al*., 1998). Figure 1.3 shows that the structure of YpSspA exists as a dimer, with the two subunits related by a 2-fold axis. Three regions have been identified to be flexible in the tertiary structure; the N and C termini and the α2 helix (Hansen *et al*., 2005a). The structure of YpSspA also revealed a surface exposed pocket at the dimer interface, which is made of hydrophobic residues. The size, shape and properties of the pocket suggest that it mediates protein-protein interactions (Hansen *et al*., 2005a). This ability to mediate protein-protein interactions by SspA proteins is important due to the role the proteins play in transcriptional gene regulation. The lack of DNA binding domain in these proteins require that they adopt a structural conformation that will mediate interactions with proteins such as the RNAP (Hansen *et al*., 2005a). The existence of such a pocket has been observed in the crystal structure of *Schistosoma japonica* GST (SjGST) which is implicated in the binding of praziquantel (McTigue *et al*., 1995), suggesting the possibility of the binding of small molecules by YpSspA.

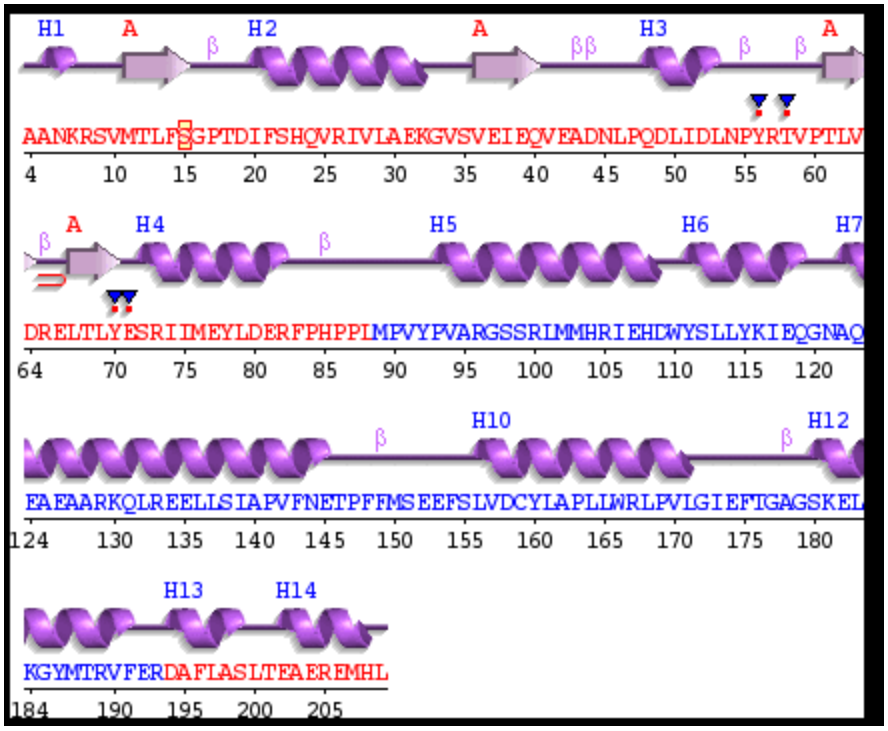


Figure 1.2: YpSspA sequence and secondary structure topology. The YpSspA sequence illustrating the proteins primary and secondary structural elements. Helices are labelled with a blue H and β-strands are labelled as β in purple. The image was adapted from PDBsum database (Laskowski *et al*., 1997).

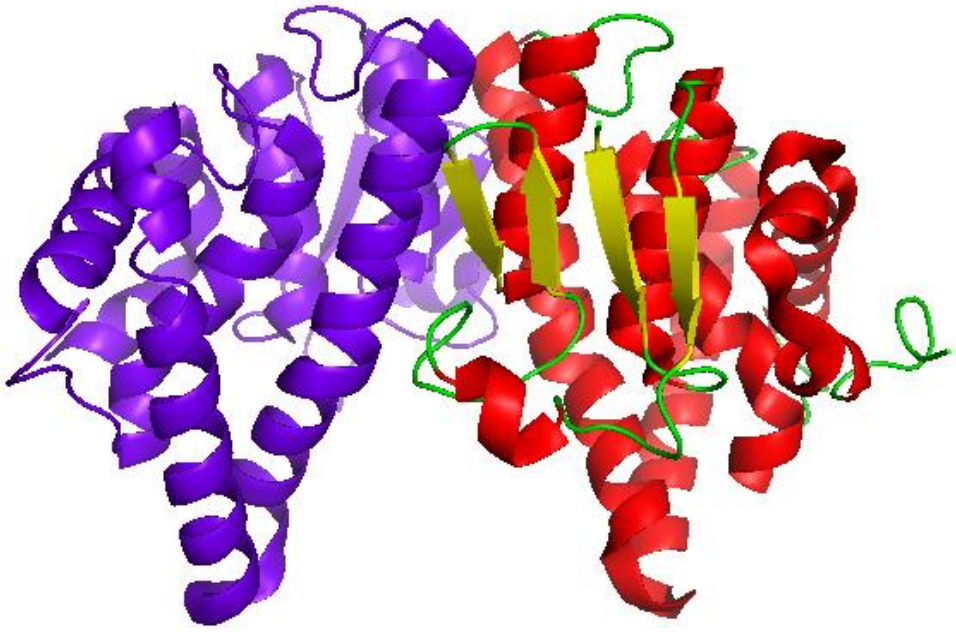


Figure 1.3: The ribbon diagram representation of dimeric YpSspA. The structure is coloured in order to show the two subunits that make up the homodimer. The subunit on the right is coloured according to the secondary structure; β-strands (yellow) and α-helices (red). Image was generated with PyMOL (DeLano, 2002) Protein Data Bank (PDB) code: 1YY7 (Hansen *et al*., 2005a).

A sequence alignment between YpSspA and EcSspA shows that the two proteins have an 83.7 % sequence similarity (Figure 1.4). A domain prediction of EcSspA based on its primary sequence shows that the protein folds into two domains (Figure 1.5); The N-terminal GST domain which adopts a characteristic thioredoxin fold (Martin, 1995) and a larger C-terminal GST domain which is predominantly alpha helical in nature.



Figure 1.4: Sequence alignment of EcSspA to that of YpSspA. The sequence alignment was performed using ClustalW (Larkin *et al*., 2007). The sequence for EcSspA and that of YpSspA are labelled.

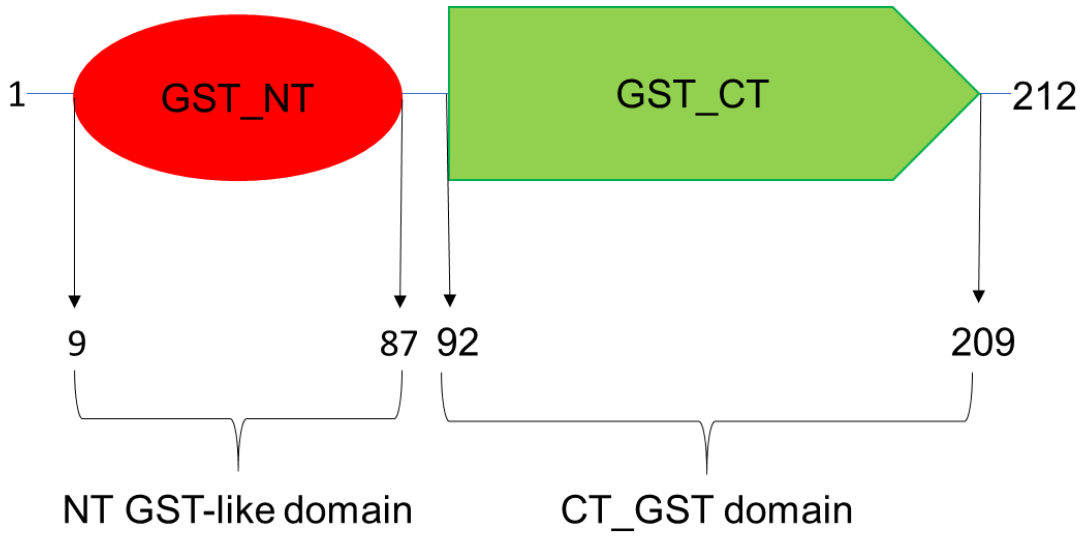


Figure 1.5: A schematic illustration of the domain structure of EcSspA as predicted by the ScanProsite server (de Castro *et al*., 2006). The numbers represent the amino acid corresponding to each domain structure.

1.7. *Escherichia coli*: Health and disease implications

Escherichia coli are Gram-negative bacteria that are commonly found in the gut of humans and other mammals. Most strains of *Escherichia coli* are non-pathogenic. However, some strains have developed the ability to cause disease. This pathogenicity can be attributed to the ability of the bacteria to survive and adapt to several environmental conditions such as acidic

conditions and adaptation to survive at low temperatures. Studies have shown that *Escherichia coli* are able to survive an exposure to acid. The assay conditions were based on those of the normal fasting stomach in which the pH is less than three and a gastric emptying of less than two hours. The ability of *Escherichia coli* to be acid resistant is defined as the percentage of cells that are able to survive under acidic conditions (Gorden and Small, 1993). A number of pathogenic strains of *Escherichia coli* have been identified and this includes; Enteropathogenic *Escherichia coli* (EPEC), Enterohaemorrhaegic *Escherichia coli* (EHEC), Enteroinvasive *Escherichia coli* (EIEC) and Enteroaggregative *Escherichia coli* (EAEC).

Escherichia coli O157:H7 strain, which is an EHEC strain, has been found to be present in the intestines of about one percent of healthy cattle. Beef may be contaminated during slaughtering and grinding of the beef whereby there is a transfer of the pathogen from the surface of the meat to the interior. Therefore, undercooking of the meat leads to bacterial survival and ingestion of the pathogen will give rise to food-borne diarrhoea (Boyce *et al*., 1995). In Africa there is a close contact between humans and large numbers of domestic animals that carry pathogenic *Escherichia coli* strains therefore posing as a threat to human health (Effler *et al*., 2001).

Enterohaemorrhagic strains of *Escherichia coli* (EHEC) result in life threatening haemorrhagic colitis and haemolytic uremic syndrome (Boyce *et al*., 1995). EHEC infections have been found to be mainly transmitted by food and water. *Escherichia coli* O157:H7 strain is an EHEC strain that has had outbreaks in numerous parts of the world resulting in severe illness and high mortality rate. In Southern Africa there was a water borne outbreak of *Escherichia coli* O157:H7 infection in Swaziland, with some cases spreading to nearby provinces of South Africa (Effler *et al*., 2001). The illness was characterised by bloody diarrhoea and vomiting. Epidemiological studies showed that the outbreak was waterborne, due to the contamination of surface waters by cattle carcasses, which could have been washed into rivers, and dams by heavy rains after a drought (Isaacson *et al*., 1993).

There is growing evidence that implicates pathogenic strains of *Escherichia coli* in diarrhoeal disease. Diarrhoegenic *Escherichia coli* strains are an emerging threat to human health in Africa and some parts of the world. Diarrhoea is one of the hallmarks of advanced human immunodeficiency virus (HIV) disease. In Africa diarrhoea has been shown to occur in 60-90% of patients with HIV infection (Bartlett *et al*., 1992). Enterotoxigenic strain of *Escherichia coli* were found to be responsible for the acute diarrhoea experienced by HIV

positive patients in Central African Republic. An association between the isolation of enteroaggregative Escherichia *coli* and persistent diarrhoea in HIV infected patients was also observed (Bartlett *et al*., 1998).

Both pathogenic and non-pathogenic *Escherichia coli* display stress response to sublethal environmental stresses. These stress responses can result in the resistance of the bacteria to the environmental stress trigger and ultimately survival of the bacteria. In some cases, the virulence of the bacteria is enhanced (Chung *et al*., 2006).

1.8. *Escherichia coli* stress response mechanisms and survival

Escherichia coli must contend with several stress triggers during their growth, survival and infection in a foreign host environment. These stress conditions result in the recruitment of proteins that lead to the transcription of genes that subsequently leads to stress tolerance and survival (Chung *et al*., 2006). The ability of a given bacteria to be stress resistant is linked to the pathogenicity of the bacteria (Testerman *et al*., 2002). Both pathogenic and non-pathogenic *Escherichia coli* have evolved mechanisms involved in the regulation of stress and stress tolerance to starvation, thermal and acid stresses (Chung *et al*., 2006).

Microorganisms possess signal transduction pathways that sense environmental stress and mediate the expression of genes involved in cellular defence mechanisms. These signal transduction pathways helps the microorganism to survive the stress conditions. They fulfil this function through the involvement of sigma factors, which are small RNA polymerase (RNAP) binding proteins. Gene expression during stress is affected by different sigma factors. In *Escherichia coli* the σ^S (RpoS) has been identified as the main regulator of the general stress response in microorganisms (Small *et al*., 1994). RpoS is induced when the cell enters stationary phase and when it encounters stresses such as high osmolarity, weak acids, starvation and others (Lange and Hengee-Aronis, 1994). *Escherichia coli* O157:H7 strain was shown to have the ability to survive a prolonged exposure to a low pH stress when in the stationary phase and acid adapted exponential phase. The acid tolerance of the bacteria is dependent upon the growth phase of the bacteria, with those in the exponential phase being less tolerant (Jordan *et al*., 1999).

When a microorganism encounters nutrient limiting or starvation conditions, it responds to these stresses by undergoing physiological changes and alters certain metabolic pathways. They then switch to starvation-induced activities that will enable them to be resistant to a

number of environmental stresses. Stringent starvation protein A (SspA) is a protein in *Escherichia coli* implicated to play a role in survival during starvation and prolonged stationary phase. Williams *et al*. (1994) have shown that EcSspA may be another protein responsible for gene expression during stationary phase. The expression of EcSspA protein increases with a decreasing growth rate, prompted by glucose, nitrogen or amino acid starvation (Williams *et al*., 1994). Indeed, an understanding of the structure of EcSspA is significant to the gain of insight into the function of the protein as a regulator of stress responses.

1.9. Role of EcSspA in stress response and pathogenesis

The EcSspA is commonly known as an RNAP associated protein which plays a significant role in lytic growth of bacteriophage P1. EcSspA has been shown to result in the expression of eleven proteins during exponential growth (Williams *et al*., 1994). The exact role of EcSspA in gene expression in *Escherichia coli* is unknown. The expression of EcSspA when the bacteria enters the stationary growth phase suggests that the protein plays an essential role in stress response of the bacteria.

EcSspA has been shown to be a transcriptional activator of bacteriophage P1 late genes (Hansen *et al*., 2003). EcSspA is important for the survival of the bacteria during an acid induced physiological stress. EcSspA inhibits the accumulation of Histone-like nucleoid structuring (H-NS) during stationary phase, which is a negative modulator of genes involved in virulence and adaptation. H-NS acts as a repressor and influences the transcription of genes through the repression of genes. Hansen *et al*. (2005b) have shown that SspA relieves acid tolerance gene from H-NS repression and that the absence of SspA results in an acid-sensitive phenotype in *Escherichia coli* cells (Hansen *et al*., 2005b). The acid tolerance of *Escherichia coli* is significant for their survival in the low pH encountered in the gastrointestinal tract during infection and therefore contributes to pathogenesis. EcSspA was shown to positively regulate the expression of Enterohemorrhagic *Escherichia coli* (EHEC) virulence factors by limiting the intracellular level of H-NS (Hansen and Jin, 2012).

1.10. Study rationale and motivation

Escherichia coli stringent starvation protein A shares sequence and structural homology with a number of GST structural family proteins. The importance of a biophysical characterisation of a prokaryotic homologue of a protein cannot be overemphasised. The information obtained

from bacterial proteins could be extrapolated to study mammalian functions given the diversity of bacteria and their ability to adapt to extreme environmental conditions. This can also be extended to an understanding of the mechanism of the behaviour of pathogens in host organism. The critical role played by EcSspA in stress response and pathogenesis as well as other suggested biological functions of the protein, necessitates an in-depth biochemical and structural studies of the protein. By further exploring the properties of EcSspA through structure-function studies and showing the similarities with other members of the protein family we can further elucidate and gain an insight into protein function. The observation that SspA adopts the GST-fold and shares a similar N-terminal motif with GST proteins, suggested the need to investigate the capacity of EcSspA to perform some of the generic functions observed in canonic GSTs such as catalytic and ligandin functions.

1.11. Aim and objectives

1.11.1. Aim

The main aim of this study is to perform structural and functional characterisation of EcSspA.

1.11.2. Objectives

- To submit the *EcSspA* gene for subcloning into the pET-28a vector at the GenScript gene synthesis lab
- Use the plasmid containing the cDNA that encodes for EcSspA to transform *Escherichia coli* strain BL21 (DE3) pLysS cells
- Overexpress the EcSspA protein in *Escherichia coli* cells and purify the His-tagged SspA using immobilised metal affinity chromatography (IMAC)
- To confirm the identity of the protein using peptide sequencing mass spectrometry
- Determine the secondary, tertiary and quaternary structure of EcSspA using circular dichroism, intrinsic fluorescence and SE-HPLC, respectively
- Determine thermal unfolding properties of EcSspA using circular dichroism (CD) spectroscopy
- To assess the hydrophobic binding pockets of EcSspA using ANS fluorescence binding assay
- Build a homology model of EcSspA and perform molecular docking of ANS to EcSspA
- To determine the thermodynamics of ANS binding using isothermal titration calorimetry (ITC)

- Perform a glutathione binding test and measurement of glutathione transferase activity using the GSH/CDNB assay

2. CHAPTER TWO

EXPERIMENTAL PROCEDURES

2.1. Materials

The pET-28a plasmid encoding SspA tagged with a hexahistidine (His_6-tag) on the N-terminal was synthesised and purchased from GenScript (Piscataway, New Jersey, USA). A 3-Colour prestained molecular mass marker (PM30-500) was purchased from Integrated Bioworks (Centurion, South Africa). GeneJet™ plasmid MiniPrep kit was purchased from Inqaba biotech (Pretoria, South Africa). Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was purchased from Sigma-Aldrich (St. Louis, MO, USA) and a 5 mL Histrap Purification column was purchased from GE Health care (Chicago, USA). BL21 (DE3) pLysS competent cells were purchased from New England Biolabs (Ontario, Canada). The pure hGSTP1-1 protein was kindly provided by Donald Mhlanga (Protein Structure Function Research Unit, University of the Witwatersrand, Johannesburg). The gel filtration standards were purchased from Biorad laboratories (Berkeley, USA). All the other reagents used were of analytical grade.

2.2. Methods

2.2.1. Bacterial transformation and cDNA sequence verification

2.2.1.2. Bacterial transformation

The pET-28a plasmid containing the cDNA insert that codes for EcSspA was used to transform *Escherichia coli* BL21 (DE3) pLysS cells. The pET-28a vector encodes for kanamycin resistance which will confer kanamycin resistance to cells transformed with pET-28a. The chosen plasmid also allows for the expression of a fusion protein fused with a His_6-tag on the N-terminal site followed by thrombin cleavage site which will aid protein purification.

Competent cells were transformed using a one-step method as previously described by Chung *et al*. (1989). Initially the lyophilised 4 µg of plasmid DNA was dissolved in 100 µL of deionised water. The competent cells were thawed on ice for 30 minutes. Both the experimental and negative control reaction tubes were prepared. To the experimental tube 2 µL of plasmid DNA was added and to the negative control 2 µL of water was added. The reaction tubes were thereafter incubated on ice for 30 minutes and heat shocked for 45

seconds on a heating block at 42 °C. The samples were immediately chilled on ice for 10 minutes. The cells were then grown by adding 900 µL of 2x YT medium (1.6 % (w/v) tryptone, 1 % (w/v) yeast extract, 0.5 % (w/v) NaCl with no antibiotics added) into each tube, followed by an incubation of the cells for one hour at 37 °C with a 230 rpm agitation on the New Brunswick benchtop incubator. The cells were then plated on Luria Broth (LB)-agar plates (1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1.0 % (w/v) NaCl, 1.5 % (w/v) agar) supplemented with 100 µg/mL kanamycin and 34 µg/µL chloramphenicol antibiotics and incubated overnight at 37 °C. Kanamycin resistance by the cells is conferred by the *Kan* gene of the pET-28a expression vector and chloramphenicol resistance arises from the pLysS plasmid which encodes the production of acetyl transferase and lysozyme.

2.2.1.2. pET-28a plasmid preparation and sequence confirmation

A single colony from the agar plate was selected to inoculate a flask of 100 mL of 2x YT liquid medium supplemented with100 µg/mL kanamycin and 34 µg/µL chloramphenicol antibiotics. The flask was incubated at 37 °C overnight (~ 16 hours) at 230 rpm agitation in the shaking incubator. The overnight culture was used for plasmid DNA purification and glycerol stocks preparation. The culture was used to make up 1:1 glycerol stocks (cell culture: 80 % glycerol stock). The pET-28a plasmid DNA was extracted and purified from the overnight culture of transformed *Escherichia coli* BL21 (DE3) pLysS cells using the GeneJet™ plasmid MiniPrep kit in accordance to the manufactures' instructions. Plasmid isolation was based on the differential solubility of the plasmid DNA and genomic DNA. The plasmid DNA was solubilised in solution by an alkaline lysis solution. The concentration of the purified DNA and purity were assessed using the NanoDrop™ ND-1000 spectrophotometer (ThemoScientific, USA). The purified plasmid DNA sample was sent to Inqaba biotech laboratories to confirm the identity of the gene and the absence of mutations thereof. The nucleotide sequence was sequenced in accordance with the DNA sequencing technique reported by Sanger *et al*. (1977).

2.2.2. Heterologous overexpression of EcSspA and protein purification

2.2.2.1. Expression trials of EcSspA

Conditions for the expression of EcSspA were optimised by performing expression trials. The expression of EcSspA in *Escherichia coli* BL21 (DE3) pLysS cells was optimised through expression trials by changing isopropyl-β-D-1-thiogalactopyranoside (IPTG) concentration,

expression time and temperature. A 2x YT medium supplemented with 100 µg/mL kanamycin and 34 µg/µL chloramphenicol antibiotics was prepared and inoculated with an overnight culture of transformed *Escherichia coli* cells at a 100-fold dilution for small scale expression trials. The medium was incubated at 37 °C in a shaking incubator until the cells reached mid-log phase (OD_{600} = 0.6). The expression of EcSspA was induced by using IPTG at three different final concentrations (0.3 mM, 0.6 mM and 1 mM). Replicate inductions were done at different temperatures (30 and 37 °C). For 30 °C expression, the cells were allowed to cool down at 4 °C prior to induction. For the 37 °C expression there was no cooling down of the cells prior to induction. A negative control of 0 mM IPTG was prepared at all temperatures. Aliquots of culture were removed at two different time intervals post induction (4 and 6 hours), and the cells harvested by centrifugation at 12000 x g at room temperature. The pellet was resuspended in deionised water, lysed through pulse sonication on ice and centrifuged at 12000 x g to separate the soluble fraction from the insoluble fraction. SDS-PAGE sample buffer was added to the samples and run on an SDS-PAGE to monitor protein expression.

2.2.2.2. Protein overexpression and purification

EcSspA was overexpressed from a litre batch culture of transformed *Escherichia coli* BL21 (DE3) pLysS cells. A glycerol stock of *Escherichia coli* BL21 (DE3) pLysS cells was used to inoculate 10 flasks of fresh 100 mL 2x YT media supplemented with kanamycin (100 µg/ml) and chloramphenicol (34 µg/µL). A 100-fold dilution of the glycerol stock was used to inoculate the fresh media. The cells were grown at 37 °C with shaking at 230 rpm until an OD_{600} of 0.6 was reached. The cells were then allowed to cool down at 4 °C. IPTG at a final concentration of 0.6 mM was added to the cultures in order to induce overexpression of the EcSspA protein. The *Escherichia coli*BL21 (DE3) pLysS cells were then incubated for a further 6 hours with shaking at 230 rpm at 30 °C in order to achieve optimum protein overexpression.

The cells were harvested by centrifugation at 4248 xg for 20 minutes. The *Escherichia coli*BL21 (DE3) pLysS cells were resuspended in 50 mM Tris-HCl, 0.5 M NaCl, 50 mM imidazole and 0.02 % sodium azide at pH 8to which phenylmethylsulfonyl fluoride (PMSF) to a final concentration of 1 mM was added. The resuspended cells were then stored at – 20 °C overnight to promote cell lysis. The cells were thawed over a gentle rotation at room temperature. DNAse and lysozyme to a final concentration of 1 mg/ml were added to the cell

suspension and incubated for 30 minutes at room temperature. The cells were then sonicated on ice for six 30 second cycles at medium intensity using a 0.5 second pulse interval. The cell lysate was then centrifuged at 23000 x g for 30 minutes at 4 °C.

Purification of EcSspA was achieved by nickel (Ni^{2+}) charged affinity chromatography. The supernatant obtained above following sonication was loaded onto the nickel charged affinity column using the AKTA fast protein liquid chromatography (FPLC) purification system. The column was pre-equilibrated with the resuspension buffer. After loading the supernatant onto the column, the column was washed with the resuspension buffer until the absorbance at 280 nm of the effluent was zero to ensure that all unbound protein contaminants were eluted from the column. The bound EcSspA protein was then eluted from the column using a 0-300 mM imidazole concentration gradient in 50 mM Tris-HCl, 0.5 M NaCl, 0.02 % sodium azide at pH 8. Purified EcSspA was dialysed and stored in a storage buffer (20 mM sodium phosphate, 1 mM EDTA, 0.02 % sodium azide and 150 mM NaCl at pH 7.4) and stored on ice at 4 °C. The whole cell extract, lysate and eluted fractions from the affinity column were analysed by SDS-PAGE (section 2.2.3).

2.2.3. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The presence and purity of EcSspA was assessed by SDS-PAGE (Laemmli, 1970) using a 15 % separating gel (30 % acrylamide, 2.7 % bisacrylamide; 10 % sodium dodecyl sulfate (SDS); 10 % ammonium persulfate and 1.5 M Tris-HCl pH 8.8) and a 4 % stacking gel (30 % acrylamide; 2.7 % bisacrylamide; 10 % SDS; 10 % ammonium persulfate). The protein samples were incubated with reducing sample buffer (125 mM Tris-HCl, 4 % (w/v) SDS, 20 % (w/v) glycerol, 10 % (v/v) β-mercaptoethanol, pH 6.8 and bromophenol blue) for 5 minutes at 95 °C prior to loading. SDS-PAGE gels were run at 180 V for 40 minutes. The gels were stained in a 0.1 % (w/v) Coomassie Blue R250 staining solution containing 10 % (v/v) glacial acetic acid, 50 % (v/v) methanol and 40 % (v/v) distilled water. The gels were destained in 10 % (v/v) glacial acetic acid, 40 % (v/v) methanol and 40 % (v/v) distilled water. The distance migrated by each of the protein bands was compared to that migrated by protein standards of known molecular mass which migrated under identical conditions. Protein separation on the SDS-PAGE was monitored using a prestained protein marker.

2.2.4. Protein concentration determination

The concentration of purified EcSspA was determined using absorbance spectroscopy. The absorbance value at 280 nm and the extinction coefficient of the protein were used to calculate the concentration. The concentration of the purified protein was subsequently determined using the Beer-Lambert law:

$$A = \varepsilon c l \quad (1)$$

where A is the absorbance at 280 nm, ε is the molar extinction coefficient of EcSspA at 280 nm (M^{-1} cm^{-1}), c is the concentration of the protein in M, and l is the path length of the cuvette (1 cm). The molar extinction coefficient of EcSspA was calculated using the extinction coefficients of tryptophan, tyrosine and cysteine residues (Mach *et al*., 1995):

$$\varepsilon = 5550\Sigma Trp + 1340\Sigma Tyr + 150\Sigma Cys \quad (2)$$

$$= 5550(3) + 1340(8) + 150(1)$$

$$= 27520\ M^{-1}\ cm^{-1}\ (55040 M^{-1}\ cm^{-1}\ \text{for dimer})$$

Absorbance at 280 nm of the undiluted protein was determined by fitting a linear regression to the absorbance at 280 nm of serially diluted EcSspA. All readings were corrected for buffer (20 mM sodium phosphate, 1 mM EDTA, 0.02 % sodium azide and 150 mM NaCl at pH 7.4).

2.2.5. Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS)

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is a highly sensitive and specific technique which allows for both the physical separation of molecules using liquid chromatography and analysis of the mass to charge ratio using mass spectrometry. An SDS-PAGE gel containing purified EcSspA protein sample was sent to the Council for Scientific and Industrial Research, (CSIR, Pretoria, South Africa) where it was subjected to LC-MS/MS in order to confirm the identity of the protein.

Initial sample preparation prior to LC-MS/MS involved trypsinisation of the sample based on a protocol described by Shevchenko *et al*. (2006). Basically, this involves the extraction of the protein gel band and destaining the protein using 50 mM ammonium bicarbonate (NH_4HCO_3)/50 % methanol followed by an alkylation step using 55 mM iodoacetamide in 25

mMNH_4HCO_3.Trypsinisation of the protein was achieved by adding trypsin to the excised gel band and incubating it overnight at 37 °C to allow for protein digestion. After trypsinisation the protein digests were analysed using a Dionex Ultimate 3000 RSLC system coupled to an AB Siex 6600 TripleTOF mass spectrometer. The acquired tandem mass spectrometry peptide sequence data were submitted to ProteinPilot™ v4.0.8404 software (https://sciex.com/products/software/proteinpilot-software) which uses a paragon search engine (Shilov *et al*., 2007) to compare the obtained MS/MS data to a custom database. This search engine identifies the peptides by inference, in order to determine the identity of the protein present in the digested peptide mixture. The protein inference works on an algorithm that takes the top ten ranked peptides hypothesised for each spectrum input and filters the peptides into proteins that can be reported as having detected at a certain level of confidence (Shilov *et al*., 2007).

2.2.6. Structural characterisation of EcSspA

2.2.6.1. Far-UV circular dichroism (CD)

Circular dichroism is a tool that is commonly employed for the quantitative analysis of the secondary structure of a protein (Greenfield, 1996). The chromophore is the amide bond of the polypeptide backbone which absorbs circularly polarised light in the far-UV region (250-190 nm) (Woody, 1995). A protein with a high alpha helical content is characterised by two troughs on the spectrum, one at 208 nm and the other at 222 nm. While the presence of β structures is characterised by a positive peak at 198 nm. Circular dichroism was used to assess the global secondary structure of EcSspA.

Far-UV CD spectral measurements were performed 20 °C using the Jasco-1500 spectropolarimeter with a spectral bandwidth of 5 nm, a data pitch of 1 nm, and a quartz cuvette with a path length of 2 mm. The spectra were recorded as an average of seven scans at a scan speed of 100 nm/min in triplicate. The protein concentration used was 2 µM in milliQ water at pH 7 to minimise the signal to noise ratio. The spectra were converted from θ (m.deg) to Θ_{MRE} using the following equation:

$$[\Theta] = \frac{100\theta}{cnl} \tag{3}$$

where Θ is the mean residual ellipticity (deg.cm^2.dmol^{-1}), θ is the measured ellipticity (mdeg), *C* is the protein concentration (mM), *n* is the number of residues in the protein and *l*

is the path length (cm) (Woody, 1995). The CD spectroscopic data were further quantitatively analysed using the Dichroweb server (Whitmore and Wallace, 2008) to determine the secondary structural content of the protein. Temperature dependent studies made use of a temperature control system to maintain the required temperature between 20 – 80 °C for thermal melting of the protein.

2.2.6.2. Intrinsic tryptophan fluorescence

Fluorescence spectroscopy is a method that is highly sensitive and is widely used for the analysis of proteins (Lackowicz, 2006). Intrinsic protein fluorescence permits the measurement of the fluorescence of aromatic amino acids that are intrinsically found in the native structure of a protein. The aromatic residues are very helpful when detecting the dynamics and changes in the tertiary structure of a protein (Lackowicz, 2006). Signals generated from fluorescence spectroscopy are environmentally sensitive and therefore play a significant role in providing information about the local tertiary structure of the protein. A fluorescence parameter that is commonly used to monitor changes in the local environment of a fluorophore is the maximum emission wavelength. This wavelength is dependent on the exposure of the excited fluorophore to its surrounding environment. Any structural changes within the vicinity of the excited fluorophore can be monitored by monitoring the wavelength at which the excited fluorophore emits energy (Royer, 1995).

All fluorescence measurements were conducted on a Jasco FP-6300 spectrofluorometer at 20 °C. The excitation and emission bandwidths were set at 5 nm. A quartz cuvette with a path length of 10 mm was used in all experiments. The protein at 0.5 µM in a 5 mM potassium phosphate buffer at pH 7.4 was excited at 295 nm for tryptophan fluorescence. All spectra were measured as an average of three accumulations at a scan speed of 200 nm/min. The spectra were plotted using the Sigmaplot® software, v12.0.

For fluorescence measurements of the denatured form of EcSspA, native EcSspA was unfolded in 8 M urea in 5 mM potassium phosphate buffer at pH 7.4. The urea concentration was confirmed using the refractive index of urea:

$$[\text{Urea}] = 117.66\Delta n + 29.753\Delta n^2 + 185.56\Delta n^3 \quad (4)$$

where Δn is the difference between the refractive indices of urea and the buffer (Pace, 1986).

2.2.6.3. Size exclusion high pressure liquid chromatography (SE-HPLC)

The oligomeric state of EcSspA in solution was determined using SE-HPLC in order to characterise the quaternary structure of the protein. The homodimeric quaternary structure of hGSTP1-1 is well documented (Reinemer *et al*., 1991; Dirr *et al*., 1994) and was used as a positive control. SE-HPLC is a technique that involves the principles of size exclusion chromatography and provides an estimate of the size of proteins within a short space of time (Tayyab *et al*., 1991).

The Shimadzu UFLC SPD-20A SE-HPLC system was used for the analysis of the proteins together with a TSK gel super SW2000 size exclusion column (resolution of 5-150 kDa). This means that only proteins within this molecular mass range can be resolved by the column. Protein samples (20 µL) were injected into the system and absorbance of the effluent monitored at an absorbance of 280 nm at a flow rate of 0.25 ml/min. EcSspA (91 µM) and hGSTP1-1 (110 µM) were loaded separately onto the column. The mobile phase in the system was a 20 mM sodium phosphate buffer, pH 7 with 500 mM NaCl. The gel filtration standards used for column calibration and construction of the standard curve were: Thyroglobulin (670 kDa), γ-globulin (154 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa).

2.2.7. Homology modelling

Homology modelling is a technique that is used to predict the three dimensional structure of proteins using experimentally solved structures as templates (Biasini *et al*., 2014). The three dimensional structure of EcSspA was generated by inputting the amino acid sequence of EcSspA into SWISS-MODEL (Arnold *et al*., 2006). SWISS MODEL is an online structure homology based modelling server which allows for the prediction of a structure of a protein using experimentally solved structures as templates, by a simple input of the amino acid sequence (Arnold *et al*., 2006).The model was generated using YpSspA (1yyz) as a template based on an 83.73 % sequence identity between the two SspA proteins.

2.2.8. Molecular docking

Binding site prediction

In order to identify potential ligand binding sites within the structure of EcSspA, the EcSspA homology model obtained from section 2.2.7 was subjected to an analysis with SiteMap incorporated within the Schrodinger Maestro (Schrödinger Release 2018-1: Maestro,

Schrödinger, LLC, New York, NY, 2018; https://www.schrodinger.com/maestro). SiteMap is a tool that provides a fast and effective means of identifying potential binding sites with a high degree of confidence (Halgren, 2007). SiteMap characterises the potential binding site by calculating a combination of properties at each site, which includes; size of the site, the degree of enclosure of the site by protein, the degree of exposure to the solvent, and the hydrophilic and hydrophobic properties of the site (Halgren, 2007). For the purpose of this study, SiteMap was applied to the entire protein structure to identify possible binding sites. The results obtained from SiteMap were then used as a basis for docking.

Protein and ligand preparation

EcSspA homology model structure was prepared for docking using protein preparation wizard in Schrodinger Maestro v11.2. This was done so as to fix common problems encountered with protein PDB files. This includes; missing side chains, incorrect protonation of a residue and steric clashes through a protein minimisation with the OPLS 2005 force field incorporated in the protein wizard. The structure of the ANS ligand was downloaded from PDB (PDB id: 2AN) and the ligand was prepared using the ligand preparation wizard, this was done so as to produce the low energy conformer of the ligand using OPLS 2005 force field.

Grid generation and ligand docking

Glide is a method used for docking ligands to the rigid three dimensional structure of a known protein receptor (Friesner *et al*., 2004). Glide uses an algorithm that approximates a search of positions, orientations and conformations of the ligand in the receptor through a series of filters. This filters range from best ligand conformation, to a diameter test, refinement of the ligand, grid minimisation to examine the torsional minima of the ligand structure and finally resulting in a top hit (Friesner *et al*., 2004). Glide was used to generate a receptor grid at the SiteMap-identified binding site of EcSspA. The molecular docking tool, Glide SP (standard precision) (Schrodinger Inc., USA) was used for ligand docking into a rigid receptor binding pocket.

2.2.8. GSH/CDNB conjugation assay

The standard GSH/CDNB conjugation assay was used to determine the enzymatic activity of EcSspA (Habig and Jakoby, 1981). The conjugation of 1-chloro-2, 4-dinitrobenzene (CDNB) to GSH occurs through a nucleophilic aromatic substitution of chlorine by the thiol group of

GSH. The substitution reaction results in the formation of the chromophore, 1-(*S*-glutathionyl)-2,4-dinitrobenzene as well as a chloride ion. Since most GST proteins are known to be catalytic towards GSH and CDNB, the ability of the EcSspA to bind GSH and catalyse the conjugation reaction was tested using the standard assay below.

Triplicate sample solutions of 10 nMhGSTP1-1 (a positive control) and EcSspA were prepared in 50 mM potassium phosphate, 1 mM EDTA and 0.02 % sodium azide at pH 6.5. Enzyme activity was measured spectrophotometrically at 20 °C using a Jasco V-550 UV/Vis spectrophotometer by monitoring the formation of 1-(*S*-glutathionyl)-2,4-dinitrobenzene at 340 nm (Habig and Jakoby, 1981). Both EcSspA and hGSTP1-1 were added separately to a 3 mL final volume solution with a final concentration of 1 mM GSH and 1 mM CDNB (dissolved in 99.9 % absolute ethanol). The enzyme activity of the reaction was determined from the slopes of linear progress curves. A non-enzymatic control reaction was also measured, to account for the spontaneous reaction.

2.2.10. GSH agarose affinity chromatography

GSH-agarose affinity chromatography is a method that is commonly used to bind GSH binding proteins such as GSTs and glutathione peroxidise (Smith and Johnson, 1988). Since EcSspA has a thioredoxin-like domain which has been shown in other proteins to bind GSH, this method was employed to establish if EcSspA can also bind GSH.hGSTP1-1, known to bind GSH and catalyse its conjugation, was used as a positive control in this study.

Purified hGSTP1-1 and EcSspA were used to test for their GSH-binding capability. The GSH-agarose beads of a 2 mL column were initially washed with 10 column volumes of equilibration buffer (50 mM Tris-HCl, 0.5 M NaCl, 50 mM imidazole and 0.02 % sodium azide at pH 8). Each protein was applied to the GSH-agarose column followed by a column wash with the equilibration buffer. Any bound protein was subsequently eluted from the column with an elution buffer (10 mM glycine-NaOH, 0.02 % sodium azide, pH 10). The effluent from the flow through, wash and desorption steps were collected and analysed for the presence of the proteins by SDS-PAGE. Protein that binds the immobilised GSH was expected to be present in the desorption step due to the high pH of the elution buffer (Smith and Johnson, 1988).

2.2.11. ANS binding studies

2.2.11.1. ANS fluorescence

8-Anilino-1-naphthalene sulfonate (ANS) is an amphipathic dye that is used to characterise hydrophobic patches on the surface of a protein (Slavik *et al*., 1982)and has been used extensively to monitor the ligandin (i.e., non-catalytic) function of GSTs (Oakley *et al*., 1999). When excited at a wavelength of 390 nm, free ANS will emit energy at 540 nm, but when the ANS molecule binds to an exposed non-polar surface its emission maximum wavelength displays a blue shift accompanied by an increase in its fluorescence intensity. The extent of the blue shift is largely dependent on the hydrophobicity of the ANS binding site (Slavik *et al*., 1982).

A 200 µM stock solution of ANS was prepared in20 mM potassium phosphate containing 0.02 % sodium azide and 1 mM EDTA at pH 6.5 and the concentration of ANS confirmed spectrophotometrically at 350 nm using the extinction coefficient of 4950 M^{-1} cm^{-1}(Nilsson and Peterson, 1975).EcSspA (2 µM) was incubated with 100 µM ANS for an hour at room temperature before analysis. The blank contained 100 µM ANS in buffer without protein. The fluorescence of ANS was determined in a Jasco FP-6300 spectrofluorometer with a 10 mm path length cuvette. ANS was excited at 390 nm with both the emission and excitation bandwidths set at 5 nm. All spectra were measured as an average of three accumulations at a scan rate of 200 nm/min. The emission spectra were measured over a 400 – 600 nm wavelength range. The spectra were recorded at 20 °C and corrected for the ANS blank.

An ANS-protein binding isotherm was generated using titrating 2 µM EcSspA with ANS over a concentration range of 0-100 µM followed by measuring the fluorescence intensity of ANS at 480 nm as described above. The titration data were fitted to a hyperbolic function using SigmaPlot v12.0:

$$F480 = \frac{Fmax[ANS]}{K\mathrm{d}+[ANS]} \quad (5)$$

where F480 is the fluorescence intensity at 480 nm, Fmax, is the maximum fluorescence intensity of ANS bound to protein, and K_d is the dissociation constant.

2.2.10.2. Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a highly sensitive technique that is commonly employed when characterising the interaction between a macromolecule and a small ligand or an interaction between two macromolecules (Ladbury and Chowdhry, 1996). The technique measures the binding energetics of the interacting molecules by measuring the heat absorbed or released during the interaction. The binding affinity (K_a), the heat of binding (ΔH) and the number of binding sites (N) are independent variables which can be derived directly from the titration. The entropy (ΔS) and free energy (ΔG) are dependent variables which are obtained by calculation (Ladbury and Chowdhry, 1996):

$$\Delta G = -RT\ln(K_a) = \Delta H - T\Delta S \quad (6)$$

where ΔG refers to Gibbs free energy (kJ/mol), R refers to the gas constant (8.3145 J/mol/K), T refers to the temperature (K), K_a refers to the association constant, ΔH refers to the enthalpy (kJ/mol), and ΔS refers to the entropy (kJ/mol).

Isothermal titration experiments were performed using a Nano-ITC instrument (TA instrument, Delaware, USA) at 20 °C. All titrations were performed in a 20 mM potassium phosphate buffer containing 1 mM EDTA at pH 6.5. The protein was first dialysed into this buffer which was subsequently used as a solvent for the ANS ligand. The ANS ligand was prepared to a final concentration of 10 mM in the dialysate buffer. This was done so as to ensure the solvent conditions for both the ligand and protein solutions are identical. The sample cell (1000 µL volume) contained 50 µM (homodimer) protein and the syringe (250 µL volume) contained 10 mM ANS. The protein was stirred at a rate of 230 rpm during the initial equilibration. Computer controlled titrations were performed by injecting an initial 2 µL injection of ANS followed by 45 injections(5µL each of ANS) into the sample cell containing the protein. The data obtained from titrating ligand into buffer without protein under the same conditions were used to correct the ligand into protein titration data for heats of dilution. The corrected titration data were fit to a single binding site independent variable model by non-linear least squares by the Nano-ITC NanoAnalyse data analysis software.

2.2.12. Crystallisation of EcSspA

Crystallisation of EcSspA was attempted using the Morpheus® crystallisation screen (Gorrec, 2009). The screening was achieved using the Mosquito crystallisation robot (TTP labtech: https://www.ttplabtech.com/) at the University of Cape Town. The crystallisation robot was

used to set up three 200 nL drop size per screen condition (96 × 3) at the following protein: precipitant ratios; 1:1, 3:1 and 1:3 and a reservoir volume of 50 µL. The protein was prepared at a concentration of 4.6 mg/ml in a 20 mM sodium phosphate buffer, 1 mM EDTA, 0.02 % sodium azide buffer at pH 7.4.The protein crystallisation was carried out with an aim of obtaining high quality crystals for structural determination.

3. CHAPTER THREE

RESULTS

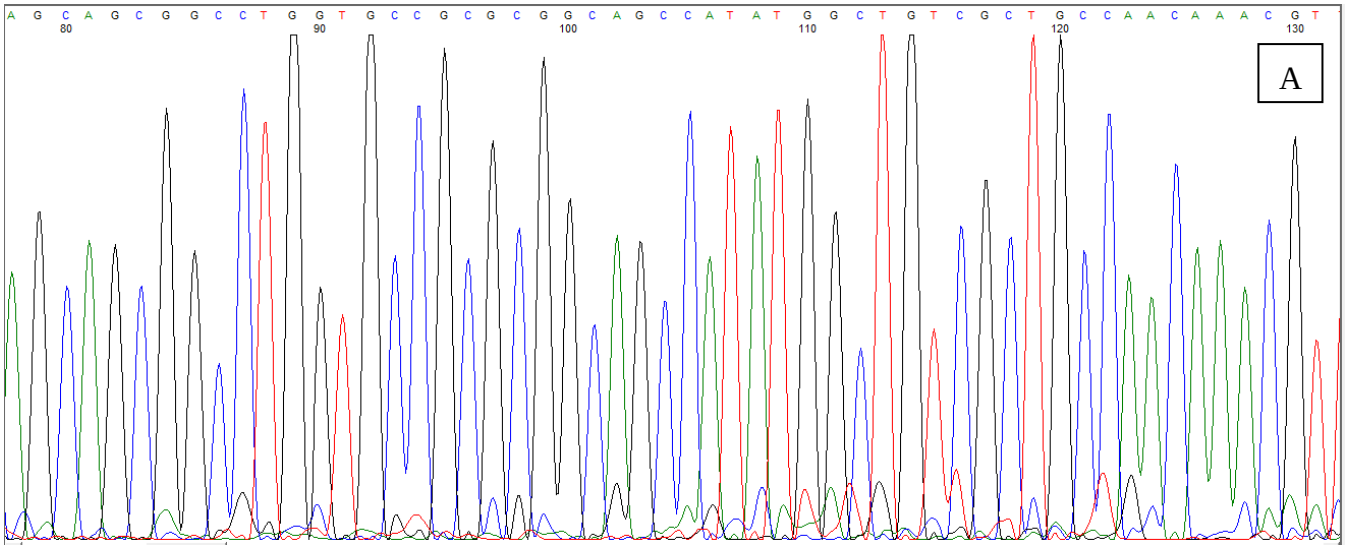
3.1. cDNA sequence verification

Sanger DNA sequencing results from Inqaba Biotech (Pretoria, South Africa) confirmed that the pET28a vector contained the nucleotide sequence encoding for SspA (Figure 3.1A). Results from translated Basic Local Alignment Search Tool (BLASTx) (Altschul *et al*., 1997), shows that the query sequence obtained from Inqaba Biotech when translated has a 100 % sequence identity to the protein sequence of SspA from *Escherichia coli* (Figure 3.1B).

3.2. Expression trials of SspA in *Escherichia coli* BL21 (DE3) pLysS cells

Expression trials of SspA were set up using *Escherichia coli* BL21 (DE3) pLysS cells at 30 °C and 37 °C (Figures 3.2 and 3.3). Varying concentrations of IPTG and induction periods were tested for the optimisation of protein overexpression. Induction was done when the cells reached an OD_{600} = 0.6 (mid-log phase of growth curve). The medium used was a 2x YT LB medium. At 37 °C the expression of SspA was observed after induction with IPTG at the three different IPTG concentrations. The protein was found in both the soluble and insoluble fraction. The bands appeared to be fainter when the protein is allowed to express for a six hours period at this temperature (Figure 3.2B).

Optimisation trials were also performed at 30 °C using the same expression cells. In this case more of the protein was expressed, as evidenced by a much thick and darker band on the SDS-PAGE (Figure 3.3). The presence of the protein in the pellet is most likely to be attributed to insufficient sonication of the cells as opposed to the formation of insoluble aggregates by the protein and this was later proven by resuspending the pellet in buffer and sonicating the suspension again. Since the intensity of the gel bands appeared to be more at 0.6 mM IPTG for both the pellet and supernatant after 4 and 6 hours (Figure 3.3A and B), these conditions (30 °C, 0.6 mM IPTG, 6 hours) were used for the expression of EcSspA



B

MULTISPECIES: stringent starvation protein A [Proteobacteria]

Sequence ID: WP_000257293.1 Length: 212 Number of Matches: 1

See 5009 more title(s)

Range 1: 1 to 212 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
437 bits(1123)	7e-154	Compositional matrix adjust.	212/212(100%)	212/212(100%)	0/212(0%)	+3

```
Query  108  MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSFEIEHVEKDNPPQDLIDLNPNQSVP  287
            MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSFEIEHVEKDNPPQDLIDLNPNQSVP
Sbjct  1    MAVAANKRSVMTLFSGPTDIYSHQVRIVLAEKGVSFEIEHVEKDNPPQDLIDLNPNQSVP  60

Query  288  TLVDRELTLWESRIIMEYLDERFPHPPLMPVYPVARGESRLYMHRIEKDWYTLMNTIING  467
            TLVDRELTLWESRIIMEYLDERFPHPPLMPVYPVARGESRLYMHRIEKDWYTLMNTIING
Sbjct  61   TLVDRELTLWESRIIMEYLDERFPHPPLMPVYPVARGESRLYMHRIEKDWYTLMNTIING  120

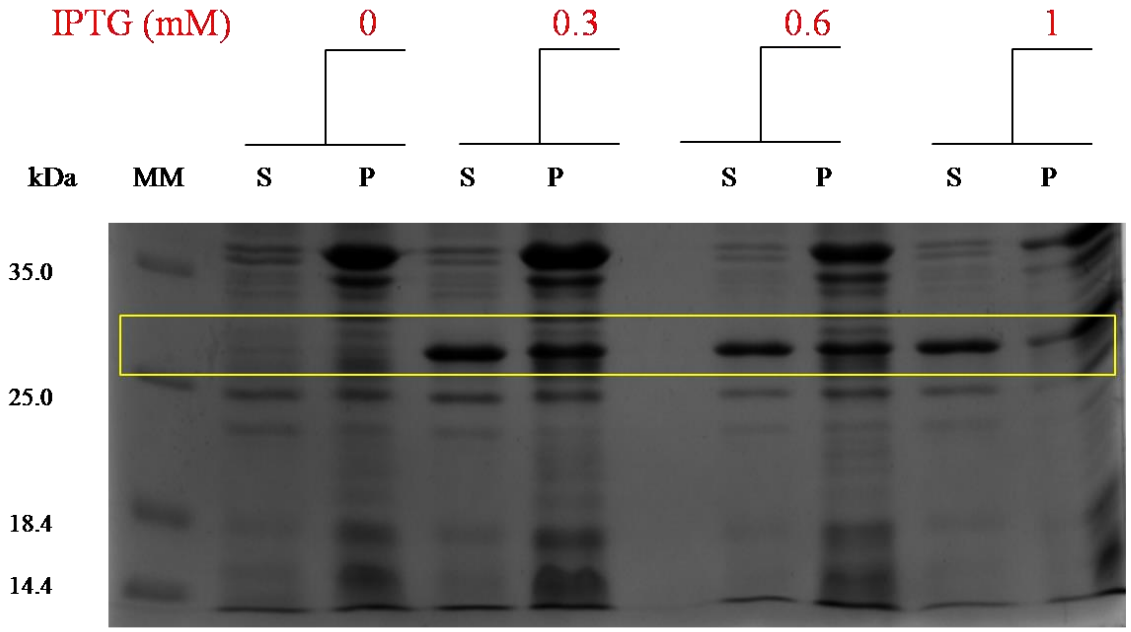
Query  468  SASEADAARKQLREELLAIAPVFGQKPYFLSDEFSLVDCYLAPLLWRLPQLGIEFSGPGA  647
            SASEADAARKQLREELLAIAPVFGQKPYFLSDEFSLVDCYLAPLLWRLPQLGIEFSGPGA
Sbjct  121  SASEADAARKQLREELLAIAPVFGQKPYFLSDEFSLVDCYLAPLLWRLPQLGIEFSGPGA  180

Query  648  KELKGYMTRVFERDSFLASLTEAEREMRLGRS  743
            KELKGYMTRVFERDSFLASLTEAEREMRLGRS
Sbjct  181  KELKGYMTRVFERDSFLASLTEAEREMRLGRS  212
```

Figure 3.1: Confirmation of EcSspA cDNA (A) A segment of the nucleotide sequencing chromatogram for SspA cDNA ORF (Open Reading Frame). Each nucleotide has a corresponding peak and it is colour coded. Black- guanine (G), green- adenine (A), blue- cytosine (C) and Red- thymine (T). The sequencing results were viewed using Flinch TV program version 1.4.0 (www.geospiza.com/FlinchTV: geospiza Inc.).

(B) BLASTx alignment of the query sequence from Inqaba Biotech against known sequences. The result indicates that the sequence corresponds to features of EcSspA. There is 100 % sequence identity on the 212 amino acids identified in the protein. A 0 % gap was identified between the subject and query sequence and the gaps are introduced where the alignment is offset.

A. 4 hours



B. 6 hours

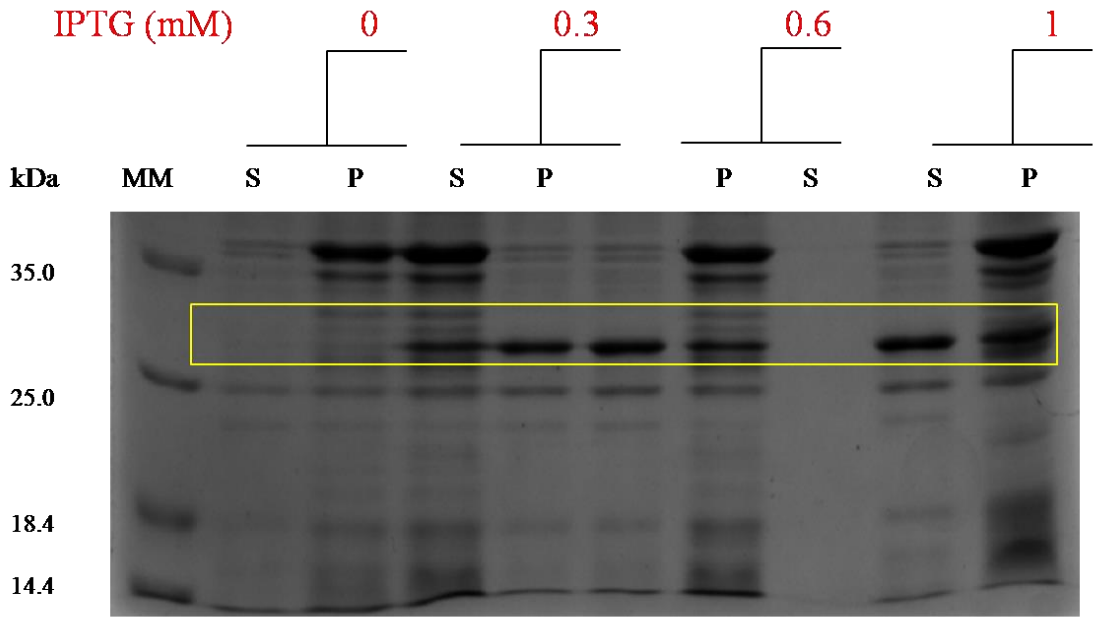
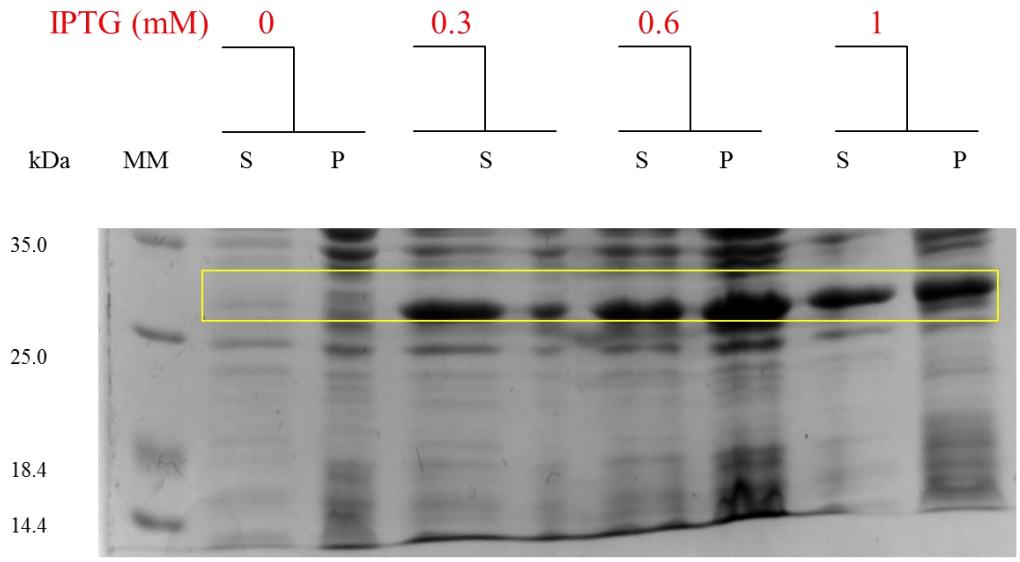


Figure 3.2: Expression trials of EcSspA in *Escherichia coli* BL21 (DE3) pLysS cells in 2x YT medium at 37 °C. Expression of SspA was induced at O. D_{600} = 0.6 with various concentrations of IPTG for 4 hours **(A)** and 6 hours **(B)**. The protein was expressed in both the soluble supernatant (S) and the insoluble pellet (P). The protein band which corresponds to the estimated size of SspA is boxed in yellow.

A. 4 hours



B. 6 hours

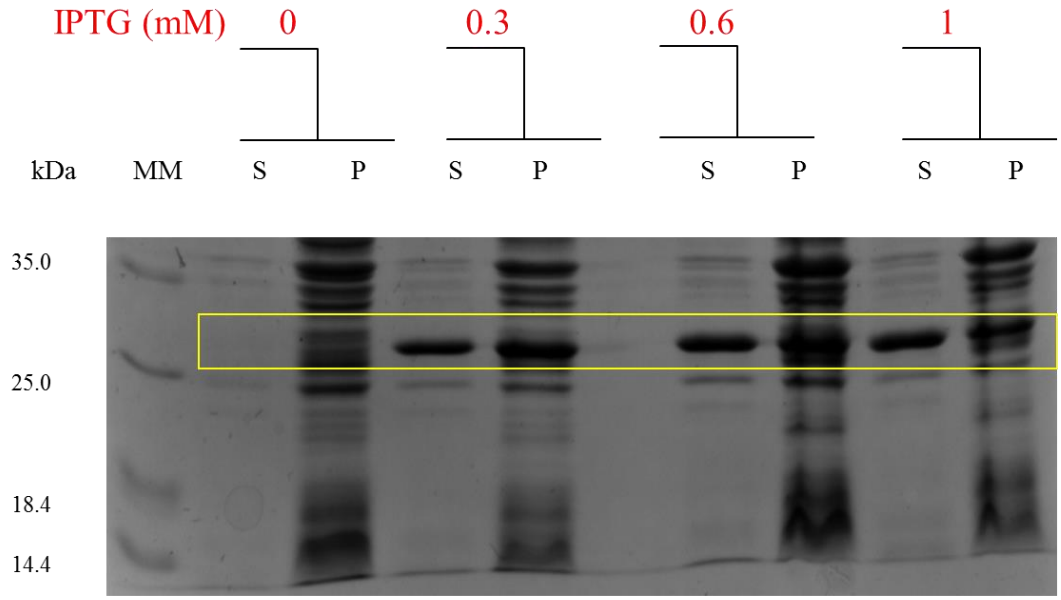


Figure 3.3: Expression trials of SspA in *Escherichia coli* BL21 (DE3) pLysS cells in 2x YT media at 30 °C. Expression of SspA was induced at OD_{600} = 0.6 with various concentrations of IPTG for 4 hours **(A)** and 6 hours **(B)**. The protein was expressed in both the soluble supernatant (S) and the insoluble pellet (P). The protein band which corresponds to the estimated size of SspA is boxed in yellow.

3.3. Protein overexpression and purification

The purification of recombinant His-tagged EcSspA was achieved using a Ni^{2+} charged IMAC column. One litre of cultured *Escherichia coli* BL21 (DE3) pLysS cells were used for the expression of EcSspA for 6 hours at 30 °C. The successful expression of the protein is evidenced by a thick protein band in lane L, as shown by SDS-PAGE (insert in Figure 3.4). The cell lysate contains other proteins which are released during cell lysis as evidenced by the presence of several protein bands in the lysate lane (Figure 3.4). The lysed cells were then subjected to centrifugation to separate the soluble (supernatant) fraction from the insoluble fraction. The protein was predominantly evidenced in the soluble fraction as shown by a prominent band on Lane S on the SDS-PAGE (Figure 3.4). The successful lysis of the cells can be observed from the first peak which shows the loading of the supernatant onto the IMAC column. Lane FT corresponds to the flow through which basically refers to proteins that were not able to bind to the column. A faint lane of protein corresponding to an estimated EcSspA size can be observed in the FT lane. This suggests that some of the His-tagged protein was unable to bind to the column, therefore suggesting that the column binding capacity might have been exceeded. The column was washed with the wash buffer to remove unbound and non-specifically interacting protein from the resin. The wash step can be evidenced by a sudden drop in absorbance as the contaminating proteins are eluted (Figure 3.4) followed by the flattening of the UV absorbance line. His-tagged EcSspA protein was then eluted from the column using an imidazole concentration gradient, which is shown to be in pure form by SDS-PAGE (lane SspA, Figure 3.4). The protein eluted at an imidazole concentration of 120 mM.

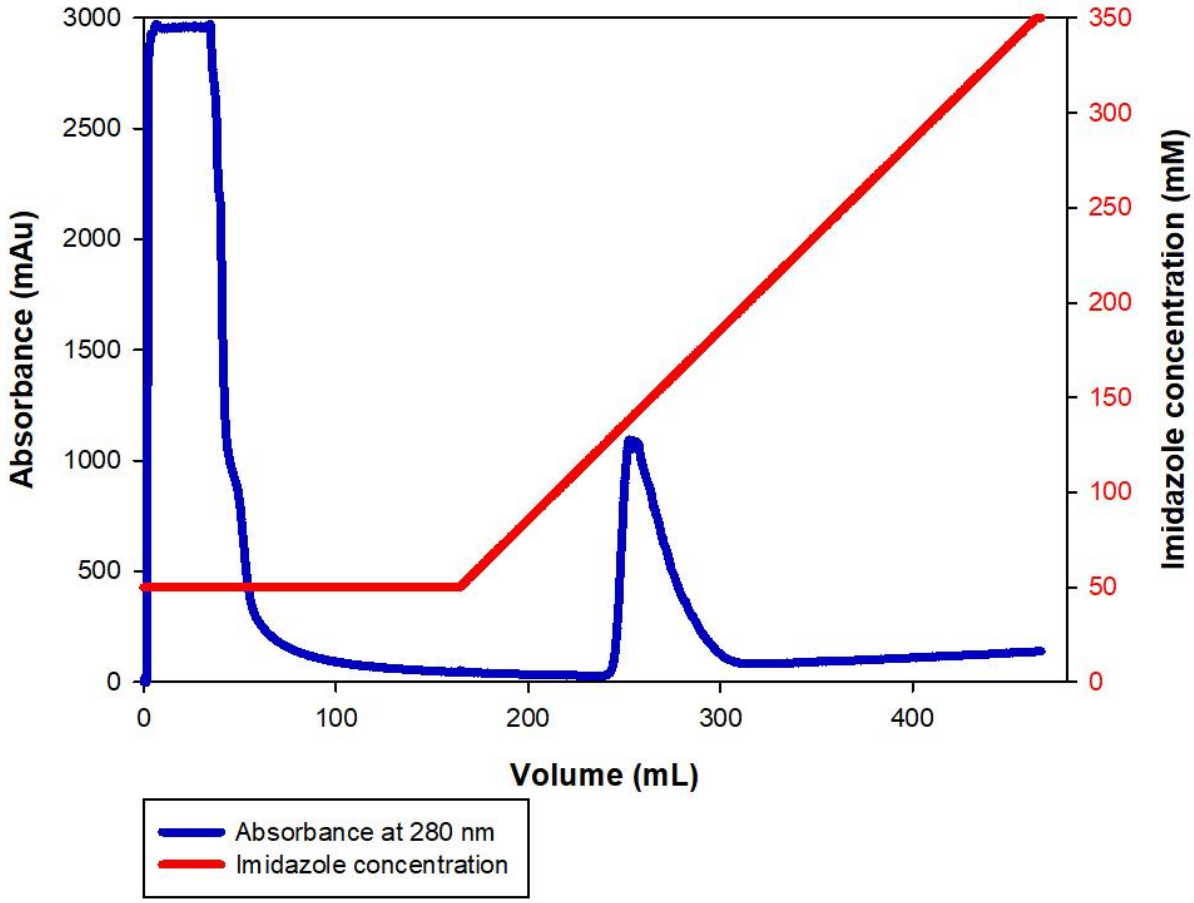


Figure 3.4: IMAC Elution profile for EcSspA purification. The elution profile of EcSspA using IMAC on the AKTA purification system. The blue profile monitors protein absorbance from supernatant loading to protein elution. The red profile monitors imidazole concentration. The protein started eluting from the column at 120 mM imidazole concentration. The insert SDS-PAGE shows the SspA purification stages. MM- molecular mass marker; L- cell lysate; S- supernatant; FT- flow through and the pure SspA protein.

3.4. Protein size estimation

Following protein purification, the size of the purified protein was estimated using SDS-PAGE. Figure 3.5A shows the band of the purified EcSspA which was run alongside the molecular mass standards. The theoretical molecular mass of EcSspA is 24 kDa based on the protein identification tool on ExPASy (Gasteiger *et al*., 2005). Based on the calibration curve in Figure 3.5B, the His-tagged EcSspA corresponds to a size of 27 kDa. The migration of the EcSspA during SDS-PAGE is slightly impeded and its larger size is due to a short sequence of 20 amino acids at the N-terminal of the protein, comprising the His-tag and the thrombin cleavage site sequence.

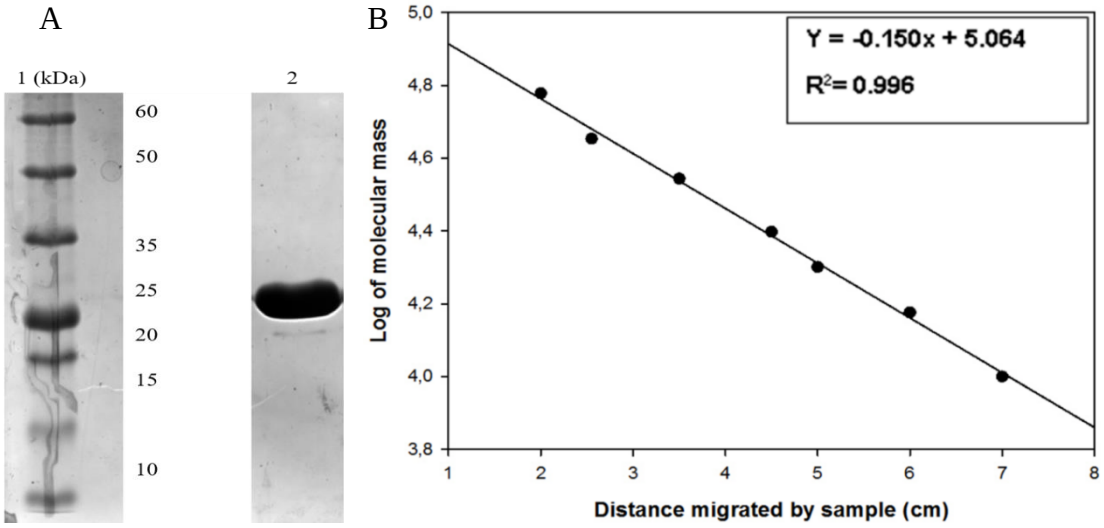


Figure 3.5: SDS-PAGE analysis of purified EcSspA (A) SDS-PAGE molecular mass calibration curve. The estimated subunit molecular mass for EcSspA is 27 kDa. The correlation coefficient is 0.99. **(B)** Lane 1 represents the molecular mass markers. Lane 2 represents the purified EcSspA protein.

3.5. Protein concentration determination and purity assessment

The concentration of the purified protein was quantified using the serial dilution curve as shown in Figure 3.6A. The concentration was determined using the Beer-Lambert law as outlined in the methods section and the protein yield was 84 μM of protein from a litre of cultured cells. After protein quantification, an additional protein purity assessment was performed by determining an absorbance spectrum as shown in Figure 3.6B. The A_{280}/A_{250} ratio was always close to 2 therefore confirming the purity of the protein (Sambrook *et al*., 1989).

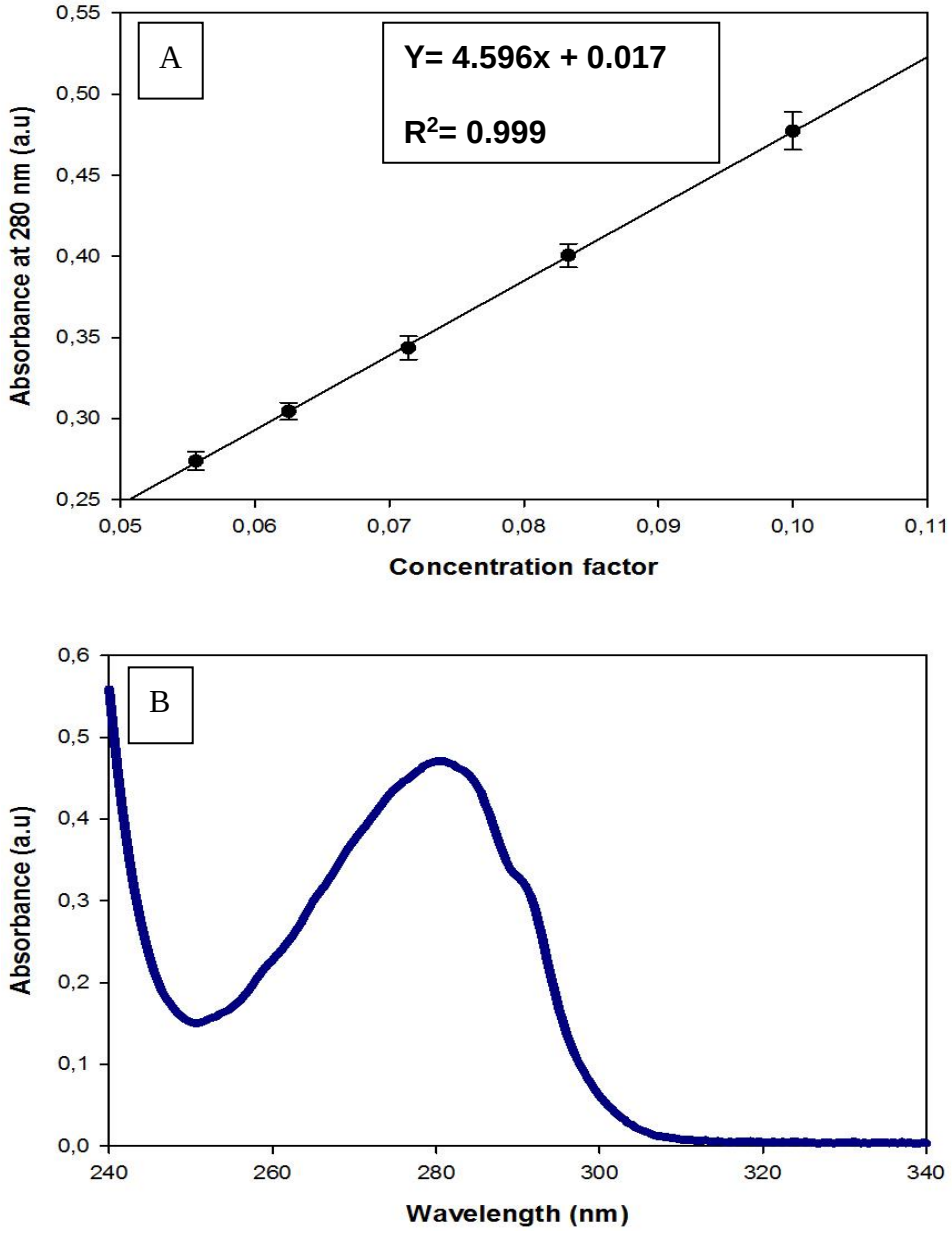


Figure 3.6: The absorbance spectrum of purified EcSspA and protein concentration determination. (A) The absorbance spectrum of EcSspA measured at wavelengths between 240 and 340 nm. The spectrum was corrected for the buffer blank. The A_{280}/A_{250} ratio is equivalent 2 therefore confirming the purity of the protein. The protein solution was diluted 10 times into the storage buffer. **(B)** A standard curve showing the serial dilution of EcSspA at 280 nm. The serial dilution was performed in triplicates and averaged to obtain the curve. Protein concentration was determined using Beer-Lambert law and the concentration of the protein was found to be 84 µM.

3.6. Confirmation of protein identity

Following protein isolation and purification, the SDS-PAGE gel with the purified sample of EcSspA was sent to the Council for Scientific and Industrial Research (Pretoria, South Africa) where the protein sample was subjected to LC-MS/MS in order to characterise and confirm its identity. The tandem MS/MS spectra provided peptide fragmentation information that was analysed using ProteinPilot Software and the data were matched to *in silico* spectra from a protein database. Figure 3.7 represents the mass spectrometer data analysis results that were obtained for the protein in the SDS-PAGE gel. Figure 3.7 shows the spectrum of one of the 268 peptides that were generated during in-gel protein digestion and were further separated using liquid chromatography. Figure 3.8 is a table illustrating some of the data for a selected peptide that were generated during fragmentation that was identified with a high level of confidence.

Figure 3.9 is an illustration of some of the data that were extracted from the ProteinPilot Software that is used for matching up the various peptide sequence data generated to a database.This will ultimately allow for a sequence coverage analysis and protein characterisation. Based on the results from the software seven proteins were identified from the extracted gel band. For each identified protein there is a percentage of the sequence covered and the number of identified peptides. However, the protein of interest was present in high abundance as evidenced by a large number of peptides that were generated. This is compared to other proteins which had fewer peptides suggesting the low abundance of the protein. The presence of this low abundance proteins could be attributed to a minimal protein contamination during the sample preparation stage for mass spectrometry and not the purification of the protein.The porcine trypsin protein sequence identified could be attributed to a trypsin sequence contaminant since trypsin was used to digest the protein into peptides by hydrolysing peptide bonds at the carboxy terminal of all lysine and arginine residues (Olsen*et al*., 2004). The different types of keratin that are listed on Table 3.1 can be due to human hair contamination during the sample preparation stage. The protein sequence coverage as shown in Table 3.1 was identified positively with a confidence equal or greater than 95 % and does correspond to the protein of interest based on the information from the protein database.

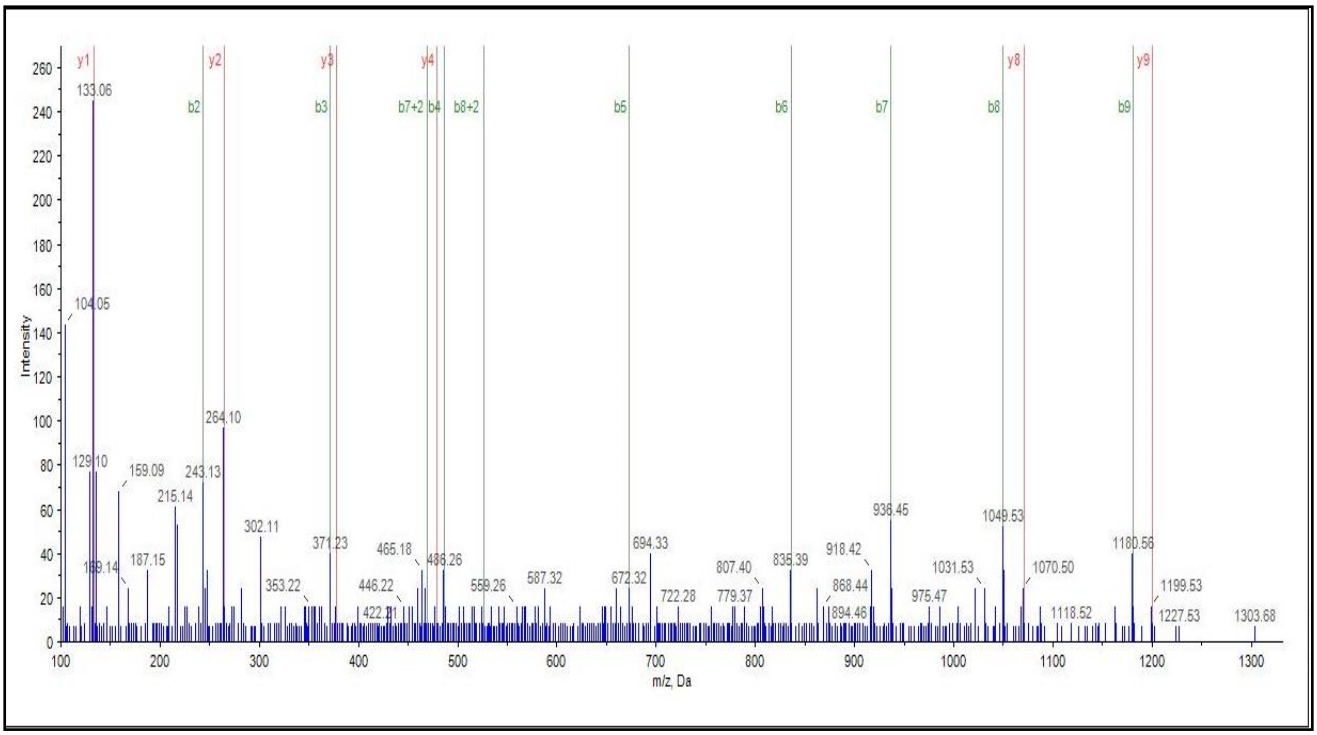


Figure 3.7: LC-MS/MS spectrum for a selected pepetide match (IEKDWYTLMN) derived from in-gel trypsin digest represented by a series of b and y ions. Following in gel protein digestion, peptides are separated by liquid chromatograhy then fragmented by Collision Induced Dissociation (CID) resulting in the above MS/MS spectrum that consists of a series of b and y ions. Fragment peaks that are generated from the amino terminus are labelled as b ions whereas peaks from the carboxyl terminus are labelled as y ions. Peptides are sequenced from both the amino and carboxyl terminus. The spectrum is analysed by ProteinPilot software against a protein sequence database for protein identification. m/z on the x axis represents the mass to charge ratio and Da is Daltons.

Fragmentation Evidence for Peptide

IEKDWYTLMN

Residue	b	b+2	y	y+2
I	114.0913	57.5493	1312.6242	656.8157
E	**243.1339**	122.0706	**1199.5401**	600.2737
K	**371.2289**	186.1181	**1070.4975**	535.7524
D	**486.2558**	243.6316	942.4026	471.7049
W	**672.3352**	336.6712	827.3756	414.1915
Y	**835.3985**	418.2029	641.2963	321.1518
T	**936.4462**	**468.7267**	**478.2330**	239.6201
L	**1049.5302**	**525.2688**	**377.1853**	189.0963
M	**1180.5707**	590.7890	**264.1013**	132.5543
N	1294.6136	647.8105	**133.0608**	67.0340

Figure 3.8: A representation of b and y ions that correspond to identified fragment masses corresponding to the amino acid sequence of a single peptide match to the EcSspA protein band. Ions that were positively identified with an equal or greater than 95 % confidence are indicated in the green blocks.

Proteins detected			
N	% of sequence coverage (*95 %)	Protein name	Number of identified peptides (*95 %)
1	92.9	Stringent starvation protein A OS-*Escherichia coli* (strain K12)	268
2	28.9	Keratin 1	22
3	30.0	Maltodextrin phosphorylase	16
4	19.2	Keratin 10	14
5	26.0	Porcine beta trypsin	11
6	5.3	Keratin type ii cytoskeletal 6E	4
7	10.3	Bovine Alpha S1-casein precursor	2

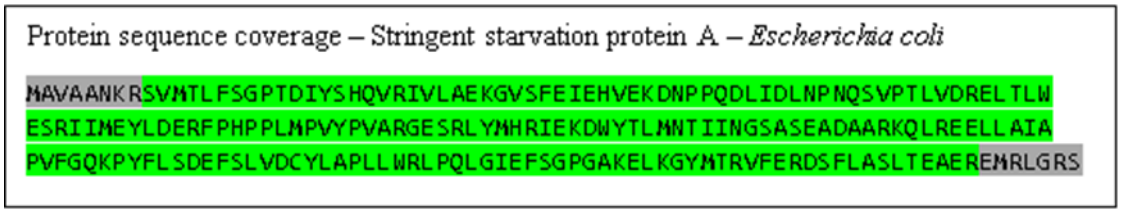


Table 3.1: Data extracted from ProteinPilot Software analysis indicating protein characterisation and sequence coverage for SDS-PAGE protein band. A gel band was excised out, then an in-gel trypsinisation was performed and then analysed by LC-MS/MS. The protein from the SDS-PAGE gel band was successfully identified by matching peptide fragment spectra from 268 peptides to *in silico* spectra from a protein database and the above information was obtained from ProteinPilotSoftware (https://sciex.com/products/software/proteinpilot-software). The percentage of sequence coverage, the name of detected proteins and the corresponding number of peptides are generated by the software. All proteins that were identified by LC-MS/MS are ranked according to the number of unique peptides that were detected for each protein. Listed proteins 2-7 were also positively identified in the sample but with fewer peptides since the proteins are in low abundance. The protein sequence coverage panel below represents the protein sequence of the identified protein from the gel band. Sequence coverage that is highlighted in green indicates a high confidence and grey highlighted sequence shows the area where no sequence coverage for the protein was obtained.* The percentage sequence coverage and the number of identified peptides with a greater than or equal to 95 % confidence was obtained from the ProteinPilot software

3.7. Structural characterisation of EcSspA

3.7.1. Structure modelling and analysis

A structural homology model of EcSspA (Figure 3.9) was built using SWISS model (Arnold *et al*., 2006). The quality of the model was estimated by using the GMQE parameter which is the global model quality estimation which measures the quality of the model. The calculated GMQE and QMEAN Values were found to be 0.94 and -0.16 respectively (Biasini *et al*., 2014) showing a relatively reliable estimation of the structure of the protein based on the template. The generated homology model was used for a prediction of the secondary structure of the protein, the structure was analysed, and an image was generated using PyMOL (DeLano, 2002).

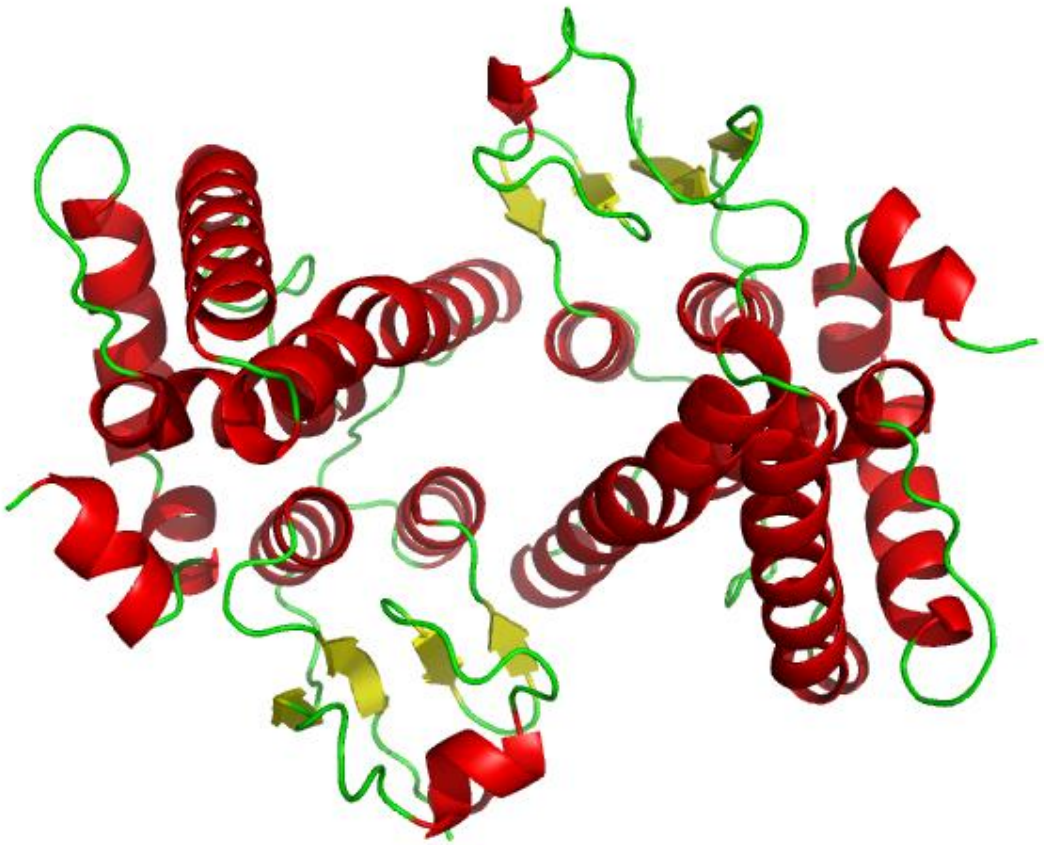


Figure 3.9: SWISS Model homology structure prediction of EcSspA. The model prediction of EcSspA was performed using PDB ID: 1YY7 structure as a template. The secondary structural elements: β-sheets (yellow), α-helices (red) and random coils (green) are shown. The image was generated using PyMOL (DeLano, 2002).

3.7.2. Far-UV circular dichroism:secondary structural characterisation

The secondary structure of EcSspA was studied and characterised using far-UV circular dichroism spectroscopy. Figure 3.10 shows the circular dichroism spectrum of EcSspA. The CD spectrum is characterised by 2 mimima troughs at 208 nm and 222 nm. The spectrum is also indicative of a maximum peak at 195 nm. The CONTILL algorithm implemented in the Dichroweb server was employed for the quantitative analysis of the circular dichroism spectroscopic data (Whitmore and Wallace, 2008). Table 3.2 shows that SspA is composed of around 96.5 % α-helical content, 0.9 % of β-strand, 2 % β-turn content and less than 1 % of the protein forms random uncoiled structures. The secondary structural content of YpSspA, which has an 87.73 % sequence identity to EcSspA was also calculated based on the crystal structure and the data is included on Table 3.1. Included on Table 3.2 also is the calculated secondary structure for the predicted SWISS Model structure which shows a 61.2 % and 7.8 % in alpha helices and beta strands, respectively.

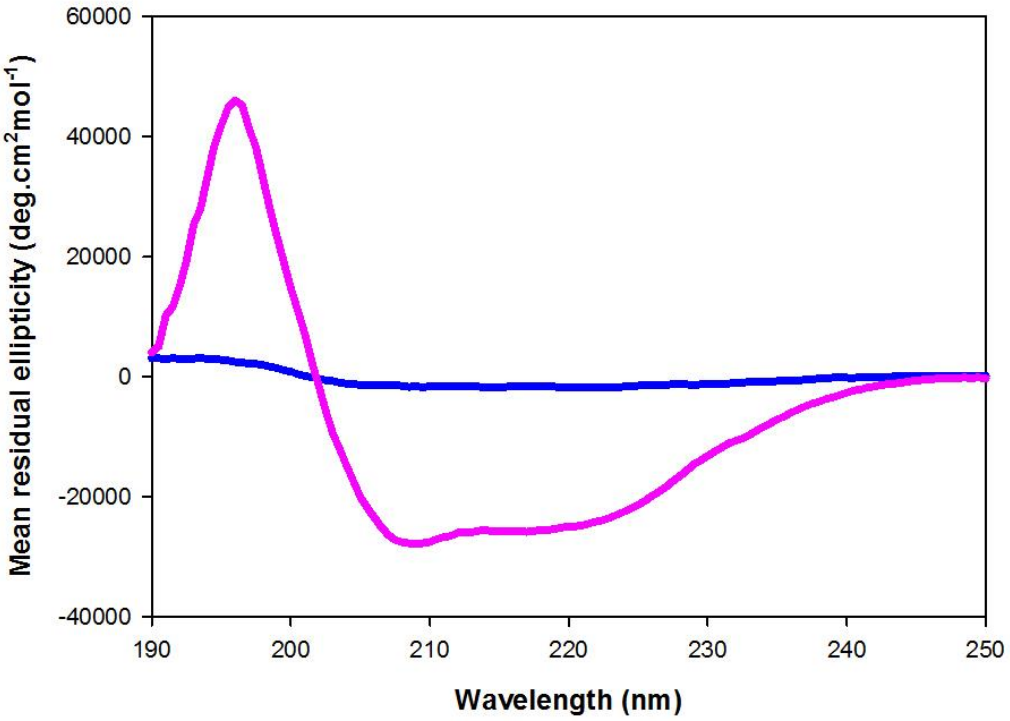


Figure 3.10: Far-UV circular dichroism spectrum of EcSspA. Circular dichroism spectrum of 2 μM SspA (pink) in milliQ water (blue). The spectrum shows a minimum at 208 nm and 222 nm which is characteristic of a predominantly α-helical protein. The spectrum was recorded at 20 °C.

Table 3.2: **The secondary structure content of SspA proteins**. Row 1: EcSspA content calculated using the CONTILL algorithm implemented in the Dichroweb server of optimised from 180-240 nm (Whitmore and Wallace, 2008). The NRMSD value for the calculated values is 0.059, and is used to indicate the goodness of fit of the calculated data to experimental data, with a value <0.1 acceptable as good fit. The data calculated based on the crystal structure of YpSspA and the generated EcSspA model are shown in rows 2 and 3, respectively.

Protein	α-helix (%)	β-strand (%)	β-turn (%)	Unordered (%)	NRMSD
EcSspA (212[a])	96.5	0.9	2.0	0.6	0.059
***YpSspA (206[a])**	69.9	8.7			
EcSspA homology model (206[a])	61.2	7.8			

[a] Number of amino acid residues corresponding to each construct

3.7.3. Intrinsic fluorescence: tertiary structural characterisation

The teriary structure of EcSspA was investigated using intrinsic fluorescence spectroscopy. This technique provides information about the local environment of the tryptophan residues in the protein (Lackowicz, 2006). EcSspA has three tryptophan residues on each subunit. Excitation of EcSspA at 280 nm results in the excitation of eight tyrosine residues in addition to the three tryptophan residues, while excitation at 295 nm results in the selective excitation of the tryptophan residues only. Therefore all the intrinsic fluorescence studies were conducted by a selective excitation of the three tryptophan residues at 295 nm. Shown in Figure 3.11 are the emission fluorescence spectra of native and unfolded species of EcSspA generated by a selective excitation of tryptophan residues at 295 nm. The fluorescence emission maxima wavelength for the native and unfolded EcSspA are 341 and 350 nm, respectively (Figure 3.11). When EcSspA is denatured in 8 M urea, the spectrum shows a red shift to 350 nm which is accompanied by an increase in quantum yield.

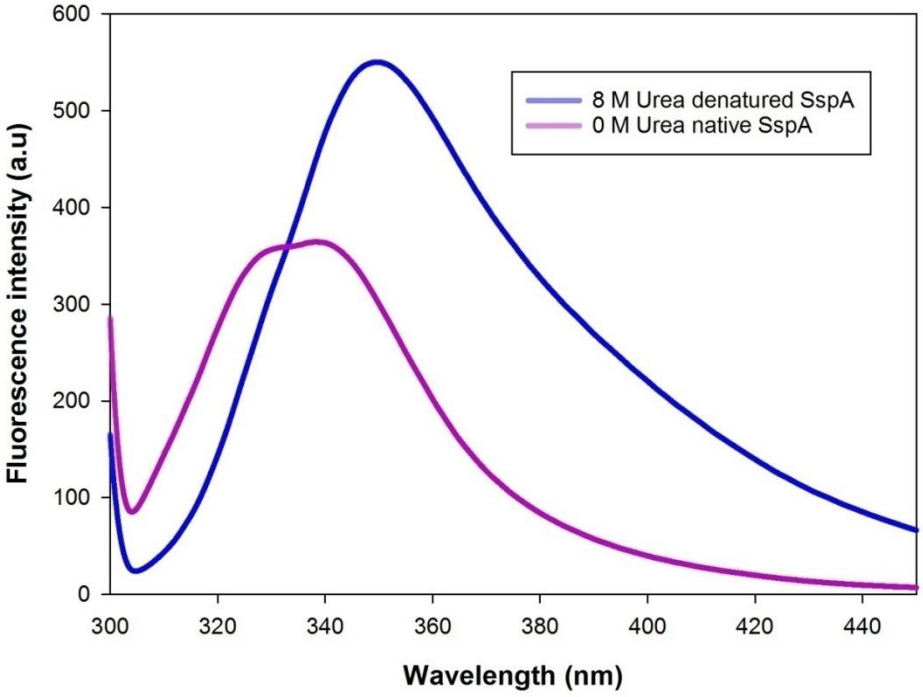


Figure 3.11: Intrinsic fluorescence emission spectrum of EcSspA. Fluorescence spectra of 0.5 µM EcSspA in 5 mM potassium phosphate at pH 7.4 in the native form (pink) and urea induced unfolding of EcSspA (blue) recorded at 20 °C. The spectra were corrected for a buffer blank. The natively folded EcSspA emits energy at a wavelength of 341 nm after excitation at 295 nm. The denatured EcSspA has an increased quantum yield and shows maximum emission wavelength of 350 nm.

3.7.4.SE-HPLC – quaternary structural characterisation

The oligomeric state of EcSspA was detemined using size exclusion high Pressure liquid chromatography (SE-HPLC). The human glutathione transferase class pi (hGSTP1-1) protein of a known oligomeric state was used as positive control for size estimation of EcSspA. The SE-HPLC chromatograms were used to determine the molecular mass of EcSspA and hGSTP1-1 (Figure 3.12). Based on the two chromatograms there is no huge difference in the elution times of the proteins which were found to be 17.46 and 17.91 minutes for hGSTP1-1 and EcSspA respectively. The resultant molecular mass for EcSspA and hGSTP1-1 were determined to be 24 kDa and 30.2 kDa, respectively. The molecular mass of the eluted proteins was determined based on the calibration curve (Figure 3.12B) constructed based on the retention times of molecular mass standards, fitted by linear regression.

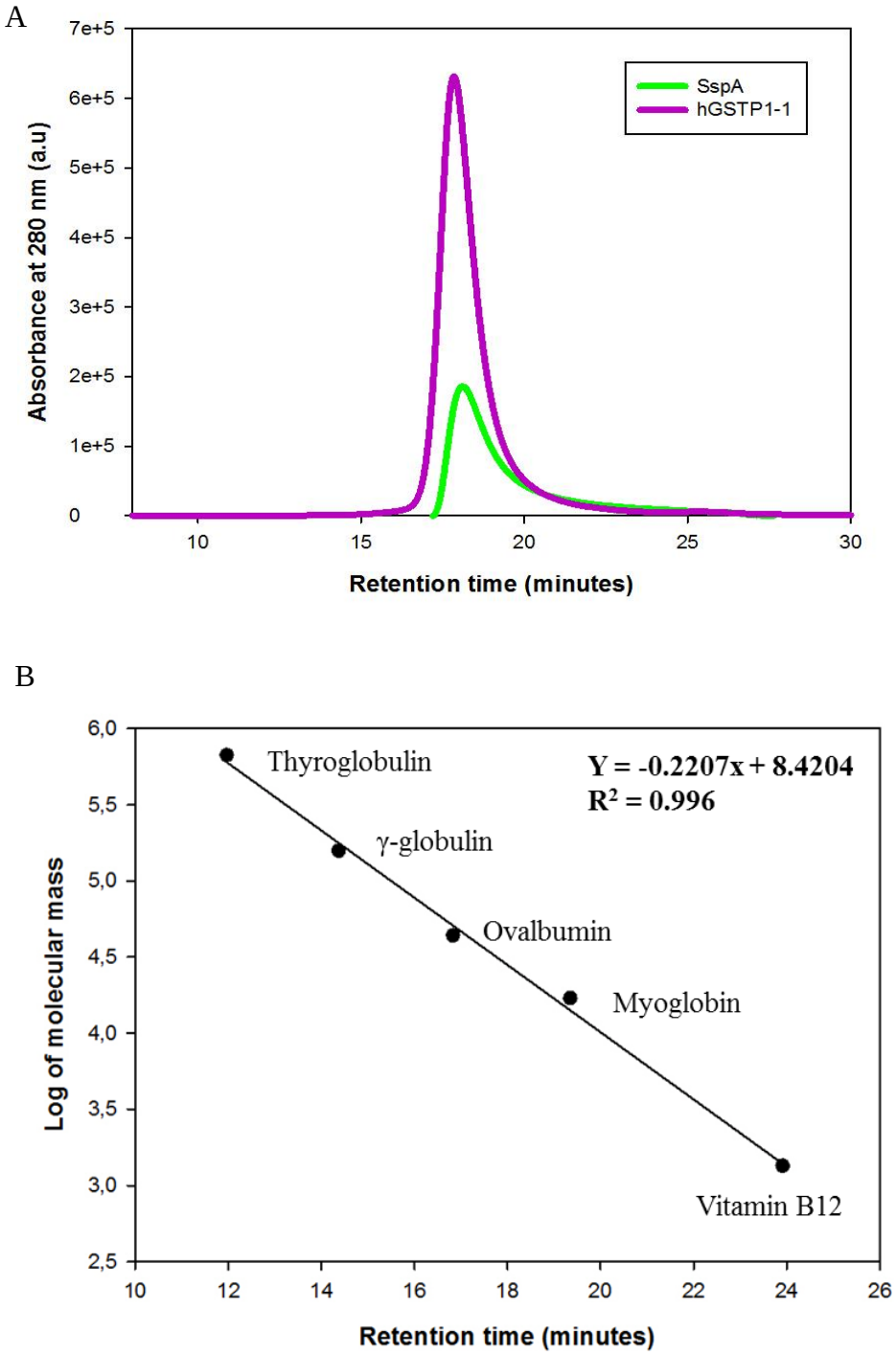


Figure 3.12: SE-HPLC elution profiles of EcSspA and hGSTP1-1. (A) The elution profiles of the proteins were recorded using absorbance at 280 nm on a Shimadzu HPLC system. The proteins were eluted in a 20 mM sodium phosphate; 500 mM NaCl at pH 7. The retention times for SspA and hGSTP1-1 were 17.46 and 17.91 minutes, respectively at a flow rate of 0.25 ml/min. **(B)** Calibration curve of gel filtration standards: thyroglobulin (670 kDa), γ-globulin (154 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa). The correlation coefficient is 0.996.

3.8. Thermal unfolding properties of EcSspA

The thermal unfolding of EcSspA was monitored using circular dichroism mean residual ellipticity at 222 nm ($MRE_{222\ nm}$) of the protein as a probe (Figure 3.13). The alteration in the structure of SspA induced by a high temperature is irreversible. The melting point (T_m) for the unfolding of the secondary structure of SspA is 60°C. An exposure to temperatures above the T_m causes SspA to precipitate. Insoluble precipitates were observed in the cuvette after the protein was exposed to temperatures above the T_m. Therefore, thermal unfolding of EcSspA is irreversible upon cooling, hence no thermodynamic parameters can be derived.

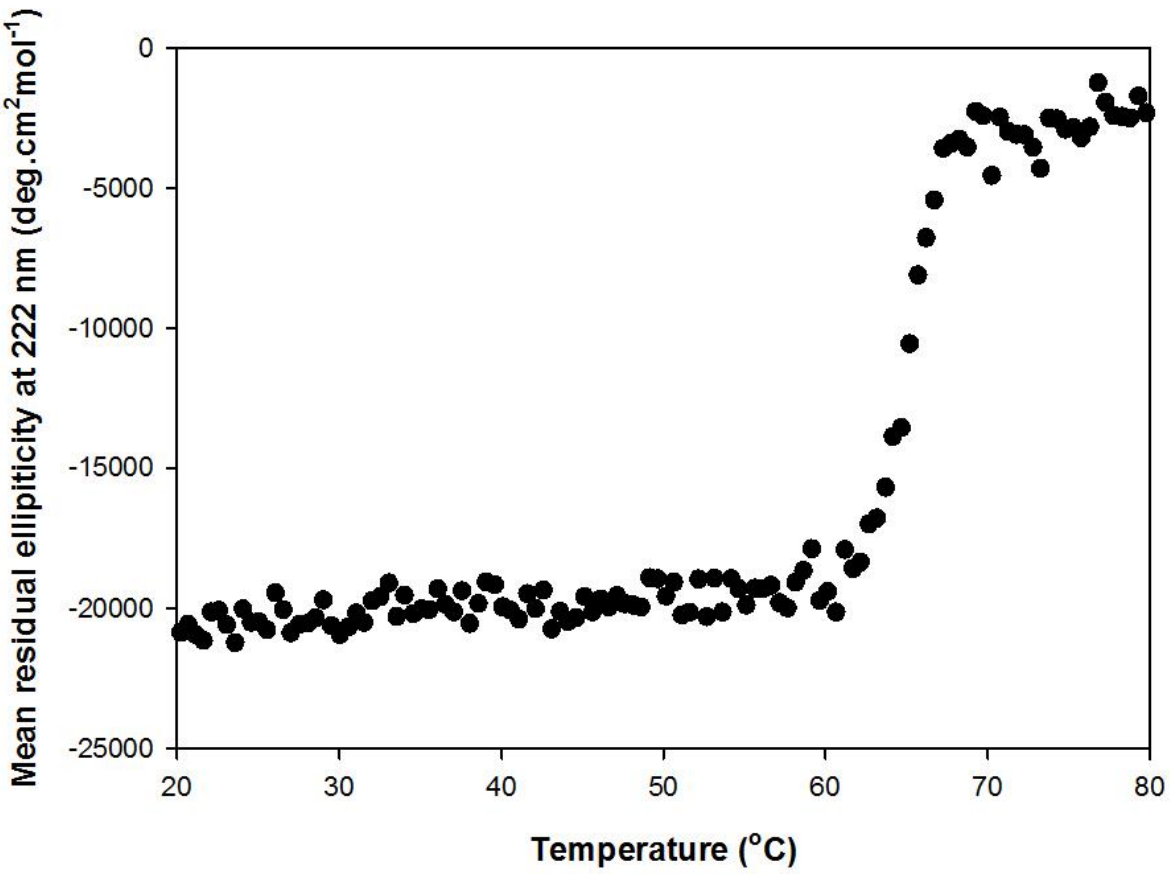


Figure 3.13: Melting curve of EcSspA using circular dichroism as a probe. Thermal unfolding of EcSspA at pH 8 was monitored using circular dichroism mean residual ellipticity at 222 nm (MRE_{222nm}) as probe. The T_m for the unfolding of the secondary structure of the protein at pH 8 is 60°C. Insoluble precipitates were observed after the protein was exposed to temperatures above 60°C.

3.9. Measurement of glutathione *S*-transferase activity and GSH binding test

In order to determine whether EcSspA has GST activity, the ability of EcSspA to catalyse the GSH/CDNB reaction resulting in the formation of a product chromophore which can be monitored at 340 nm was performed as proposed by Habig *et al*. (1974). A non-enzymatic control for the spontaneous reaction was performed to correct from the expected enzymatic reactions. Progress curves of the GSH/CDNB catalysis assay for all the controls and experimental reactions are shown in Figure 3.14. The enzyme activity can be derived from the slope of the progress curves measured in absorbance units at 340 nm per minute. The presented curves shown were not corrected for the non-enzymatic control. hGSTP1-1 was used as a positive control to show an expected GST-catalysed reaction. This is evidenced by a very steep positive sloped curve show in Figure 3.14. EcSspA showed no enzymatic response at the two different concentrations of the protein.

EcSspA was subsequently tested for the ability to bind GSH, by performing a glutathione binding test using glutathione agarose affinity chromatography. The ability of EcSspA to bind GSH was checked by passing it through a GSH-column, followed by analysing the eluted protein by SDS-PAGE (Figure 3.15). hGSTP1-1 was used as a positive control that does bind to the GSH agarose column as expected. This is evidenced by the absence of the protein in the flow through (Lane FT1) and a minimal presence in the wash solution (W1) in Figure 3.15. This suggests that the protein binds to the column and some of the excess protein came off during the column wash step. SspA does not bind GSH. This is evidenced by a large presence of the protein in the flow through and wash solution (Lane FT2 and W2). There was no protein in the effluent solution which was protein elution using a high GSH concentration elution buffer.

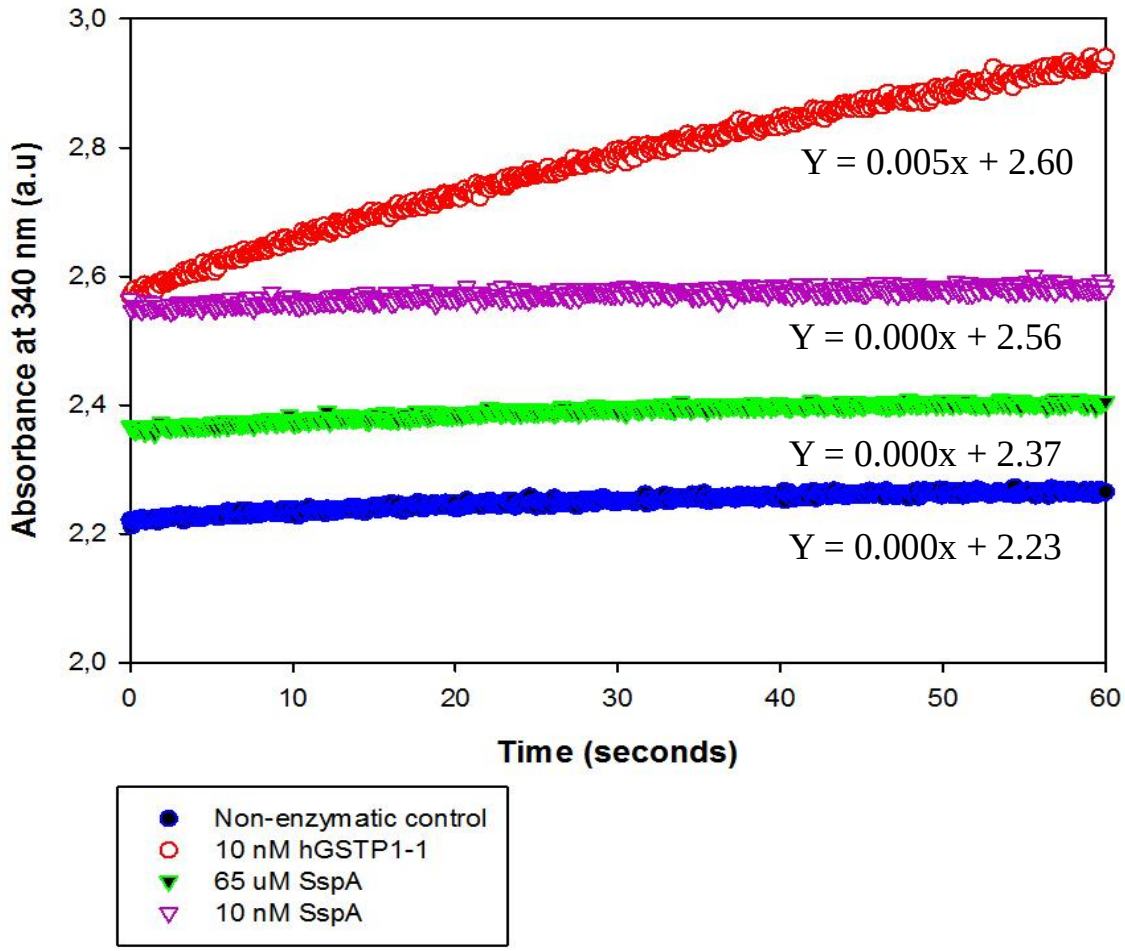


Figure 3.14: Progress curves of the CDNB/GSH catalysis. The progress curves show the enzymatic reaction carried out using hGSTP1-1 and EcSspA in 50 mM potassium phosphate, 1 mM EDTA and 0.02 % sodium azide at pH 6.5. The curve for the non-enzymatic control is included in the curves. The curves for proteins were not corrected for the non-enzymatic control. The final concentration of hGSTP1-1 used was 10 nM to obtain the linear progress curve. The final concentrations of EcSspA used were 65 µM protein and 10 nM diluted SspA. No enzymatic response was observed in the presence of EcSspA irrespective of the protein concentration.

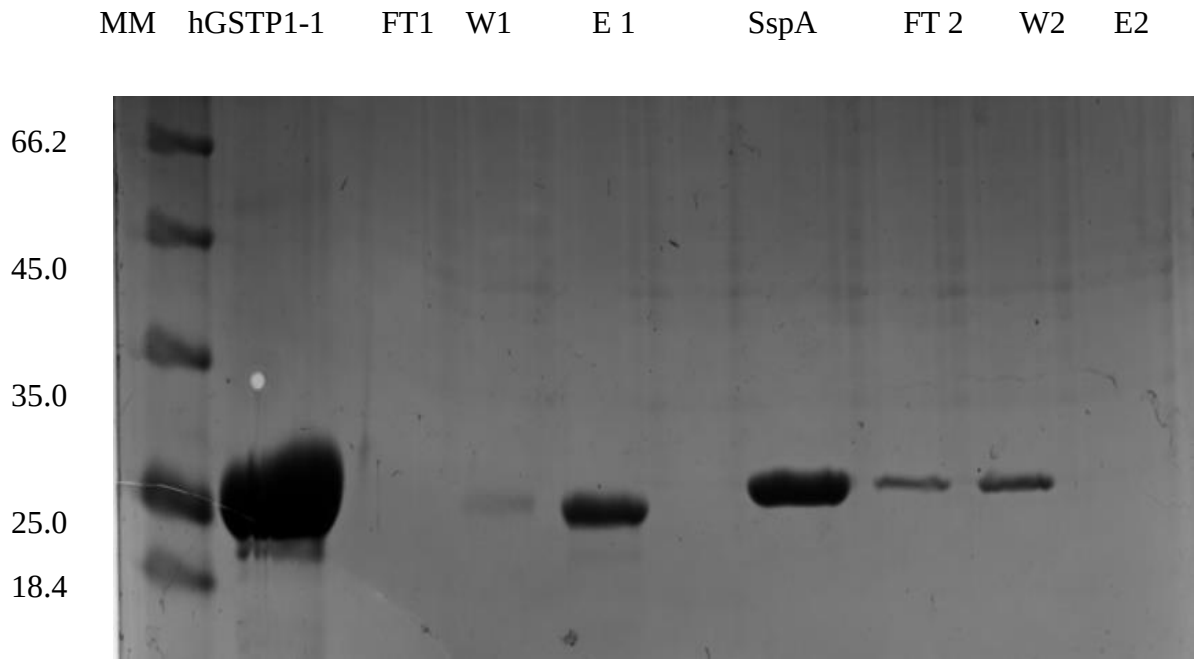


Figure 3.15: An SDS-PAGE showing the binding capability of pure EcSspA and hGSTP1-1 on a GSH-agarose column. FT= flow through W= wash solution E= Efluent. It can be seen from the gel that hGSTP1-1 binds to the column as none of the protein comes out during the flow through and wash step the protein comes out in the eluent when eluted using a 10 mM glutathione elution buffer. With EcSspA no binding of the protein to the resin was observed as most of the protein was observed in the flow through and wash step solution and nothing comes out during elution thus suggesting thatEcSspA does not bind GSH.

3.10. ANS binding studies

ANS is an amphipathic anionic dye which serves as a valuable molecular tool to probe for the presence of hydrophobic pockets in a protein (Abdalla *et al*., 2006). The binding of ANS molecule within a hydrophobic environment results in an increase in the quantum yield of ANS fluorescence upon excitation at 390 nm followed by a shift of the emission maximum of ANS to a lower wavelength when it is bound to a hydrophobic surface (Slavik *et al*., 1982). The emission spectrum of free ANS in the absence of EcSspA peaks at 520 nm (Figure 3.16). ANS showed a significant increase in the quantum yield when excited at 390 nm in the presence of EcSspA and a blue shift in the maximum emission wavelength from 520 nm to 480 nm (Figure 3.16). The results are in a range typical of the binding of ANS ligand to other proteins (see Slavik *et al*., 1982). When the fluorescence intensity of ANS fluorescence signal was measured as a function of ANS concentration a saturation curve was obtained (Figure 3.17), which shows that there is a substantial increase in the fluorescence of ANS upon an increase in ANS concentration until a point of saturation is reached. This ANS binding isotherm curve was used to estimate the apparent equilibrium dissociation constant (K_d) and was estimated to be 50 µM.

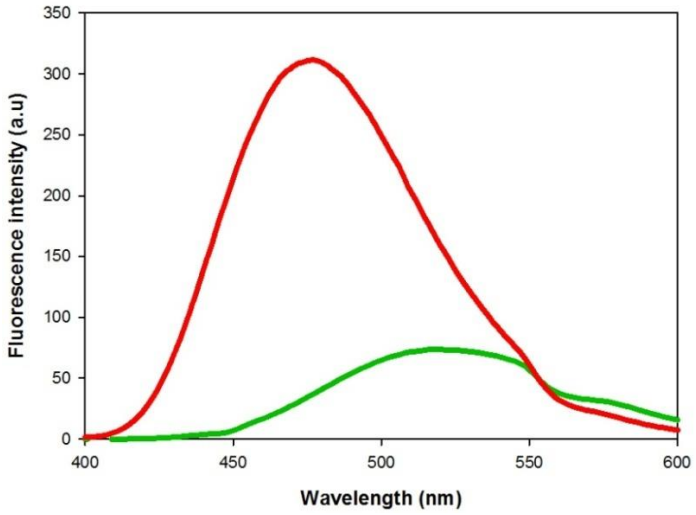


Figure 3.16: Fluorescence emission spectra of ANS bound to EcSspA. Fluorescence emission spectrum of free ANS at pH 6.5 (green) and ANS bound to EcSspA (red). The fluorescence spectrum of ANS bound to EcSspA was corrected for free ANS. The experimental data were obtained by incubating 2 µM SspA in 20 mM potassium phosphate pH 6.5 with ANS at a final concentration of 100 µM. ANS was selectively excited at 390 nm and emission spectrum determined from 400 to 600 nm. The maximum emission wavelength for free ANS and ANS bound to SspA were 520 nm and 480 nm, respectively.

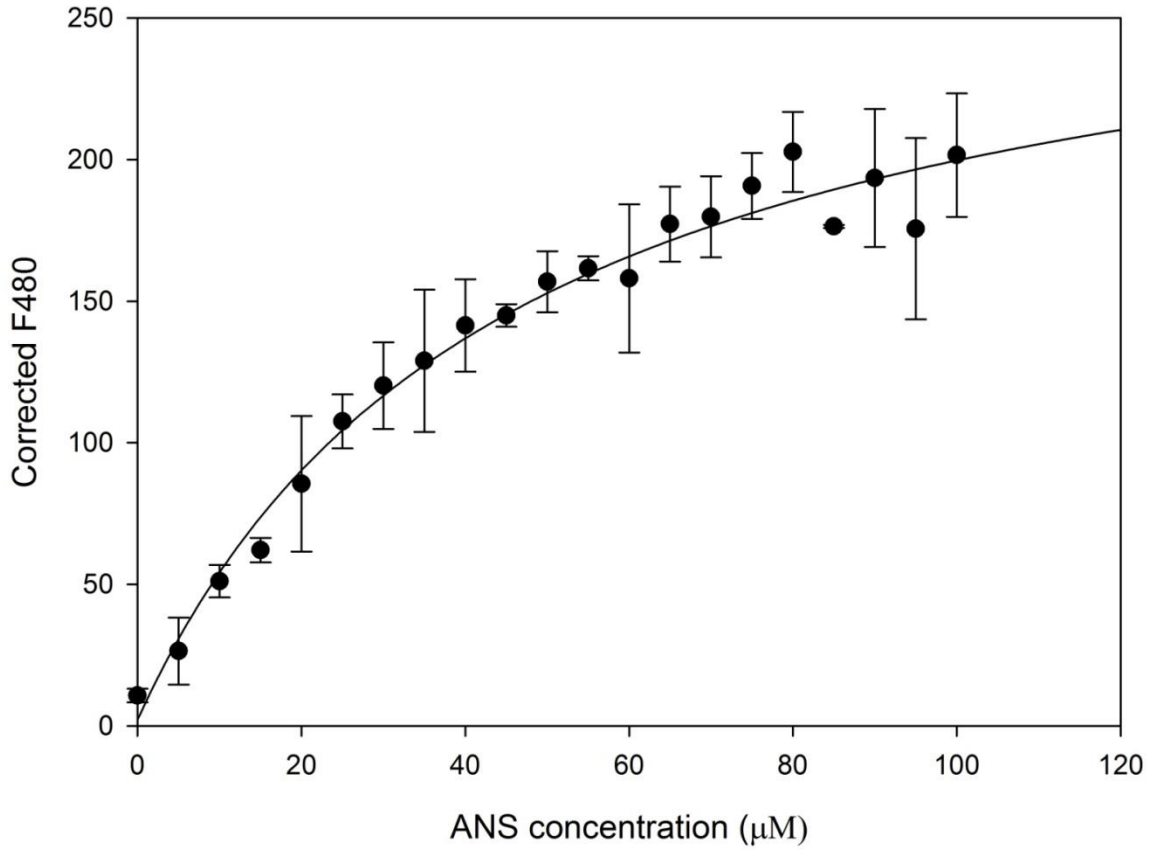


Figure 3.17: An isotherm showing the binding of ANS to EcSspA obtained using fluorescence spectroscopy at pH 6.5 at 20 °C. The extent of binding, as measured by the fluorescence intensity of bound ANS at 480 nm is plotted against increasing concentration of ANS. The dimeric protein concentration was 2 μM and the buffer solution was 20 mM potassium phosphate containing 1 mM EDTA. The solid line gives the best fit to the data for a non-cooperative model and yields a K_d value of 50 μM at an R^2 = 0.96 and was plotted using Sigma Plot software.

3.11. Thermodynamics of ANS binding to EcSspA

The thermodynamics of ANS binding to EcSspA were studied using isothermal titration calorimetry. The thermodynamics of ANS-SspA binding was obtained by measuring the change in enthalpy (ΔH), the dissociation constant (K_d), the binding stoichiometry (N), the change in entropy (ΔS) and the Gibbs free energy (ΔG). Figure 3.18 displays both the corrected heat data (3.18A) and the normalised integral of the data (3.18B). The corrected heat data was corrected for the heat of dilution of the ANS into the buffer and is shown as an insert in Figure 3.18A. The binding isotherm and the corresponding fitted data (fitted using a single independent binding site model) (Figure 3.18) displays a profile of a typical exothermic reaction. The thermodynamic parameters of the binding of ANS to EcSspA are outlined in Table 3.3. The stoichiometry of binding of ANS to EcSspA was observed to be 10 molecules of ANS per dimeric protein. The entropy of binding of ANS was observed to be a positive value of 45.79 J/mol/K. The enthalpy obtained is -5.18 kJ/mol resulting in ΔG = -19.64 kJ/mol.

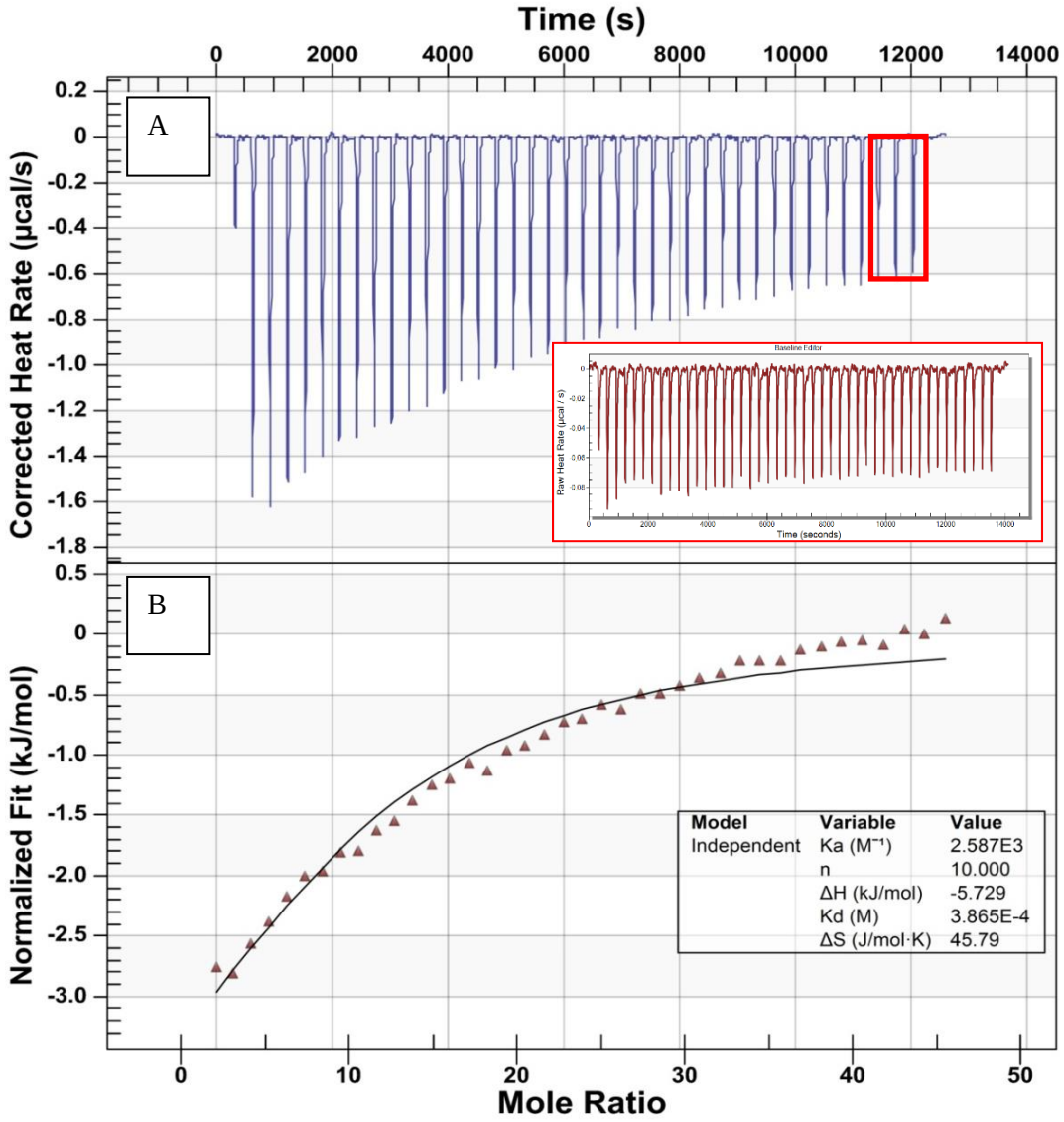


Figure 3.18: Representative calorimetric titration profile of the binding of the anionic dye ANS to EcSspA protein. The experiment was performed at 20 °C. The conditions were 0.05 mM protein and 10 mM ANS in 20 mM potassium phosphate buffer containing 1 mM EDTA at pH 6.5 **(A)** Exothermic heat effects associated with the injection of ANS into the ITC sample cell containing EcSspA protein. The insert panel shows the heats of dilution of ANS into buffer and is boxed in red to show corresponding heats on (A). **(B)** Binding isotherm corrected for heats of dilution corresponding to data in (A). The solid line through the data represents the line of best fit obtained using single independent binding site model.

Table 3.3: Thermodynamics parameters of EcSspA-ANS interaction titration outlined.

K_d (µM)	Δ*H* (kJ/mol)	-TΔ*S* (kJ/mol)	Δ*G* (kJ/mol)
337.9 ± 24.58	**-5.18 ± 0.47**	**-14.46 ± 0.71**	**- 19.64 ± 0.25**

3.12. Molecular docking studies of ANS binding to EcSspA

Molecular docking studies of ANS to EcSspA homology model were performed to predict the binding interactions. Initial potential binding site identification were performed using SiteMap. Five potential binding sites were identified on the protein structure and site 1 (best predicted site) was selected for docking. Figure 3.19 shows that the predicted ligand binding site is at the dimer interface. Tyr21, Glu27, Arg73 and Pro60 are the residues found within the binding site and form weak non-covalent interactions with the ligand. This includes; salt bridges and hydrogen bonds. The interaction between ANS and the protein is shown in Figure 3.20.



Figure 3.19: Homology model of EcSspA in complex with ANS. The image shows dimeric EcSspA with the two chains coloured in purple and green. The ANS ligand bound to the dimer interface is encircled in yellow. The image was generated using Schrodinger PyMOL molecular graphics system.

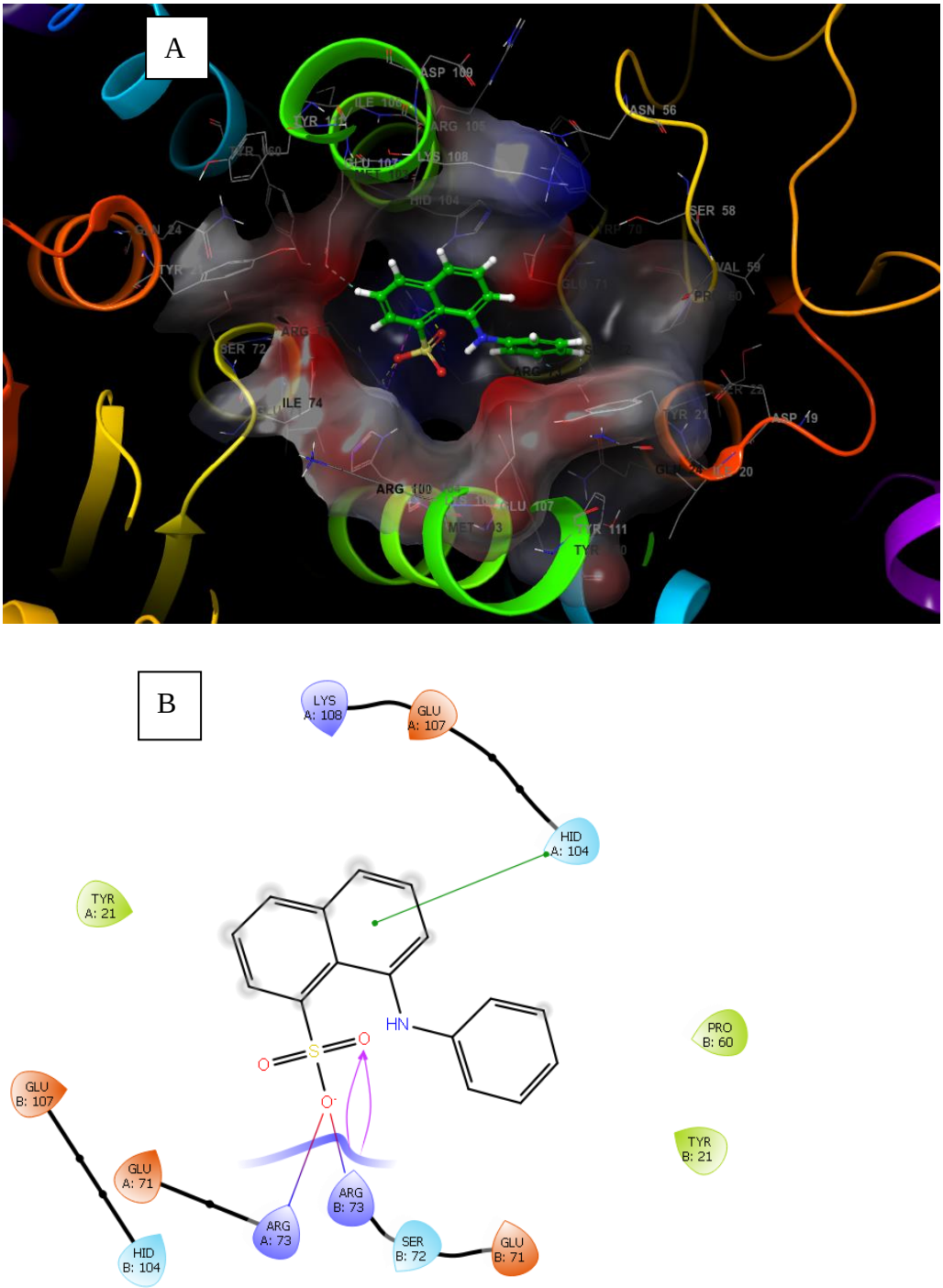


Figure 3.20: Simulated binding of ANS to EcSspA homology model. (A) Green ball and stick representation of ANS bound at the dimer interface simulated using glide docking. EcSspA binding site is represented by the molecular surface and residues found within the binding site are labelled. EcSspA-ANS interactions are shown by; pink broken lines (salt bridge), yellow broken line (H-bonds), and blue (aromatic H-bonds). **(B)** A 2D ligand interaction diagram with nearby residues at the binding site (dimer interface).

3.13. Crystallographic studies

Figure 3.19 shows representative three-dimensional crystals of size ~ 100 µm of the EcSspA protein. Previous attempts at crystallising EcSspA resulted in crystals of dimensions of 0.3 × 0.2 × 0.06 mm of which the best diffraction observed was at a very low resolution of 17 Å (Andrykovitch *et al*., 2003). The crystals shown in Figure 3.19 were too small for crystallographic studies. Precipitates were formed under other conditions used. In the future similar crystals may be used as seeds for subsequent optimisation experiments to get the best crystals. The crystals will then be harvested and taken for three-dimension structure determination of EcSspA. This will provide a deepened understanding of the structure of the protein.

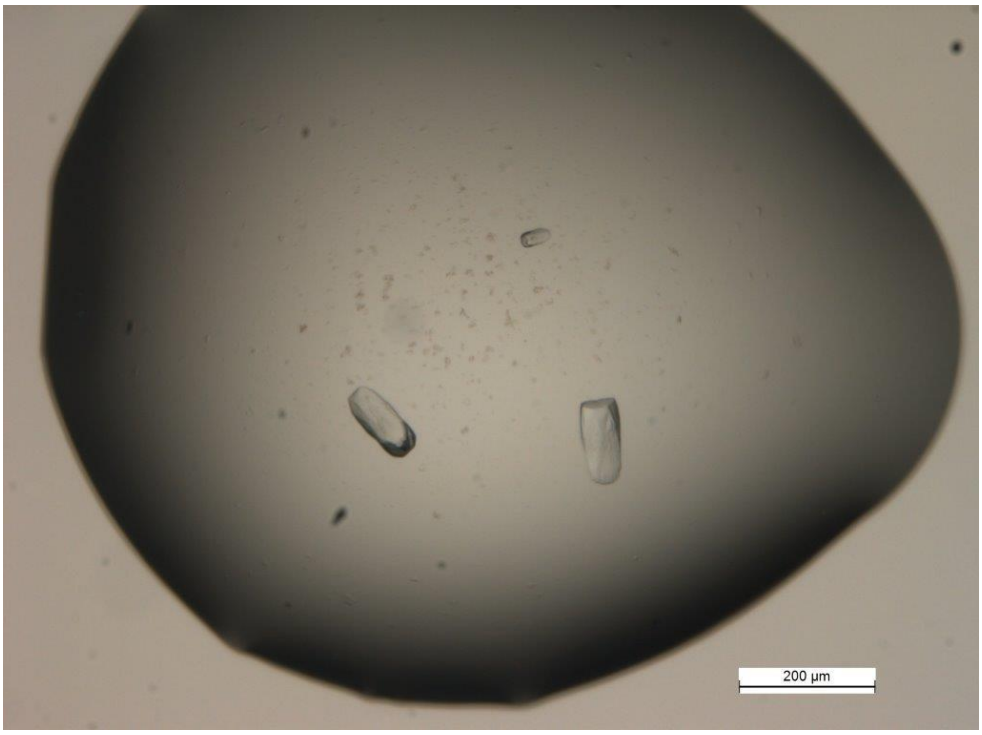


Figure 3.21: Three-dimensional protein crystals of size ~ 100 µm of EcSspA. The protein was prepared at 4.65 mg/ml in a pH buffer solution containing 20 mM sodium phosphate, 1 mM EDTA, 0.02 % sodium azide, 150 mM sodium chloride. Crystallisation drops were prepared by mixing the protein and precipitant at ratios of 1:1, 3:1 and 1:3 in a sitting drop plate with a reservoir volume of 50 µl at 20 °C.

4. CHAPTER FOUR

DISCUSSION

EcSspA is a bacterial homologue of several eukaryotic GST enzymes. The protein is classified as belonging to the GST superfamily of proteins that is characterised with a conserved N-terminal theoredoxin fold motif and a globular alpha helical C-terminal domain (Marchler-Bauer *et al*., 2016). Most of the eukaryotic GST enzymes that have been characterised thus far are known to possess both the enzymatic and ligandin function. The main aim of this study was to investigate whether these functions of GSTs are conserved within a bacterial protein with a GST-like fold. This further elucidates the importance of the relationship between protein structure and function. Knowledge of the structure and function of currently characterised GST fold proteins in literature served as a hypothetical basis on the current approach for the structural and functional characterisation of EcSspA in this study.

4.1. EcSspA is a predominantly alpha helical protein

Based on the far-UV CD data it was confirmed that EcSspA is a predominantly alpha helical protein. The observed CD data results are consistent with the crystal structure of EcSspA orthologue from *Yersinia pestis* with an 83.73 % sequence identity (Hansen *et al*., 2005a). However, the estimated secondary structure content of EcSspA based on circular dichroism analysis assumes a rather significantly lower presence of β-strands when compared to the calculated structural content of YpSspA (Table 2.2). It should be noted though, that the CD spectra and the calculated secondary structural content of EcSspA corresponds to the one typically observed in GST-related proteins, which are known to be predominantly alpha helical proteins with a few β-strands (Kolawole *et al*., 2011; Fanucchi *et al*., 2008; Masino *et al*., 2002). The high percentage of alpha helical content calculated from Dichroweb does correspond to values that were previously observed by Achilonu *et al*. for the full length heEF1γ GST-related protein (Achilonu *et al*., 2014). Overestimation of the secondary structure of proteins with no solved crystal structure seems to be a challenge with Dichroweb database. Therefore, despite the high confidence (NRMSD) in the data calculated by Dichroweb, the data should nevertheless be treated with great caution.

4.2. EcSspA tryptophan residues are less solvent accessible

The fact that the indole ring of tryptophan residues is solvent sensitive can reveal so much about the localised tertiary structure and dynamics of a protein. The dimeric structure of EcSspA consists of six tryptophan residues with three on each subunit. With the use of both intrinsic tryptophan fluorescence and molecular modelling it can be concluded that most of the tryptophan residues are hidden within the interior of the EcSspA protein structure except for Trp70 which is on the surface.

A red shift in the tryptophan fluorescence emission maxima was observed from 341 nm to 350 nm when the native protein was denatured (Figure 3.11). An exposed tryptophan residue will emit at a longer wavelength compared to the same residue in the interior of a protein (Lackowicz, 2006). The emission maximum wavelength of a tryptophan residue in an aqueous buffer solution is shown to be 350 nm and is highly dependent on the local environment (Lackowicz, 2006). Similar emission maximum wavelength was observed when EcSspA was denatured in this study, therefore confirming that the residues are more solvent exposed compared to when the protein is in its native conformation. It has been shown that there is a relationship between the secondary structure of a protein and the accessibility of the residues that make up the secondary structure (Lins *et al*., 2003). The distribution of the tryptophan residues across the surface of EcSspA protein and secondary structure was further analysed to support the claim that the tryptophan residues are less solvent accessible (Figure 4.1B). Based on the calculated protein molecular surface (Figure 4.1A) which was performed in order to locate the presence of tryptophan residues on the surface, only Trp70 was shown to be exposed to the environment. To try and further locate the location of the residues on the secondary structure, a ribbon diagram labelled with the tryptophan residues was generated (Figure 4.1B) which shows that the residues are mainly on alpha helical folds. This explains the reduced accessibility of the residue in the native conformation and is consistent with the claim that residues on helical folds have an intermediate solvent accessibility (Lins *et al*., 2003).

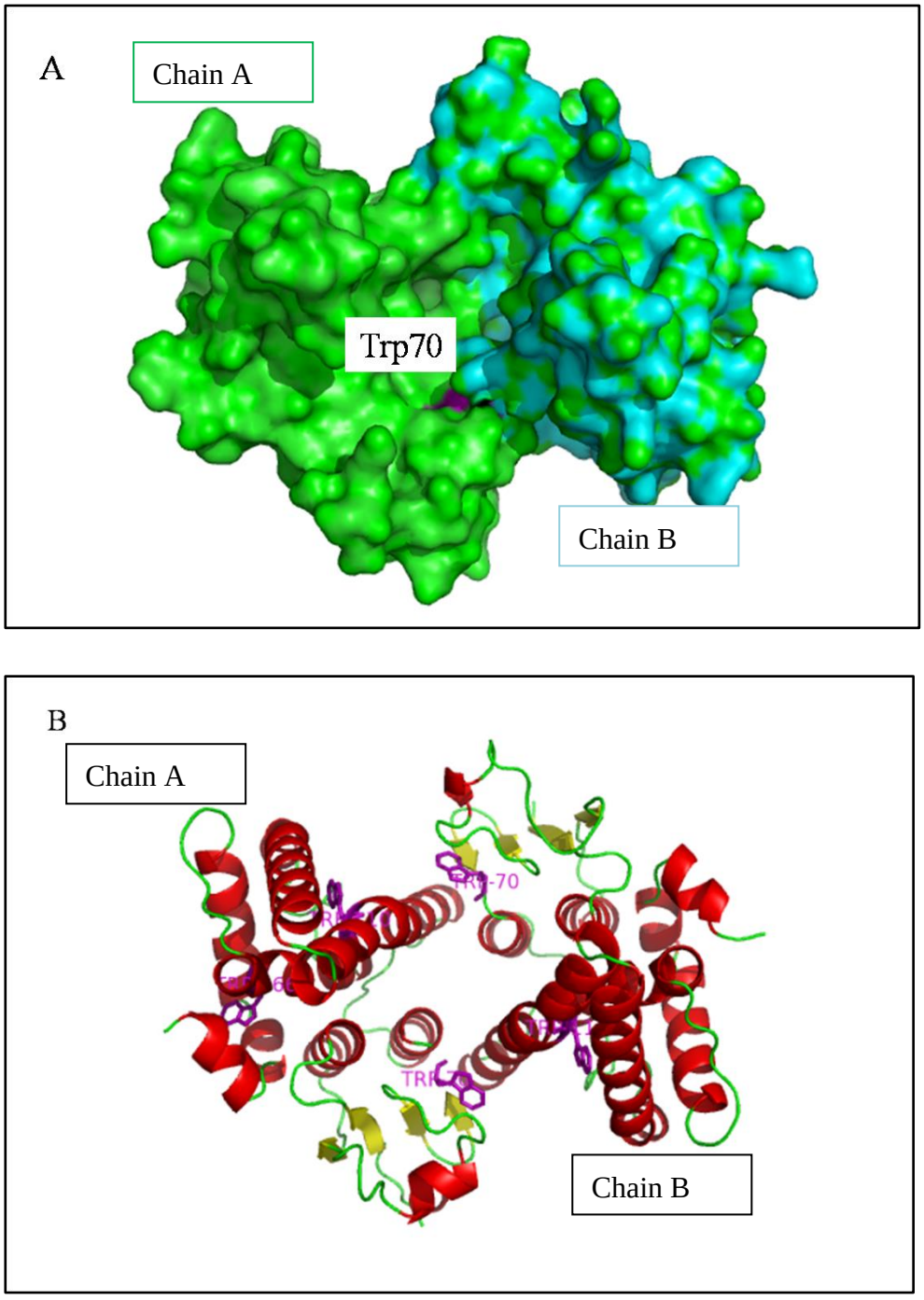


Figure 4.1: Tryptophan residue distribution in dimeric EcSspA. (A) Molecular surface of EcSspA showing the distribution of a single tryptophan residue at the surface. Of the three residues found in each subunit only a patch representing Trp70 appears on the surface of the protein. All the other residues are masked in the interior. **(B)** The distribution of tryptophan residues on the secondary structures of EcSspA. Red (α-helices), yellow (β-structures) and yellow (loops). The tryptophan residues are mainly located on α-helices. The residues are coloured purple. The images were generated using PyMOL and are in the same orientation (DeLano, 2002).

4.3. Oligomeric state of EcSspA

SE-HPLC was performed in order to characterise the oligomeric state of EcSspA in solution. hGSTP1-1 which is an obligate dimeric GST was used as a control and is known to elute as a dimer under experimental conditions. SE-HPLC chromatographic profiles shows peaks which corresponded to monomeric (~24 kDa) species of EcSspA and 30.2 kDa for the dimeric hGSTP1-1. The calculated theoretical molecular masses of a His-tagged dimeric EcSspA and dimeric hGSTP1-1 are 50 kDa and 48 kDa, respectively. The significant difference between the experimentally determined and theoretically calculated molecular masses for the proteins could be attributed to the anomalous behaviour of the proteins in size-exclusion columns and the choice of gel filtration standards. Proteins have been shown to behave anomalously when subjected to size exclusion therefore posing a challenge to the molecular mass and oligomeric state determination of a protein in solution (Nedelkov *et al*., 2006).

Based on the results as shown in Figure 3.12B the elution profiles of the two proteins are aligned and there is no great distinction in terms of the elution times, therefore confirming that both proteins are dimeric. Previous characterisation of rGSTM1-1 using size exclusion has shown that a clear distinction in the elution peaks of a monomeric GST and a dimeric GST can be resolved (Thompson *et al*., 2006). Therefore, rejecting the suggestion by the experimentally determined mass in this study that EcSspA is monomeric. One of the greatest limitations of using size exclusion is that it takes no account of the molecular shape of the biomolecule being separated especially when non-globular molecules are used (Tayyab *et al*., 1991). Therefore, a great caution should be paid when analysing data from SE-HPLC. Given the limitations of SE-HPLC, future studies on an attempt to determine the absolute molecular mass of the protein should consider size exclusion coupled with light scattering as previously done by Thompson *et al*. (2006). This technique will allow for an accurate estimation of the size of the protein irrespective of its shape and conformation.

4.4. EcSspA does not bind GSH and shows no catalytic activity for GSH/CDNB conjugation

This study has shown that EcSspA displays no catalytic activity for the GSH/CDNB conjugation and does not interact with GSH immobilised on agarose beads. The crystal structure of YpSspA has shown that the protein assumes the canonical GST like fold (Hansen *et al*., 2005a). This then led to the hypothesis that proteins of similar structures have similar

functions; therefore, EcSspA could have some GST activity. According to this study EcSspA has no activity. The results do not support the initial hypothesis.

Studies by Hansen *et al*. have also shown that SspA is impaired in GST activity and binding of GSH (Hansen *et al*., 2005a). A structure-based sequence alignment between GSTB1-1 and EcSspA has shown that the catalytic residue Cys10 in GSTB1-1 is replaced with Tyr21 in EcSspA (Hansen *et al*., 2005a). In this study a structural based sequence alignment was performed between the subunits of EcSspA and the GSH bound hGSTP1-1. This revealed that the conformation of these subunits differs significantly and superimpose with a rmsd value of 4.71 (see Figure 4.2). The difference in the conformation of the proteins is largely due to a lack of amino acid sequence identity with a 20.2 % sequence identity (48.7 % sequence similarity) (Gasteiger *et al*., 2005). SspA protein lacks some of the key residues that play a significant role in GSH binding. The lack of GSH binding can therefore be attributed to the absence of correctly positioned binding and catalytic residues at the active which is supported by lack of binding of the protein to a GSH immobilised on agarose and a lack of GSH/CDNB catalytic activity. This is further supported by an attempt by Hansen *et al*. (2005a) at soaking crystals of YpSspA with GSH and its analogues and no bound ligands were identified. Previous studies on mglA (macrophage locus protein A); a protein similar to the SspA protein family has also demonstrated that SspA proteins have no GST activity (Cuthbert *et al*., 2015).

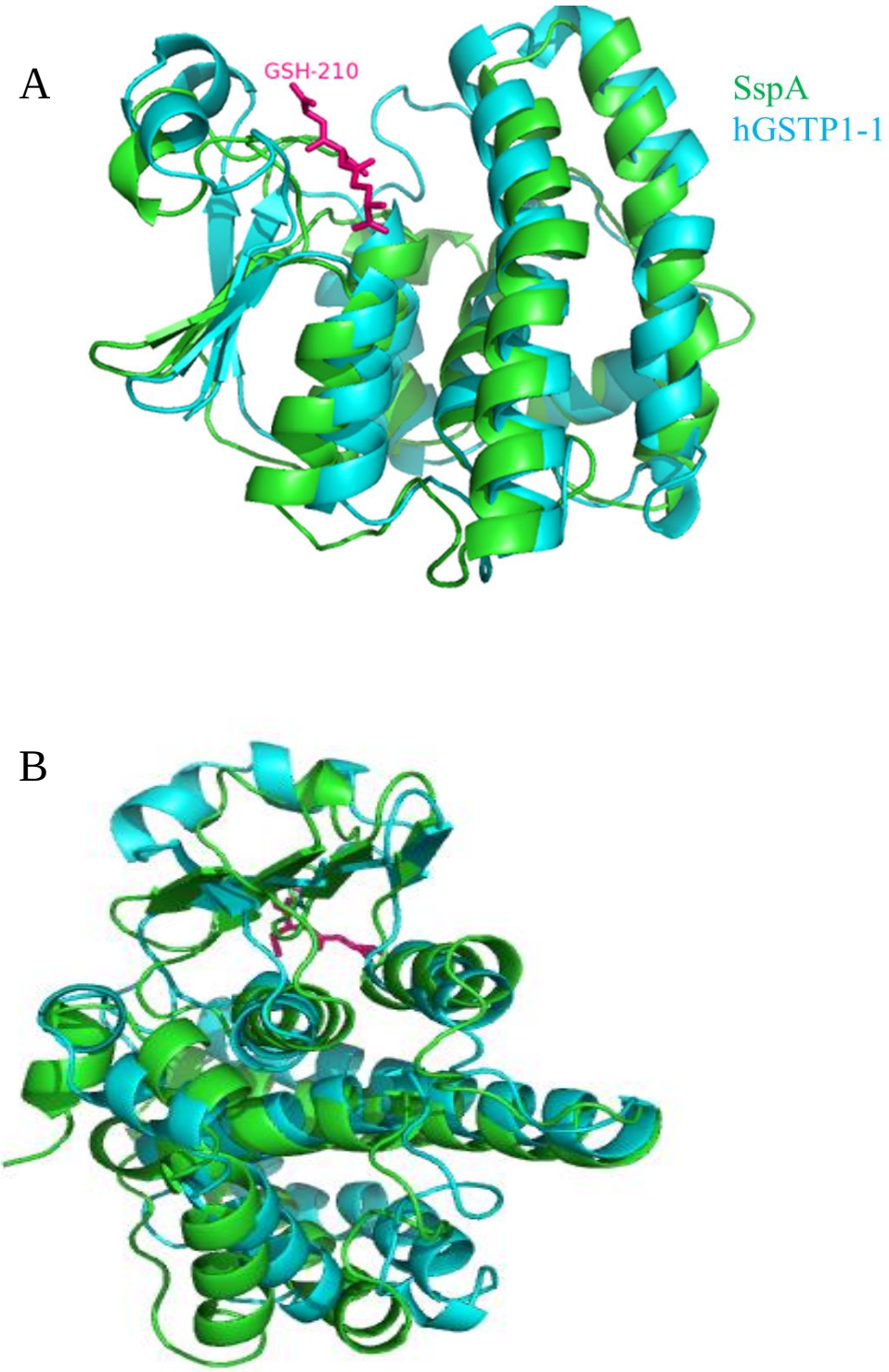


Figure 4.2: Structural comparison of EcSspA and hGSTP1-1. **(A)**Superimposition of an EcSspA subunit onto that of hGSTP1-1 bound to a GSH molecule (PDB ID: 6GSS). **(B)**Superimposition of the same molecules in a different orientation. The subunits superimpose with a rmsd of 4.71 Å for 161 corresponding Cα atoms. The proteins are illustrated as ribbon diagrams (helices as spirals, β-strands as arrows and loops as pipes). The pink molecule is GSH. The images were generated using PyMOL (DeLano, 2002).

4.5. EcSspA possesses ligandin function towards ANS

The substantial increase in the fluorescence intensity accompanied by a significant blue shift of the emission maximum wavelength of ANS bound in the presence of EcSspA, together with the results from isothermal titration calorimetry, indicate the binding of ANS to the protein. The degree of the blue shift in the emission maximum wavelength is dependent upon the presence of water in the ANS binding site (Kirk *et al*., 1996). The emission maximum wavelength for ANS bound to hGSTA1-1 is 475 nm (Sayed *et al*., 2002), ~ 480 nm in ANS bound to 1-FABP (Kirk *et al*., 1996), ~ 480 nm in ANS bound to NT-heEF1γ (Achilonu *et al*., 2014) and in this study it was found to be ~ 480 nm when bound to EcSspA suggesting that the polarity of the ANS binding site in these proteins is similar. This is based on the comparison of the emission maximum wavelengths of 545 nm for free ANS in water and of ~ 454 nm for ANS bound to the highly hydrophobic site in apomyoglobin (Stryer, 1965). The fluorescence intensity of ANS bound to a protein is dependent on the accessibility of the binding cleft, the more solvent exposed the lower the fluorescence intensity (Kirk *et al*., 1996; Matulis and lovrien, 1998). The fluorescence enhancement of ANS bound EcSspA is lower than that observed for lipid binding proteins, indicating some exposure of EcSspA-bound ANS to the solvent. This is consistent with results from several studies which have shown that GST-like proteins do have a solvent accessible ANS-binding site in their native state conformation (Kinsley *et al*., 2008; Dirr *et al*., 2005; Sayed *et al*., 2002; Bico *et al*., 1995).

Moreover, the saturation curve of ANS fluorescence signal as a function of ANS concentration (see Figure 3.17) may indicate that the ANS interaction with EcSspA is site specific. The resultant K_d (50 µM) is comparable to that of other GST-related proteins; NT-heEF1γ (70 µM) and hGSTA1-1 (65 µM) determined by fluorescence spectroscopy and isothermal titration calorimetry, respectively (Achilonu *et al*., 2014; Sayed *et al*., 2002). This value is however, not comparable to the one determined by isothermal titration calorimetry in this study (337.9 ± 24.58µM) which is 6-fold greater. Furthermore, a stoichiometric *N* value of 10 molecules of ANS bound per dimer was obtained from ITC, which is overestimated, and such a value has never been observed before for ANS binding GST proteins. Based on the EcSspA homology model and ANS docking data, it was predicted that EcSspA can only bind one ANS molecule at the subunit interface of dimeric EcSspA (Figure 3.19), this together with the work done by Yassin *et al*. which have shown that *Schistosoma japonicum*

GST binds one ANS molecule per mole of the dimeric enzyme through an equilibrium dialysis (Yassin *et al*., 2004) shows that the *N* value of 10 ANS molecules per dimeric EcSspA by ITC is an overestimation.

The Wiseman constant (c-value) is an important experimental condition that needs to be considered when performing ITC-ligand protein experiments. This c-value is the product of the binding association constant, the total protein concentration in the ITC cell and the stoichiometry parameter at the start of the experiment ($c = K_a \times [protein] \times N$) and is required to be $20 < c < 100$ (Myszka *et al*., 2003) for an accurate estimation of binding thermodynamic parameters. Unfortunately, the conditions used in this study were not favourable; the protein concentration used resulted in a c-value of 1 based on the K_d determined from fluorescence experiment and an *N* value of 1. The reason being that given a K_d value of 50 µM obtained from fluorescence studies (Figure 3.17), in order to get a c-value of 20, a protein concentration of 1000 µM will be required. However, protein concentrations higher than the one used in this study results in aggregation of the protein solution during protein preparation and stirring of the protein sample during calorimetric titration, resulting in an unstable instrumental baseline.

When considering the data shown in Figure 4.3A, lower c-values result in shallow titration curves which results in an unreliable estimate of the three parameters (*N*, K_d and ΔH). Figure 4.3B shows that a sigmoidal curve is required for an accurate estimation of the parameters. This therefore makes the *N* value and K_d obtained under conditions of low protein concentration used in this study (i.e., c-value) less accurate when compared to the K_d obtained from fluorescence enhancement study. Fluorescence studies allowed for a use of a protein concentration [EcSspA] that is way lower than the total ligand concentration $[ANS]_t$ and at high ANS concentration EcSspA was completely saturated. A hyperbolic data showing a complete saturation of the EcSspA by ANS (Figure 3.17) allowed for a fit of the data to a hyperbolic equation which allowed for an estimation of K_d at a correlation coefficient of 0.96, this together with the populated data points when [ANS] is less than K_d (Figure 3.17) gives a more reliable estimate when compared to the ITC value.

Molecular docking studies have further shown that the interaction between ANS and EcSspA is weak and is characterised by presence of weak non-covalent interactions. A salt bridge and hydrogen bonds have been found to influence this protein ligand interaction (Figure 3.20). The ligand binding site of EcSspA is characterised by the presence of a number of uncharged

non-polar residues. These non-polar residues contribute to the hydrophobic nature of the ANS binding site and are in close proximity with the ligand, interacting with the hydrophobic aniline and naphthyl rings of the ANS. The calorimetric titration isotherm for the binding of ANS to EcSspA (Figure 3.18) shows that the complex formation is exothermic and enthalpically driven. Electrostatic interactions between sulfonate group of ANS and arginine of hGSTA1-1 has exothermic contributions to the enthalpic nature of the interaction (Dirr *et al*., 2005). Arg73 found in the dimer interface is responsible for an electrostatic interaction via ion pairing of ANS and arginine side chain and could therefore be used to explain the exothermic heats observed for ANS-EcSspA interaction. The ion pairing between arginine side chain and ANS is observed in a number of GSTs and ANS binding proteins (Schönbrunn *et al*., 2000; Ory *et al*., 1999; Matulis and Lovrien, 1998). The hydrophobic nature of the ANS binding site in EcSspA is consistent with the increase in emission intensity and blue shift observed from spectroscopic studies (Figure 3.16).

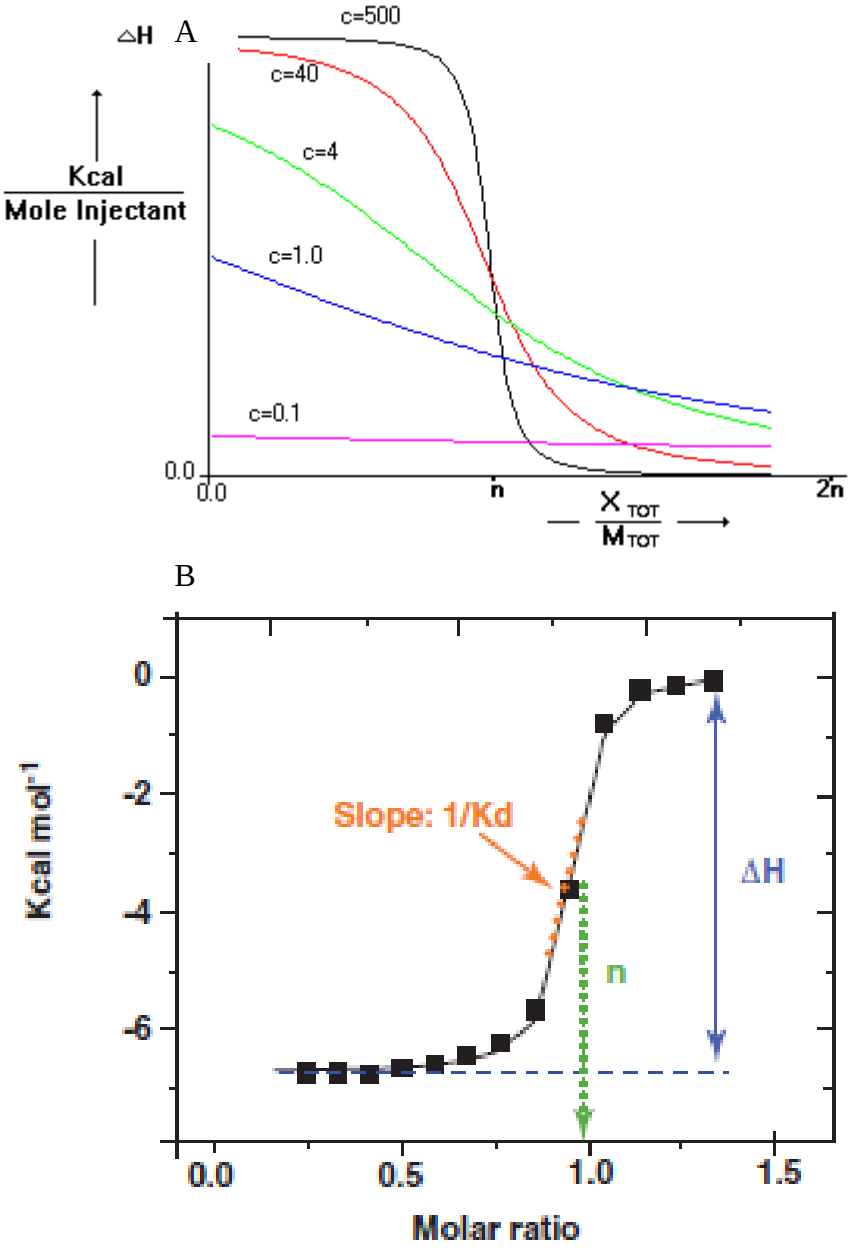


Figure 4.3: Simulated binding isotherms for various c-value parameters and extraction of the binding parameters. (A) The different coloured binding isotherms show that a higher c-value will result in a titration curve that is too steep to resolve *K* accurately (while *N* and Δ*H* are well resolved). Lower c-values results in shallow titration curves from which all three parameters are poorly resolved. The image was adapted from Wiseman *et al*. (1989). **(B)** A titration profile showing raw thermodynamic data extracted from a typical ITC experiment. The data is plotted based on a suitable binding model. The change in enthalpy Δ*H*, stoichiometry *N*, and the association constant K_a can be directly determined from the plot. The change in enthalpy is determined from the pre and post transition regions of the curve. The stoichiometry of binding is determined from the inflection point of the curve. The slope at the transition region is used to determine the association constant. This image was adapted from Nunez *et al*. (2002).

5. REFERENCES

Abdalla, A. M. and Hamed, R. R. (2006) Multiple unfolding states of glutathione transferase from *Physaacuta* (Gastropada: Physidae). *Biochem. Biophys. Res. Commun*. **340**, 625-632

Achilonu, I., Siganunu, T. P. and Dirr, H. W. (2014) Purification and characterisation of recombinant human eukaryotic elongation factor 1 gamma. *Protein. Expr. Purif*. **99**, 70-77

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389-3402

Andrykovitch, M., Routzahn, K. M., Li, M., Gu, Y., Waugh, D. S. and Ji, X. (2003) Characterisation of four orthologues of stringent starvation protein A. *Crystallogr D. Biol. Crystallogr. Acta*. **59**, 881-886

Armstrong, R. N. (1991) Glutathione *S*-transferases: reaction mechanism, structure, and function. *Chem. Res. Toxico*. **4**, 131-140

Aslund, F., Ehn, B., Miranda-Vizuete, A., Pueyo, C. and Holmgren, A. (1994) Two additional glutaredoxins exist in *Escherichia coli*: glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant. *Proc. Natl. Acad. Sci*. **91**, 9813-9817

Bartlett, John G., Peter C. Belitsos, and Cynthia L. Sears. (1992) "AIDS enteropathy". *Clin. Infect Dis.* **15**, 726-735

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T. and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. **42**, 252-258

Bico, P., Erhardt, J., Kaplan, W. and Dirr, H. (1995) Porcine class π glutathione *S*-transferase: Anionic ligand binding and conformational analysis. *Biochim .Biophys. Acta. Rev*, **1247**. 225-230

Boyce, T. G., Swerdlow, D. L. and Griffin, P. M. (1995) *Escherichia coli* O157: H7 and the haemolytic–uremic syndrome.*N. Engl. J. Med*. **333**, 364-368

Caccuri, A. M., Aceto, A., Piemonte, F., Ilio, C., Rosato, N. and Federici, G. (1990) Interaction of haemin with placental glutathione transferase. *FEBS J*. **189**, 493-497

Chung, C. T., Niemela, S. L. and Miller, R. H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci*. **86**, 2172-2175

Chung, H. J., Bang, W. and Drake, M. A. (2006) Stress response of *Escherichia coli*. *Compr. Rev. Food Sci. Food Saf*.**5**, 52-64

Coschigano, P. W. and Magasanik, B. (1991) The URE2 gene product of *Saccharomyces cerevisiae* plays an important role in the cellular response to the nitrogen source and has homology to glutathione *S*-transferases. *Mol. Cell. Biol*. **11**, 822-832

Cuthbert, B. J., Brennan, R. G. and Schumacher, M. A. (2015) Structural and Biochemical Characterisation of the *Francisella tularensis* Pathogenicity Regulator, Macrophage Locus Protein A (MglA). *PloS one*. **10**, e0128225

De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E. and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. **34**, 362-365

DeLano, W. L. (2002) The PyMOL molecular graphics system. *http://pymol. Org*

De Reuse, H. and Taha, M. K. (1997) RegF, an SspA homologue, regulates the expression of the *Neisseria gonorrhoeae* pilE gene. *Res. Microbiol*. **148**, 289-303

Dirr, H., Reinemer, P. and Huber, R. (1994) X ray crystal structures of cytosolic glutathione *S*-transferases. *FEBS J*. **220**, 645-661

Dirr, H. W., Little, T., Kuhnert, D. C. and Sayed, Y. (2005) A conserved N-capping motif contributes significantly to the stabilisation and dynamics of the C-terminal region of class Alpha glutathione S-transferases. *J. Biol. Chem*. **280**, 19480-19487

Effler, E., Isaäcson, M., Arntzen, L., Heenan, R., Canter, P., Barrett, T. and Griffin, P. M. (2001) Factors contributing to the emergence of *Escherichia coli* O157 in Africa. *Emerg. Infect. Diseases.* **7**, 812

Fanucchi, S., Adamson, R. J. and Dirr, H. W. (2008) Formation of an unfolding intermediate state of soluble chloride intracellular channel protein CLIC1 at acidic pH. *Biochem*. **47**, 11674-11681

Fernandes, A. P., Fladvad, M., Berndt, C., Andrésen, C., Lillig, C. H., Neubauer, P. and Vlamis-Gardikas, A. (2005) A novel monothiol glutaredoxin (Grx4) from *Escherichia coli* can serve as a substrate for thioredoxin reductase. *J. Biol. Chem*. **280**, 24544-24552

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T. and Shaw, D. E. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem*. **47**, 1739-1749

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. 571-607. Humana Press

Gorden, J. and Small, P. (1993) Acid resistance in enteric bacteria. *Infect. Immun*. **61**, 364-367

Gorrec, F. (2009) The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr*. **42**, 1035-1042

Greenfield, N. J. (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem*. **235**, 1-10

Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974) Glutathione *S*-transferases the first enzymatic step in mercapturic acid formation.*J. Biol. Chem*. **249**, 7130-7139

Habig, W. H. and Jakoby, W. B. (1981) Assays for differentiation of glutathione *S*-Transferases. *Methods Enzymol*. **77**, 398-405

Halgren, T. (2007) New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des*. **69**, 146-148

Hansen, A. M., Lehnherr, H., Wang, X., Mobley, V. and Jin, D. J. (2003*) Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. *Mol. Microbiol*. **48**, 1621-1631

Hansen, A. M., Gu, Y., Li, M., Andrykovitch, M., Waugh, D. S., Jin, D. J. and Ji, X. (2005a) Structural basis for the function of stringent starvation protein a as a transcription factor. *J. Biol. Chem*. **280**, 17380-17391

Hansen, A. M., Qiu, Y., Yeh, N., Blattner, F. R., Durfee, T. and Jin, D. J. (2005b) SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. *Mol. Microbiol*. **56**, 719-734

Hansen, A. M. and Jin, D. J. (2012) SspA up-regulates gene expression of the LEE pathogenicity island by decreasing H-NS levels in enterohemorrhagic *Escherichia coli*. *BMC Microbiol*. **12**, 1

Harrop, S. J., DeMaere, M. Z., Fairlie, W. D., Reztsova, T., Valenzuela, S. M., Mazzanti, M. and Bauskin, A. R. (2001) Crystal structure of a soluble form of the intracellular chloride ion channel CLIC1 (NCC27) at 1.4-Å resolution. *J. Biol. Chem*. **276**, 44993-45000

Hayes, J. D. and Pulford, D. J. (1995) The glutathione *S*-transferase supergene family: regulation of GST and the contribution of the lsoenzymes to cancer chemoprotection and drug resistance part II. *Crit. Rev. Biochem. Mol. Biol*. **30**, 521-600

Holmgren, A., Söderberg, B. O., Eklund, H. and Brändén, C. I. (1975) Three-dimensional structure of *Escherichia coli* thioredoxin-S2 to 2.8 A resolution. *Proc. Natl. Acad. Sci*. **72**, 2305-2309

Ishihama, A. and Saitoh, T. (1979) Subunits of RNA polymerase in function and structure: IX. Regulation of RNA polymerase activity by stringent starvation protein (SSP). *J. Mol. Biol*. **129**, 517-530

Isaäcson, M., Canter, P. H., Effler, P., Arntzen, L., Bomans, P. and Heenan, R. (1993) Haemorrhagic colitis epidemic in Africa. *The Lancet*, **341**, 961

Jordan, K. N., Oxford, L. and O'Byrne, C. P. (1999) Survival of low-pH stress by *Escherichia coli* O157: H7: correlation between alterations in the cell envelope and increased acid tolerance. *Appl. Environ. Microbiol*. **65**, 3048-3055

Ketterer, B., Carne, T. and Tipping, E. (1978) Ligandin and protein A: intracellular binding proteins. *Transport by proteins, Walter de Gruyter, Berlin*, 69-94

Kirk, W. R., Kurian, E. and Prendergast, F. G. (1996) Characterisation of the sources of protein-ligand affinity: 1-sulfonato-8-(1') aniline naphthalene binding to intestinal fatty acid binding protein. *Biophys .J*. **70**, 69-83

Kinsley, N., Sayed, Y., Mosebi, S., Armstrong, R. N. and Dirr, H. W. (2008) Characterisation of the binding of 8-anilinonaphthalene sulfonate to rat class Mu GST M1-1. *Biophys. Chem*. **137**, 100-104

Kolawole, A. O., Ajele, J. O. and Sirdeshmuhk, R. (2011) Studies on glutathione transferase of cowpea storage bruchid, *Callosobrochusmaculatus*F. *Pest. Biochem. Physiol*. **100**, 212-220

Koonin, E. V., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Mushegian, A. R., Bork, P. and Valencia, A. (1994) Eukaryotic translation elongation factor 1γ contains a glutathione transferase domain—Study of a diverse, ancient protein super family using motif search and structural modelling. *Protein Sci*. **3**, 2045-2055

Lackowicz, J. R. Principle of Fluorescence Spectroscopy, 2006. *Springer Science and Business Media, LLC, New York ISBN*. **13**, 978-0

Ladbury, J. E. and Chowdhry, B. Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol*. **3**, 791-801

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

Lange, R. and Hengge-Aronis, R. (1994) The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev*. **8**, 1600-1612

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. and Thompson, J. D. (2007) ClustalW and ClustalX version 2.0. *Bioinformatics*. **23**, 2947-2948

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. (1997) PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**, 488-490

Lins, L., Thomas, A. and Brasseur, R. (2003) Analysis of accessible surface of residues in proteins. *Protein Sci*. **12**, 1406-1417

Litwack, G., Ketterer, B. and Arias, I. M. (1971) Ligandin: a hepatic protein which binds steroids, bilirubin, carcinogens and a number of exogenous organic anions. *Nature*. **234**, 466

Mach, H., Volkin, D. B., Burke, C. J. and Russell Middaugh, C. (1995) Ultraviolet absorption spectroscopy. *Protein Stability and Folding: Theory and Practice*, 91-114

Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I. and Koonin, E. V. (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res*. **9**, 608-628

Masai, E., Katayama, Y., Kubota, S., Kawai, S., Yamasaki, M. and Morohoshi, N. (1993) A bacterial enzyme degrading the model lignin compound β-etherase is a member of the glutathione *S*-transferase superfamily. *FEBS lett*. **323**, 135-140

Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jörnvall, H. (1985) Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. *Proc. Natl. Acad. Sci*. **82**, 7202-7206

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S. andGwadz, M. (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. **45**, 200-203

Masino, L., Kelly, G., Leonard, K., Trottier, Y. and Pastore, A. (2002) Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS letters*. **513**, 267-272

Matulis, D. and Lovrien, R. (1998) 1-Anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. *Biophys. J*. **74**, 422-429

McTigue, M. A., Williams, D. R. and Tainer, J. A. (1995) Crystal Structures of a *Schistosomal* Drug and Vaccine Target: Glutathione *S*-Transferase from *Schistosomajaponica* and its Complex with the Leading Antischistomal Drug Praziquantel.*J. Mol. Biol*.**246**, 21-27

Myszka, D. G., Abdiche, Y. N., Arisaka, F., Byron, O., Eisenstein, E., Hensley, P. and Doyle, M. L. (2003) The ABRF-MIRG'02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction. *J. Biomol. Tech*. **14**, 247

Nedelkov, D., Nelson, R. W. and Folta-Stogniew, E. (2006) Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *New and Emerging Proteomic Techniques*, 97-112

Nilsson, S. F., Rask, L. and Peterson, P. A. (1975) Studies on thyroid hormone-binding proteins. II. Binding of thyroid hormones, retinol-binding protein, and fluorescent probes to prealbumin and effects of thyroxine on prealbumin subunit self association. *J. Biol. Chem*. **250**, 8554-8563

Nishida, M., Harada, S., Noguchi, S., Satow, Y., Inoue, H. and Takahashi, K. (1998) Three-dimensional structure of *Escherichia coli* glutathione *S*-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106. *J. Mol. Biol*. **281**, 135-147

Núñez, S., Venhorst, J. and Kruse, C. G. (2012) Target–drug interactions: first principles and their application to drug discovery. *Drug Discov. Today*. **17**, 10-22

Oakley, A. J., Bello, M. L., Nuccetelli, M., Mazzetti, A. P. and Parker, M. W. (1999) The ligandin (non-substrate) binding site of human pi class glutathione transferase is located in the electrophile binding site (H-site) 1. *J. Mol. Biol*. **291**, 913-926

Olsen, J. V., Ong, S. E. and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics*. **3**, 608-614

Ory, J. J. and Banaszak, L. J. (1999) Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate. *Biophys. J*. **77**, 1107-1116

Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol*. **131**, 266-280

Reinemer, P., Dirr, H. W., Ladenstein, R., Schäffer, J., Gallay, O. and Huber, R. (1991) The three-dimensional structure of class pi glutathione *S*-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *EMBO J*. **10**, 1997-2005

Royer, C. A. (1995) Approaches to teaching fluorescence spectroscopy. *Biophys J*. **68**, 1191-1195

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Ed. 2. Cold spring harbor laboratory press.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci*. **74**, 5463-5467

Sayed, Y., Hornby, J. A., Lopez, M. and Dirr, H. (2002) Thermodynamics of the ligandin function of human class Alpha glutathione transferase A1-1: energetics of organic anion ligand binding. *Biochem. J*. **363**, 341-346

Schönbrunn, E., Eschenburg, S., Luger, K., Kabsch, W. and Amrhein, N. (2000) Structural basis for the interaction of the fluorescence probe 8-anilino-1-naphthalene sulfonate (ANS) with the antibiotic target MurA. *Proc. Natl. Acad. Sci*. **97**, 6345-6349

Sheehan, D., Meade, G., Foley, V. M. and Dowd, C. A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J*. **360**, 1-16

Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc*. **1**, 2856-2860

Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P.and Schaeffer, D. A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics*. **6**, 1638-1655

Slavik, J., Horák, J.,Říhová, L. andKotyk, A. (1982) Anilinonaphthalene sulfonate fluorescence and amino acid transport in yeast. *J. Membr. Biol*. **64**, 175-179

Small, P., Blankenhorn, D., Welty, D., Zinser, E. and Slonczewski, J. L. (1994) Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of rpoS and growth pH. *J.Bacteriol*. **176**, 1729-1737

Smith, D. B. and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene*. **67**, 31-40

Stryer, L. (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin: A fluorescent probe of non-polar binding sites. *J. Mol. Biol*. **13**, 482-495

Testerman, T. L., Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Libby, S. J. and Fang, F. C. (2002) The alternative sigma factor σE controls antioxidant defences required for *Salmonella* virulence and stationary phase survival. *Mol .Microbiol*. **43**, 771-782

Tew, K. D., Manevich, Y., Grek, C., Xiong, Y., Uys, J. and Townsend, D. M. (2011) The role of glutathione *S*-transferase P in signalling pathways and *S*-glutathionylation in cancer. *Free Radical Biol. Med*. **51**, 299-313

Tomarev, S. I. and Zinovieva, R. D. (1988) Squid major lens polypeptides are homologous to glutathione *S*-transferases subunits. *Nature*. ***336***, 86

Tomarev, S. I., Chung, S. and Piatigorsky, J. (1995) Glutathione *S*-transferase and *S*-crystallins of cephalopods: evolution from active enzyme to lens-refractive proteins. *J. Mol. Evol.* **41**, 1048-1056

Toung, Y. P. S. and Tu, C. P. D. (1992) *Drosophila* glutathione *S*-transferases have sequence homology to the stringent starvation protein of *Escherichia coli*. *Biochem. Biophys. Res. Comm*. **182**, 355-360

Whitmore, L. and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*. **89**, 392-400

Wiseman, T., Williston, S., Brandts, J. F. and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem*. **179**, 131-137

Williams, M. D., Ouyang, T. X. and Flickinger, M. C. (1994) Starvation-induced expression of SspA and SspB: the effects of a null mutation in sspA on *Escherichia coli* protein synthesis and survival during growth and prolonged starvation. *Mol. Microbiol*. **11**, 1029-1043

Woody, R. W. (1995) Circular dichroism. *Methods Enzymol*. **246**, 34-71

Wrench, A. P., Gardner, C. L., Siegel, S. D., Pagliai, F. A., Malekiha, M., Gonzalez, C. F. and Lorca, G. L. (2013) MglA/SspA complex interactions are modulated by inorganic polyphosphate. *PLoS One*. ***8***, e76428

Yassin, Z., Ortiz-Salmerón, E., García-Maroto, F., Barón, C. and García-Fuentes, L. (2004) Implications of the ligandin binding site on the binding of non-substrate ligands to *Schistosoma japonicum*-glutathione transferase. *Biochim. Biophys. Acta Rev*. **1698**, 227-237