Waterston, Claire Louise (2009) *Structural studies of putative general stress and related proteins from Deinococcus radiodurans.* PhD thesis.

http://theses.gla.ac.uk/804/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

Structural Studies of Putative General Stress and Related Proteins from *Deinococcus radiodurans*

by

Claire Louise Waterston

A thesis submitted for the degree of Doctor of Philosophy in the Faculty of Physical Sciences, University of Glasgow

Declaration

This thesis has been written in accordance with the University of Glasgow regulations, has not been presented for a degree at any other university and is original except where indicated otherwise by reference in the text. The work contained within is the author's own except where work was done in collaboration as indicated.

© Claire Louise Waterston

Signed	Date
Claire Louise Waterston	21st May 2009

Dedicated to my Grandfather, James Arthur Anderson (1922-2008) who tirelessly supported and inspired my education from an early age.

Abstract

This study describes the cloning, expression, purification, biophysical characterisation and crystallisation of DR_1146; a putative general stress protein from the extremophilic bacterium *Deinococcus radiodurans* (R1).

The extraordinary ability of *D. radiodurans* to resist mutation or apoptosis on exposure to high does of ionising radiation has formed the basis of a structural genomics project underway at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The work presented in this study forms part of the ESRF's *D. radiodurans* initiative, and was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and the ESRF as an Industrial Cooperative Award in Science and Engineering (CASE) PhD studentship. A period of one-year was spent on secondment at the ESRF, working within the Macromolecular Crystallography Group.

Several constructs of the *dr_1146* gene have been successfully overexpressed in *E. coli* cells to give high yields of target protein. Purification by immobilised metal affinity chromatography (IMAC) was facilitated by the incorporation of a 6xHis tag and supplemented by a final gel filtration step. Although high purity levels were achieved, imaging by SDS-PAGE analysis identified that DR_1146 was susceptible to stringent proteolysis. It is thought that initial crystallisation trials were unsuccessful due to inhomogeneity of the sample caused by reported degradation of the target protein.

Biophysical characterisation of DR_1146 by isothermal titration calorimetry (ITC) and fluorescence spectroscopy (FS) identified a moderate affinity of 4-11 μM for the flavin molecules, riboflavin, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Differential scanning calorimetry (DSC) and circular dichroism (CD) experiments demonstrated an increase in chemical and thermal stability of the protein on binding to the flavin molecule, FMN.

Analytical ultracentrifugation (AUC) and Nuclear magnetic resonance (NMR) spectroscopy were employed to investigate the solution behaviour of DR_1146 in the presence of FMN. AUC results uncovered a monomer-dimer equilibrium; with DR_1146 self-associating to form a dimer at a concentration of 7.67 μM. NMR

spectroscopy depicted that global changes occur within the structure of DR_1146 on binding to FMN. The high quality of spectra obtained showed potential for 3-D structure determination by NMR if ordered crystals could not be obtained for X-ray diffraction. Interestingly, analysis of NMR spectra proved to be integral to identifying a homogenous sample for successful crystallisation of DR_1146. By monitoring chemical shifts it was possible to determine the time needed for degradation of DR_1146 to cease, and the amount of FMN needed to ensure saturation of binding sites. From this particular sample, a stable 28 kDa fragment was isolated by gel filtration. Automated sitting-drop vapour-diffusion experiments resulted in the growth of yellow DR_1146-FMN crystals for which, although poor in quality, X-ray diffraction was obtained. Overall this study reflects the importance and advantage of incorporating information gained from biophysical characterisation into the strategies employed for successful protein crystallisation.

The characterisation of DR_1146 as a flavoprotein points towards a possible role in electron transfer due to the extensive redox capacity of flavin. This could implicate the protein in the production of damaging reactive oxygen species (ROS) as a result of irradiation, contributing to oxidative stress levels. Alternatively, if DR_1146 is identified as a FMN-binding pyridoxine 5′-phosphate oxidase (PNPOx) enzyme, as sequence homology suggests, it could play a role in detoxification and stress response through production of pyridoxal 5′-phosphate (PLP), a known scavenger of ROS. Only further characterisation and elucidation of a 3-D structure would confirm or dispel these functional hypotheses and ultimately provide a greater understanding of how *D. radiodurans* is able to deal with such oxidising conditions.

Simultaneously, experiments were carried out on other soluble and membrane protein targets from *D. radiodurans* and their corresponding homologues from *Streptococcus pneumoniae* (TIGR4). The aim of comparable studies was to identify key structural or functional differences between the two Gram-positive bacterial strains. Identification of features unique to *D. radiodurans*, but unconserved in *S. pneumoniae*, could contribute to further understanding of bacterial radioresistance.

SP_1651 is a thiol peroxidase which forms part of the Mn-ABC transport system in *S. pneumonia*. Its homologue from *D. radiodurans*, DR_2242 is a putative thiol-specific antioxidant protein, the structure of which has been solved by Dr. Dave Hall as part of the ESRF's structural genomics project (unpublished). The aim of this part of the project was to elucidate the structure of SP_1651 so that a comparison with DR_2242 could be made.

The *sp_1651* gene (*psaD*) was successfully expressed and purified to homogeneity by IMAC and gel filtration. After the proteolytic removal of a 6xHis tag, the purified protein was crystallised by sitting-drop vapour-diffusion. Preliminary diffraction with a resolution limit of 3.2 Å was obtained, however data showed high mosaic spread. Unfortunately, attempts to reproduce initial crystals failed and hence, structural comparisons with DR_2242 could not be made.

DR_0463 is a 108 kDa maltooligosyltrehalose synthase (MTSase) which has been shown to catalyse the breakdown of maltooligosaccharide (or starch) into the disaccharide, trehalose. The full length gene was expressed in BL21(DE3)pLysS cells, producing large yields of insoluble target protein. DR_0463 was solubilised with 8 M Urea and then purified by IMAC in the presence of the denaturant. The low affinity of DR_0463 for the Ni^{2+} matrix of the HisTrap column proved to be problematic when trying to obtain homogeneity. However, by sequentially repeating IMAC purification up to three times with the same protein sample, a large proportion of impurities were removed.

SP_1648 (PsaB) is an ATP-binding protein that forms part of the Mn-ATP transport system in *S. pneumoniae* and its homologue from *D. radiodurans*, DR_2284 is predicted to share similar function. Purification of soluble SP_1648, expressed in B834(DE3) cells, was complicated by an inability to bind the protein to the column matrix for IMAC. In the case of DR_2284, expression trials yielded only a minute amount of insoluble protein in BL21-AI competent cells. The bottlenecks in early expression and purification stages provided valuable experience in dealing with problematic proteins.

As an introduction to molecular cloning, two genes predicted to encode integral membrane proteins from *D. radiodurans*, were cloned for preliminary expression trials. This work was carried out at the ESRF and contributed to an extension of

the structural genomics project, to incorporate membrane protein targets from *D. radiodurans*. Full length forms of the genes thought to encode an undecaprenyl diphosphatase (UDP) and a diacylglycerol kinase (DGKA) were successfully cloned in to pET-28b, with incorporation of separate N- and C-terminal 6xHis tags.

Acknowledgements

I would like to thank my supervisors Dr. Andy Freer (University of Glasgow) and Dr. Dave Hall (previously at the ESRF, Grenoble, France) for their continued guidance, support and encouragement throughout this project. In addition I would like to thank members of the Protein Crystallography Group (Glasgow), in particular Dr. Alan Riboldi-Tunnicliffe for kindly providing cloned constructs from *S. pneumoniae*.

Thanks are due to members of the Macromolecular Crystallography Group (Grenoble) for their help and kindness during my time spent in France. A special thanks to Dr. Rana Roy for his guidance in the molecular cloning of membrane proteins from *D. radiodurans*. Initial clones of *dr_1146* and *dr_0463* in pDEST17 vectors were kindly provided by Protein'eXpert, Grenoble, France.

I would especially like to thank my collaborators at the University of Glasgow, involved in the biophysical characterisation of DR_1146. ITC, DSC and FS experiments and data analysis were performed by Margaret Nutley and Prof. Alan Cooper, BBSRC/EPSRC Biological Microcalorimetry Facility. CD experiments and data analysis were carried out by Dr. Sharon Kelly, Protein and Nucleic Acid Characterisation Facility. AUC experiments and data analysis were carried out in collaboration with Dr. Olwyn Byron, Faculty of Biomedical and Life Sciences, Division of Infection and Immunity and NMR spectroscopy was performed and analysed by Dr. Brian Smith, Faculty of Biomedical and Life Sciences, Structural Biology.

MALDI-TOF and ESI-MS were performed by Kenneth Beattie, Finger Prints Proteomics, University of Dundee. Edman degradation was carried out by David Campbell, also at the University of Dundee. Small molecule ESI-MS was carried out by Jim Tweedie, University of Glasgow.

I am extremely grateful to my family, especially my parents, Catherine and John Carr, and Colin Waterston, for their constant love and support during this project and before. I would also like to say a heartfelt thanks to Mark for his love, patience and encouragement over the past four years.

Abbreviations

ABC	Adenosine triphosphate binding cassette
AMP	Adenosine monophosphate
mAU	Milli-absorbance units
arb	Arbitrary units
ATP	Adenosine triphosphate
AUC	Analytical ultracentifugation
bacA	Gene encoding UDP in *D.radiodurans*
BLAST	Basic Local Alignment Search Tool
BME	β-mercaptoethanol
CD	Circular dichroism
clvd	Protein sample with the 6xHis tag removed
CMR	Comprehensive Microbial Resource
CV	Column volume
Da	Dalton
DGKA	Diacylglycerol kinase from *D. radiodurans*
dgkA	Gene encoding DGKA in *D. radiodurans*
DLS	Dynamic light scattering
DNA	Deoxy-ribonucleic acid
dNTP	Deoxy-ribonucleotide-triphosphate
DR_0463*	Maltooligosyltrehalose synthase from *D. radiodurans*
DR_1146*	Putative general stress protein from *D. radiodurans*
DR_2242*	Putative ATP-binding protein (Mn-ABC transporter) from *D.radiodurans*
D. radiodurans	*Deinococcus radiodurans* (Strain R1)
DSBs	Double strand breaks
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray ionisation mass spectroscopy
ESDSA	Extended synthesis-dependent strand annealing
ESRF	European Synchrotron Radiation Facility
FAD	Flavin adenine dinucleotide (oxidised)
FMN	Flavin mononuclotide (Riboflavin 5'-phosphate) (oxidised)
$FMNH^{\cdot}$	Flavin mononucleotide semiquinone intermediate
$FMNH_2$	Flavin mononuclotide (reduced)
FS	Fluorescence spectroscopy

GuHCl	Guanidine Hydrochloride
HRV C3	Human rhinoviral 3C protease
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-ß-D-thiogalactoside
ITC	Isothermal titration calorimetry
LB	Luria-Bertani media
LDS	Lithium dodecyl sulphate
MALDI-TOF	Matrix assisted laser desorption/ionisation time-of-flight
MES	2-(N-morpholino) ethane sulphonic acid
MTHase	maltooligosyltrehalose trehalohyrdolase from *D. radiodurans*
MTSase	Maltooligosyltrehalose synthase from *D. radiodurans*
MW	Molecular weight
MWCO	Molecular weight cut off
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PCT	Pre-crystallization Test
PDB	Protein Data Bank
PDB ID	Protein Data Bank Identification
PEG	Polyethylene glycol
PLP	Pyridoxal 5'-phospate
PNP	Pyridoxine 5'-phosphate
PNPOx	Pyridoxine 5'-phosphate oxidase
PMP	Pyridoxamine 5'-phosphate
PMPOx	Pyridoxamine 5'-phosphate oxidase
rmsd	Root mean square deviation
ROS	reactive oxygen species
sd	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SE	Sedimentation equilibrium
S. oneidensis	*Shewanella oneidensis*
SV	Sedimentation velocity
treY	Gene encoding MTSase in *D. radiodurans*
T_DR_1146*	DR_1146 with N-terminal truncation
c_t_dr_1146	Gene encoding DR_1146 with N- and C- terminal truncation
SP_1648 (PsaB)*	ATP-binding protein (Mn-ABC transporter) from *S. pneumoniae*
SP_1651 (PsaD)*	Thiol peroxidase from *S. pneumoniae*
S. pneumoniae	*Streptococcus pneumoniae* (Strain TIGR4)

UDP	Undecaprenyl diphosphatase from *D. radiodurans*

*Unless otherwise stated, all gene symbols are lower case italic replicas of the corresponding proteins locus number, for example, *dr_1146* is the gene which encodes the protein DR_1146 in *D. radiodurans* and the gene *psaD* encodes the protein PsaD.

List of Contents

1 Introduction

1.1 *Deinococcus radiodurans* R1

1.1.1 Structural genomics

Deinococcus radiodurans (Strain R1) is a non-pathogenic, red pigmented, gram-positive bacterium [1] with an exceptional ability to survive acute exposure to oxidising, DNA-damaging conditions such as ionising radiation, desiccation, UV radiation and cellular H_2O_2 [2].

This study is part of a structural genomics project underway at the European Synchrotron Radiation Facility (ESRF), Grenoble, which aims to elucidate the 3-D crystal structures of proteins encoded from the genome of *D. radiodurans* in order to gain further understanding of the mechanisms employed by this extremophillic bacterium. The essence of this project is to progress from the gene sequence through to structure *via* production, purification and crystallisation of the target protein. The challenge then is to assign biochemical and biological function of that protein within the organism [3]. It is the hope that structures of target proteins will reveal functions relevant to the defence of *D. radiodurans* against mutation and cell death that will provide basis for radiation recovery targets.

The publication of the complete genome sequence of *D. radiodurans* R1 [4, 5], which is annotated on the Comprehensive Microbial Resource (CMR) database [6, 7], has enabled the ESRF's structural genomic initiative to take place. It is reported that the 3.28 Mb chromosome of *D. radiodurans* is segmented into a 2.64-Mb chromosome (chromosome I), a 0.41-Mb chromosome (chromosome II), a 0.18-Mb megaplasmid and a 0.045-Mb plasmid [8]. Each cell has between 4 to 10 genome copies, depending on the phase of bacterial growth [9] and the chromosome has an unusual ring-like structure [10].

All of the gene targets investigated in this study belong to chromosome 1 of the *D. radiodurans* genome [6, 7].

1.1.2 General characteristics and morphology

D. radiodurans was discovered in 1956 by A. W. Anderson whilst trying to find a method of sterilising food using γ-radiation. This novel vegetative bacterium, found on samples of ground meat, survived exposure to 4000 Gray (Gy) of irradiation [1], a dose hundreds of times higher than that deemed fatal for most other organisms.

Since its initial discovery, *D. radiodurans* has been isolated from contrasting locations worldwide including nutrient rich environments such as soil and animal faeces, in addition to nutrient-depleted environments including dry Antarctic valleys and arid desert soil [5, 11-13].

D. radiodurans belongs to the Deinococcaceae family which is comprised of 11 members, all with the genus name *Deinococcus* [14-16]. All of the Deinococci are non-motile, do not form spores and are mostly spherical, existing as single cells, pairs or tetrads. Figure 1.1 illustrates the tetrad morphology of *D. radiodurans*.

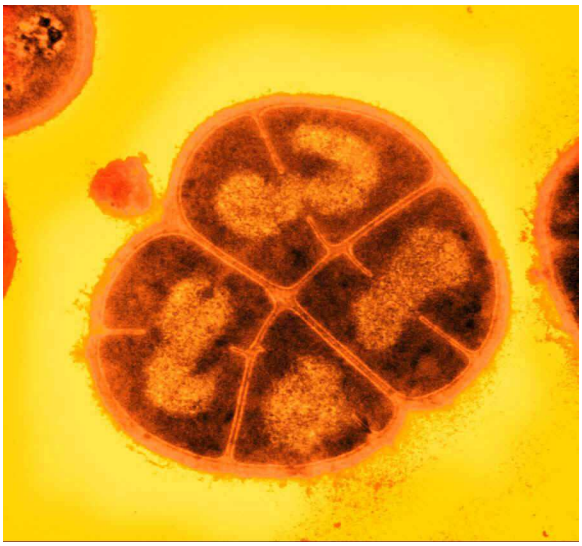


Figure 1.1: Electron micrographic image of *D. radiodurans* displaying tetrad morphology. Taken from the web page of M.J. Daly (http://www.usuhs.mil/pat/deinococcus/index_20.htm) [17].

In contrast to its thermophilic relatives, *D. geothermalis* and *D. murrayi* [18], *D. radiodurans* is a mesophile which exhibits optimum cellular growth between 30-37°C [19].

The origin of this fascinating species and its evolution to possess such abnormal levels of radioresistance remains unknown although many speculative theories have been proposed. Pavlov *et al*. (2006) [20] postulated that *D. radiodurans* could be of extra-terrestrial origin, transported through space to earth by debris from Mars. Although exciting, this arguably eccentric theory is unlikely, as the species shows genetic and biochemical characteristics similar to terrestrial bacteria.

1.1.3 Repair of cellular damage caused by oxidising conditions

Like other microorganisms, the genome of *D. radiodurans* is shattered into hundreds of short DNA fragments on high exposure to oxidising conditions. However, researchers claim that it is the unique ability of this bacterium to accurately reassemble into a functional genome, which differentiates it from others [2, 21, 22].

The most severe form of genomic damage [23] double strand breaks (DSBs) occur as a result of oxidative stress [22]. This condition is caused by a build up of reactive oxygen species (ROS) to levels which exceed the capacity of the scavengers needed to neutralise them; leaving the cell susceptible to damage [24, 25]. DNA damaging hydroxyl radicals (OH·) are generated by radiolysis of water by direct γ-irradiation and by energy metabolism during recovery from radiation [24, 25].

ROS cause damaging DSBs in the DNA of both *E. coli* and *D. radiodurans* at comparable rates [10, 26]. However, *D. radiodurans* is able to tolerate 30 times more DSBs before cell death compared to *E. coli* [23, 27]. In fact, DR shows resistance up to 5000 Gy without loss of viability or cell death [22, 28]. This confirms that DR is not able to prevent cellular damage occurring, but that it has a novel method of preventing the lethal effects of DSBs by effectively repairing the damage [19].

Although extensively researched, the mechanisms by which *D. radiodurans* is able to avoid cell death after damage caused by oxidising conditions are not wholly understood. It is thought that such extremophillic behaviour can be attributed to a combination of non-enzymatic and enzymatic defences which

allow conventional pathways, present in other bacteria, to function more efficiently [10]. New methodologies are regularly developed and dismissed as the search for the true pathways continue. Two recently proposed mechanisms include: (a) an enzymatic-dependent DNA repair system involving a two-stage process for genomic reassembly [29] and (b) non-enzymatic protection from protein oxidation by detoxification of ROS by cellular Mn(II) [10, 30]. In addition, the genome of DR has unusually high numbers of detoxification genes, such as superoxide dismutases (SODs) and catalases [31, 32], whose levels of expression are greatly increased after irradiation [33].

1.1.3.1 DNA repair by extended synthesis-dependent strand annealing (ESDSA)

A study by Zahradka *et al*. (2006) [29] claims to exclude previously identified mechanisms of genome reassembly in favour of a two-stage DNA repair pathway consisting of extended synthesis-dependent strand annealing (ESDSA) followed and completed by crossovers. To allow this process to occur effectively, two copies of the genome with random DNA breakages are required. EDSA involves a nucleotide-rich polymerase chain reaction (PCR), using chromosomal fragments (with overlapping homologies) as both primers and template DNA. This polymerase I-dependent amplification reaction produces new complementary single-stranded extensions that then anneal with high precision to form double stranded intermediates. These intermediates then undergo homologous recombination into circular chromosomes, facilitated by the protein RecA. The final result is a ‘double-stranded patchwork’ of DNA blocks synthesised before radiation, connected by DNA blocks synthesised after radiation [29].

1.1.3.2 Protection from protein oxidation by high cellular Mn(II) levels

Research carried out by Daly *et al*. (2007) [30] shifts focus away from DNA repair mechanisms by relating the high radioresistance of *D. radiodurans* with its ability to prevent radiation induced protein damage. It has been proposed that the capacity to protect cellular proteins from oxidation is equally or more critically important for maintaining viability than DNA repair. This hypothesis could explain why other organisms, with similar repair genes to *D. radiodurans*, are

susceptible to apoptosis on low levels of radiation and low instances of DSBs [30, 34, 35].

Protein damage is caused by oxidation of amino acid side chains by ROS and H_2O_2, which are generated during irradiation, desiccation [22] and energy metabolism [24]. The extent of protein damage is established by the level of irreversible and irreparable carbonyl group production, which occurs as a result of amino acid oxidation [36, 37]. Resistant bacteria, like Gram-positive *D. radiodurans*, show lower levels of protein oxidation in comparison with radiation sensitive bacteria like Gram-negative *Shewanella oneidensis*.

Such resistance to oxidative radiation and desiccation has been quantifiably liked with high levels of Mn(II) and low levels Fe(II) in the cell [30]. For example, *D. radiodurans* has a relatively high Mn(II)/Fe(II) concentration ratio of 0.24 and is susceptible to 10 % cell death on exposure to 10-12 kGy of ionising radiation and more than 30 days of desiccation [10, 38]. In contrast, highly sensitive *S. oneidensis* has a much lower Mn(II)/Fe(II) ratio of less than 0.001 and undergoes 10 % cell death with only 0.07 kGy of ionising radiation and 1 day of desiccation [10, 38]. This suggests that the potential antioxidant capability of Mn(II) could play a role in protein protection. The link between radiosensitivity and high Fe(II) levels is thought to lie in the Fe(II)-dependence of damaging ROS produced during metabolic recovery. As Fe(II) is released from proteins as a result of irradiation, Fe(II)-rich bacteria are particularly susceptible to Fe(II)-induced oxidative stress [38].

Although the mechanism of protein protection remains unknown, it has been proposed that a non-enzymatic Mn(II)-based protection system, existent in radioresistant bacteria like *D. radiodurans*, facilitates survival by scavenging Fe(II)-dependent ROS [10, 39]. While Mn(II) has no influence on DNA damaging OH· (primary product of radiolysis), it is able to scavenge endogenously induced peroxyl radicals which may minimise protein damage by preventing propagation of secondary reactions that ultimately damage proteins [24, 40, 41]. It is thought that this may protect other enzymatic detoxifying and repair systems allowing them to survive and function with greater efficiency [30, 38]. The impact of Mn(II) in the stress response of *D. radiodurans* is corroborated by an observed

increase in radiation sensitivity and protein oxidation on restriction of Mn(II) during growth [10, 26, 42].

Interestingly, there is also a positive correlation between high cellular Mn(II)/Fe(II) concentration ratios and capacity to survive desiccation [10, 38]. This adds to the debate that mechanisms contributing to radioresistance and control of oxidative stress may be an evolutionary consequence of the robustness needed to survive desiccation [22, 43].

1.1.4 Current and proposed applications

It is thought that a more complete understanding of the repair and protection systems employed by *D. radiodurans* may lead to the discovery of novel mechanisms that could aid understanding of responses in higher organisms providing an insight that could be utilised clinically. For example, by understanding how Mn(II)/Fe(II) levels contribute to radiation resistance, approaches could be developed to control recovery from radiation injury [44]. In addition, further understanding of the ability of *D. radiodurans* to regain integrity by supply of water and ions, after extensive desiccation [45], could inspire research into anti-ageing and regenerative medicine [29].

A current application of DR has been investigated by the US Department of Energy, who have funded work into using genetically modified forms of the bacterium as a bioremediation agent, in an effort to address the problem of highly toxic radioactive nuclear waste sites which are leaking into the surrounding areas [46]. Such sites contaminated with radionucleotides, heavy metals and harmful solvents [47], would require billions of dollars to clean up by physiochemical methods [48] and thus bioremediation technologies are highly favourable to avoid such expense. Many of the experiments carried out with other possible bacterial [49-52] and plant candidates [53, 54], have failed due to their sensitivity to ionising radiation. As a solvent tolerant, soil bacterium, which can grow and express foreign genes in the presence of radiation levels typical of toxic waste sites [47, 55], *D. radiodurans* has proved to be a good target for such studies. Researchers have generated strains of *D. radiodurans* that can express a cloned mercury resistant gene *mer*A [56] which could be used to reduce toxic Hg (II) to less toxic, more volatile, elemental mercury in problematic areas [57].

1.1.5 Putative general stress protein, DR_1146

The primary focus of this study is the expression, purification and crystallisation of an un-characterised gene from *D. radiodurans*, with the locus name *dr_1146*. It is annotated on the CMR database [6, 7] as putative general stress protein 26 and is predicted to exhibit the characteristics listed in Table 1.1. The nucleotide and amino acid sequence of DR_1146 are included in Appendix 1.

Locus name	DR_1146
Putative identification	General stress protein 26, putative
Coordinates	1155760-1156341
Gene length	582 bp
Protein length	193 aa
Molecular weight	21,043.23 Da
PI	4.58
% GC content	61.14
TIGR cellular role category	Cellular processes: adaptations to atypical conditions

Table 1.1: Predicted characteristics of the *dr_1146* gene as annotated on the CMR database [6, 7].

A NCBI BLAST search [58, 59], using the amino acid sequence of DR_1146, identified sequence homology with other bacterial proteins putatively associated with cellular general stress responses. In addition, the search identified close sequence homology (43-58 %) with putative pyridoxamine 5'-phosphate oxidase (PMPOx) related FMN binding proteins, including a protein from *Deinococcus geothermalis*, another member of the Deinococcaceae family.

1.1.5.1 Pyridoxamine 5'-phosphate oxidases

The PNPOx family of enzymes have been shown to catalyse the terminal step in the *de novo* biosynthesis of the cofactor pyridoxal 5'-phosphate (PLP) in microorganisms and plants. They also function as a constituent in the recycling of PLP in higher organisms that require vitamin B6 [60]. Failure to provide adequate supplies of PLP in mammals is usually manifested as neurological disorders [61]. PNPOx enzymes have been identified and characterised in bacteria [62], humans [63], animals [60, 64, 65], yeast [66] and plants [67]. The 3-D crystal structures for human and *E. coli* PNPOx have been determined along with probable mechanistic pathways [61, 68].

PLP, the biochemically active form of Vitamin B6, is known as one of the most versatile cofactors in nature, necessary for over 100 enzymatic reactions, the

majority of which are involved in the metabolism of amino acids [69-71]. Recently, PLP was uncovered as an antioxidant, with potency equivalent to that of vitamins C and E [72-74]. As an effective singlet oxygen quencher [75] it may play a role in *D. radiodurans* cellular response to oxidative stress.

In *E. coli*, PLP has been shown to be produced by the FMN-dependent oxidation of the substrates, PNP (pyridoxine 5'- phosphate) or PMP (pyridoxamine 5'-phosphate) whilst associated to PNPOx [61, 68]. This reaction requires that the FMN cofactor maintains strict structural requirements with only a few alterations in the isoalloxazine ring being permitted [76]. A pair of electrons are transferred from C4' of PNP (or PMP) to the N5' of a tightly bound molecule of FMN, forming $FMNH_2$. This is thought to occur by either direct hydride transfer or by removal of a proton by a basic amino acid to form a C4' carbanion which attacks N5' of FMN. The high potential electrons are then transferred to molecular oxygen, thus regenerating FMN and forming H_2O_2 [77] (Figure 1.2).

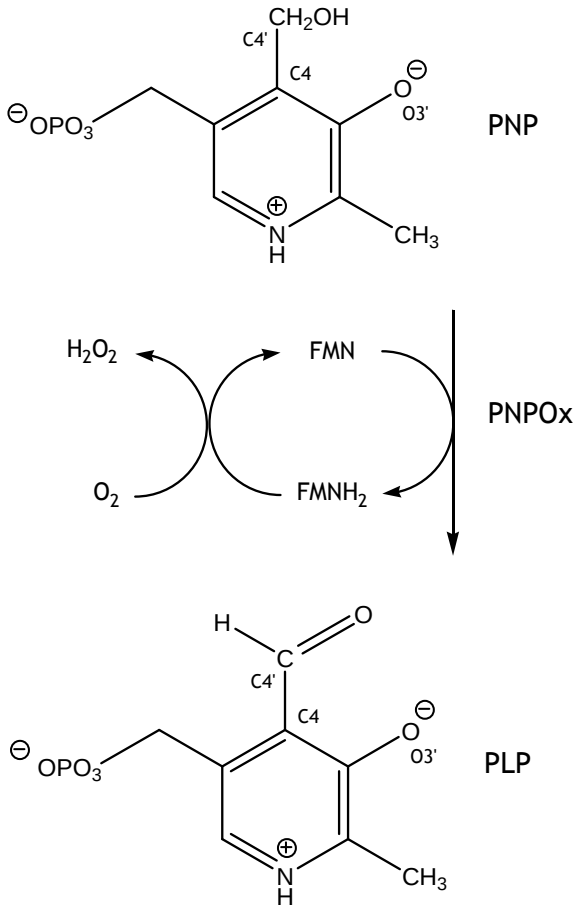


Figure 1.2: Schematic diagram illustrating the involvement of FMN in the PNPOx catalysed conversion of PNP into PLP in *E. coli*. From Safo *et al.* (2000) [61].

In addition to its production by PNPOx, PLP has been shown as an effective reaction inhibitor when complexed with the enzyme. It has greater affinity for the active site than either PNP or PMP substrates suggesting that product inhibition is a mode of regulation for the reaction [60, 76]. Although this reaction results in the production of H_2O_2, it has been proposed that *D. radiodurans* is able to deal with this strong oxidant by exhibiting extremely high catalase activity [10, 78].

The PNPOx enzymes are the smallest members of the flavin-containing oxidase class with molecular weights between 22-29 kDa. They are normally found as a dimer and have highly conserved motifs that include amino acids involved in FMN and PLP binding [61].

1.1.5.2 Putative general stress protein (COG3871) from *Nostic punctiforme*

A search of the Protein Data Bank [79, 80] using the amino acid sequence of DR_1146, found homology (32 % identity) with a structure of an uncharacterised putative general stress protein (COG3871) from *Nostic punctiforme* PCC 73102 (PDB ID: 2I02) [81]. Like PNPOx, this dimeric protein was crystallised in the presence of a single FMN molecule. Interestingly, the crystal structure of this homologue does not include any bound PLP cofactors in contrast to those of the PNPOx enzymes including *E. coli* PNPOx.

The FMN binding ability of the homologues discussed, fuelled investigation into the possible flavin-binding capacity of DR_1146 and the subsequent effects on protein stabilisation, structure and crystallisation.

1.1.6 Flavin chemistry and flavoproteins

Since their discovery in the 1930s, flavins have been labelled one of the most important cofactors involved in enzymatic functions [82-87]. Flavoproteins are ubiquitous proteins that are tightly complexed with one or more flavin prosthetic groups [88]. A large number of flavoproteins have been characterised over the past 60 years, with the researcher Vincent Massey at the forefront of this work [89].

The two most commonly occurring flavins, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) [90], are derivatives of riboflavin, otherwise known as vitamin B1. The structures of FMN and FAD are based upon the heterocyclic isoalloxazine ring system of riboflavin, from which extends either a ribityl phosphate prosthetic group (FMN) or a ribityl adenine diphosphate prosthetic group (FAD) (Figure 1.3) [90].

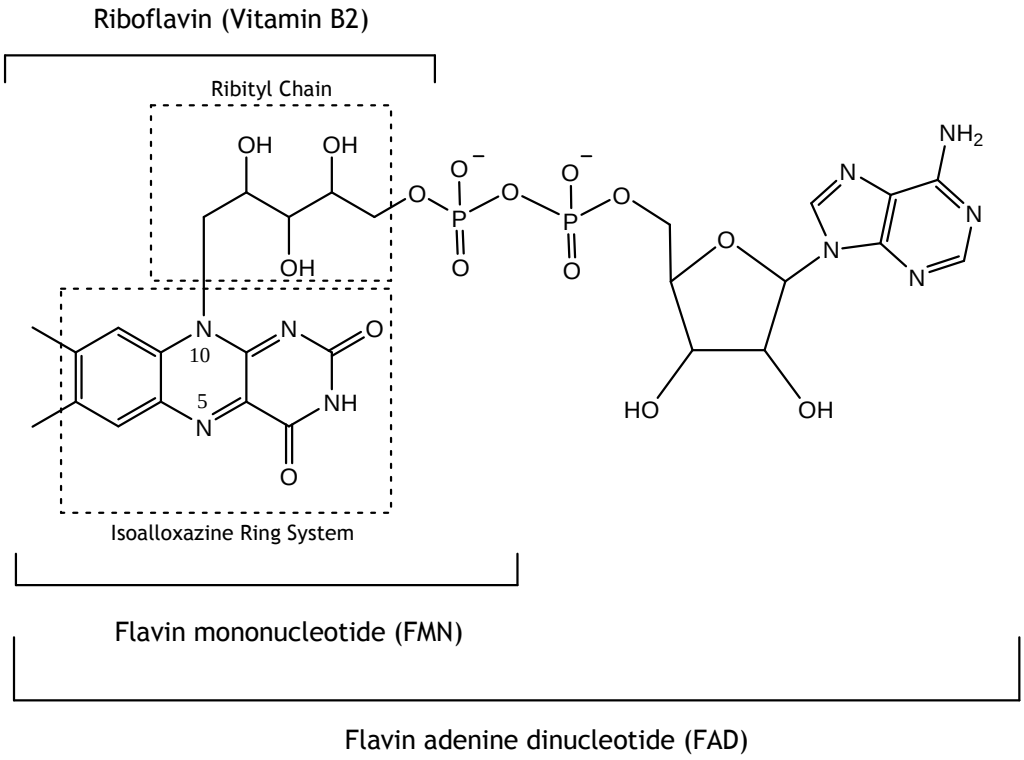


Figure 1.3: Structures of riboflavin, FMN and FAD in their oxidised state [90].

In vivo synthesis occurs *via* ATP-phosphorylation of the ribityl chain of riboflavin to produce riboflavin 5'-phosphate (FMN), followed by the transfer of an additional AMP moiety to form FAD [78]. This two-step process is facilitated by the enzymes riboflavin kinase [91, 92] and FAD synthase [93].

The conjugated isoalloxazine ring component of flavins is accountable for many of their unique physical and chemical properties [85]. It gives rise to a bright yellow colour and a characteristic absorption spectrum consisting of two peaks between 300-500 nm when in an oxidised redox state [94]. The aromatic ring system is a centre for extensive redox chemistry [85], which enables flavin molecules to catalyse a wide variety of biochemical and biological activities when complexed with proteins [95, 96]. Flavins are perhaps best known for their

role in aerobic metabolism [85], however, their activities extend to include cell apoptosis [97, 98], detoxification [99, 100] and light driven DNA repair [101, 102].

The determination of over 40 flavoprotein crystal structures has given insight into the nature of interactions between the flavin cofactor and the protein [85]. In most flavoproteins, the flavin prosthetic group is non-covalently bound. However, there are some reports of the flavin molecule being covalently attached to the polypeptide chain [103]. The majority of flavin-protein intermolecular interactions, independent of nature, are with the ribityl side chain and prosthetic groups of either FMN or FAD [85]. Although many flavoproteins show specificity for either FMN or FAD, some contain both FMN and FAD binding domains [88], which is the case for the enzyme NADPH-cytochrome P450 reductase [104]. Interestingly, characterisation of PNPOx enzymes shows high affinity and specificity for FMN [105], however the literature does not describe or quantify binding affinities for FAD. The affinity of flavoproteins for flavin molecules has been facilitated by the evolution of various binding folds [84, 106, 107]. Structural motifs within these characteristic folds, including the TIM barrel and the Rossmann fold, interact with specific moieties of the flavin prosthetic group [88].

Flavins can exist in three different redox states; an oxidised state, a semiquinone radical intermediate, and a fully reduced state [83]. Changes in their oxidation state are reflected in absorbance spectra as demonstrated in Figure 1.4 which illustrates spectral changes for the flavoprotein, glucose oxidase, on gaining electrons to form a reduced state [85, 108].

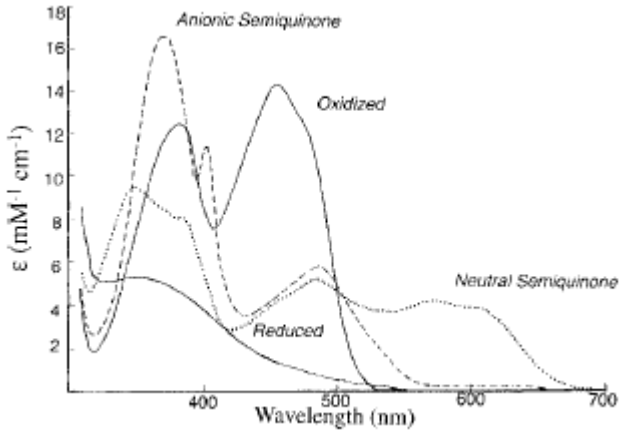


Figure 1.4: Absorbance spectra of the flavoprotein, glucose oxidase, in the oxidised, semiquinone (neutral and ionic) and fully reduced states. Figure taken from Massey 2000 [85], based on data from Massey 1996 [108].

The versatility of flavoproteins is shown by the capability to undergo both one- and two-electron transfer processes [85]. This is demonstrated by the role of the FMN in the multi-redox oxidation of NADH during cellular respiration [78] (Figure 1.5).

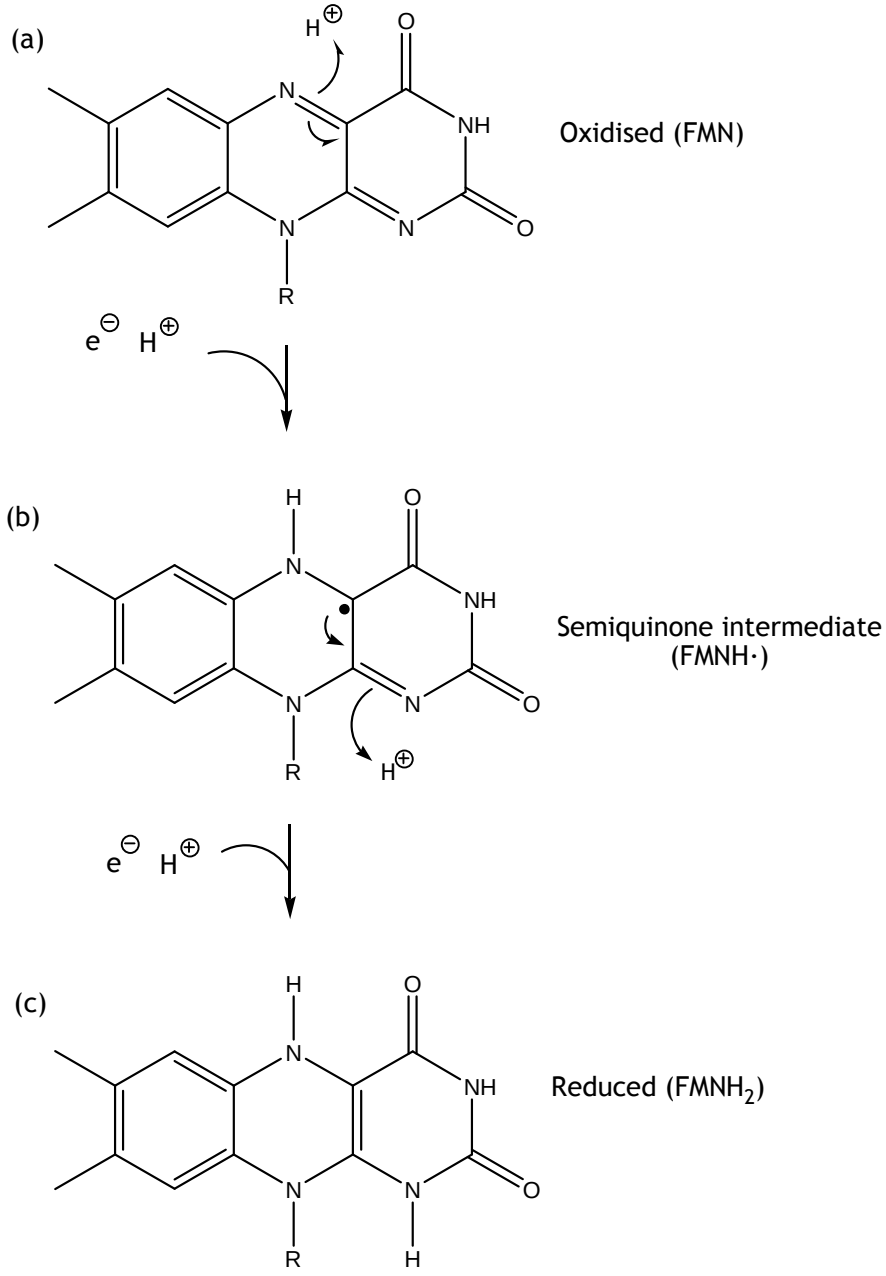


Figure 1.5: FMN facilitated single-electron transfer for the oxidation of NADH in the respiratory chain [78].

During this process, NADH binds to a dehydrogenase, which in turn has FMN bound. Two high potential electrons are transferred from NADH to FMN to give reduced $FMNH_2$ (Figure 1.5a-c). This two-electron dehydrogenation, involves the formation of a semiquinone radical intermediate (FMNH·) (Figure 1.5b). The electrons are then transferred, *via* a series of single-electron steps, to iron-sulfur clusters; a process by which the reduced $FMNH_2$ is re-oxidised to FMN [85].

Reduced flavins show high reactivity with molecular oxygen, upon which a one-electron reaction takes place. This results in the reduction of O_2, yielding a

flavin radical and superoxide [85]. Subsequently, these two radicals become involved in several chemical pathways [109-113] which include the formation of hydroxyl radicals through reaction of superoxide with peroxide [114]. The production of hydroxyl radicals is commonly believed to be one of the major sources of DNA and protein damage [24, 25, 85].

It has been proposed that there is a relationship between the amount of ROS-producing intracellular flavins and the proclivity of cells to generate oxidative stress. This in turn, leads to a correlation between the number of flavins and the ability to survive radiation and other oxidising conditions [115]. For example, radioresistant *D. radiodurans* encodes only 39 flavoproteins [5, 10] in comparison with the radiosensitive bacterium *S. oneidensis,* which encodes 59 [10, 116]. Also, the rates at which different flavins react with oxygen to produce ROS, vary over several orders of magnitude depending on the extent of flavin exposure [38].

The ability of DR_1146 to bind flavins could implicate this protein in the production of metabolic ROS as a result of irradiation. Characterisation and elucidation of a 3-D structure may contribute to a greater understanding of how *D. radiodurans* is able to deal with such oxidising conditions in order to maintain viability and prevent cell death.

Alternatively, if DR_1146 is identified as a PNPOx enzyme, as sequence homology suggests, it may play a role in stress response through scavenging of superoxide by its reaction product PLP. Such action may prevent the production of cell damaging hydroxyl radicals and thus contribute to the radioresistance of *Deinococcus radiodurans*.

1.1.7 DR_1146 project aims

The primary focus of this study was to express and purify DR_1146 in order to gain diffraction quality crystals for 3-D structure determination using crystallographic methods. Elucidation of the structure would enable comparison of fold [117, 118] and structural motifs [119] with structural databases, which could reveal similarities, from which information about biochemical and biological function could be inferred [3].

It is well documented that crystallisation is a major bottleneck in X-ray crystallographic structure determination [120, 121] and that there are no systematic methods to ensure that ordered 3-D crystals will be obtained.

It is however, believed that a critical factor for success is the homogeneity, monodispersity and stability of the protein preparation itself [120]. This study utilises biophysical methods of characterisation to aid protein preparation and as a result produce protein crystals that are well ordered and will subsequently generate a high resolution structure. In particular, the capacity of DR_1146 to bind flavin molecules is investigated. Structural information gained from protein-ligand complexes are often very useful in gaining functional insight as they show the nature of the ligand, where and how it is bound and which residues form the active site from which a catalytic mechanism could be postulated [3]. In the absence of a structure, information about cofactor binding can aid efforts to grow crystals as saturation of protein binding sites may produce a homogenous sample favourable for crystallisation.

1.1.8 Maltooligosyltrehalose synthase (DR_0463)

D. radiodurans encodes a large 108 kDa enzyme known as maltooligosyltrehalose synthase (MTSase) from the gene *treY* (locus name *dr_0463*) [6, 7]. The putative characteristics of the gene and protein, as annotated on the CMR database, are summarised in Table 1.2.

Locus name	DR_0463
Putative identification	Maltooligosyltrehalose synthase
Gene symbol	Trey
Coordinates	461679-458743
Gene length	2937 bp
Protein length	978 aa
Molecular weight	108,738.59 Da
pI	6.34
% GC content	68.75 %
TIGR cellular role category	Cell envelope: Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides

Table 1.2: Predicted characteristics of the *treY* gene as annotated on the TIGR CMR database [6, 7].

MTSase forms one-half of the two-step metabolism of trehalose in *D. radiodurans*. Together with its partner protein, maltooligosyltrehalose

trehalohyrdolase (MTHase), they catalyse the breakdown of maltooligosaccharide (or starch) into trehalose [122].

Trehalose, or α-D-glucopyranosyl- 1,1-α-D-glucopyranose, is a non-reducing disaccharide which contains an α-1,1-linkage between two glucose molecules. The sugar has been shown to protect cell membranes and proteins against oxidative stresses including drought, heat and osmotic changes [123-126] by serving as a carbohydrate reserve, water substitute [127] and a free radical scavenger [128]. It can therefore be postulated that for *D. radiodurans*, trehalose may facilitate protection or recovery from the desiccating conditions of arid dessert environments. In addition, trehalose has many industrial applications including use as a sweetener, preservative or stabiliser in food, cosmetics, vaccines and medicine [129].

The process by which MTSase and MTHase produce trehalose is summarised in Figure 1.6 [122].

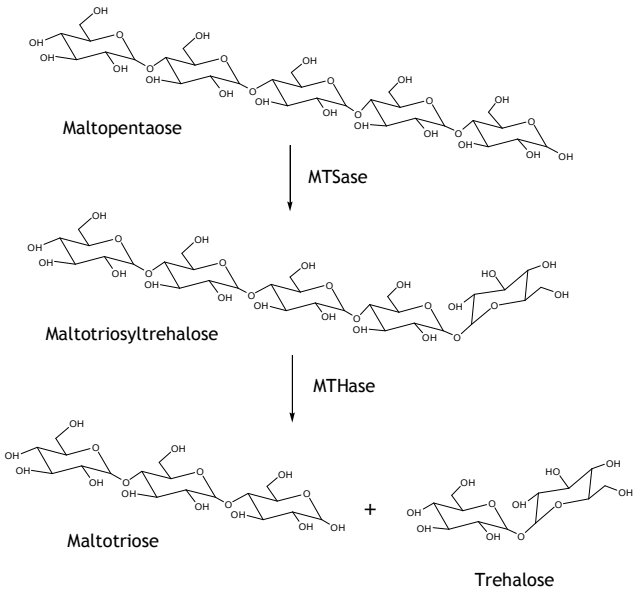


Figure 1.6: Two-step breakdown of maltopentaose into trehalose and maltotriose catalysed by MTSase and MTHase [122].

Firstly, MTSase converts the glycosidic bond, between the last two glucose moieties of starch, from an α-1,4-linkage to an α-1,1-linkage, thus producing a glycosyl trehaloside. MTHase then cleaves the α-1,4-linkage adjacent to the α-1,1- glycosidic bond to release free trehalose and maltotriose as a side product [122]. Solution of the 3-D structure of MTHase from *D. radiodurans* has given

some insight into the mechanism by which the second part of the enzymatic reaction occurs including the involvement of active site amino acids. Elucidation of the structure of MTSase would complete understanding and thus may uncover a vital role in stress response through trehalose production.

This study reports the expression of *treY* in the insoluble fraction of *E. coli* cells and the subsequent solubilisation and purification of MTSase by immobilised metal affinity chromatography (IMAC).

1.2 *Streptococcus pneumoniae* TIGR4

Like *D. radiodurans*, *Streptococcus pneumoniae* is a Gram-positive, non-motile bacterium [130] which exhibits some resistance to oxidative stress [131-133]. However, where *D. radiodurans* is a relatively harmless microorganism, *S. pneumoniae* has earned a reputation as a major human pathogen associated with life threatening diseases such as meningitis and pneumonia [134, 135]. Pneumococcal infection is most prevalent in young, elderly and immunocompromised populations [136-138], with high mortality rates maintained by acquired resistance to widely used antibiotics, including penicillin [139]. In 1997, nearly half of the elderly population of the United States were immunised with a 23-valent polysaccharide vaccine [140]. Although the efficiency of this vaccine has been extensively proven [141, 142], over 60,000 cases of pneumococcal disease continue to be diagnosed every year in the US [143]. Hence, the search continues to find novel vaccine candidates and to understand fully, the pathogenic mechanisms employed by *S. pneumonia*.

1.2.1 Manganese ABC transport system in S. pneumoniae

S. pneumoniae has been shown to encode an ABC-type transport system from the gene locus, *psaBCA*. There is strong evidence to suggest that this system facilitates the transfer of divalent manganese across the cytoplasmic membrane [144]. The ABC transport complex in *S. pneumonia* consists of three protein components: PsaA, an extracellular solute-binding protein, PsaC, a hydrophobic transmembrane protein responsible for solute transport [145], and PsaB, a cytoplasmic ATP-binding protein. Immediately downstream of the *psaBCA* operon

is *psaD*, a gene which encodes a periplasmic thiol peroxidase protein [144]. Although not strictly part of the Mn(II) transport system, *psaD* is thought to be cotranscribed by the *psaBCA* Promoter [139].

The most extensively studied component of the pneumococcal Mn(II)-ABC-permease is the surface protein PsaA which has been identified as a potential drug target and vaccine candidate [134]. The solute-binding protein has been shown to play an indirect role in pathogenesis as $psaA^-$ mutants are unable to produce important virulence factors like adhesins. It is proposed that reduced levels of Mn(II) importation in the absence of PsaA could prevent the manganese regulated expression of proteins associated with virulence [134].

As discussed earlier, the essential role of Mn(II) in detoxification and stress response in *D. radiodurans* has been reported in depth [30, 38]. It is hence not surprising that compromised Mn(II) influx has also been associated with oxidative stress and growth retardation in *S. pneumoniae* [132, 139]. Individual $psaA^-$, $psaD^-$, $psaB^-$ and $psaC^-$ pneumococcal mutants have shown requirement for added manganese to maintain cell growth rates. In addition, $psaA^-$ and $psaD^-$ mutants show significant susceptibility to oxidative conditions in comparison to wild type strains [131]. This has lead to the postulation that regulation of Mn(II) transport by PsaA and PsaD in *S. pneumoniae* could influence the response of this bacterium to oxidative stress [131].

The 3-D crystal structure of PsaA from *S. pneumoniae* presents major differences in domain architecture and folding topology on comparison with Gram-negative ABC-type proteins [146]. This could infer a novel mode of Mn(II) uptake and release in Gram-positive bacteria like *S. pneumonia* and *D. radiodurans*. Hence, structural elucidation and comparison of the other ABC-components from the two bacteria could contribute to further understanding of the role of manganese transport in pathogenesis and response to oxidative stress.

1.2.2 Putative thiol peroxidase, SP_1651

This study presents the conditions for recombinant expression of *psaD* from *S. pneumonia* (Strain TIGR4), and reports subsequent purification and crystallisation of the encoded protein SP_1651. The initial aim of experiments

was to solve the 3-D crystal structure of SP_1651 so that a comparison could be made with a homologue from *D. radiodurans*. DR_2242 is a putative thiol-specific antioxidant protein, which exhibits 26 % sequence homology with SP_1651. The structure of DR_2242 has been solved by Dr. Dave Hall as part of the ESRF's structural genomics project but remains unpublished and uncharacterised to date. Although Mn(II) is known to play a pivotal role in preventing oxidative protein damage in *D. radiodurans* [10, 30, 38], research into the structure of the Mn-ABC transport system remains undocumented. Hence, structural comparison with SP_1651 could provide information about the biochemical function of DR_2242, and perhaps aid understanding of the importance of Mn(II)-ABC type transfer in *D. radiodurans*.

1.2.3 ATP-binding protein involved in Mn(II) transport, SP_1648

This study describes the recombinant expression of *psaB* and purification of the encoded protein SP_1648 from *S. pneumoniae*. The homologue from *D. radiodurans* (DR_2284) shares 34 % identity with SP_1648 and hence is thought to display similar function through ATP-binding. Again, investigations were carried out in the hope of producing purified protein for comparable studies of SP_1648 and DR_2284.

2 Materials and methods

Chemicals, biochemicals and materials of the highest quality were purchased from Sigma-Aldrich Company, Gillingham, UK or Fisher Scientific, Loughborough, UK unless stated otherwise.

General techniques were carried out as described in Sambrook *et al*. (1989) [147] unless stated otherwise.

2.1 Provided constructs for gene expression

2.1.1 Full length dr_1146 gene in pDEST17

A clone of the full length *dr_1146* gene in a pDEST™17 vector (Table 2.1) was kindly provided by Protein'eXpert, Grenoble, France. This plasmid DNA was used in the PCR step of the cloning of a full length and an N-terminal truncated form of the *dr_1146* gene into expression vector pET-15b.

Vector	Key features	Manufacturer & References
pDEST™17	T7 promoter for expression in *E. coli*. Ampicillin/carbenicillin resistance. Non-cleavable 6xHis tag.	Invitrogen™ Gateway® Cloning Rosenberg *et al.*, 1987 [148] Studier & Moffat, 1986 [149] Studier *et al.*, 1990 [150]
POPINF	T7 promoter for expression in *E. coli*. Ampicillin/carbenicillin resistance. HRV-3C protease cleavable N-terminal 6xHis tag.	Developed by Oxford Protein Production Facility using In-Fusion™ enzyme (Clontech)
pET-15b	T7 promoter for expression in *E. coli*. Ampicillin/carbenicillin resistance. Thrombin cleavable N-terminal 6xHis tag GSH – remaining residues after cleavage.	Novagen® (Merck Biosciences) Rosenberg *et al.*, 1987 [148] Studier & Moffat, 1986 [149] Studier *et al.*, 1990 [150]
PET151/D-TOPO®	T7 promoter for expression in *E. coli*. TEV cleavable N-terminal 6xHis tag. Ampicillin resistance.	Rosenberg *et al.*, 1987 [148] Studier & Moffat, 1986 [149] Studier *et al.*, 1990 [150]
pcR®-Blunt II-TOPO®	Linearised cloning vector. Kanamycin resistance. Direct insertion of blunt ended PCR products.	Invitrogen TOPO® cloning Shuman, 1991 [151]
PLysS	Encodes T7 lysozyme. Chloramphenicol resistance.	Grodberg & Dunn, 1988[152] Moffat & Studier, 1987[153]

Table 2.1: Summary of vectors for cloning and expression

2.1.2 Full length sp_1651 gene in pOPINF

A clone of the full length *sp_1651* gene in the pOPINF vector (Table 2.1) was kindly provided by Dr. Alan Riboldi-Tunnicliffe, University of Glasgow, Glasgow, UK. Cloning was carried out at the Oxford Protein Production Facility (OPPF), Oxford, UK.

2.2 Molecular cloning using restriction enzymes

Full length and N-terminal truncated versions of the *dr_1146* gene were cloned into the vector pET-15b (Table 2.1) using restrictions enzymes.

2.2.1 Oligonucleotide primers

Oligonucleotide primers were designed by hand with a preferential length of 22-30 bp, a melting temperature of 55-80°C and a GC content between 50-60 %. Appropriate restriction sites were incorporated at the 5'-terminals. Oligonuclotides were synthesised by Sigma-Proligo and were provided at a concentration of 100 µM and stored at –20°C. Oligonuclotides for full length and truncated constructs are depicted in Table 2.2.

Primer name	Sequence	Comment
Full length DR_1146 fwd	5'-CAT ATG TGT GGT CTG GCC CAT TCT CCG GA-3'	*Nde*I (CA∇TATG) restriction site including ATG start codon.
Full length DR_1146 rev	5'-GGA TCC TTA GAG ATT AAC CGT G-3'	*Bam*HI (G∇GATCC) restriction site and stop codon.
N term truncated DR_1146 fwd	5'-CATATG AGC CGT GAC GAA GCC GTG-3'	*Nde*I (CA∇TATG) restriction site including ATG start codon.
N term truncated DR_1146 rev	5'-GGA TCC TTA GAG ATT AAC CGT G-3'	*Bam*HI (G∇GATCC) restriction site and stop codon.
N+C term truncated DR_1146 fwd	5'-AA GTT CTG TTT CAG GGC CCG AGC CGT GAC GAA GCC GTG AAG ACG-3'	5'-AA GTT CTG TTT CAG GGC CCG ……3' extension corresponding to pOPINF.
N+C term truncated DR_1146 rev	5'-ATG GTC TAG AAA GCT TTA GCC GCT CTC CCA GTA TTC GGC GCC-3'	5'-ATG GTC TAG AAA GCT TTA …..3' extension corresponding to pOPINF.

Table 2.2: Oligonucleotide primers for PCR

2.2.2 Enzymes

*Bam*HI and *Nde*I restriction enzymes and their corresponding buffers were purchased from New England Biolabs, Hitchin, UK. *Pfu* DNA polymerase (with10x $MgSO_4$ buffer) was purchased from Promega, Southampton, UK. T4 DNA Ligase was purchased from New England Biolabs. All enzymes were stored at –20°C.

2.2.3 Polymerase Chain Reaction (PCR)

PCR was carried out using a Techgene PCR machine (Techne) and 200 µL thin wall PCR tubes. The reaction master mix was composed on ice with reagents added in the order listed in Table 2.3. The reaction was initiated with a 2 min initial denaturing step at 95°C. The cycle consisting of a denaturing step at 95°C for 30 s, an annealing step at 68°C for 30 s and an extension step at 72°C for 1 min, was repeated 30 times. A final extension step was carried out at 72°C for 5 mins and the reaction was held at 4°C. The annealing temperature was varied depending on the melting temperature of the primers. PCR products were separated by agarose gel electrophoresis.

Reagent	Volume of stock solution used	Final Concentration
dNTP mix	5 µL of 2 mM stock	200 µM each dNTP
Forward primer	2.5 µL from 10 µM stock	0.5 µM
Reverse primer	2.5 µL from 10 µM stock	0.5 µM
Genomic DNA	1 µL of 50 ng/µL stock	1 ng/µL (50 ng in 50 µL)
10 x Pfu buffer with $MgSO_4$	5 µL	1 x
Sterile deionised H_2O	33.5 µL	-
Pfu DNA Polymerase	0.5 µL of 3 units/µL stock	1.25 units/50 µL
Total reaction volume	50 µL	

Table 2.3: Master mix for PCR

2.2.4 Agarose gel electrophoresis and gel purification

Gels of between 0.5 – 2.0 % (w/v) agarose in 1x TAE buffer (Table 2.4), supplemented with 0.1 µg/mL ethidium bromide were used for the analysis, separation and purification of linear and circular DNA. DNA samples were diluted 5:1 with Blue/Orange 6x loading Dye (Promega) and ran alongside a 1 kb DNA ladder (Promega) in order to estimate band size. Gels were run at 100 V for 40 minutes and imaged under UV light using an Alphalimager™ System (Alpha Innotech). If needed, the desired DNA fragment was excised from the gel and

purified using a QIAquick Gel Extraction Kit (Qiagen) [154] following the manufacturer's protocol.

Buffer	Composition	Preparation
50 x TAE	2M Tris Base 50 mM EDTA pH 8 5.71% (v/v) glacial acetic acid	pH should be approx 8.5. Diluted to 1x for use.
1 x TE	10 mM Tris-HCl pH 7.5 1 mM EDTA	pH adjusted to 8.0 before use.
5 x M9 salt solution	33.9 g Na_2HPO_4 15.0 g KH_2PO_4 2.5 g NaCl	Make up to 1 L with deionised H_2O and autoclaved. Used at 1x concentration

Table 2.4: Molecular biology buffers

2.2.5 Blunt ended insertion of PCR products into pCR® -Blunt II-TOPO® vector

1 µL of pCR® -Blunt II-TOPO® vector and 2 µL sterile H_2O was added to 3 µL of purified PCR product. The reaction was mixed gently and incubated for 5 minutes at room temperature then quenched on ice.

2.2.6 Transformation of DH5α competent cells

1-2 µL of plasmid DNA was used to transform *E. coli* Library Efficiency® DH5α ™ competent cells (Invitrogen) following the manufacturer's protocol [155-157]. DH5α cells were used for plasmid storage and maintenance. The transformed cells were then plated onto LB Agar plates with appropriate antibiotic selection (Table 2.1) and incubated overnight (20 hrs) at 37°C. The desired result would be an even spread of single colonies.

2.2.7 Isolation of plasmid DNA

A 5 mL culture of Luria-Bertani (LB) media (Table 2.5) containing appropriate antibiotic selection was inoculated with a single colony of the desired strain. This was incubated overnight at 37°C with shaking. Plasmid extraction and purification was carried out using a QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturer's protocol [158]. The concentration and yield of DNA was estimated by A_{260} in conjunction with the Online DNA Concentration Calculator available at http://www.pubquizhelp.com/other/dnacalculator.html.

Media	Composition	Comment
Luria-Bertani (LB) media	10 g/L tryptone 10 g/L NaCl 5 g/L yeast extract	Sterilised by autoclaving before use.
LB-Agar	LB media 1.5% (w/v) micro-agar	Sterilised by autoclaving before use.
SOC	20 g/L tryptone 5 g/L yeast extract 10 mM NaCl 2.5 mM KCl 20 mM $MgCl_2$ 20 mM glucose	Purchased from Invitrogen®. Stored at 4°C.
Overnight Express™ Autoinduction media		Purchased from Novagen® (Merck Biosciences)
M9 Minimal media	1 x M9 salt solution 2 mM $MgSO_4$ 0.1 mM $CaCl_2$ 0.4% glucose solution 20 mg/L thiamine 2 µM $FeCl_2$ 1 g/L $^{15}NH_4SO_4$	1 x M9 salts were prepared from 5x stock (Table 2-5). Other components were made from filtered stock solutions. 1 g/L $^{15}NH_4SO_4$ was added directly. Solution was made up in 500 mL volumes using sterile deionised H_2O.

Table 2.5: Growth media composition

2.2.8 Double restriction digestion

Approximately 200 ng of DNA (10 µL) was digested using 10 units (1 µL) of *Bam*HI and 10 units (1 µL) of *Nde*I with the addition of 2 µL buffer D and 6 µL sterile H_2O. The 20 µL reaction mixture was mixed gently and incubated at 37°C for 3 hrs. When digesting vector DNA the reaction was incubated overnight at 37°C. The resulting product was analysed by agarose gel electrophoresis and purified as described in Section 2.2.4.

2.2.9 Ligation

T4 DNA Ligase (New England Biolabs) was used to ligate the digested vector and insert DNA in a molar ration of 7:1. A typical reaction mixture comprised 1 µL vector, 7 µL insert, 1 µL 10 x T4 DNA Ligase Reaction Buffer and finally 1 µL (40 units) of T4 DNA Ligase. The 10 µL reaction was incubated overnight (20 hrs) at 16°C. The resulting plasmid DNA was used to transform DH5α cells and was isolated as described in Section 2.2.7. The concentration and yield of plasmid DNA was estimated by A_{260}.

2.2.10 DNA sequencing

Final plasmid DNA was sequenced using T7 forward and reverse primers by The Sequencing Service, University of Dundee, UK.

2.3 Molecular cloning by In-Fusion™ Reaction

An attempt to clone an N and C-terminal truncated form of *dr_1146* into a pOPINF vector was made using the In-Fusion ™ PCR Reaction system (Clontech).

2.3.1 Deinococcus radiodurans genomic DNA

Deinococcus radiodurans R1 genomic DNA at a concentration of 250 ng/µL was kindly provided by the Macromolecular Crystallography Group, ESRF, Grenoble, France.

2.3.2 pOPINF vector

The Linearised pOPINF vector was kindly provided by the OPPF, Oxford, UK at a stock concentration of 100 ng/µL.

2.3.3 Oligonucleotides

Oligonucleotide primers were designed as described in Section 2.2.1. The primer extensions required for insertion into pOPINF are shown in Table 2.2. Primers were synthesised by Sigma Genosys in a pelleted form. They were reconstituted with sterile H_2O to a concentration of 100 mM and stored at –20°C.

2.3.4 In-Fusion™ PCR cloning reaction

The required DNA insert was amplified by PCR as described in Section 2.2.3. A reaction mixture containing 100 ng purified insert, 100 ng linearised and purified pOPINF vector was made up to 10 µL with sterile H_2O and added to a well of the dry-down In-Fusion™ plate. The contents were mixed briefly, sealed and incubated for 30 minutes at 42°C. The reaction was immediately quenched with

40 µL TE buffer (Table 2.4). 5 µL was used to transform a vial of DH5α competent cells.

2.4 Gene Expression

2.4.1 Bacterial cell lines

A summary of the cell lines used is given in Table 2.6. When investigating the optimum conditions for gene expression, the desired plasmid DNA was transformed into several cell lines.

Cell line	Genotype	Manufacturer/ reference
Library Efficiency® DH5α ™	F- φ80*lac*ZΔM15 Δ.(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Invitrogen™ Hanahan, 1983[155] Tartof & Hobbs, 1987[157]
BL21 Star™(DE3)	F- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm rne131* (DE3)	Invitrogen™ Studier & Moffat, 1986[149] Grodberg & Dunn, 1988[152] Grunberg-Manago, 1999[159]
BL21(DE3)pLysS	F- *ompT hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3) pLysS (Cam^R)	Invitrogen™ Studier & Moffat, 1986[149] Grodberg & Dunn, 1988[152] Moffatt & Studier, 1987[153]
BL21-AI™	F- *ompT hsdS*$_B$(r_B^-, m_B^-) *gal dcm araB::T7 RNAP-tetA*	Invitrogen™ Studier & Moffat, 1986[149] Grodberg & Dunn, 1988[152] Miyada *et al.*, 1984[160] Lee, 1980[161] Lee *et al.*,[162]
Rosetta™(DE3)pLysS	F– *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE (Cam^R)	Novagen® (Merck Biosciences) Studier & Moffat, 1986[149] Grodberg & Dunn, 1988[152] Novy *et al.*, 2001[163] Moffatt & Studier, 1987[153] Baca & Hol, 2000[164]
Rosetta™ 2(DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)	Novagen® (Merck Biosciences) Studier & Moffat, 1986[149] Grodberg & Dunn, 1988[152] Novy *et al.*, 2001[163] Baca & Hol, 2000[164]
B834(DE3)	F– *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm met*	Novagen® (Merck Biosciences) Wood, W.B., 1966[165] Leahy *et al.*, 1992[166]

Table 2.6: Summary of bacterial cell lines

2.4.2 Antibiotics

Antibiotics were selected depending on choice of vector and cell line (Tables 2.1 and 2.6). 1000 x stock solutions were created at the appropriate strength, with the appropriate solvent as outlined in Table 2.7. All stock solutions were filter sterilised (0.2 µm filter) and stored at -20°C until needed.

Antibiotic	Stock concentration	Final concentration	Solvent
Ampicillin	50 mg/mL	50 µg/mL	H_2O
Carbenicillin	50 mg/mL	50 µg/mL	H_2O
Chloramphenicol	32 mg/mL	32 µg/mL	EtOH
Kanamycin	50 mg/mL	50 µg/mL	H_2O

Table 2.7: Antibiotic stock solutions and preparation

2.4.3 Isopropyl-β-D-thiogalactopyranoside (IPTG)

A 0.4M stock (1000 x final concentration) of IPTG was filter sterilised (0.2 µm filter), divided into 1 mL aliquots and stored at –20°C until needed.

2.4.4 Small scale expression trials in LB media

A 200 mL culture of LB media (Table 2.5) with the appropriate antibiotic selection (Tables 2.1 and 2.7) was inoculated with a single colony from the transformation of an *E. coli* cell line. This was done for each cell line transformed. The inoculated culture was incubated at 37°C, with shaking, until an A_{600} of 0.6 was reached. A 0.5 mL sample was taken for SDS-PAGE analysis (Section 2.4.8) and the culture divided into two equal 100 mL portions. Protein production was induced by the addition of 0.4 mM (final concentration) IPTG. One flask was incubated, with shaking, at 22°C and the other, with shaking, at 37°C for 20 hours. 0.5 mL samples of each culture were taken throughout incubation and were analysed by SDS-PAGE to monitor the expression of soluble protein. In the case of low expression or inconclusive SDS-PAGE, a western blot (Section 2.4.9) was used to identify and analyse the expression of target protein.

2.4.5 Large scale expression in LB media

The variables (cell line and temperature) resulting in the largest amount of soluble protein were selected for large scale experiments in order to gain sufficient yields for crystallisation experiments and biophysical characterisation.

A 60 mL starter culture, with appropriate antibiotic selection, was inoculated with a single colony from a transformation of the successful cell line from initial trials. The starter culture was incubated, with shaking, for 20 hours at 37°C until an A_{600} of approximately 3.5 was reached. 5 mL volumes of the starter culture were used to inoculate 10 x 0.5 L of LB media (Table 2.5) held in 2 L baffled flasks with the appropriate antibiotic selection. The 2 L flasks were incubated, with shaking, at 37°C. When an A_{600} of 0.6 was reached, protein production was induced by the addition of 0.4 mM IPTG (final concentration). The temperature was adjusted to the optimum setting and the cultures were incubated for 20 hrs or for the optimum time established by initial trials.

2.4.6 Small scale expression trials and large scale expression in autoinduction media

Recombinant expression of *sp_1651* was carried out using Overnight Express™ Autoinduction media (Novagen) (Table 2.5). Experiments were carried out in the same way as those described for LB media (Sections 2.4.4 and 2.4.5) with exception of incubation and induction of protein production by the addition of IPTG. After inoculation of cultures with either single colonies or starter culture, the cultures were incubated at 4 hours at 37°C, with shaking. The temperature was then adjusted to the optimum setting and cultures were incubated for 20 hrs or for the optimum time established by initial trials.

2.4.7 Harvesting and storage of cells

E. coli cells were harvested by centrifugation at 5000 rcf for 20 minutes. The resulting pellets were weighed and then stored at -80°C until needed. Typically batches of cell pellet harvest from 2 L of cells would be used to gain enough protein for further experiments but this varied depending on the construct used.

2.4.8 SDS-PAGE analysis of whole cell pellets

0.5 mL time point samples taken during initial expression trials were analysed using the NuPAGE® Pre-Cast Gel System based on the Laemmli SDS-PAGE method [167, 168] to establish the presence of expressed target protein. The samples were pelleted by centrifugation at 10,000rpm for 10 mins. The LB media was removed and the soluble and insoluble cell fractions were isolated and normalised using B-PER® Bacterial Protein Extraction Reagent (PIERCE), following the manufacturer's protocol [169]. 20 µL aliquots of both the soluble and insoluble cell fractions were added to 6 x LDS loading buffer (1 x final concentration) and 0.3 mM DTT (final concentration). The resulting samples were incubated at 70°C for 10 minutes then briefly centrifuged before loading. 20 µL of each sample were loaded on to a 10 well NuPAGE® Novex Bis-Tris gel (Invitrogen) alongside SeeBlue Plus2 Pre-Stained Standard (Invitrogen) to estimate the molecular weight of the resulting bands. Gels were run at 200V in the presence of 1 x MES running buffer (Invitrogen) for 42 minutes. After separation, gels were washed briefly with deionised H_2O then stained with Coomassie Brilliant Blue R250 (0.1 % in 50 % Ethanol, 10 % acetic acid) for 20 minutes. Gels were destained by brief incubation in destain (40 % Ethanol, 10% acetic acid), followed by overnight shaking in deionised H_2O.

2.4.9 Western blot analysis

If results of SDS-PAGE analysis were inconclusive, as was often the case with whole cell pellets or poorly expressing proteins, further analysis by western blotting was carried out to establish target protein presence and purity

After running a fresh SDS-PAGE gel as described in Section 2.4.8, the protein was transferred from the gel onto Immobilon Polyvinylidene difluoride (PVDF) membrane (Millipore) following the NuPAGE® Novex Bis-Tris Gel Western Transfer Protocol (Invitrogen) [167]. This involved the use of the Xcell II™ Blot Module and 1x NuPAGE® Transfer Buffer (Invitrogen). Detection of the 6xHis tagged protein was carried out using the SuperSignal® West HisProb™ Kit (Pierce) following the manufacturer's protocol [170]. When following the SuperSignal®

West HisProb™ Kit protocol, film (Kodak) was developed using a X-OMAT SRX-101A system (Konica Minolta).

2.5 Protein purification

2.5.1 Buffer preparation

The pH of buffers was measured and adjusted using a Corning pH meter 240 and the appropriate acid or base. The pH meter was calibrated before use according to the manufacturer's instructions using standard buffers of pH 4.0, 7.0 and 10.0. The pH of buffers for purification was varied depending on the isoelectric point (pI) of the target protein but was typically pH 7.5.

All purification buffers were filter sterilised (0.2 µm filter) and degassed under vacuum before use. The composition of all buffers used is shown in Table 2.8.

Buffer	Composition
Cell lysis buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 20 mM imidazole pH 7.5 5 mM $MgCl_2$ DNase 1 Complete EDTA-Free Protease Inhibitor Cocktail tablet
Solubilisation buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 8 M urea or 6 M guanidine hydrochloride
IMAC wash buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 20 mM imidazole pH 7.5
IMAC elution buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 500 mM imidazole pH 7.5
IEX buffer A	20 mM Tris-HCl pH 7.5
IEX buffer B	20 mM Tris-HCl pH 7.5 1M NaCl
Gel filtration buffer	50 mM Tris-HCl pH 7.5 200 mM NaCl Optional addition of 0.01% NaN_3, BME or DTT

Table 2.8: Composition of purification buffers

2.5.2 Preparation for dialysis

10 kDa MWCO dialysis membrane (Medicell International Ltd) was boiled in deionised H_2O for 3 minutes then rinsed with chilled deionised H_2O. Typically, a protein sample would be dialysed overnight at 4°C against 5 L of the desired buffer.

2.5.3 Mechanical cell lysis

The required volume of cell pellet was thawed in a 25°C water bath and then re-suspended in lysis buffer (Table 2.8) with the addition of 1 Complete EDTA-Free Protease Inhibitor Cocktail Tablet (Roche Applied Science). Typically 5 mL of lysis buffer was used per 1 g of cell pellet. The resuspended pellet was homogenised then mechanically lysed using either the French Press Cell disrupter (Thermo), with 3 passes at 950 psi, or the Constant Cell Disruption System, with 2 passes at 20 kpsi, depending on equipment availability and sample volume. Throughout cell lysis, the sample was kept on ice in an attempt to minimise proteolysis. The resulting lysate was centrifuged at 5700 rcf for 20 minutes followed by 34,900 rcf for 40 minutes at 4°C. The soluble supernatant and insoluble pellet were separated and the soluble supernatant was filtered (0.2 µm filter).

If required for purification, the insoluble pellet was solubilised with solubilisation buffer (typically 20 mL), which contained 8 M urea (Table 2.8). This procedure required continuous stirring at 4°C for 1 hour after which the sample was centrifuged for a further 30 minutes at 34,900 rcf. The solubilised supernatant was then separated from remaining cell debris and filtered (0.2 µm filter).

2.5.4 Immobilised metal affinity chromatography (IMAC)

The cell lysate was loaded onto a 5 mL HisTrap HR column (GE Healthcare) which was pre-equilibrated with IMAC wash buffer (Table 2.8). This was done using a peristaltic pump at a speed of 3 mL/min at 4°C. The column was connected to an AKTA Purifier FPLC (GE Healthcare) and the unbound lysate was eluted with IMAC wash buffer at 3 mL/min until A_{280} reached zero. The protein

was eluted using a step gradient program which was run at 3 mL/min at 4°C, collecting 5 mL fractions. The program included 8-14 column volume steps of increasing imidazole concentrations, typically 10, 20 and 100 % IMAC elution buffer (Table 2-8). The presence and purity of the target protein was established by SDS-PAGE analysis of fractions collected.

2.5.5 Concentration and yield measurement by UV

Protein concentration and subsequent protein yields were estimated using the Beer-Lambert Law. To do this, protein molecular weights and extinction coefficient values were estimated using the ProtParam tool [171] from the ExPASy proteomics server [172]. The absorbance of a purified protein solution in a 1cm quartz cuvette was measured using a Lambda 40 spectrophotometer (Perkin Elmer). The estimated protein concentration was then used to estimate the yield of purified protein depending on the volume eluted after purification. Extinction coefficients (ε), molecular weights (MW) and isoelectric point (pI) values estimated for the proteins studied, including vector residues and tags are listed in Table 2.9.

Construct	pI	MW (Da)	ε (cm^{-1} M^{-1})
Full length DR_1146 in pDEST17 with 6xHis tag	5.49	23,400.90	28,420
Full length DR_1146 in pET-15b with 6xHis tag	5.57	23,206.60	23,590
Full length DR_1146 in pET-15b without 6xHis tag	4.90	21,324.55	23,590
N term truncated DR_1146 in pET-15b with 6xHis tag	5.72	20,319.57	23,470
N term truncated DR_1146 in pET-15b without 6xHis tag	4.97	18,437.52	23,470
SP_1651 in pOPINF with 6xHis tag	5.27	21,336.10	18,575
SP_1651 in pOPINF without 6xHis tag	4.73	19,349.90	18,575
DR_0463 in pDEST17 with 6xHis tag	6.13	111,096.50	150,245
SP_1648 in pOPINF with 6xHis tag	6.90	28,984.60	16,055
DR_2284 in pET151/D TOPO with 6xHis tag	7.17	29,954.10	36,565

Table 2.9: pI, MW and extinction coefficient values for target constructs including contributions by tag and expression vector residues. Calculated using ProtParam tool [171] available on the ExPASy proteomics server [172].

2.5.6 Ion exchange chromatography (IEX)

The IMAC purified target protein was dialysed against 5 L of IEX buffer A (Table 2.8) to remove NaCl and Imidazole before IEX. A 5 mL QTrap HP anion exchange column was equilibrated with IEX buffer A and the dialysed protein sample loaded by peristaltic pump at 3 mL/min at 4°C. The column was attached to the AKTA Purifier FPLC and the protein was eluted with a linear gradient between 0-70% IEX buffer B (Table 2-8) over 30 column volumes. 5 mL fractions were collected and analysed by SDS-PAGE. The resulting purified target protein was dialysed against 5 L of gel filtration buffer (Table 2.8) and was then concentrated by centrifugation using 10 kDa MWCO concentrators (Vivascience) to 10 mg/mL before loading onto a gel filtration column for size exclusion chromatography (SEC).

2.5.7 Size exclusion chromatography (SEC)

A S200 10/300 HR global gel filtration column (GE Healthcare) attached to an AKTA Purifier FPLC at 4°C was equilibrated with gel filtration buffer (Table 2-8). The concentrated target protein sample was centrifuged for 10 minutes at 13,000 rpm and a volume of 500 µL (or less depending on the amount of protein available) was injected onto the column. The protein was eluted with 1.5 column volumes of gel filtration buffer at a speed of 0.5 mL/min. 0.5 mL fractions were analysed by SDS-PAGE and pooled accordingly. The purified protein was concentrated by centrifugation to the desired concentration for crystallisation experiments or biophysical characterisation.

2.5.8 Removal of 6xHis tag by thrombin cleavage

Thrombin cleavage was required to remove the 6xHis tag from protein expressed by a pET-15b vector. This step was carried out either after IMAC or gel filtration depending on requirement. 0.26 units of thrombin (Novagen) per mg tagged protein were added to the sample which was then dialysed against 5 L 20 mM Tris pH 7.5, 200 mM NaCl. To remove the thrombin, the sample was loaded by peristaltic pump at 1 ml/min, in 5 mL batches, onto a 1mL benzamidine column (GE Healthcare) which was pre-equilibrated with BC elution buffer (Table 2.8).

The protein was eluted with 10 column volumes of BC elution buffer at 1 ml/min. It was then loaded onto a 1 mL HisTrap HR column (GE Healthcare), pre-equilibrated with IMAC wash buffer, to remove the cleaved 6xHis tag. The cleaved protein, with thrombin removed, was then eluted by 10 CV IMAC wash buffer and analysed by SDS-PAGE. The cleaved protein was pooled together and dialysed against 5 L IEX buffer A and then purified further by IEX to ensure that the thrombin was completely removed.

2.5.9 Removal of 6xHis tag by HRV-3C protease

Tagged HRV-3C protease was required for removal of 6xHis tag from protein expressed from the OPINF vector. This step was carried out either after IMAC or gel filtration depending on requirement. 5 µL of HRV-3C (from a 1 mg/mL stock) was used per mg of protein for efficient tag cleavage. The protease was added along with 10 mM DTT and 1 mM EDTA and the sample was dialysed against 5L 20 mM Tris pH 7.5, 200 mM NaCl. It was then loaded on to a pre-equilibrated 1 mL HisTrap HR column by peristaltic pump at 1 ml/min, 4°C. Protein, with the HRV-3C protease and 6xHis tag removed, was eluted with 10 column volumes of IMAC wash buffer and analysed by SDS-PAGE.

2.5.10 Mass spectrometry

MALDI-TOF and ESI mass spectrometry was performed by Kenneth Beattie, Finger Prints Proteomics, University of Dundee, UK. Purified protein was provided as a 1 mg/mL solution in 20 mM Tris-HCl pH 7.5, 50 mM NaCl. Additional small molecule ESI-MS was carried out by Jim Tweedie, University of Glasgow, UK.

2.5.11 Cofactor extraction

Approximately 15 mg of protein was concentrated to a volume of 200 µL (over 100 mg/mL). 100 µL MeOH was added resulting in the precipitation of the protein. The sample was centrifuged at 13,000 rpm for 30 minutes and the supernatant was extracted. This supernatant was analysed by ESI-MS.

2.5.12 Amino acid sequence determination by Edman degradation

Edman degradation was carried out by David Campbell, University of Dundee, UK. In preparation, the purified protein was transferred onto PVDF membrane by the method described in Section 2.4.8 and stained with Coomassie R-250 for 1 min. The membrane was then de-stained and air-dried at room temperature. The single band was excised and placed in a sterile eppendorf tube for analysis.

2.6 Crystallisation experiments

2.6.1 JBS solubility screen and dynamic light scattering (DLS) experiments

Optimisation of the final protein purification buffer composition prior to crystallisation trials was investigated using the JBS Solubility kit (Jena Bioscience) [173]. The manufacturer's protocol was followed [173] to investigate the effect of 24 different buffers and 14 different additives on protein solubility and aggregation behaviour. The monodispersity of each sample was analysed by dynamic light scattering using a Dynapro-801 (Protein Solutions). All samples were centrifuged at 13,000 rpm and filtered through a 0.2 µm filter before testing. The protocol suggested a sample size of 20 µL, however, the Dynamo-801 needed a minimum sample size of 200 µL and therefore the experiments were scaled up as follows.

Purified protein was concentrated to over 20 mg/mL by centrifugation. 40 µL sitting drops (20 µL protein: 20 µL reservoir buffer) were set up in a 24 well sitting drop tray with 0.5 mL of each buffer solution (24) in individual reservoirs. The vapour diffusion experiment was incubated overnight at 20°C. After 24 hours, the drops were investigated for precipitation under a light microscope. The clear drops (no precipitation) were pipetted into individual eppendorfs and diluted to 200 µL with the corresponding buffers resulting in a concentration of approximately 1 mg/mL. Each sample was centrifuged at 13,000 rpm for 10 minutes before DLS experiments were carried out at room temperature. The buffer which yielded the most monodispersed DLS reading was selected and the

protein (at 2-3 mg/mL) was dialysed into a 50 mM concentration of this buffer. 150 µL of protein was added to each of 14 additives, mixed, and then incubated for 2 hours at room temperature. DLS experiments were carried out and the best additive was selected. The protein was dialysed against the selected buffer and additive and concentrated for crystallisation experiments.

2.6.2 Pre-crystallization test (PCT™)

After purification to greater than 90 % homogeneity, as estimated by SDS-PAGE, and in some cases, determination of the optimum crystallisation buffer, target protein was concentrated to 20 mg/mL by centrifugation. In order to determine the optimum protein concentration for crystallisation experiments a pre crystallization test (PCT™) (Hampton Research) was carried out. This was done for concentrations of 20, 15, 10 and 5 mg/ml following the manufacturer's protocol [174]. The concentration of the stock target protein was adjusted according to PCT™ results.

2.6.3 Initial sparse matrix crystallisation screening

Initial sparse matrix crystallisation screening was carried out at 20°C using the sitting drop vapour diffusion method. The MICROLAB® STAR liquid handler (Hamilton Life Science Robotics) was used to distribute 50 µL reservoir volumes of commercial screens into 96 well MRC plates (Innovadyne). The Cartesian™ Dispensing System (Genomic Solutions®) was then used to aliquot a protein drop volume of 0.5 µL followed by a 0.5 µL reservoir drop for each of the 96 conditions in the plate. Fifteen commercial sparse matrix screens were used and are listed with (Table 2.10). Each tray set up was bar-coded and stored in a Rhombix™ Imaging System (Thermo Electron Corporation) at 20°C which allowed crystal growth to be monitored for three months. Photos were taken every day for the first week of incubation and once per week thereafter.

Crystal screen name	Manufacturer
Crystal Screens 1 and 2	Hampton Research
Cryo Screens 1 and 2	Emerald Biosystems
PACT Screens 1 and 2	Molecular Dimensions
JCSG plus Screens 1 and 2	Molecular Dimensions
Peg/Ion Screens 1 and 2	Hampton Research
Structure Screens 1 and 2	Molecular Dimensions
Wizard Screens 1, 2 and 3	Emerald Biosystems

Table 2.10: Commercial sparse matrix crystallisation screens

2.6.4 Testing X-ray diffraction quality of initial crystals

Before testing, the crystals were soaked in a range of cryoprotectants. Crystals were initially tested in-house on a MicroMax™ - 007 X-ray generator (Rigaku) equipped with a Cryo-cooling system (Oxford Cryo Systems) and a Mar345 detector. Synchrotron radiation sources (SRS) at ESRF, Grenoble, France and Diamond, Oxford, UK were also used to assess crystal diffraction. CrystalCap Copper Magnetic™ bases and Mounted CryoLoops™ were purchased from Hampton Research.

2.6.5 Optimisation of initial crystal hits

Attempts at reproducing and optimising initial crystallisation conditions were carried out by varying the pH and precipitant/salt type and concentration. Optimisation experiments were designed using Screen Designer software and executed using the MICROLAB®STAR liquid handler. Drop set up and tray incubation was carried out as describe in Section 2.6.3.

2.7 Biophysical characterisation

2.7.1 Cofactors used for biophysical characterisation

The cofactors riboflavin, riboflavin 5'-monophosphate sodium salt dihydrate (FMN), and flavin adenine dinucleotide disodium salt hydrate (FAD) were purchased form Sigma Aldrich and were used in biophysical characterisation experiments without further purification. Their structures are depicted in Figure 2.1.

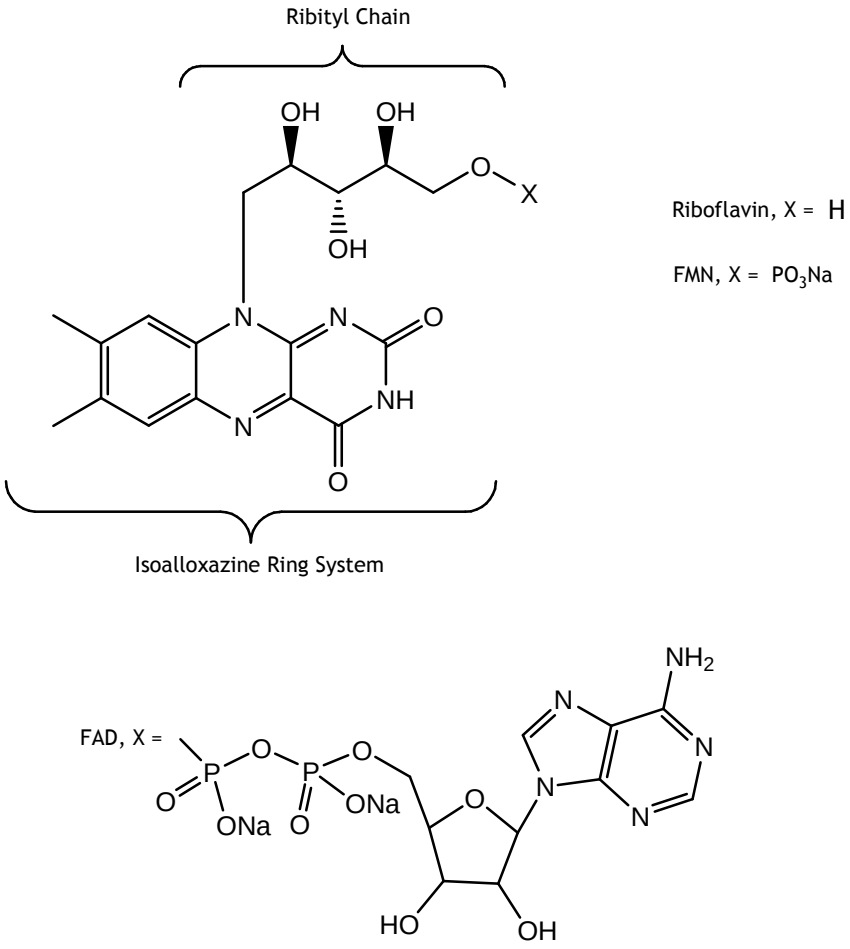


Figure 2.1: Structures of cofactors used for ITC experiments: riboflavin, riboflavin 5'-phosphate sodium salt (FMN) and flavin adenine diphosphate disodium salt (FAD)

2.7.2 Isothermal Titration Calorimetry (ITC)

ITC was performed by Margaret Nutley, BBSRC/EPSRC Biological Microcalorimetry Facility, University of Glasgow, UK. ITC experiments to measure protein – cofactor complexation thermodynamics were done at 25°C using a VP-ITC titration microcalorimeter (MicroCal) following standard instrumental procedures [175, 176]. Purified protein was dialysed extensively against 20 mM Tris pH 7.5, 150 mM NaCl. The same buffer was used to prepare the cofactors. Samples were degassed gently immediately before use. A typical binding experiment involved an initial 1 µL pre-injection followed by 28 x 10 µL sequential injections of cofactor solution (typically 0.2-0.4 mM concentration, determined using ε = 12,200 $M^{-1}cm^{-1}$ at 445 nm) into the ITC cell (1.4 mL active volume, 310 rpm stirring) containing protein solution (typically 18-21 µM). Control experiments were performed under identical conditions by injection of cofactor into buffer

alone to correct for heats of cofactor dilution. Integrated heat effects, after correction for heats of dilution, were analysed by non-linear regression in terms of a sequential single-site binding model using the standard Origin 5.0™ (MicroCal) software package in order to determine the stoichiometry of binding (N), the association/dissociation constants ($K_A = 1/ K_D$) and the enthalpy of binding (ΔH°). Other thermodynamic quantities were calculated using standard expressions shown in Equation 2.1.

$\Delta G^\circ = -RT.\ln K = \Delta H^\circ - T.\Delta S^\circ$ (1 cal = 4.184 J). **Equation 2.1**

2.7.3 Fluorescence Spectroscopy (FS)

Fluorescence emission and excitation spectra were determined on protein-FMN samples in an aqueous buffer of 20 mM Tris-HCl, 150 mM NaCl. A Spex Fluoromax instrument with 5 nm slits was used with 1 cm pathlength quartz cuvettes, thermostatted at 25°C. Excitation was induced at 445 nm, with emission recorded at 520 nm. The FMN cofactor was prepared in buffer at a concentration of 10 μM, giving an absorbance of less than 0.12 at $\lambda\, exc$ to avoid inner filter effects. Fluorescence titration curves were obtained by adding 10 μL volumes of protein stock solutions (made up in the cofactor solution to avoid dilution effects) to 2.5 mL of FMN. The dissociation constant (K_D) and number of binding sites per mole of protein in solution (N) for the protein-FMN complex was determined using the quenching of FMN fluorescence upon binding to the protein. Data were fitted to standard hyperbolic binding curves using Origin 5.0™ software. Data analysis was carried out by Professor Alan Cooper, Department of Chemistry, University of Glasgow.

Fluorescence emission and excitation spectra for the addition of protein to 10 μM FMN solution were recorded as described above.

2.7.4 Differential Scanning Calorimetry (DSC)

DSC experiments were carried out by Margaret Nutley, Biophysical Chemistry, University of Glasgow. DSC experiments on protein-cofactor complexes were conducted on a VP-DSC calorimeter and thermal transitions were analysed using MicroCal Origin 5.0™ software following the subtraction of the instrumental

buffer-buffer baseline. The sample cell in each experiment contained a pre-mixed solution of 92 μM protein and varying concentrations of cofactor (0-2 mM) prepared in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, with identical buffer (ensured by dialysis) in the reference cell. Samples were briefly degassed before loading and scanned over a 20-120°C temperature range at a scan rate of 60°C/h.

To assess how well the protein refolds after temperature denaturation, samples were cooled to room temperature and the DSC scan was repeated. The resulting data were compared with that gained from initial experiments.

2.7.5 Circular Dichroism (CD)

CD experiments and data analysis were carried out by Dr. Sharon Kelly, Protein and Nucleic Acid Characterisation Facility, University of Glasgow, UK.

CD spectra were recorded for 6xHis-tagged and untagged DR_1146 in 50 mM phosphate buffer pH 7.5, 150 mM NaCl, in the presence and absence of FMN, using a Jasco 810 spectropolarimeter at 20°C. Spectra were recorded in the far UV region (187-260 nm) at a typical concentration of 0.5 mg/mL using 0.02 cm path length quartz cuvettes. Experiments in the near UV region (320-250 nm) were recorded using 0.5cm pathlength cuvettes and a protein concentration of 1 mg/mL. Data were collected using the following instrument parameters: bandwidth 1 nm; scanning rate 50 nm/min; response time 0.5 s and data pitch 0.2 nm.

CD peptide backbone conformations can be estimated from spectral contributions in the far UV region (240-180 nm). Data were submitted to Dichroweb, an online server hosting various secondary structure deconvolution algorithms [177]. Estimates of secondary structure content were obtained using both the CONTIN procedure and the CDSSTR algorithm [178].

The unfolding of DR_1146, using guanidine hydrochloride (0 to 6 M) or a temperature gradient (5 to 80°C), in the presence and absence of two equimolar amounts of FMN, was monitored by CD in the near and far UV regions. The experimental set up, sample preparation and procedure was identical to that for

initial CD experiments (described above). Graphs displaying experimental data were produced using Origin software.

2.7.6 Analytical Ultracentrifugation (AUC)

Analytical ultracentrifugation (AUC) experiments and data analysis were carried out in collaboration with Dr. Olwyn Byron, Division of Infection and Immunity, Faculty of Biomedical and Life Science, University of Glasgow, UK.

Purified protein, at a concentration of 35 mg/mL, was dialysed against a buffer of 5 mM FMN, 50 mM Tris-HCl pH 7.5, 200 mM NaCl. The dialysate was used in the solvent channels of the AUC cells. The concentration dependent oligomerisation behaviour of the protein was determined using sedimentation equilibrium (SE) and sedimentation velocity (SV) AUC.

2.7.6.1 Sedimentation velocity (SV)

A SV experiment was performed using a Beckman-Coulter (Palo Alto, CA, USA) Optima XL-I AUC and an AN-60 Ti rotor. 380 µL samples (0.35, 1, 3.5, 10, 22 and 35 mg/mL) were loaded into 12-mm double-sector centrepieces and the dialysate was loaded into the second sector. Experiments were performed at 4°C, at a rotor speed of 49 krpm. A series of 480 scans, 1 minute apart, was obtained for each sample, using interference optics. Sedimentation profiles were analysed with SEDFIT software [179], which allows the subtraction of systematic noise and direct modelling of the sedimentation boundary as a continuous distribution of discrete non-interacting species (c(s) analysis). The sedimentation coefficients that best fitted the Lamm equation were found by finite element analysis in SEDFIT [179].

Buffer density (ρ), viscosity (η) and the partial specific volume ($\bar{v}$) of the protein at 4°C and 20°C were calculated using SEDNTERP software [180].

Apparent sedimentation coefficients ($s_{4,b}$) for protein species at 4°C in buffer, were determined by integration of corresponding peaks. Subsequent true sedimentation coefficients ($s_{20,w}$), corrected for solvent viscosity and density were calculated using Equation 2.2:

$$s_{20,w} = \frac{(1 - \bar{v}\rho)_{20,w}}{(1 - \bar{v}\rho)_{4,b}} \times \frac{\eta_{4,b}}{\eta_{20,w}} \times s_{4,b}$$ **Equation 2.2**

where, 20 = 20°C, w = water, 4 = 4°C and b = buffer.

Values of $s_{20,w}$ were then plotted against protein concentrations (mg/mL) to establish the apparent sedimentation coefficient for a given protein species at zero concentration ($s^{o}_{20,w}$).

A theoretical estimate of sedimentation coefficient for the protein, if assumed to be a hydrated sphere, was calculated using the following sequence of equations (Equation 2.3 to 2.7).

Anhydrous volume $V = \frac{M\bar{v}}{N_A}$ **Equation 2.3**

If the protein is spherical, spherical radius R can be determined from rearrangement of

$$V = \frac{4}{3}\pi R^3$$ **Equation 2.4**

The frictional coefficient, f_0, of the anhydrous sphere of radius R

$$f_0 = 6\pi\eta R$$ **Equation 2.5**

The sedimentation coefficient for an anhydrous sphere (s_{anhyd}) with the same molar mass as the protein is calculated from

$$s = \frac{M(1 - \bar{v}\rho)}{N_A f_0}$$ **Equation 2.6**

Using a conversion factor of 0.86 (determined using buffer density, partial specific volume and a hydration level of 0.4 g/g protein), the sedimentation coefficient for a hydrated sphere (s_{hyd}) is equal to

$$s_{hyd} = 0.86(s_{anhyd})$$ **Equation 2.7**

Comparison of s_{hyd} with experimentally determined $s^{o}_{20,w}$ can indicate if the protein is more extended in shape than a regular sphere.

2.7.6.2 Sedimentation equilibrium (SE)

A SE experiment was carried out in a Beckman-Coulter (Palo Alto, CA, USA) Optima XL-I analytical ultracentrifuge, using interference optics. Experiments were conducted at 4°C and at rotor speeds of 20, 30 and 40 krpm. Scans were obtained at each speed until satisfactory overlay of traces separated by 3 h was obtained, indicating attainment of equilibrium. 80 µL of the same seven samples and buffer used for the SV experiment were loaded into the same counterpieces. The data collected were analysed by global modelling using SEDPHAT [181] software (http://www.analyticalultracentrifugation.com/sedphat/).

The apparent weight-average molecular weights ($M_{w,app}$) were determined by fitting data obtained at 20, 30 and 40 krpm both individually and globally with a species analysis model.

To determine association constants (K_A) for the self-association of the protein datasets were fitted with several different self-association models. SE data acquired with interference optics generate K_A values with units of fringes^{-1}. It was essential to convert this to units of M^{-1} using a numerical conversion factor established by plotting fringes (obtained from SV analysis) against molar protein concentration to give a linear gradient (Figure 2.2 and Table 2.11).

[Protein] (mg/mL)	[Protein] (µM)	Fringes
0.35	19.0	1.5
1	54.2	3
2	108.5	4
3.5	189.8	7
10	549.8	11
22	1193.2	36
35	1898.3	60

Table 2.11: The number of interference fringes generated by the seven protein concentrations used in the SV experiment.

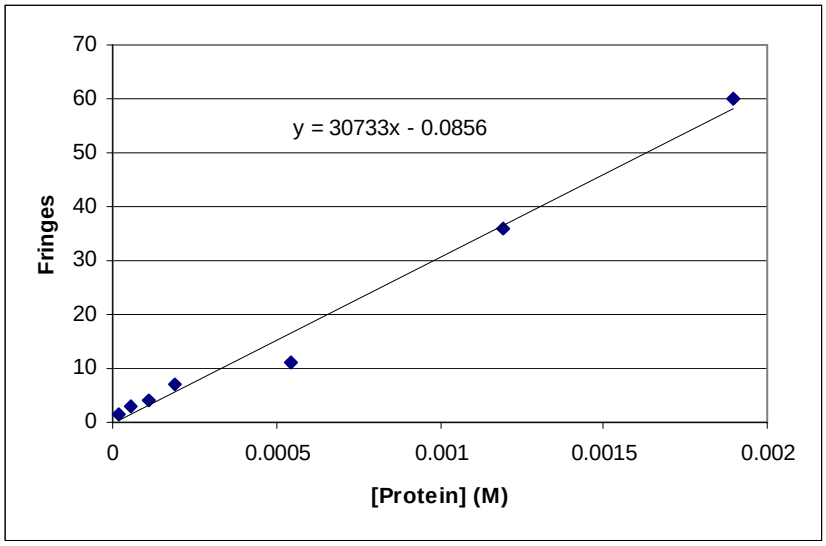


Figure 2.2: Plot of fringes against molar protein concentration to give a numerical conversion factor of 30733 fringes M^{-1}.

The subsequent relationship between fringes and molar concentration is described in Equations 2.8, 2.9 and 2.10.

$$\text{fringes} = 30733M \qquad \textbf{Equation 2.8}$$

$$\text{fringes}^{-1} = \left(3.25/10^{-5}\right)M^{-1} \qquad \textbf{Equation 2.9}$$

$$\text{So } K_A\ (M^{-1}) = K_A\ (\text{fringes}^{-1})/(3.25/10^{-5}) \qquad \textbf{Equation 2.10}$$

Hence, to convert from units of fringes^{-1} to M^{-1}, the value of K_A (in fringes^{-1}) was divided by $3.25/10^{-5}$. K_A in units of M^{-1} was then used to determine the molar K_D for self-association ($K_A = 1/K_D$).

2.7.7 Nuclear Magnetic Resonance (NMR)

NMR experiments and data analysis were carried out by Dr. Brian Smith, Structural Biology, University of Glasgow, UK.

2.7.7.1 One-dimensional, non-labelled NMR experiments

Recombinant, T_ DR_1146 was expressed in LB media (Table 2.5) from a pET-15b vector in *E. coli* BL21 Star(DE3) cells as described in Section 2.4.5. The protein was purified as in Section 2.5.4 – 2.5.6 and samples, with and without the 6xHis tag, were eluted from a S200 10/300 HR Global gel filtration column in a buffer of 50 mM Tris pH 7.5, 150 mM NaCl, 0.01% NaN_3. An additional T_DR_1146_clvd

sample was purified in the presence of 1 μM FMN. The samples were typically concentrated to 0.6 mM in a volume of 600 μL.

All NMR measurements were performed using a 600 MHz Bruker Avance spectrometer equipped with a TCI CryoProbe™ (Bruker BioSpin), at a proton frequency of 14.1 T and a temperature of 25 °C.

For non-labelled experiments the sweep width, number of complex points and number of scans were 8993 Hz, 4608 points and 8 scans respectively. A method of excitation sculpting for water suppression was employed [182].

2.7.7.2 Two-dimensional ^{15}N-^{1}H Heteronuclear Single Quantum Correlation Spectrum (HSQC)

^{15}N labelled T_DR_1146_clvd was produced by expression in M9 minimal media (Table 2.5) using $^{15}NH_4SO_4$ glucose (Spectra Stable Isotopes) as the sole nitrogen source. The sample was purified as in Sections 2.5.4 – 2.5.6 with a final gel filtration step into a buffer of 50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.01% NaN_3. The purified ^{15}N labelled protein was concentrated to 0.9 mM and a sample volume of 600μL.

^{15}N-^{1}H HSQC experimental parameters included a sweep width of 8389 Hz and 896 complex points for ^{15}N F1, and a sweep width of 1642 Hz and 64 complex points for 1H F2.

After initial HSQC experiments on the T_DR_1146_clvd sample, aliquots of FMN from a 50 mM stock solution (high concentration to avoid dilution affects) were added to investigate the amount needed to saturate the protein.

3 Cloning, expression and purification of DR_1146

3.1 Expression and purification of *dr_1146* from pDEST17

3.1.1 Construct design

A full length form of the *dr_1146* gene (nucleotide residues 1 to 582), cloned into a pDEST17 vector (Table 2.1), was kindly provided by Protein'eXpert, Grenoble, France. The nucleotide and putative amino acid sequence for DR_1146 is annotated on the Comprehensive Microbial Resources database [6] which can be accessed on the web page of The J. Craig Venter Institute [7].

3.1.2 Recombinant expression of dr_1146 from pDEST17

1 µL plasmid DNA (*dr_1146* in pDEST17) was used to transform BL21 Star(DE3) competent cells by the method described in Section 2.2.6. 20 µL of the transformation reaction mixture gave an even spread of single colonies after 20 hrs incubation at 37°C.

Recombinant *dr_1146* was expressed in LB media (Table 2.5) from the pDEST17 vector (Table 2.1) as described in Section 2.4.5. A 100 x dilution of starter culture was used to initiate cell growth in a 500 mL culture which was incubated at 37 °C until an OD_{600} of 0.6 was reached. The cells were then induced with IPTG and incubated for 20 hrs at 22°C. SDS-PAGE analysis of cell extracts taken after this incubation period determined that DR_1146 was expressed in both soluble and insoluble forms in equal amounts.

Expression was scaled up to 4 x 500 mL cultures as described in Section 2.4.5. This resulted in sufficient yield of soluble protein in order to set up crystallisation experiments, thus no efforts were made to reduce the amount of insoluble protein production.

3.1.3 Optimised IMAC purification of DR_1146

Cell pellet from 2 L of culture was re-suspended in 30 mL of lysis buffer (Table 2.8) and the cells were lysed by French press as described Section 2.5.3. After centrifugation the soluble lysate was loaded on to a 5 mL HiTrap HR column as described in Section 2.5.4. and flushed with 10 CV of IMAC wash buffer to remove unbound proteins.

IMAC was carried out on an AKTA Purifier FPLC as described in Section 2.5.4.and the program employed is depicted in Table 3.1.

Step number	Number of column volumes (1 CV = 5 mL)	% IMAC Wash Buffer (20 mM imidazole)	% IMAC Elution Buffer (500 mM imidazole)	Effective imidazole conc. (mM)
1	10	100	0	20
2	20	90	10	68
3	12	80	20	116
4	10	0	100	500

Table 3.1: Optimised IMAC protocol

Figure 3.1 shows the resultant A_{280} elution profile for the optimised purification of DR_1146. Peaks 1, 2 and 3 correspond to protein eluted with 10, 20 and 100 % IMAC elution buffer respectively.

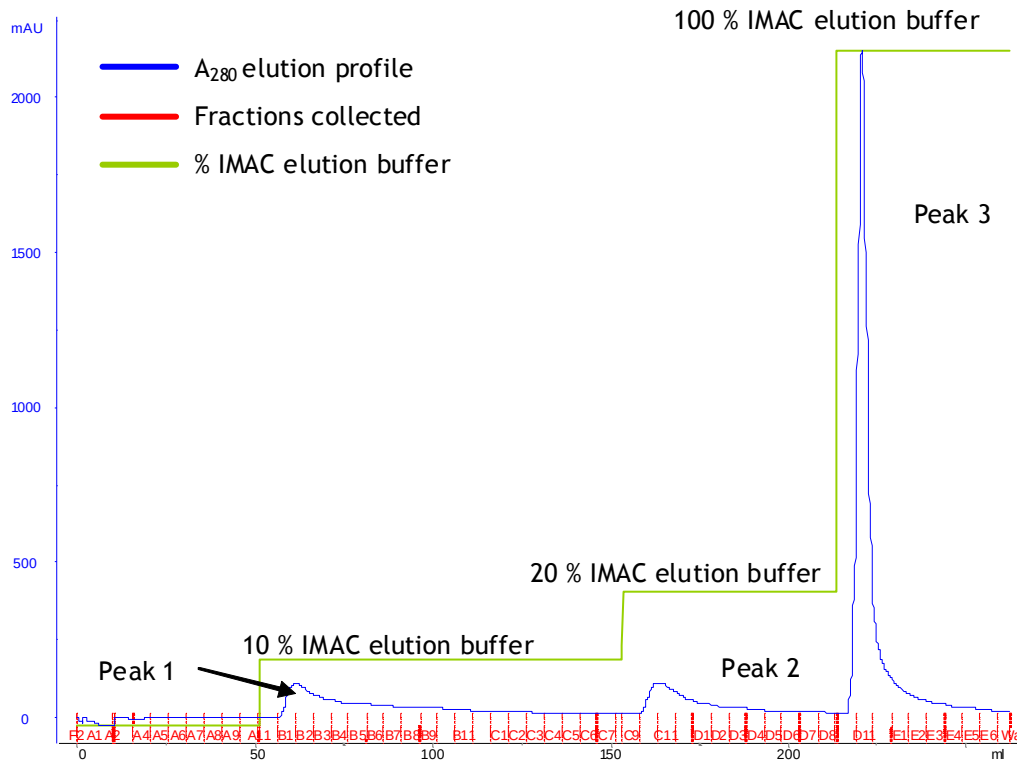


Figure 3.1: A_{280} elution profile for optimised IMAC purification of DR_1146. Contaminant proteins are eluted with 10 and 20 % IMAC elution buffer steps. Peak 3 corresponds to the elution of SP_1651 with 100 % IMAC elution buffer.

An SDS-PAGE analysis of the fractions collected for each peak is shown in Figure 3.2. A large proportion of impurities present in the initial lysate (lane 1) did not bind to the column and thus were eluted in the column flow through (lane 2). Subsequent 0, 10 and 20 % IMAC elution buffer washes (lanes 3, 4 and 5 respectively) removed further impurities. A strong 24 kDa band which corresponds to DR_1146 (lane 6) was eluted in the large peak 3 (2400 mAU) of Figure 3.1. The strong affinity of the 6xHis tag for the Ni^{2+} column matrix allowed efficient purification of the target protein by IMAC.

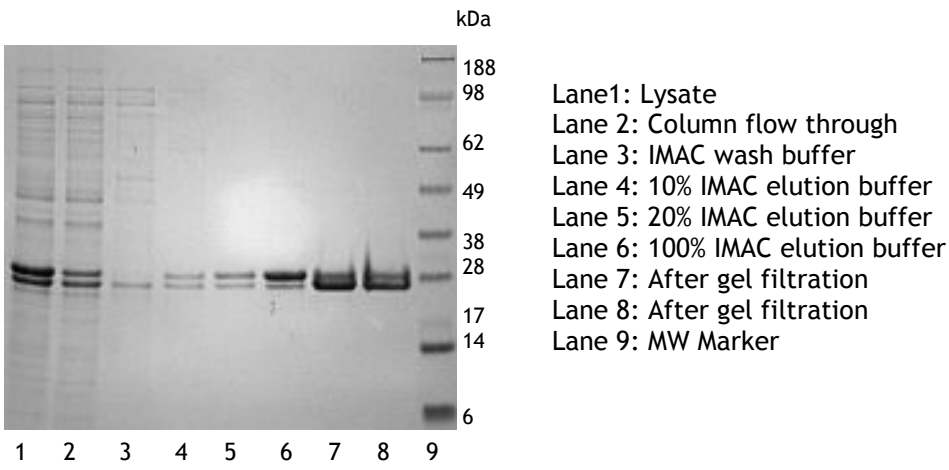


Figure 3.2: SDS-PAGE analysis of purification of DR_1146 by IMAC and SEC

When pooling together fractions D10 to E3 (corresponding to peak 3 of Figure 3.1) for further purification, a pale yellow colour was observed. A yield of 50 mg protein was calculated *via* measurement of absorption at 280 nm (Section 2.5.5).

A faint band below the main DR_1146 band (lane 6) in Figure 3.2 was observed from the time of cell lysis (lane1). During initial purification trials, an ion exchange purification step (Section 2.5.6) was carried out in an attempt to separate these two entities. Although partial separation was possible, continued monitoring of the higher molecular weight separated band showed a decrease in intensity with the reappearance of the lower molecular weight band. This observation confirmed that proteolysis of DR_1146 was occurring.

3.1.4 Gel filtration of DR_1146 and subsequent crystallisation experiments

The IMAC purified DR_1146 was dialysed against a gel filtration buffer of 50 mM Tris-HCl pH 7.5, 200 mM NaCl and was concentrated by centrifugation to 10 mg/mL. It was then purified further by size exclusion chromatography (SEC) using an analytical S200 10/300 HR Global gel filtration column. This was carried out in 500 µL batches as described in Section 2.5.7. Figure 3.3 shows the resultant A_{280} elution profile from the SEC experiment. The large single peak 2 was eluted in a buffer volume of 14.93 mL. Comparison with the column calibration profile suggests a molecular weight of 46.5 kDa, which indicates that DR_1146 exists as a dimer in solution. Although not within the column void volume (7.77 mL) a very small peak seen at 12.61 mL was thought to be due to aggregation of the concentrated protein.

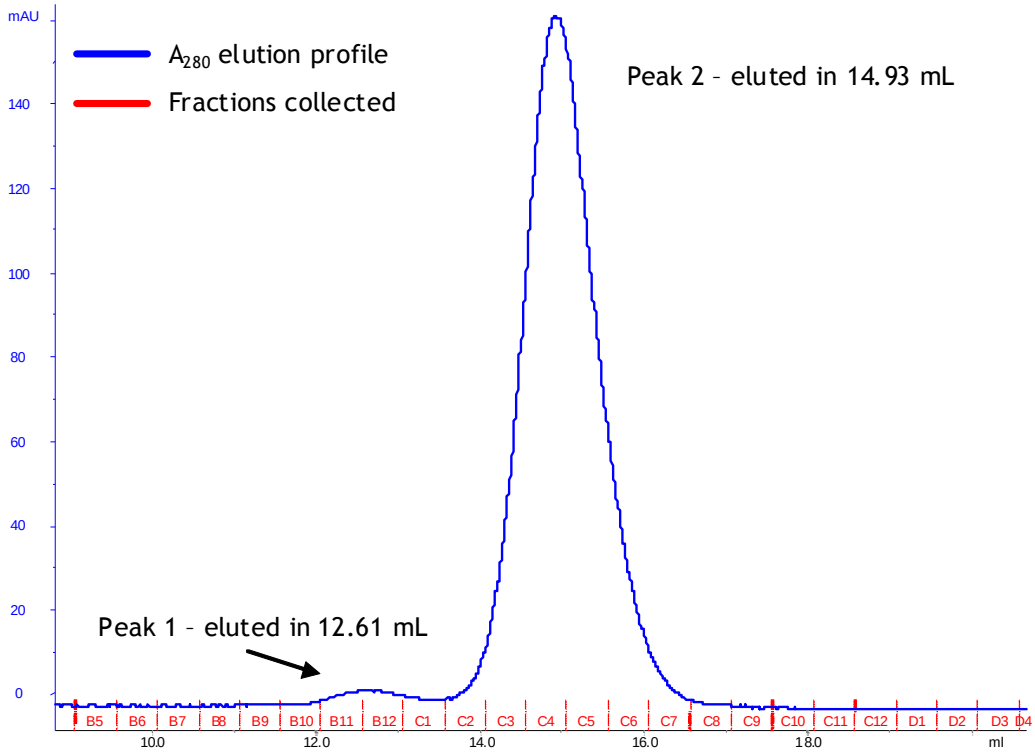


Figure 3.3: A_{280} elution profile for analytical gel filtration experiment to increase purity and determine oligomeric state of DR_1146. It shows a peak eluted in a buffer volume of 14.93 mL suggesting a dimeric conformation in solution. The small peak at 12.61 mL could be attributed to protein aggregation.

While lanes 7 and 8 of Figure 3.2 illustrate further purification of DR_1146 by gel filtration, the increase in intensity of the lower molecular weight band shows the extent of protein degradation. This experiment was carried out two days after cell lysis.

The resulting elution from gel filtration (lanes 7 and 8 of Figure 3.2) was concentrated to 20 mg/mL and crystallisation experiments with all commercial and in-house screens listed in Table 2.10 were set up as described in Section 2.6.3. Trays were stored at 20°C and monitored using a Rhombix Imaging System for duration of three months. The complete experimental process, consisting of IMAC, gel filtration and crystallisation set up was performed in a minimum time period of two days in an attempt to minimise the extent of protein degradation.

After a three-month period of monitoring crystallisation experiments, no protein crystal growth occurred. It was thought that this may be due to either (a) the presence of the 6xHis tag (b) the proteolysis of the target protein or (c) possible aggregation during concentration for crystal trials.

On concentration of the protein for crystal trials an increase in intensity of the yellow colour of DR_1146 was observed.

3.1.5 Non-reduced SDS-PAGE analysis of DR_1146

Lanes 2 and 4 of Figure 3.4 show SDS-PAGE analysis of non-reduced DR_1146 samples, prepared without the addition of 0.3 mM DTT. In comparison with reduced samples (lanes 3 and 5) a strong band at approximately 48 kDa can be seen. The intensity of this band is significantly reduced on addition of 0.3 mM DTT. In combination with the gel filtration result previously described, this mass shift suggests that DR_1146 exists as a dimer.

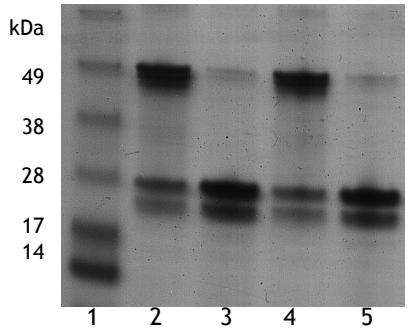


Figure 3.4: SDS-PAGE analysis of non-reduced and reduced samples of purified DR_1146

MALDI-TOF analysis of these samples was inconclusive due to degradation of DR_1146. There were a number of masses in the 23000 Da region that could give rise to a dimer with a molecular mass of about 47000 Da.

3.1.6 Investigating degradation of DR_1146

In an attempt to slow or prevent the observed degradation, the following modifications to the purification protocol were made: (a) additional Complete EDTA-free Protease Inhibitor Cocktail Tablets were added both before and after cell lysis and (b) EDTA was added after IMAC. Neither of these changes resulted in retardation of degradation and subsequently had no impact on crystallisation.

20 µL samples of purified DR_1146 were incubated at 37°C to investigate the extent of proteolysis. The resultant samples were analysed by SDS-PAGE and western blot.

In total, five degraded bands were imaged by SDS-PAGE. Three of the five bands could be seen on both the SDS-PAGE gel and the developed film from the western blot. For these bands, this result suggested that the N-terminal 6xHis tag was still intact after degradation. This implied that the proteolysis reaction was occurring at the C-terminus.

The other two degraded bands could be seen only on the SDS-PAGE gel suggesting that the 6xHis tag needed for detection by western blot was no longer intact. This indicated that some proteolysis was also occurring at the N-terminal of the construct resulting in the removal of the 6xHis tag.

In conclusion proteolysis is probably occurring at both the N and C-terminal of DR_1146.

3.2 Re-cloning of *dr_1146* into pET-15b

3.2.1 Aims and reasoning behind construct design

In order to investigate the impact of controlled 6xHis tag removal on protein crystal growth a full length (FL) and truncated (T) form of the *dr_1146* gene were cloned into a pET-15b vector with a thrombin cleavable, N-terminal 6xHis tag.

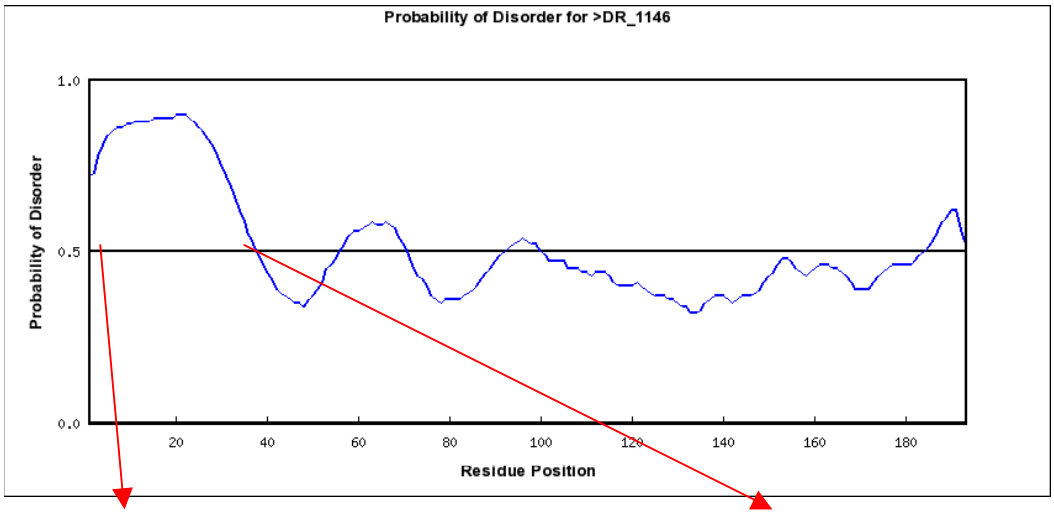
Throughout this chapter, protein encoded from the full length gene with the N-terminal 6xHis tag still intact will be referred to as FL_DR_1146. The protein with the 6xHis tag cleaved using thrombin will be referred to as FL_DR_1146_clvd.

The full length insert (*fl_dr_1146*) consisted of nucleotide residues 1 to 582, taken from the Comprehensive Microbial Resources database (CMR) [6, 7] and is shown in Appendix 1.

An N-terminal truncated version of the *dr_1146* gene (*t_dr_1146*) was created in an attempt to diminish possible occurrence of N-terminal proteolysis observed for the pDEST17 construct. This gene consisted of nucleotide residues 88 to 582 taken from the CMR database [6] and is shown in Appendix 2. The reasoning

behind the design of this insert was based on the analysis of the following results and observations:

(a) The RONN server [183-185] predicted disorder in the first 37 residues of the amino acid sequence of full length DR_1146. Figure 3.5 shows the trace generated by the RONN server, which maps the predicted probability of disorder in the amino acid sequence. The probability of disorder for the first 37 residues (marked in red) is displayed by a probability greater than 0.5. As disorder is considered to be unfavourable for ordered nucleation and subsequent protein crystal growth, it was decided to remove this mobile region.



1 MCGLAHSPDM SDNKNAGQSSDSQGNGAPQMSRDEAVKTVAGLVKGIKFAM

51 LVSVNDAGHIHSRPMTTQETEFDGDIWFIGAKDSEVVHDVRSRPQVNVSY

101 ADTGSNNYVSVHGTAELIEDRAKLDELWSDMYNMYFEGGKEDPNVQLIKI

151 NAEGAEYWESGGKVRTLFAFAKNLIPGQRADASELGKNDTVNL

Figure 3.5: RONN [183-185] disorder prediction and amino acid sequence of DR_1146. RONN plot shows that the first 37 amino acid residues (red) are predicted to be disordered (probability > 0.5). Residue Met30 (blue) is thought to encode the true nucleotide sequence start codon.

(b) Figure 3.5 also shows the predicted full length amino acid sequence of DR_1146, taken from the CMR database [6, 7]. It can be seen that residue 30 of the amino acid sequence is a methionine (marked in blue). It is a possibility this methionine could be encoded by the true start codon of the nucleotide sequence, with the original full length sequence being wrongly annotated on the CMR database [6].

(c) A NCBI BLAST [58, 59] search was carried out in an effort to identify proteins with amino acid sequence homology to DR_1146. Comparisons of the top five homologous proteins begin at around residue 30 of the full length sequence of DR_1146. This information contributes to the hypothesis that the sequence depicted by the CMR database [6, 7] may have been wrongly annotated.

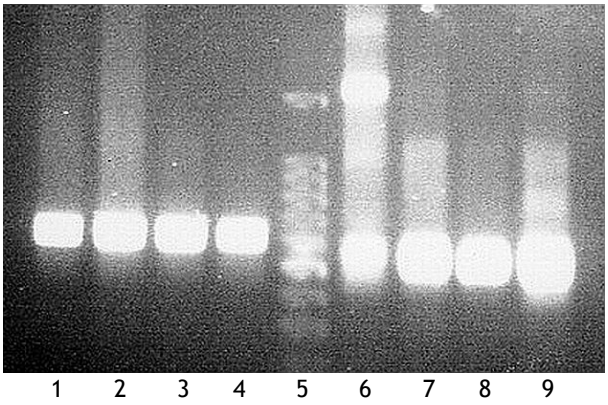
As a result of the above findings, primers were designed to produce an N-terminal truncated construct that would facilitate transcription at the ATG codon (nucleotide residues 88, 89 and 90) encoding amino acid residue 30, the third methionine of the full length sequence.

An *Nde*I recognition site was incorporated into the FL and T forward primers, and a *Bam*HI recognition site was included into the FL and T reverse primers in order to allow insertion into a pET-15b vector. The complete composition of the designed FL and T forward and reverse primers are shown in Table 2.2.

3.2.2 Cloning of full length and truncated dr_1146 into pET-15b using restriction enzymes

The construct DNA of *dr_1146* in pDEST17 was used as the DNA template in the PCR step of the re-cloning experiment. The master mix recipe and temperature cycle used are shown in Section 2.2.3 and Table 2.3.

An optimum annealing temperature of 68°C was established by preliminary experiments resulting in successful amplification of both *fl_dr_1146* and *t_dr_1146* as shown in Figure 3.6. The strong bands in lanes 1-4 correspond to successful amplification of the full length gene fragment of 582 bp and those in lanes 6-9 to that of the N-terminal truncated gene fragment of 495 bp.



Lanes 1-4: *fl_dr_1146* (582 bp)

Lane 5: BenchTop 1 kb DNA ladder

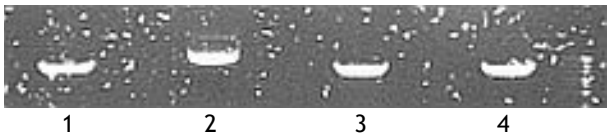
Lanes 6-9: *t_dr_1146* (495 bp)

Figure 3.6: PCR amplification of *fl* and *t_ dr_1146* gene fragments imaged by agarose gel electrophoresis.

These bands were extracted and purified then were inserted into the pcR-Blunt II-TOPO vector (Table 2.1) *via* the protocol outlined in Section 2.2.5 [151]. The success of the insertion reaction was confirmed by an even spread of colonies on an LB-Agar plate after transformation of DH5α cells. Two single colonies for each construct were picked and prepared by Miniprep.

Double digestions of both the TOPO reaction products and the pET-15b vector were carried out following the protocol outlined in Section 2.2.8. Agarose gel electrophoresis confirmed successful digestion of the TOPO reaction products and successful linearisation of pET-15b (Figures 3.8 and 3.9).

Successful linearisation of pET-15b can be seen in lane 1 of Figure 3.7. This is shown by a decrease in size compared to an undigested pET-15b sample in lane 2. It was extracted from the gel and purified as described in Section 2.2.4.



Lane 1: Linearised pET-15b with *Bam*HI and *Nde*I
Lane 2: Negative control – pET-15b with no *Bam*HI or *Nde*I
Lane 3: pET-15b with *Bam*HI only
Lane 4: pET-15b with *Nde*I only

Figure 3.7: Linearisation of pET-15b with *Bam*HI and *Nde*I restriction enzymes.

Lanes 1-3 of Figure 3.8 show the successful release of the full length fragments from the TOPO reaction product after digestion with *Nde*I and *Bam*HI restriction enzymes.

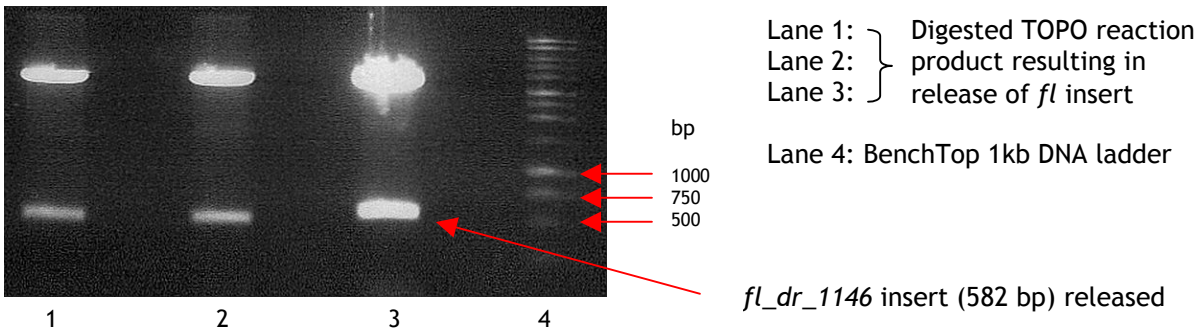


Figure 3.8: Restriction digestion of full length TOPO reaction product resulting in release of *fl_dr_1146* insert.

Lanes 1 and 3 of Figure 3.9 show release of the truncated fragment from the TOPO reaction product after digestion with *Nde*I and *Bam*HI restriction enzymes.

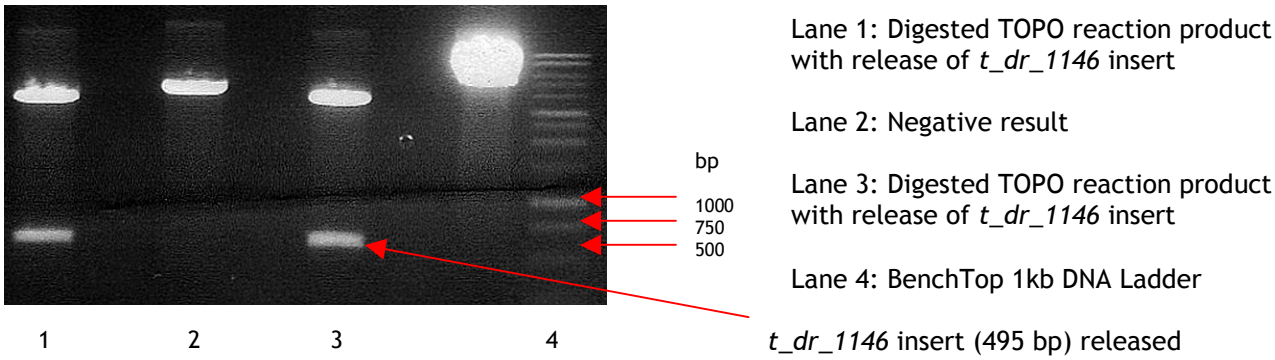


Figure 3.9: Restriction digestion of truncated TOPO reaction product resulting in release of *t_dr_1146* insert

The two inserts for each construct were extracted and purified then ligated into the linearised pET-15b vector as described in Section 2.2.9. All four reactions gave single colonies after transformation into DH5α cells confirming a positive ligation reaction. The newly cloned full length and truncated DNA constructs were prepared by Miniprep.

A successful restriction digestion with *Nde*I and *Bam*HI restriction enzymes confirmed the presence of inserts in the vector for both full length and truncated constructs. DNA sequencing and comparison of results with the CMR database [6, 7] sequence (Appendices 1 and 2) confirmed the correct identity and codon distribution of the inserts.

3.2.3 Expression of fl_ dr_1146 from pET-15b

1 µL *fl_dr_1146* plasmid DNA (*fl_dr_1146* in pET-15b) was used to transform BL21 Star(DE3) competent cells and resulted in an even spread of single colonies after a 20 hr incubation at 37°C.

A 100 x dilution of starter culture was used to initiate cell growth in two 500 mL cultures which were incubated at 37°C until an OD_{600} of 0.66 and 0.7 was reached. The cells were then induced with IPTG and one of the flasks was incubated for at 37°C and one at 22°C for duration of 20 hrs. At 2, 4, 5, 6 and 20 hr time points, an OD_{600} reading was taken and a 1 mL sample from each culture was taken. Theses samples were then normalised to an OD_{600} of 5 and the soluble fractions were analysed by SDS-PAGE as shown in Figure 3.10. Comparison of the bands in lane 6 indicates that a larger amount of protein was produced after 20 hr expression at 22°C rather than at 37°C.

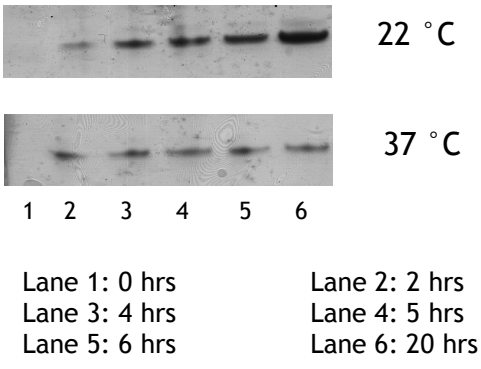


Figure 3.10: SDS-PAGE analysis of time point samples taken during expression trials for FL_DR_1146.

A 3.45 g pellet was harvested from the 22°C 500 mL culture by centrifugation. Expression was scaled up using 500 mL LB cultures in 2 L flasks with 20 hr incubation at 22°C after induction with IPTG.

3.2.4 Optimised purification of FL_DR_1146

Cells from 2 L of culture were lysed by French press and the resulting soluble lysate purified by IMAC using a 5 mL HiTrap HR column. This was attached to the AKTA purifier FPLC as described in Section 2.5.4 and an analogous protocol to that outlined in Table 3.1 was followed. To ensure optimum removal of

impurities with each increase of % IMAC elution buffer, the gradient was held until the A_{280} elution profile returned to baseline after each peak.

A high yield of 70 mg of protein was eluted with 100 % IMAC elution buffer resulting in a large, broad peak of 2500 mAU. As with purified DR_1146 expressed from pDEST17, a yellow colour was observed. Analysis of the protein eluted by SDS-PAGE showed a strong single band corresponding to FL_DR_1146 at approximately 24-25 kDa.

During initial purification trials it was established that exceeding the binding capacity of the 5 mL HiTrap HR column greatly improved the purity of target protein by reducing the amount of non-specific binding of competing proteins. Binding lysate from 2 L of cell culture resulted in a high level of purity conducive to crystallisation trials.

The purified protein was concentrated by centrifugation and a PCT test was carried out as described in Section 2.6.2. As a result, a full set of commercial and in-house crystallisation screens (Table 2.10) were set up at 20 and 10 mg/mL. They were incubated at 20°C and monitored by the Rhombix Imaging System.

After a period of three months, no protein crystals were found and therefore an additional SEC purification step was employed. This was carried out in 500 µL batches using a S200 10/200 HR Global column following the method described in Section 2.5.7. Figure 3.11 is an SDS-PAGE analysis of FL_DR_1146 elutant from the gel filtration run.

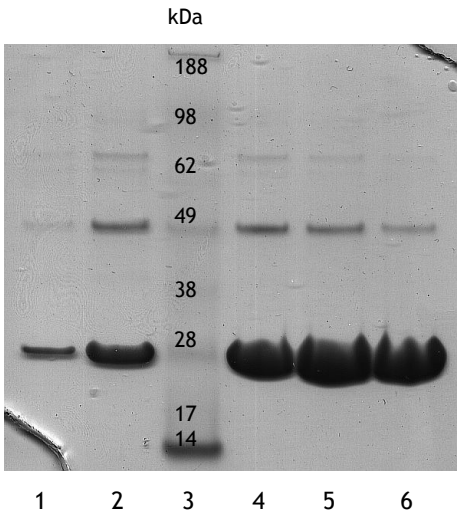


Figure 3.11: SDS-PAGE analysis of purification of DR_1146 by SEC.

Although the gel is highly overloaded, it highlights the high level of purity obtained. The band that can be seen at approximately 49 kDa represents the protein as a dimer; also described for the pDEST17 construct.

The A_{280} profile of the gel filtration purification showed a single peak eluted in a volume of 14.95 mL. Comparison with the column calibration profile suggests a molecular weight of 46.1 kDa, which indicates that FL_DR1146, like the pDEST17 construct, exists as a dimer in solution. In contrast to analytical SEC experiments for the pDEST17 construct, the previously observed aggregation peak was absent. This suggests that the occurrence of protein aggregation before SEC may have been eliminated by changing the expression vector from pDEST17 to pET15b.

Although the purity of FL_DR_1146 was very high, subsequent crystallisation experiments were still unsuccessful. Although less rapid than for the FL_DR_1146 pDEST17 construct, proteolysis was still evident after a period of 3 days.

3.2.5 Removal of N-terminal 6xHis tag by Thrombin cleavage

Small-scale experiments established that 0.25 units of thrombin were sufficient to effectively cleave 1 mg of fusion protein after overnight incubation at 4°C.

The cleavage reaction was carried out as outlined in Section 2.5.8. Small scale experiments showed that the presence of imidazole resulted in a retardation of

the cleavage reaction. In order to overcome this problem, the cleavage reaction mixture was dialysed against 20 mM Tris pH 7.5, 200 mM NaCl overnight.

Lane 2 of Figure 3.12 shows a decrease in the MW of the FL_DR_1146 band as a result of a successful cleavage reaction. The cleaved 6xHis tag can be seen in lane 2 at approximately 4 kDa.

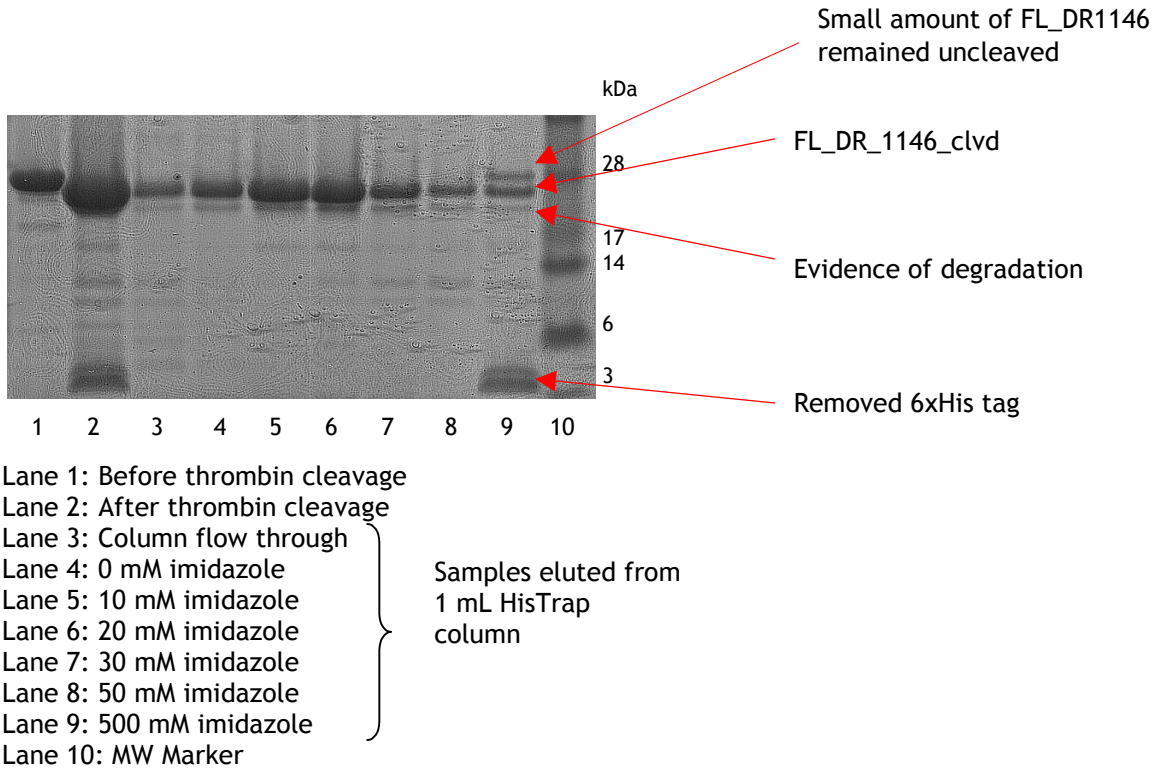


Figure 3.12: SDS-PAGE analysis of FL_DR1146_clvd purified by IMAC.

After the completion of the cleavage reaction, a Complete EDTA-Free Protease Inhibitor Cocktail Tablet was added to the cleavage mixture to stop further thrombin proteolysis. In order to remove the cleaved tag, it was passed through a 1 mL HisTrap column as described in Section 2.5.4. Figure 3.12 shows that the majority of cleaved FL_DR_1146 (FL_DR_1146_clvd) was eluted from the column after 0, 10, 20 and 30 mM imidazole washes. The 6xHis tag, traces of un-cleaved protein and impurities (lane 9) were successfully separated from the target protein by this procedure.

After dialysis to remove NaCl and traces of imidazole, the eluted FL_DR_1146_clvd was purified further by ion exchange chromatography to remove the remaining thrombin. This technique (Section 2.5.6) relied on a

difference in isoelectric point (pI) between FL_DR_1146_clvd (predicted pI of 4.90) and that of thrombin (pI = 7.1).

A PCT and subsequent crystallisation experiments were carried out on the purified FL_DR_1146_clvd as described in previous Sections. However, again no credible protein crystals were obtained.

The FL_DR_1146_clvd bands in Figure 3.12 again show evidence of degradation after tag removal. Attempts to prevent or slow this, as described for the FL_DR_1146 pDEST 17 construct, were unsuccessful.

It was concluded that the change of vector and subsequent removal of 6xHis tag had no impact on the crystallisation of FL_DR_1146. This lead to the hypothesis that either (a) proteolysis was continuing to occur at the N-terminal of the protein after 6xHis tag removal, (b) it was occurring at the C-terminal of the protein or (c) a combination of proteolysis at both terminals was occurring.

3.2.6 Expression and purification of t_ DR_1146 from pET-15b

After successful transformation of BL21 Star(DE3) cells using *t_dr_1146* pET15-b construct DNA, soluble protein was expressed following the protocol established for the *fl_dr_1146* pET-15b construct. This included 20 hr incubation at 22°C, in LB media, after induction with IPTG.

2 L of cell pellet was lysed by French press and the resulting lysate was purified by IMAC following the same method employed for the full length protein expressed from pET-15b. However, 6xHis tagged T_DR1146 had a lower affinity of the Ni^{2+} matrix incorporated in the 5 mL HiTrap column. This resulted in elution of the majority of the tagged protein with 10 and 20 % IMAC elution buffer rather than with 100 % IMAC elution buffer as seen for the previous construct. This is shown in Figure 3.13 where peaks 1 (2400 mAU) and 2 (950 mAU) are very broad in comparison with peak 3.

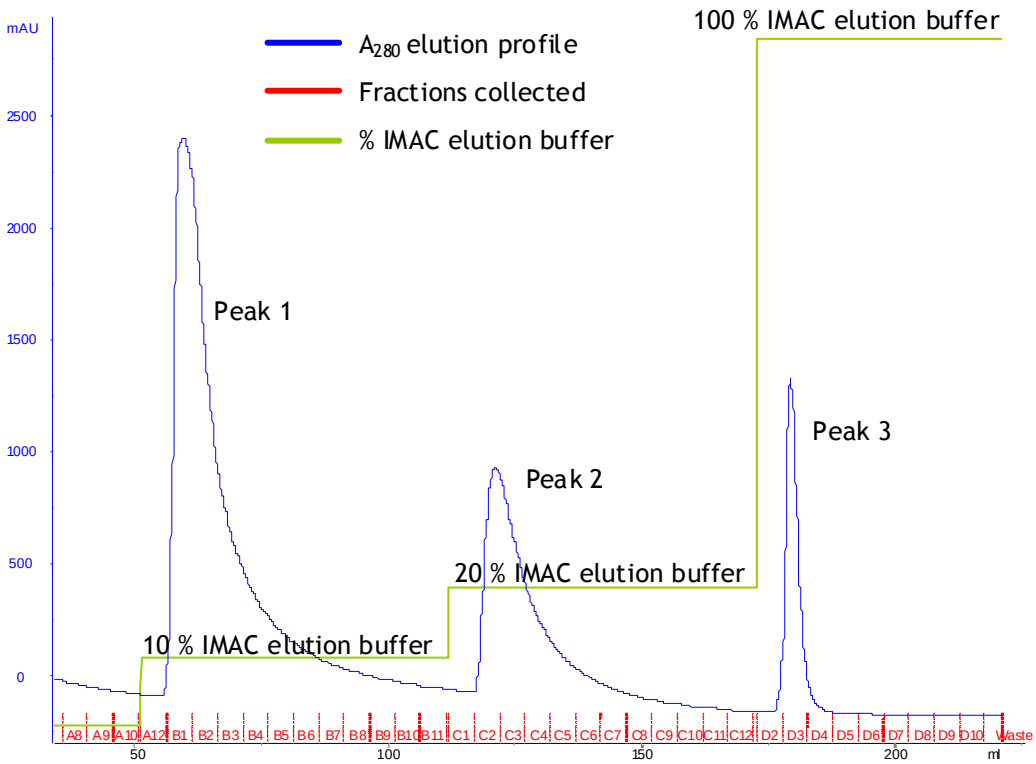
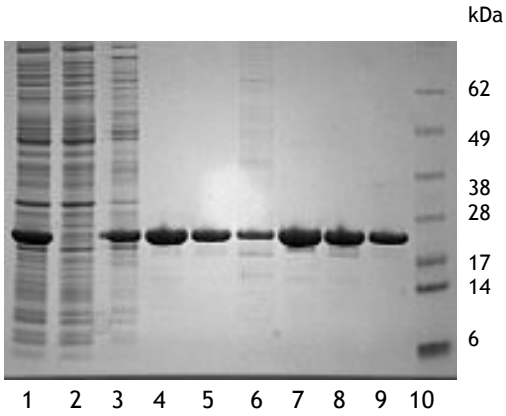


Figure 3.13: A_{280} elution profile for optimised IMAC purification of T_DR_1146. Peaks 1 and 2 correspond to the elution of T_DR_1146 with 10 and 20 % IMAC elution buffer.

SDS-PAGE analysis (Figure 3.14, lanes 4 and 5) confirmed the strong presence of T_DR_1146 in peaks 1 and 2 of Figure 3.13.



Lane 1: Lysate
Lane 2: Column flow through
Lane 3: IMAC wash buffer
Lane 4: 10 % IMAC elution buffer
Lane 5: 20 % IMAC elution buffer
Lane 6: 100 % IMAC elution buffer
Lane 7: After gel filtration
Lane 8: After gel filtration
Lane 9: After gel filtration
Lane 10: MW Marker

Figure 3.14: SDS-PAGE analysis of purification of T_DR_1146 by IMAC and SEC.

By exceeding the binding capacity of the 5 mL HiTrap HR column, non-specific binding was prevented, resulting in very pure protein elution for protein corresponding to peaks 1 and 2. Unlike previous IMAC purifications, the T_DR_1146 protein eluted in peak 3 (lane 6) was less abundant and impure.

When pooled, the protein shown in lanes 4 and 5 of Figure 3.14, gave a yellow solution. A total yield of 120 mg of protein was calculated by measuring absorbance at 280 nm.

An additional SEC step was employed in an effort to further increase purity. The resultant protein purity is illustrated in lanes 7, 8 and 9 of Figure 3.14. The resulting A_{280} profile (Figure 3.15) shows a single peak eluted with 15.13 mL of gel filtration buffer. Comparison with the column calibration profile suggests a molecular weight of 42.2 kDa, which again indicates that T_DR_1146 exists as a dimer in solution.

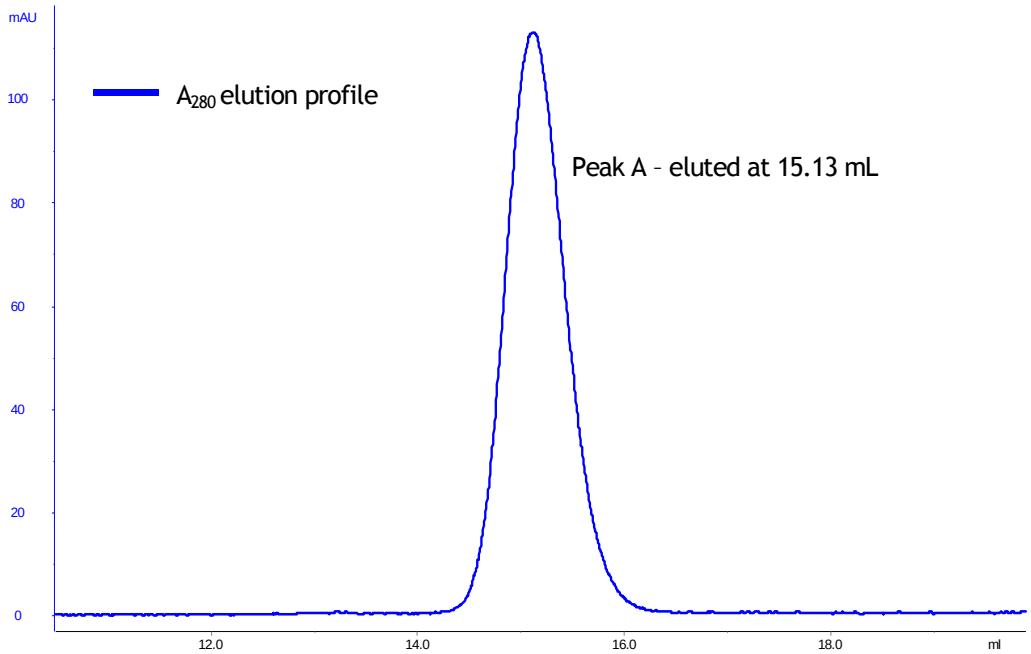


Figure 3.15: A_{280} elution profile for analytical gel filtration experiment to determine oligomeric state of T_DR_1146 shows a single peak eluted in a buffer volume of 15.13 mL suggesting that the protein exists as a dimer in solution.

It is evident from lanes 7, 8 and 9 of Figure 3.14 that a small amount of degradation had occurred after a three-day period. Although still a problem, it appeared to be less rapid than for the FL constructs.

Following PCT, crystallisation experiments were set up for two concentrations of 28 and 9 mg/mL. They were stored and monitored as described in Section 2.6.3.

After three months no protein crystal were found in any of the wells. However, precipitation appeared more crystalline and less amorphous at both concentrations compared to experiments set up for the FL protein.

3.2.7 Removal of 6xHis tag from T_DR_1146

The 6xHis tag was cleaved following the same method established for the FL protein. Again 0.25 units of thrombin per mg of fusion protein enabled efficient tag removal after overnight dialysis at 4°C.

It was decided that the second IMAC step included in the purification of FL_DR_1146_clvd was unnecessary and therefore both the tag and thrombin were removed by ion exchange.

The difference in molecular weight between the bands in lanes 2 and 3 of Figure 3.16 demonstrates effective removal of the 6XHis tag after the overnight incubation at 4 °C. In lanes 3 and 4, the cleaved 6xHis tag is represented by two close bands seen at approximately 4 kDa. Lanes 6-9 show the T_DR_1146_clvd protein after ion exchange and the absence of the double band at 4 kDa confirms successful removal of the 6xHis tag by this method.

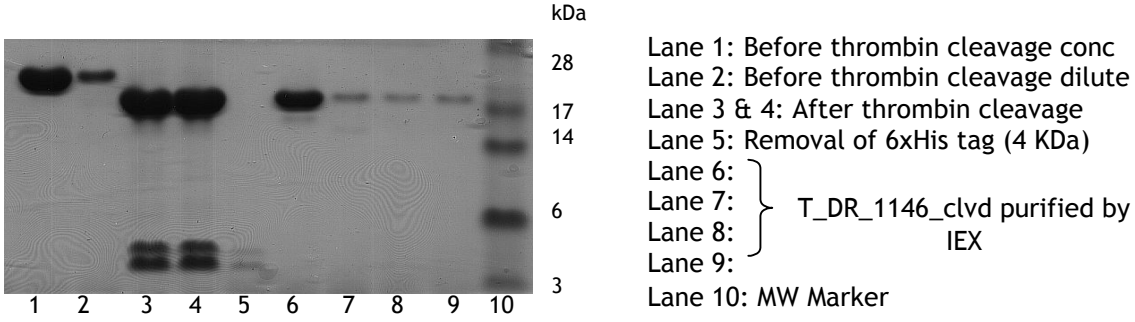


Figure 3.16: SDS-PAGE analysis of thrombin cleavage of T_DR_1146 and removal of 6xHis tag and thrombin by ion exchange chromatography (IEX).

Crystal trials were again compiled, using 18 and 9 mg/mL protein concentrations. After the three months there was no evidence of protein crystal growth, however, as with the tagged T_DR_1146, the precipitate, where it occurred in the crystallisation drops, was more crystalline and less amorphous than for previous experiments. The N-terminal truncation appeared more favourable in terms of protein crystallisation and nucleation.

After five days of incubation at 4°C, faint signs of degradation of the T_DR_1146_clvd protein was observed. Although, the onset of proteolysis appeared to be less rapid than for the other constructs, its occurrence still

proved problematic for successful crystallisation. It was proposed that proteolysis was occurring at either or both the C and N-terminals after tag removal.

As T_DR_1146 appeared to undergo less degradation than FL_DR_1146, the truncated construct was selected for biophysical characterisation.

4 Biophysical characterisation of T_DR_1146

Throughout the remaining Chapters, protein encoded from the *t_dr_1146* gene with the N-terminal 6xHis tag still intact will be referred to as T_DR_1146. Protein samples with the 6xHis tag removed using thrombin will be referred to as T_DR_1146_clvd. Protein samples with FMN added during purification will be referred to as 'holo' and those without, as 'apo'.

4.1 Introduction and research aims

Throughout the purification processes discussed in Chapter 3, DR_1146 was observed to be pale yellow in colour. It was therefore postulated that the protein was binding to a yellow prosthetic group that was present in the LB media or alternatively, that was produced by either *E. coli* enzymes or the target DR_1146 protein during expression.

A UV absorption spectrum of yellow T_DR_1146_clvd at 0.275 mg/mL shows small absorbance maxima at 375 and 461 nm (Figure 4.1) which are indicative of flavin as a substrate in flavin binding proteins [94].

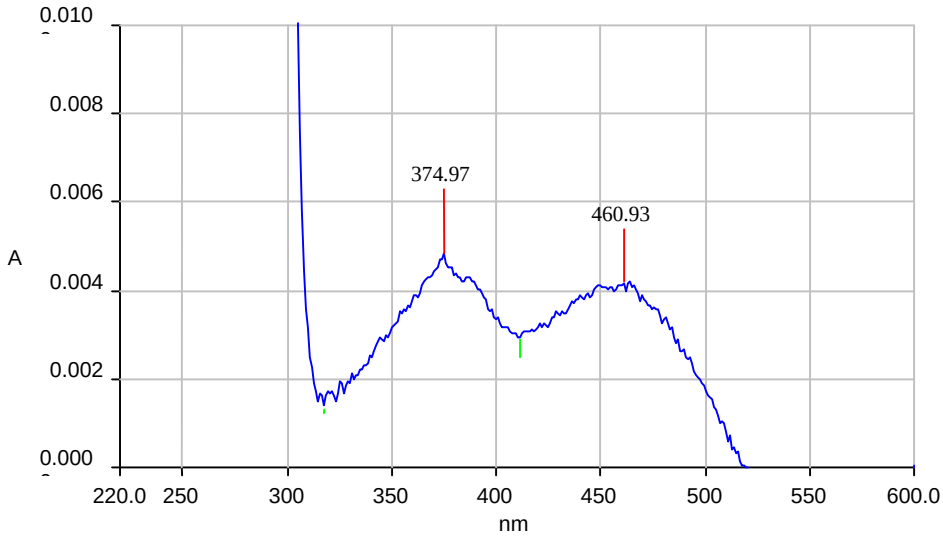


Figure 4.1: UV absorption spectrum for T_DR_1146_clvd (at 0.275 mg/mL) showing small absorbance maxima at 375 and 461 nm, which are characteristic of flavin present in a flavin binding proteins [94].

Addition of an excess of the flavin compound riboflavin 5'-phosphate dihydrate sodium salt to a sample of T_DR_1146_clvd resulted in a bright yellow solution.

After overnight dialysis and gel filtration, fractions identified as containing T_DR_1146_clvd remained bright yellow. Absorbance spectra of the gel filtered sample showed an increase in maxima at 375 and 461 nm, and a shift in spectral maxima from 279 nm to 272 nm. These observations suggested that (a) the flavin was binding to T_DR_1146_clvd and (b) that in the native state the protein was not fully saturated by flavin.

In an attempt to confirm the identity of the yellow compound bound to the protein during expression, the chromophore was extracted as described in Section 2.5.11. Analysis by electrospray ionisation mass spectroscopy (ESI-MS) showed a peak at 508.30 m/z (Figure 4.2).

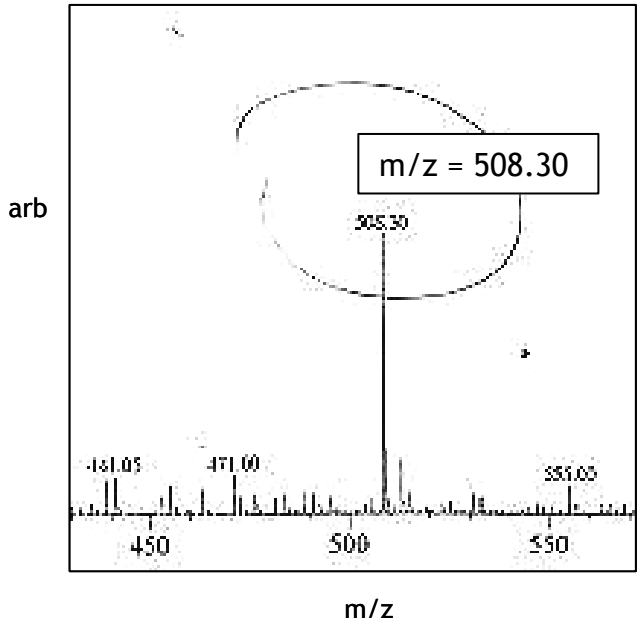


Figure 4.2: Positive ion ESI-MS of yellow prosthetic group extracted from T_DR_1146 showing a peak at 508.30 m/z.

In comparison with molecular weights of various flavin compounds (Table 4.1) the m/z value was closest to that of riboflavin 5'-phosphate dihydrate sodium salt at 514.36 g/mol. A sample of this compound was prepared using methanol and analysed by ESI-MS for comparison. The spectra (not shown) showed a comparable peak at 508 m/z suggesting that the extracted chromophore to be more than likely a hydrated form of riboflavin 5'-phosphate sodium salt.

Flavin compound	Molecular weight (g/mol)
Riboflavin	376.36
Riboflavin 5'-phosphate	456.34
Riboflavin 5'-phosphate sodium salt	478.33
Riboflavin 5'-phosphate dihydrate sodium salt	514.36
Flavin adenine dinucleotide	785.55
Flavin adenine diphosphate disodium salt	829.51

Table 4.1: Molecular weights of various flavin compounds.

Throughout this study the abbreviations FMN and FAD will be used to represent Riboflavin 5'-phosphate dihydrate sodium salt and Flavin adenine diphosphate disodium salt respectively.

A BLAST search [58, 59], using the predicted amino acid sequence of T_DR_1146 from the CMR database [6, 7], identified 58 % sequence homology with a putative pyridoximine 5'-phosphate oxidase related FMN binding protein from *Deinococcus geothermalis*. A search of the Protein Data Bank [79, 80] found the structure of a putative general stress protein (COG3871) from *Nostic punctiforme* PCC 73102 (PDB ID: 2I02) [81] which shares 32 % sequence identity with T_DR_1146. This dimeric protein was crystallised in the presence of a single FMN molecule bound in a hydrophobic pocket at the dimer interface (discussed in more detail in Chapter 5). The FMN binding ability of these homologues fuelled further investigation into the flavin binding capacity of T_DR_1146, and the subsequent effects on protein stabilisation, structure and crystallisation.

In this Chapter the investigation of flavin binding to T_DR_1146 and T_DR_1146_clvd using isothermal titration calorimetry (ITC) and fluorescence spectrometry is discussed. The effect of FMN on thermal stability was assessed using differential scanning calorimetry (DSC), and circular dichroism (CD) experiments were used to monitor the effects of FMN binding on chemical stability and protein structure. The oligomerisation behaviour of holo T_DR_1146_clvd in solution was investigated by analytical centrifugation (AUC) and preliminary NMR experiments were carried out to assess the feasibility of 3-D structure determination in solution.

4.2 Isothermal Titration Calorimetry (ITC)

4.2.1 Experimental aims

The aims of ITC experiments carried out on T_DR_1146 and T_DR_1146_clvd were (a) to confirm that the protein binds to a flavin cofactor (FMN), (b) to investigate the effect of 6xHis tag removal on association (K_A) and dissociation (K_D) constants and (c) to compare binding affinities and stoichiometries (N values) for three flavin compounds, riboflavin, FMN and FAD.

4.2.2 ITC experiments on T_DR_1146 and T_DR_1146_clvd with FMN

Experiments were carried out as described in Section 2.7.2 with injections of T_DR_1146 (18.5 μM) and T_DR_1146_clvd (18.7 μM) samples into 0.35 mM FMN buffer solutions held in the ITC cell. The resulting data showed typical heat pulses consistent with exothermic binding of the FMN cofactor to the protein (Figure 4.3).

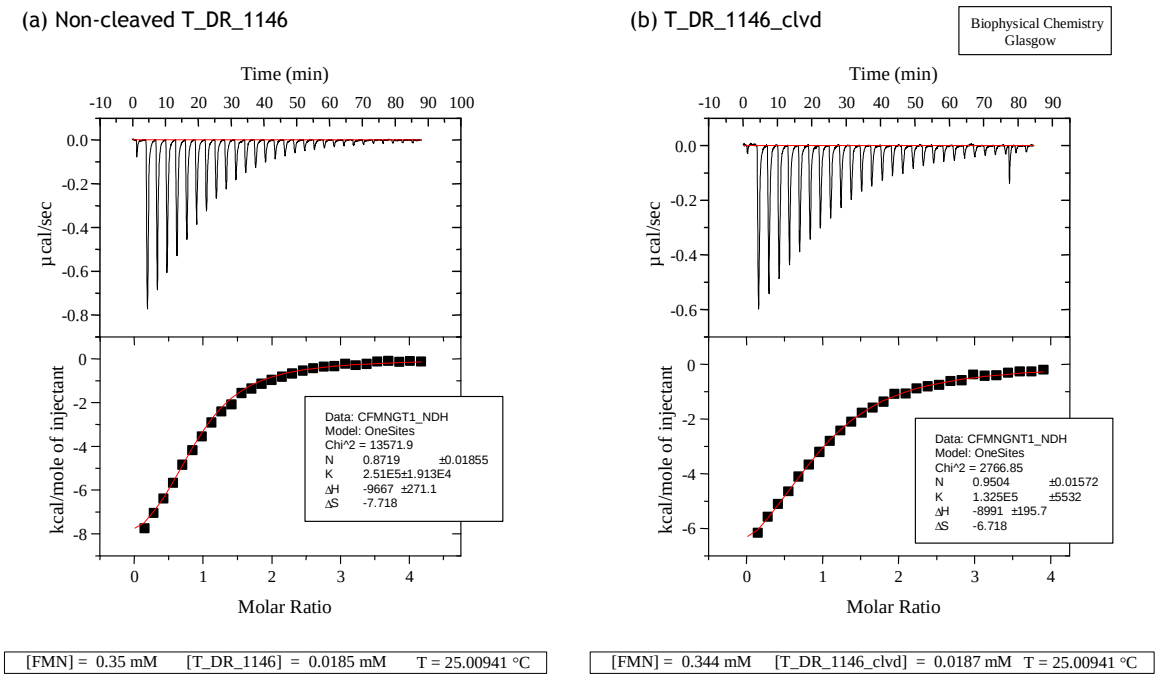


Figure 4.3: ITC data for binding of (a) non-cleaved T_DR_1146 and (b) T_DR_1146_clvd to FMN at 25°C. The upper panel shows the raw data (exothermic heat pulses) for protein-cofactor binding. The lower panel shows the differential binding curve derived from the integration of heat effects (with dilution heats subtracted), fitted to a standard one-site binding model.

The binding parameters determined for both T_DR_1146 and T_DR_1146_clvd are listed in Table 4.2.

Parameters	T_DR_1146	T_DR_1146_clvd
N (no. of sites per mole of protein)	0.87 ± 0.02	0.95 ± 0.02
$K_A/10^5\ M^{-1}$	2.51 ± 0.19	1.33 ± 0.06
$K_D/10^{-6}\ M$	3.98 ± 0.30	7.55 ± 0.34
ΔH° kcal/mol	-9.67 ± 0.27	-8.99 ± 0.20

Table 4.2: Comparison of ITC binding parameters for the addition of T_DR_1146 and T_DR_1146_clvd to FMN (0.35 μM) in 20 mM Tris pH 7.5, 150 mM NaCl buffer at 25 °C.

The data extracted for both T_DR_1146 and T_DR_1146_clvd fitted well to a standard one-site binding model giving no significant difference in K_D values suggesting that the presence of the 6xHis tag causes no significant change in binding affinity.

N *v*alues of 0.87 ± 0.02 and 0.95 ± 0.02 for non-cleaved and cleaved samples respectively, are consistent with 1:1 binding stoichiometry with one FMN molecule bound per monomer (two per dimer). This result suggests that, unlike the homologous dimeric structure of 2I02 (PDB ID) [81] which has only one FMN molecule bound at the dimer interface, T_DR_1146 has two FMN molecules bound per dimer.

4.2.3 ITC experiments on T_DR_1146_clvd with riboflavin, FMN and FAD

After establishing that FMN binds to T_DR_1146, it was decided to investigate binding to other flavin compounds, namely riboflavin and FAD. Experiments were carried out as described in Section 2.7.2, with separate injections of FMN, Riboflavin and FAD (dissolved in a buffer of 20 mM Tris pH 7.5, 150 mM NaCl) into samples of T_DR_1146_clvd held in the cell. The experimental concentrations of protein and cofactors used are shown in Table 4.3. Due to low aqueous solubility, the concentration of riboflavin (and that of the protein sample) had to be halved. This allowed comparisons to be made with results gained for FMN and FAD.

Parameters/cofactor	Riboflavin	FMN	FAD
T_DR_1146_clvd concentration (mM)	0.01	0.02	0.02
Cofactor concentration (mM)	0.27	0.38	0.43
N (no. of sites per mole of protein)	1.24 ± 0.03	0.95 ± 0.03	1.02 ± 0.06
K_A /10^5 M^{-1}	2.10 ± 0.13	1.17 ± 0.09	1.21 ± 0.20
K_D/10^{-6} M	4.75 ± 0.30	8.57 ± 0.66	8.24 ± 1.37
ΔH° kcal/mol	-11.14 ± 0.34	-11.72 ± 0.46	-7.49 ± 0.60

Table 4.3: Comparison of ITC binding parameters for the addition of Riboflavin, FMN and FAD to T_DR_1146_clvd in 20 mM Tris pH 7.5, 150 mM NaCl buffer at 25 °C.

The ITC results for riboflavin and FAD showed very similar heat pulse patterns (consistent with exothermic binding) to that of FMN (Figure 4.4).

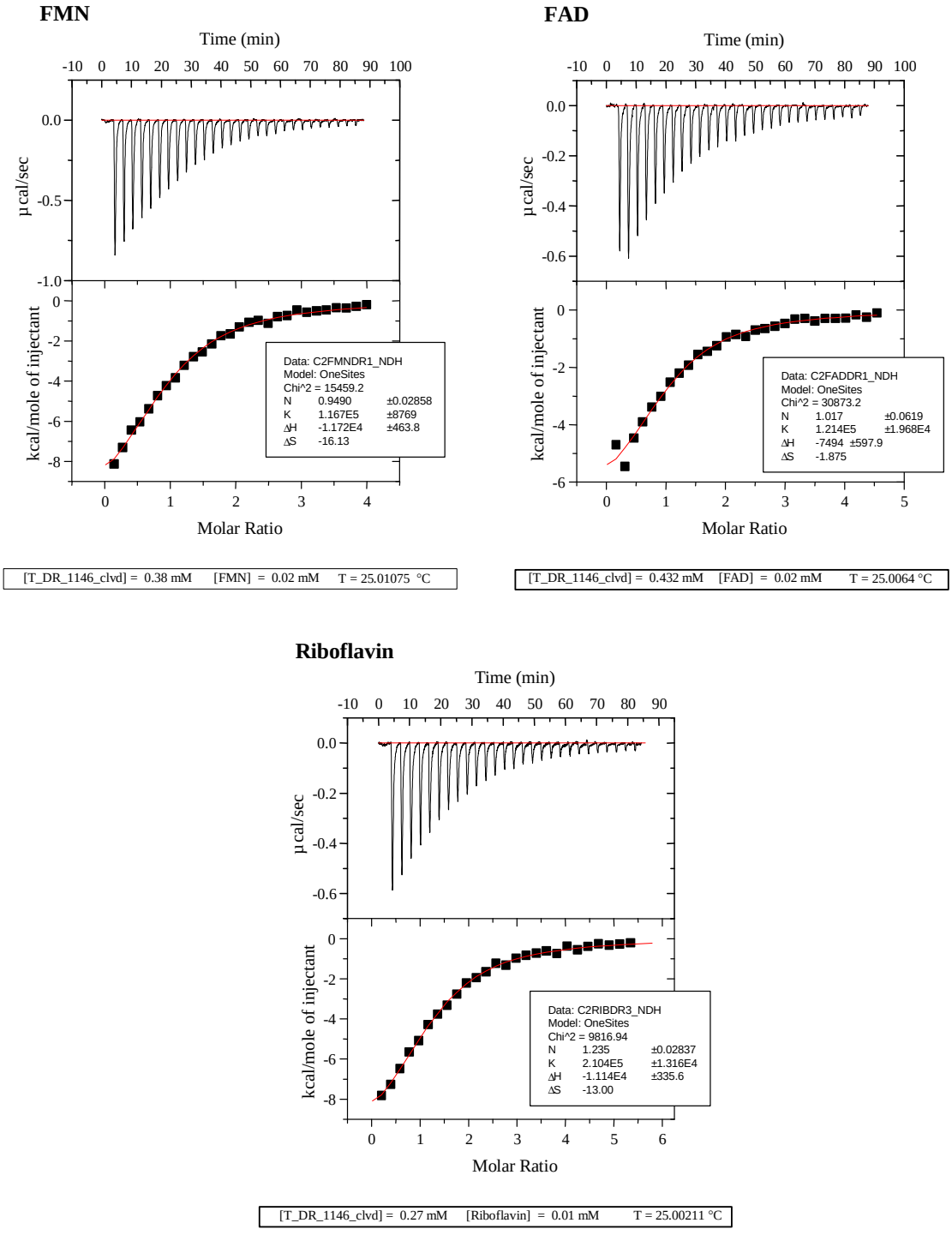


Figure 4.4: ITC data for binding of FMN, FAD and riboflavin cofactors to T_DR_1146_clvd at 25°. The upper panel shows the raw data (exothermic heat pulses) for cofactor-protein binding. The lower panel shows the differential binding curve derived from the integration of heat effects (with dilution heats subtracted), fitted to a standard one-site binding model with parameters shown in Table 4-3. Patterns of heat pulses, binding curves and calculated parameters are very similar for all three flavin cofactors suggesting there is no significant difference in binding affinity for T_DR_1146_clvd.

The values of K_A, K_D and N determined for each flavin cofactor are compared in Table 4.3. K_D parameters for all three flavins were very close within the micromolar range thus no significant difference in binding affinity could be

concluded. The K_D value determined for FMN was very close that determined in initial experiments thus showing consistency for different T_DR_1146_clvd preparations.

Values of N for all three flavins are close to 1, again confirming 1:1 binding stoichiometry.

The structures of all three flavins consist of an aromatic isoalloxazine ring system and a ribityl hydroxyl chain ending with a prosthetic group (-X), which differs for each cofactor (Figure 4.5).

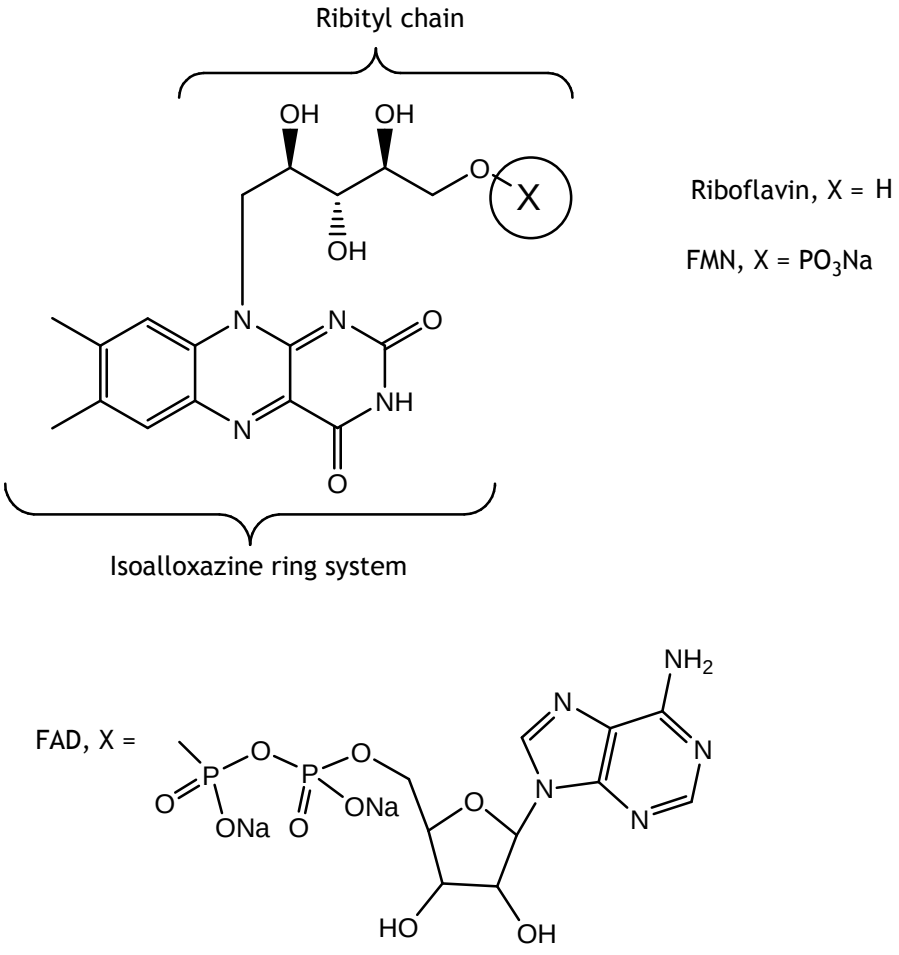


Figure 4.5: Structures of riboflavin, FMN and FAD cofactors used for ITC experiments on T_DR_1146_clvd. All three structures include an isoalloxazine ring system and ribityl chain with a prosthetic group (–X) for each flavin.

ITC experiments indicated that there is no significant discrimination between K_D values determined for riboflavin, FMN and FAD, which suggests that the presence, identity and size of the prosthetic –X group has no effect on binding affinity. It can therefore be postulated that the protein may have a binding

pocket in which the isoalloxazine ring and ribityl hydroxyl components of the flavins are bound, whilst the prosthetic –X group remains free.

4.2.4 ITC experiments on T_DR_1146_clvd with PLP

It has been shown that the pyridoximine 5'-phosphate oxidase protein from *E. coli* is able to bind one or two molecules of pyridoxal 5'—phosphate (PLP) in both its apo and holo (with FMN) forms [186]. As T_DR_1146_clvd shares amino acid sequence homology with a putative pyridoximine 5'-phosphate oxidase FMN binding protein (discussed in Section 1-1), it was decided to investigate PLP binding in order to give an insight into functionality.

In a separate ITC experiment, 0.344 mM PLP buffer solution was injected into 18.7 μM apo T_DR_1146 as described in Section 2.7.2. Figure 4.6 shows the raw data extracted from the experiment.

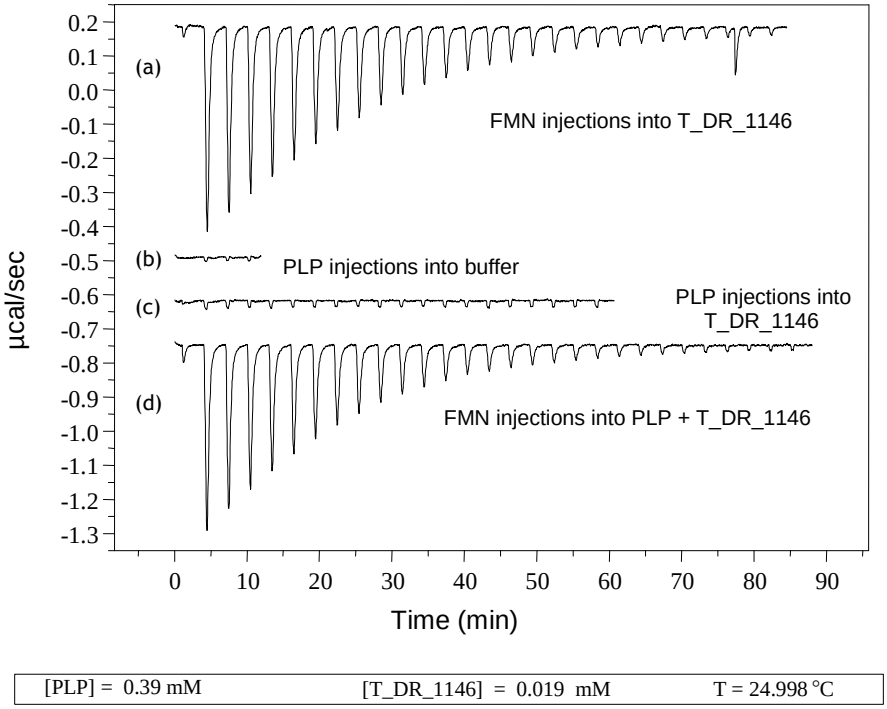


Figure 4.6: ITC data for addition of (a) FMN injected into T_DR_1146, (b) PLP injected into buffer, (c) PLP injected into T_DR_1146 and (d) FMN injected into PLP + T_DR_1146. Results indicate that apo T_DR_1146 does not bind PLP and that the presence of PLP does not affect FMN binding.

In contrast to the pattern of exothermic heat pulses gained for binding of FMN to T_DR_1146 (Figure 4.6a), addition of PLP to the protein (Figure 4.6c) resulted in

a pattern similar to that gained from injection into buffer only (Figure 4.6b). It was therefore concluded that unlike *E. coli* pyridoximine 5'-phosphate oxidase, apo T_DR_1146 does not bind PLP. This suggests that T_DR_1146 does not function as a pyridoximine 5'-phosphate oxidase-like protein, although a functional assay would be required to corroborate this. The protein sample with PLP added was then titrated against FMN (Figure 4.6d). A typical exothermic binding pattern was obtained suggesting that the presence of PLP does not affect FMN binding.

4.3 Fluorescence Spectroscopy (FS)

4.3.1 Experimental aims

The aims of FS experiments on T_DR_1146 and T_DR_1146_clvd samples were (a) to confirm binding to FMN and compare K_D and N values with those gained from ITC experiments and (b) to establish if the pale yellow colour of the apo protein observed during initial purification trials (before the addition of excess FMN) is due to a bound flavin or 'FMN-like' compound.

4.3.2 Fluorescence quenching of FMN on binding to T_DR_1146

Fluorescence emission and excitation spectra for the addition of T_DR_1146 and T_DR_1146_clvd to 10 μM FMN solution were recorded as described in Section 2.7.3.

After recording the intrinsic fluorescence of the FMN solution alone, six 10 μL aliquots of T_DR_1146 stock solution (1.14 mM) were added to the solution. The immediate effect on addition of T_DR_1146, was a decrease or quenching of FMN fluorescence intensity confirming that the non-cleaved protein binds to the flavin cofactor (Figure 4.7).

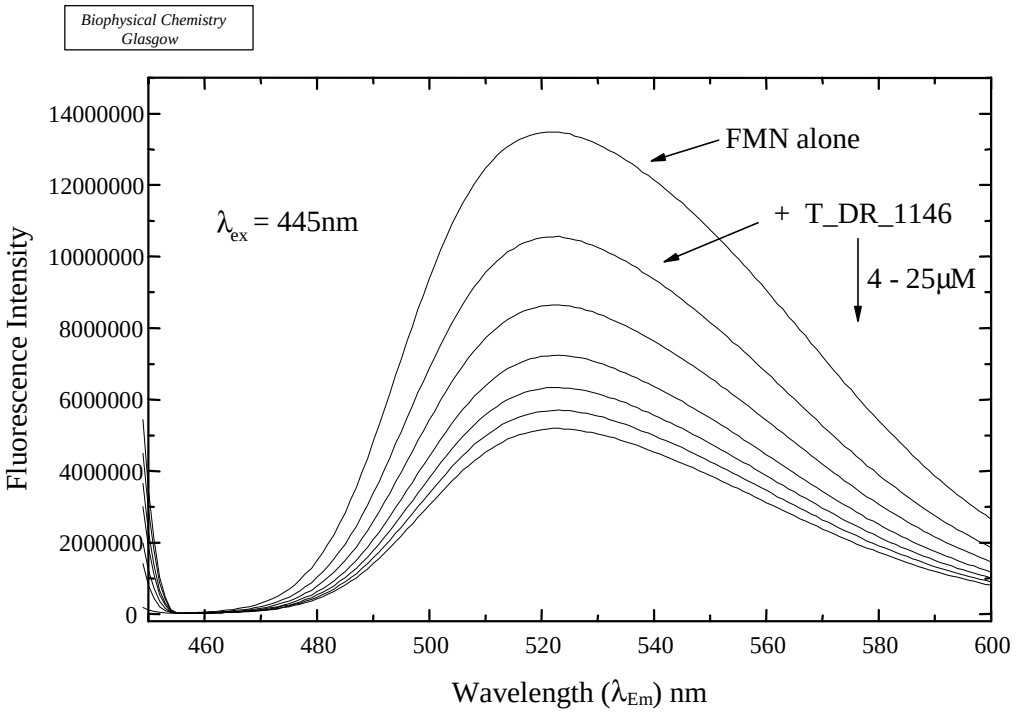


Figure 4.7: Quenching of FMN intrinsic fluorescence (excitation wavelength = 445nm) by increasing concentrations of T_DR_1146. 10 µl aliquots of protein (1.14 mM) added to 2.5ml FMN (10 µM) in 20 mM Tris-HCl, 150 mM NaCl buffer at room temperature.

A K_D of 7.95 µM $\pm$1.48 was determined by fitting data to standard hyperbolic binding curve (Figure 4.8). The number of binding sites (N) was established as 0.65$\pm$0.14, which is lower than the expected 1:1 molar binding ratio obtained from ITC experiments.

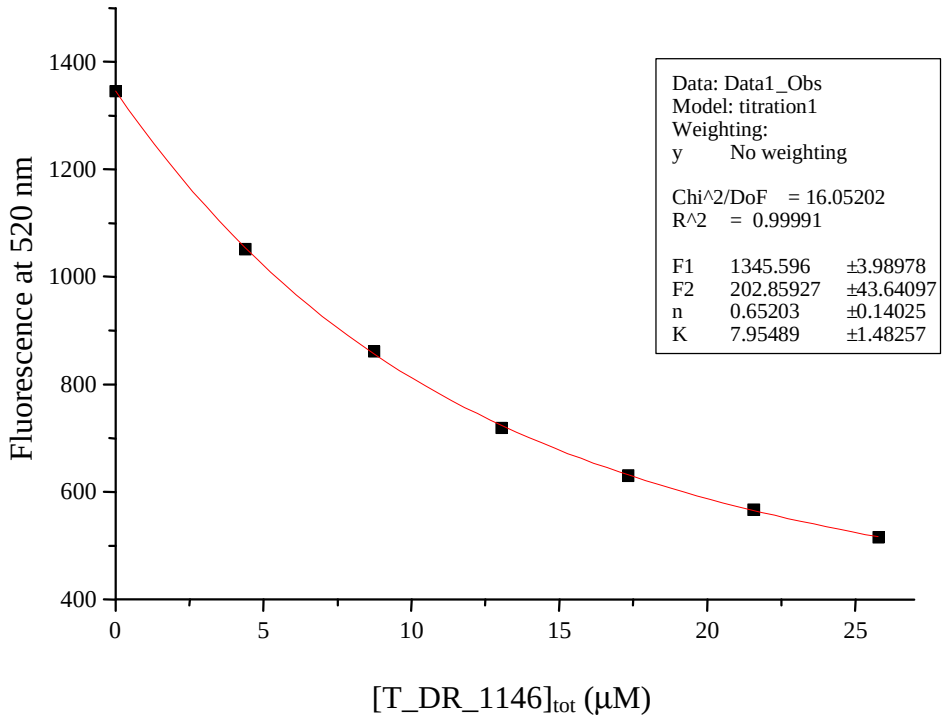


Figure 4.8: Data fitted to standard hyperbolic binding curve to determine K_D (7.95 µM ± 1.48) and N (0.65 ± 0.14) values for quenching of FMN fluorescence by increasing concentrations of T_DR_1146.

4.3.3 Fluorescence quenching of FMN on binding to T_DR_1146_clvd

Two 10 µL aliquots, followed by seven 50 µL aliquots of T_DR_1146_clvd stock solution (0.25 mM) were added to 10 µM FMN solution to investigate the effect of 6xHis tag removal on FMN binding. As for the non-cleaved protein, quenching of the FMN fluorescence intensity was observed after addition of T_DR_1146_clvd samples confirming binding to the flavin cofactor (Figure 4.9).

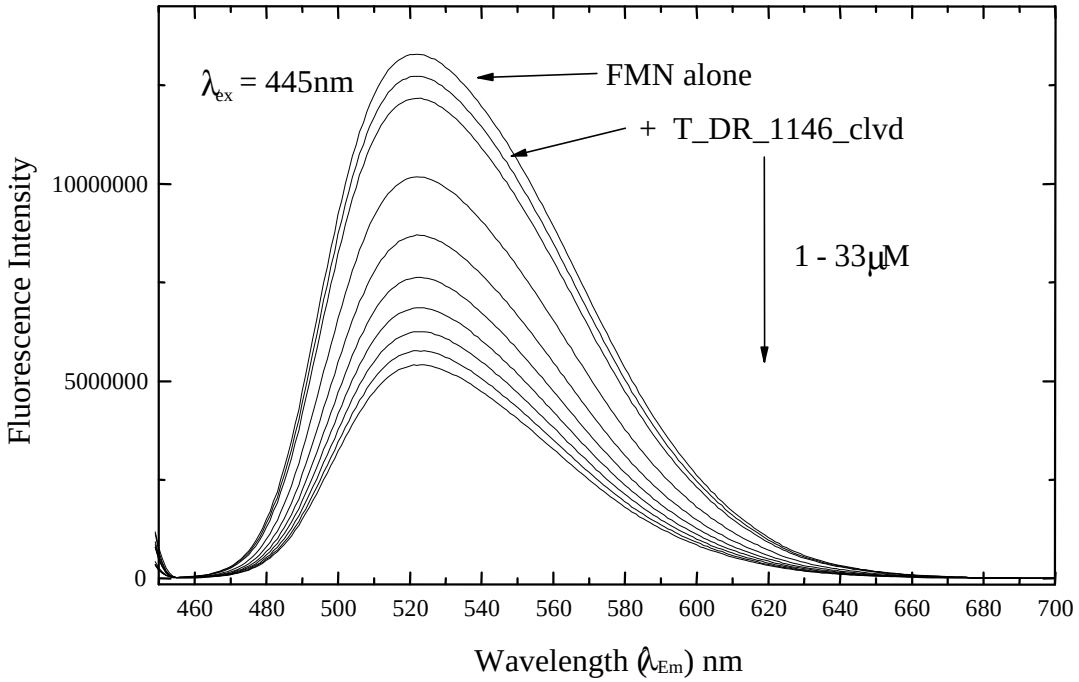


Figure 4.9: Quenching of FMN intrinsic fluorescence (excitation wavelength = 445nm) by increasing concentrations of T_DR_1146_clvd. 10 or 50μl aliquots of protein (0.25 mM) added to 2.5ml FMN (10 μM) in 20 mM Tris-HCl, 150 mM NaCl buffer at room temperature.

The constants K_D and N for the cleaved protein were determined as 11.48 μM ± 1.86 and 0.84 ± 0.17 respectively (Figure 4-10). As the binding parameters determined for T_DR_1146 and T_DR_1146_clvd are in the micromolar range, they are comparable to those determined by ITC. It was therefore concluded that the protein binds flavin cofactors within a range of 4-11.5 μM at a ratio of 1 molecule of FMN per monomer of protein.

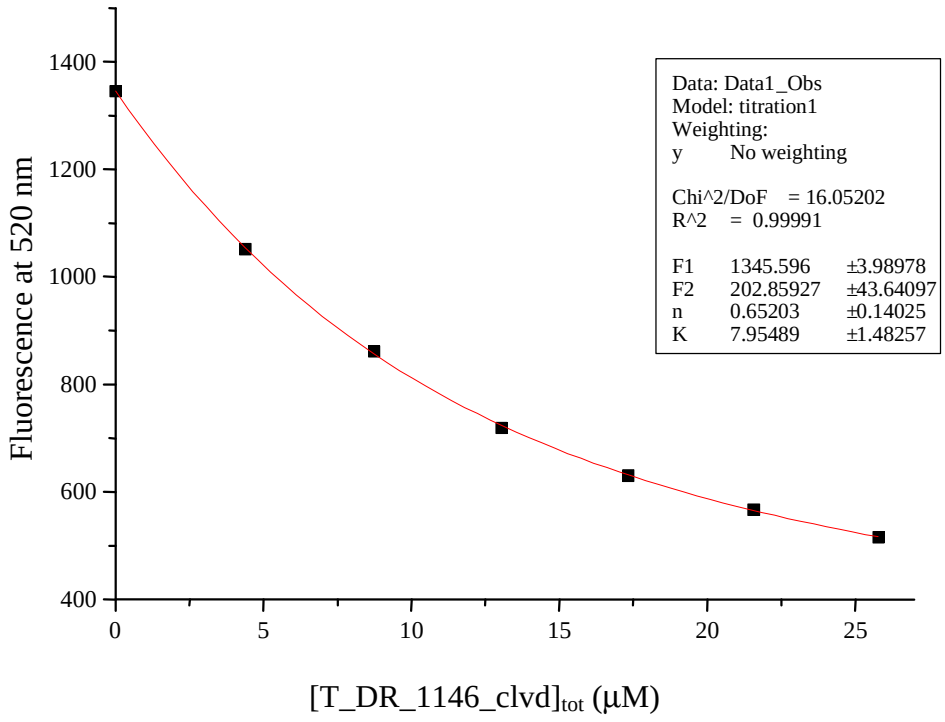


Figure 4.10: Data fitted to standard hyperbolic binding curve to determine K_D (11.48 µM ± 1.86) and N (0.84 ± 0.17) values for quenching of FMN fluorescence by increasing concentrations of T_DR_1146_clvd.

4.3.4 Fluorescence spectra for non-cleaved and cleaved T_DR_1146 without the addition of FMN

As described in Section 3.1.3, during purification of DR_1146, a pale yellow colour was observed. As ITC and fluorescence results confirmed that the protein binds flavins, it was thought that the observed yellow colour could be due to binding to trace amounts of a flavin cofactor during expression. To investigate this hypothesis, fluorescence emission spectra (445 nm excitation) for the initial T_DR_1146 and T_DR_1146_clvd samples were recorded without the addition of excess FMN.

In Figure 4.11, a small fluorescence peak at approximately 520nm can be seen for both T_DR_1146 and T_DR_1146_clvd samples. This suggests the presence of an "FMN-like" species in the recombinant protein sample.

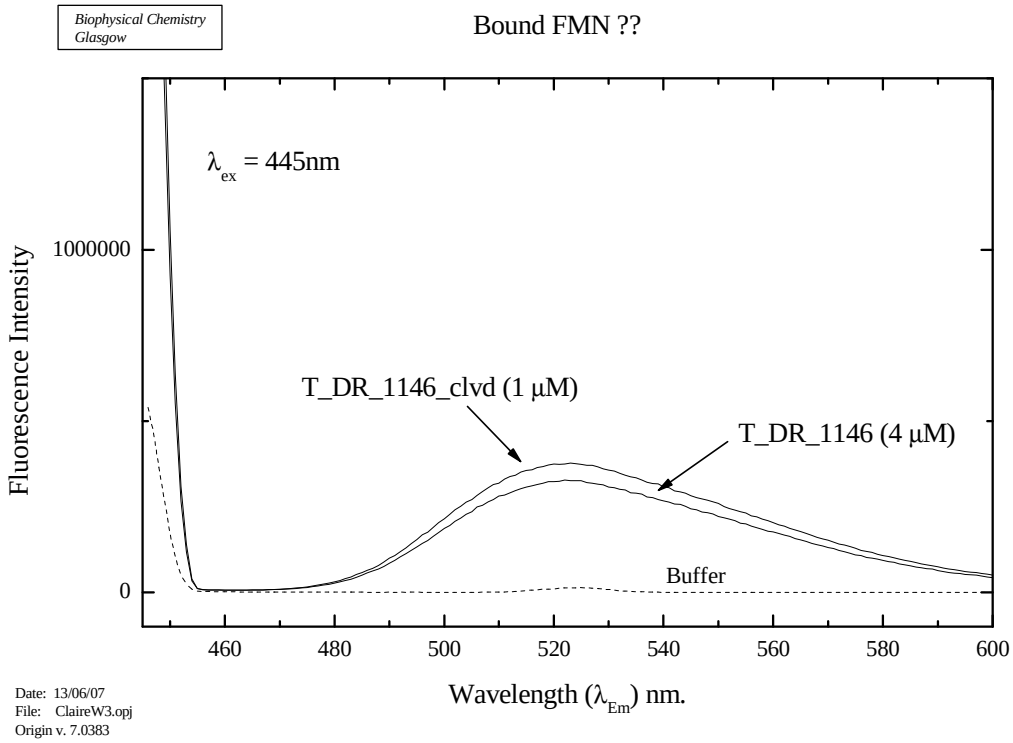


Figure 4.11: Fluorescence emission spectra (445nm excitation) of T_DR_1146 and T_DR_1146_clvd *before* addition of FMN. The small fluorescence peaks at about 520nm suggest the presence of bound "FMN-like" species in the recombinant protein.

4.3.5 Comparison of T_DR_1146_clvd dissociation constants with other flavin binding proteins

Table 4.4 compares K_D values gained for T_DR_1146_clvd with those reported for other recombinant flavin-binding proteins.

Protein	Method	K_D Riboflavin (M)	K_D FMN (M)	K_D FAD (M)
T_DR_1146_clvd	ITC	4.75×10^{-6}	8.57×10^{-6}	8.24×10^{-6}
General NAD(P)H-Flavin Oxidoreductase FRG/FRase I from *Vibrio fischeri* [187]	FS	3.70×10^{-5}	5.00×10^{-7}	1.20×10^{-5}
WrbA from *Escherichia coli* [188]	FS	-	2.00×10^{-6} (approx.)	-
Flavodoxin from *Desulfovibrio desulfuricans* [189]	FS	-	1.00×10^{-10}	-
Dodecin from *Thermus thermophilus* [190]	FS	2.33×10^{-7}	3.11×10^{-7}	5.89×10^{-7}
Riboflavin Transporter RibU from *Lactococcus lactis* [191]	ITC	1.80×10^{-9}	-	-
	FS	6.00×10^{-10}	3.60×10^{-8}	No binding
PNPOx from rabbit liver [105]	FS	-	$1.10\text{-}3.00 \times 10^{-8}$	-

Table 4.4: Comparison of K_D values established for T_DR_1146_clvd with those reported for other flavin binding bacterial proteins show some binding in the same micromolar range and some stronger binding in the nanomolar range.

The General NAD(P)H-Flavin Oxidoreductase FRG/FRase I from *Vibrio fischeri* binds Riboflavin, FMN and FAD in the same micromolar range as T_DR_1146_clvd. WrbA, a flavodoxin-like protein from *Escherichia coli* also binds FMN within this range. The K_D values for three other bacterial flavo-proteins and a PNPOx from rabbit liver are within the nanomolar range signifying stronger flavin binding than for T_DR_1146_clvd.

4.4 Differential Scanning Calorimetry (DSC)

4.4.1 Experimental aims

The aims of DSC experiments were (a) to determine T_m values for the apo forms of T_DR_1146 and T_DR_1146_clvd and therefore investigating the effect of tag removal on thermal stability (T_m values) and (b) to investigate the effect of FMN binding on thermal stability.

4.4.2 DSC experiments to assess the thermal stability of T_DR_1146 and T_DR_1146_clvd

Experiments were carried out as described in Section 2.7.4 with 19 μM concentrations of apo non-cleaved and apo cleaved protein in a buffer of 20 mM Tris-HCl pH 7.5, 150 mM NaCl.

The normalised experimental data in Figure 4.12 shows sharp endothermic transition peaks for both T_DR_1146 and T_DR_1146_clvd protein samples corresponding to similar T_m values of 50.7 and 51.35°C respectively. This suggests that removal of the 6xHis tag has negligible effect on thermal stability.

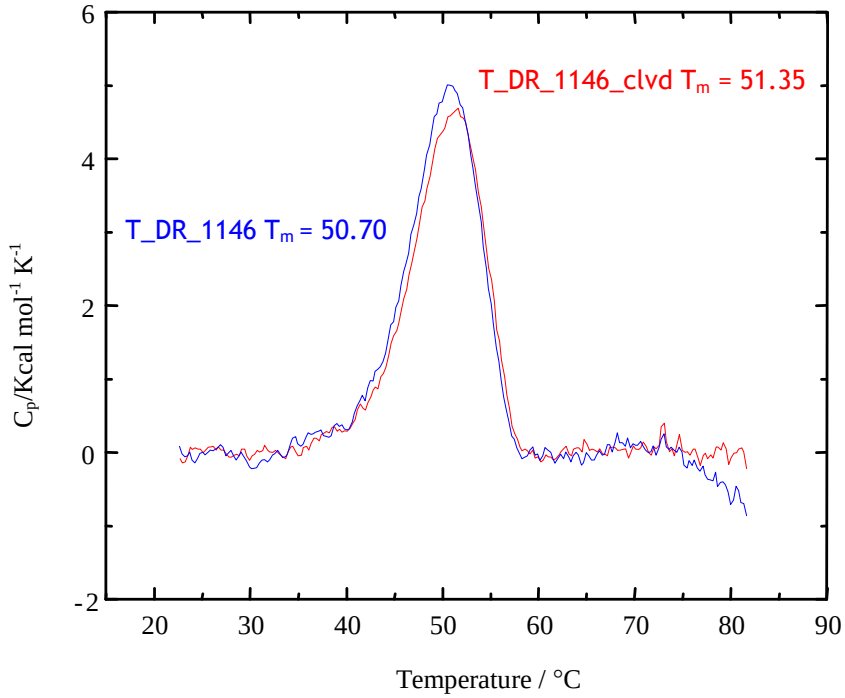


Figure 4.12: DSC thermogram showing T_m values of 50.70°C and 51.35°C determined for non-cleaved T_DR_1146 and T_DR_1146_clvd respectively indicating no significant difference in thermal stability on 6xHis tag removal.

After complete denaturation, the two samples were cooled back down to room temperature and the DSC experiments were repeated. In Figure 4.13 it can be seen that for the rescan of the non-cleaved sample the sharpness and intensity of the transition peak was greatly reduced with frequent fluctuations. This suggests that on cooling the protein sample did not refold very well possibly due to aggregation on initial heating. In contrast, the rescan of the cleaved protein

sample showed only a small decrease in transition intensity and sharpness suggesting that it refolded relatively well on cooling. It was concluded from this result that the 6xHis tag might obstruct refolding.

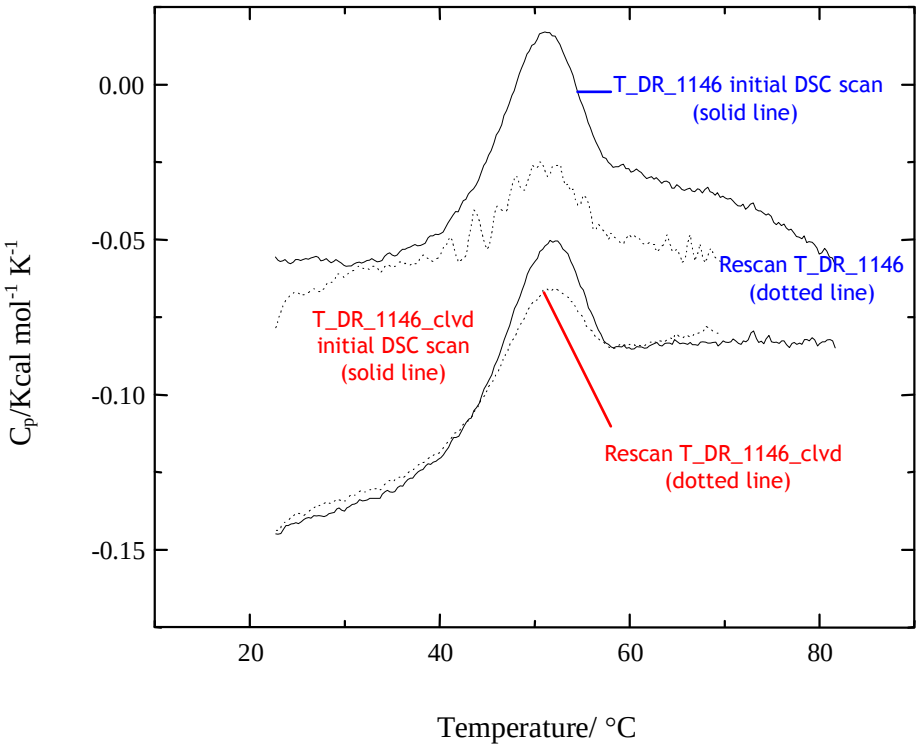


Figure 4.13: DSC thermogram comparing transition peaks from repeat DSC scans (dotted lines) with initial DSC data (solid lines) for T_DR_1146 (upper section) and T_DR_1146_clvd (lower section). The rescan for T_DR_1146_clvd is much closer to the initial DSC data suggesting that for non-cleaved T_DR_1146 the 6xHis tag may obstruct the refolding process.

4.4.3 DSC experiments to assess the increase in thermal stability of T_DR_1146_clvd on binding to FMN

Further DSC experiments were carried out to investigate the effect on thermal stability on addition of increasing concentrations of FMN. The experiments were carried out as described in Section 2.7.4 using 0.092 mM samples of T_DR_1146_clvd and 0.02, 0.2 and 2.0 mM concentrations of FMN in 20 mM Tris-HCl pH 7.5, 150 mM NaCl buffer.

The normalised DSC thermogram (Figure 4.14) shows a large increase in T_m on addition of FMN confirming that binding of the ligand to the protein increases thermal stability.

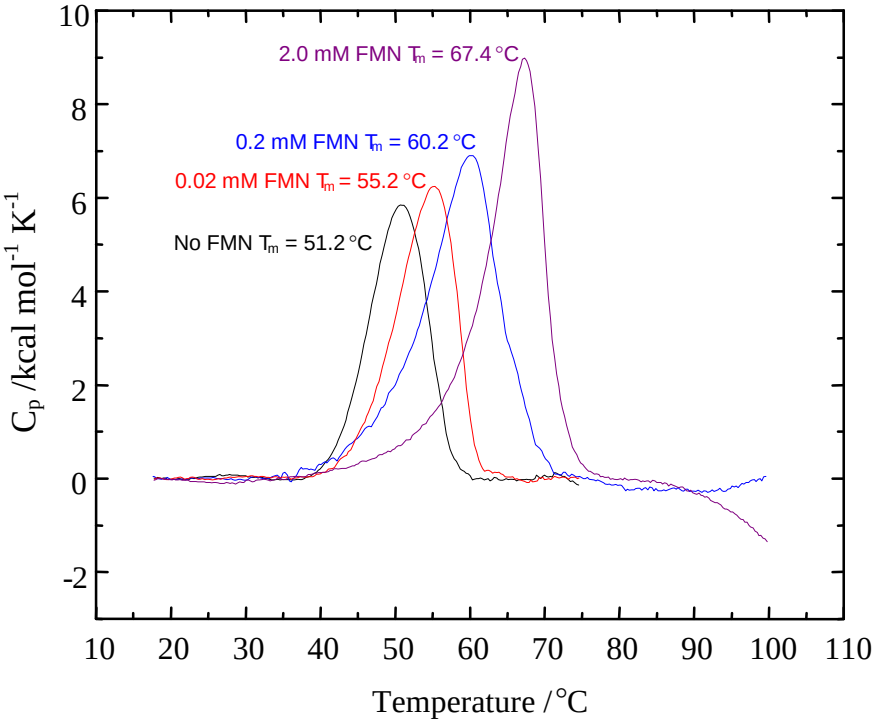


Figure 4.14: DSC thermogram for the addition of increased concentrations of FMN (0-2.0 mM) to a 0.092 mM sample of T_DR_1146_clvd shows increased intensity/sharpness of melting transitions and T_m values on increasing concentrations of FMN.

A 2:1 ratio of FMN to protein increased the T_m from 51.2°C to 60.2°C. A further increase of FMN concentration (20:1) gave a higher T_m of 67.4°C, with a melting transition of increased intensity and sharpness. In conclusion, the binding of FMN to T_DR_1146_clvd results in greater thermal stability and stabilisation of the protein.

The overall results of the DSC experiments suggest that (a) the 6xHis tag may obstruct protein refolding and thus may affect overall stability and (b) addition of a two fold molar excess of FMN increases thermal stability significantly. Due to the correlation between protein stability and crystal growth, it was decided that for future crystallisation experiments, the 6xHis tag would be removed and that at least a two equimolar excess of FMN would be added during purification.

4.5 Circular Dichroism (CD)

4.5.1 Experimental aims

The aims of the CD experiments were (a) to establish whether T_DR_1146 and T_DR_1146_clvd are folded in the presence and absence of FMN, (b) to monitor any changes in secondary and tertiary structure on binding to FMN, (c) to estimate the secondary structure of apo and holo T_DR_1146_clvd, (d) to monitor the temperature denaturation of apo and holo T_DR_1146_clvd, (e) to monitor the chemical unfolding of apo and holo T_DR_1146_clvd using increasing concentrations of guanidine hydrochloride and (f) to establish if the FMN cofactor remains bound to holo T_DR_1146_clvd after unfolding.

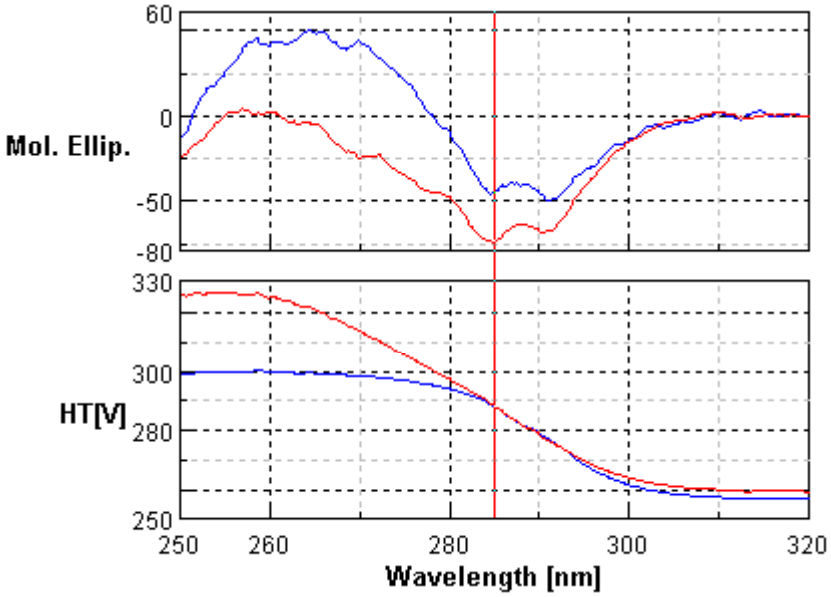
4.5.2 CD experiments to assess the structural effects of FMN binding on T_DR_1146 and T_DR_1146_clvd

CD experiments were carried out as described in Section 2.7.5 using T_DR_1146 and T_DR_1146_clvd samples in the presence and absence of 1μM FMN.

Overall, results show changes in protein conformation on binding FMN as judged by spectra obtained in the far UV (260-190 nm) and the near UV (320-250 nm) CD regions, which correspond to secondary and tertiary elements respectively.

Changes in intensity around 300-250 nm in the near UV regions for both T_DR_1146 and T_DR_1146_clvd (Figure 4.15) suggest that the environments of the aromatic amino acids change as a result of FMN binding. It is possible that the structure of the protein becomes more rigid as flavin binds or that the aromatics are stacking with the flavin rings giving increased signal intensity in this spectral region. Higher resolution techniques would be required to ascertain the atomic detail of these spectral changes since CD is insufficient to resolve such fine detail.

(a) non-cleaved T_DR_1146



(b) T_DR_1146_clvd

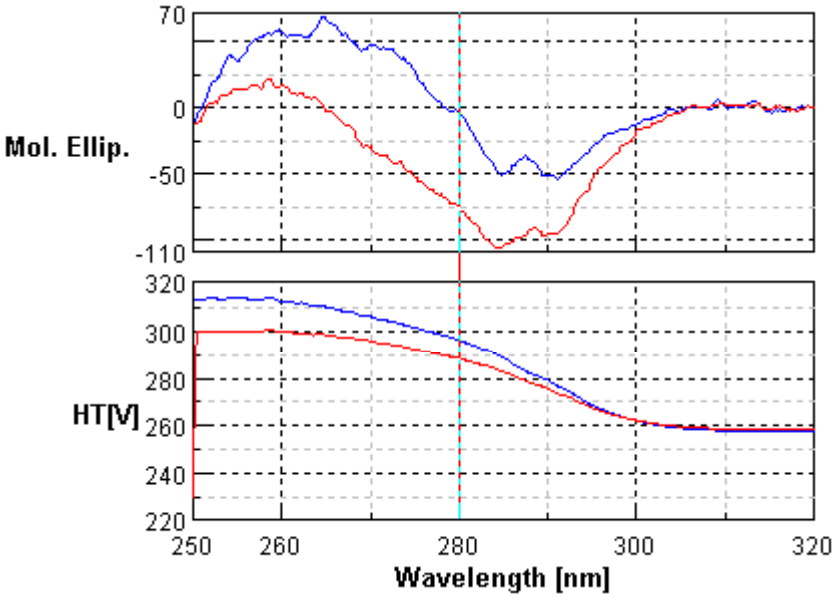
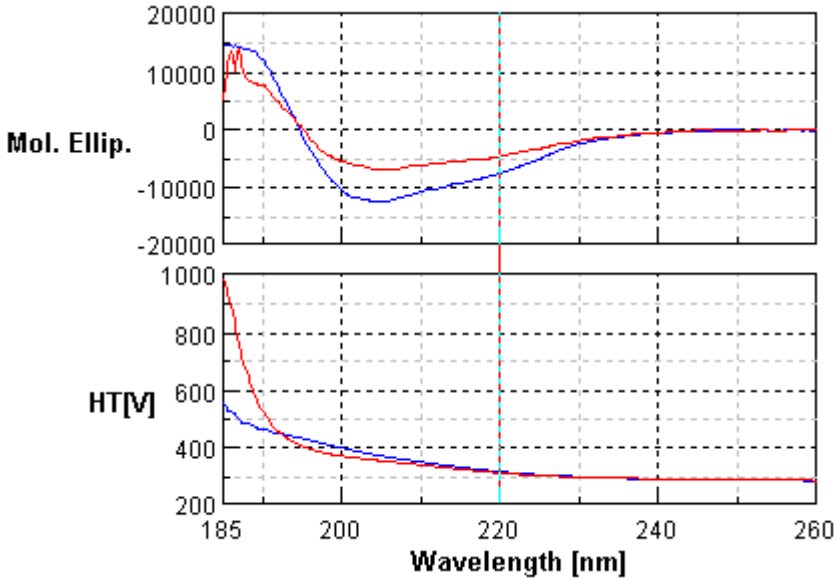


Figure 4.15: Near UV spectra of apo (blue) and holo (red) (a) non-cleaved T_DR_1146 and (b) T_DR_1146_clvd. Both in 50 mM phosphate buffer pH 7.5 with 150 mM NaCl.

In the far UV region (Figure 4.16), corresponding to secondary structural conformation, there is a marked decrease in spectral intensity following binding of FMN suggesting changes in the overall secondary structure of the protein upon ligand binding. These results are consistent with the changes observed in the near UV region. It is probable that the protein undergoes gross secondary and tertiary conformational changes in the presence of FMN. Some of the differences

observed in the far UV region may also reflect contributions from the aromatic groups. When aromatic rings align, such contributions are often positive around 220nm which would result in a decrease in the overall negative ellipticity [192].

(a) non-cleaved T_DR_1146



(b) T_DR_1146_clvd

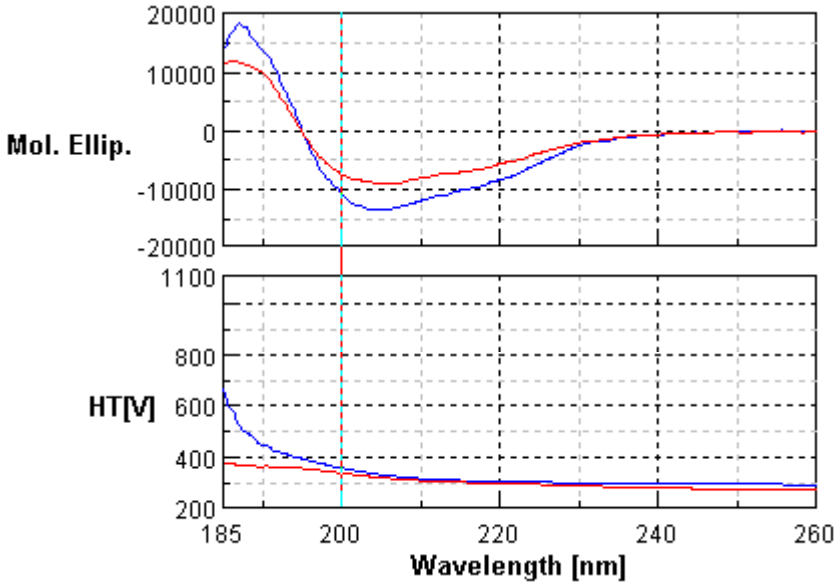


Figure 4.16: Far UV spectra of apo (blue) and holo (red) (a) non-cleaved T_DR_1146 and (b) T_DR_1146_clvd in a buffer of 50 mM phosphate buffer pH 7.5 with 150 mM NaCl.

4.5.3 Secondary structure prediction

The secondary structure compositions for full length apo and holo T_DR_1146 and T_DR_1146_clvd and holo T_DR_1146_clvd were estimated using data obtained from the far UV CD spectral region (240-190nm). These data were submitted to Dichroweb, an online server hosting various secondary structure deconvolution algorithms [177]. Average values obtained from the CONTIN and the CDSSTR algorithms [178] were used to estimate secondary structural content. The estimates obtained for percentage contribution from helices, sheets and turns are shown in Table 4.5.

Protein	% Helix	% Sheet	% Turns	% Unordered	Total %
Apo T_DR_1146	16	29	21	34	100
Holo T_DR_1146	10	34	23	33	100
Apo T_DR_1146_clvd	16	28	21	35	100
Holo T_DR_1146_clvd	10	33	23	34	100

Table 4.5: CD secondary structure estimate for apo and holo T_DR_1146 and T_DR_1146_clvd.

On analysis of the tabulated data, it was concluded that T_DR_1146_clvd is a mixed alpha-beta protein and that FMN binding causes significant changes in secondary structure. No notable difference can be seen as a result of 6xHis tag removal.

4.5.4 CD experiments to monitor the thermal unfolding of T_DR_1146_clvd using a temperature gradient

Figure 4.17 shows the decrease in percentage of native ellipticity at 208 nm of apo and holo T_DR_1146_clvd over a temperature gradient of 5-80°C. At approximately 50°C the ellipticity of the apo protein begins to decline more rapidly than that of the holo protein. This data correlates approximately with a T_m value of 51.35°C established for the apo protein by DSC (Section 4.4.2). At a temperature of 80°C the final percentage of remaining native ellipticity for the apo protein (46.2 %) was much lower than that of the holo protein (69.3 %) confirming that the presence of FMN increases thermal stability.

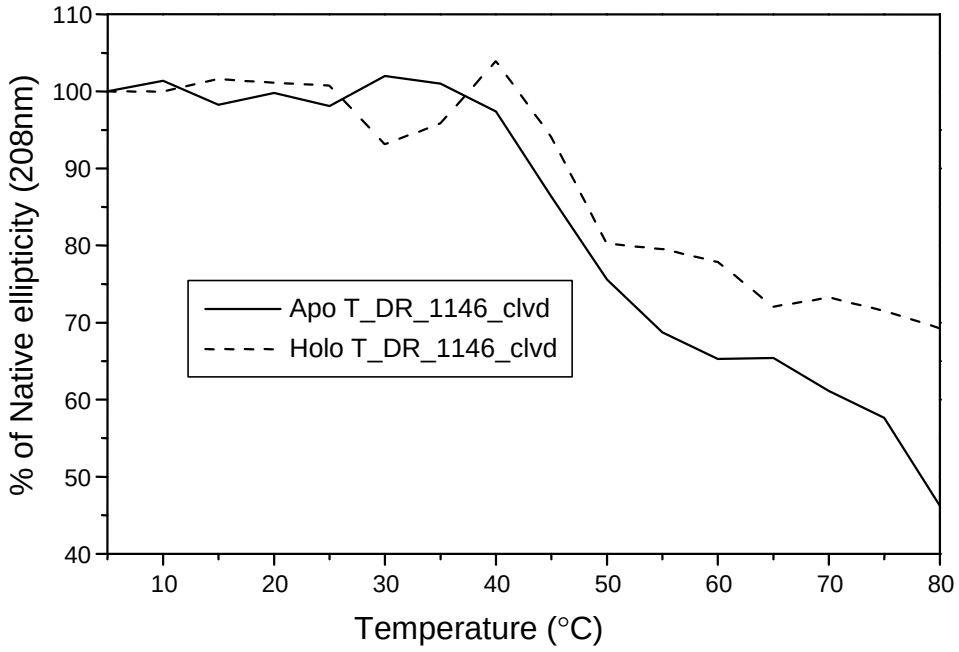


Figure 4.17: CD data collected from the thermal unfolding of apo (unbroken line) and holo (broken line) T_DR_1146_clvd monitored at 208 nm.

4.5.5 CD experiments to monitor the chemical unfolding of T_DR_1146_clvd using guanidine hydrochloride

Figure 4.18 shows the decrease in percentage of native ellipticity at 222 nm of apo and holo T_DR_1146_clvd as a result of adding increasing concentrations (0-6 M) of guanidine hydrochloride (GuHCl).

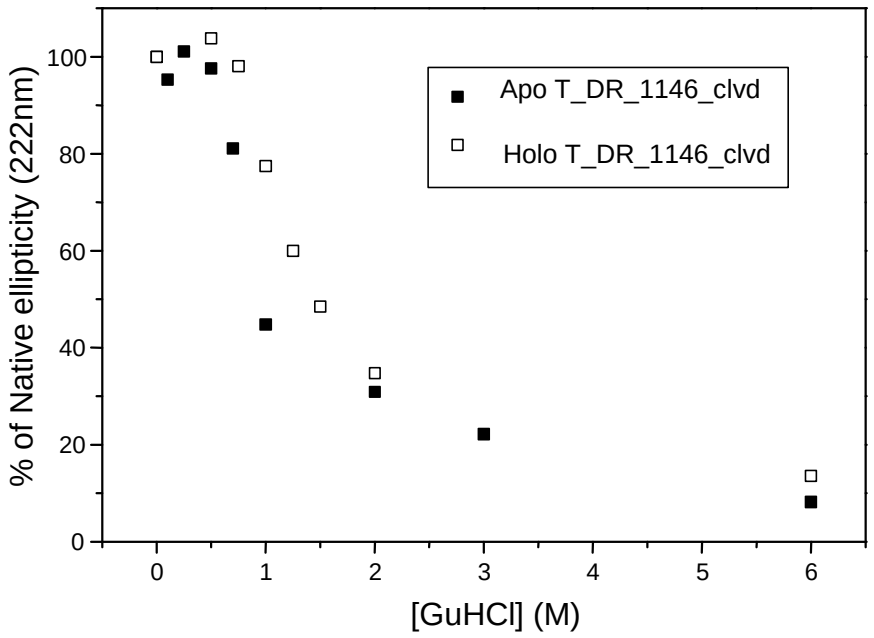


Figure 4.18: CD data collected from the chemical unfolding of apo and holo T_DR_1146_clvd monitored at 222 nm using increasing concentrations of GuHCl.

The apo protein was found to be slightly less stable than the holo protein. Incubation of apo protein in 1M GuHCl resulted in an ellipticity value which was 44.7 % of the native protein whereas the holo protein remained at 77.5 %. This shows that T_DR_1146_clvd is more chemically stable when FMN is bound.

Estimates of the midpoint of denaturation for apo and holo T_DR_1146_clvd were gained by plotting the total percentage change in native ellipticity against GuHCl concentration (Figure 4.19).

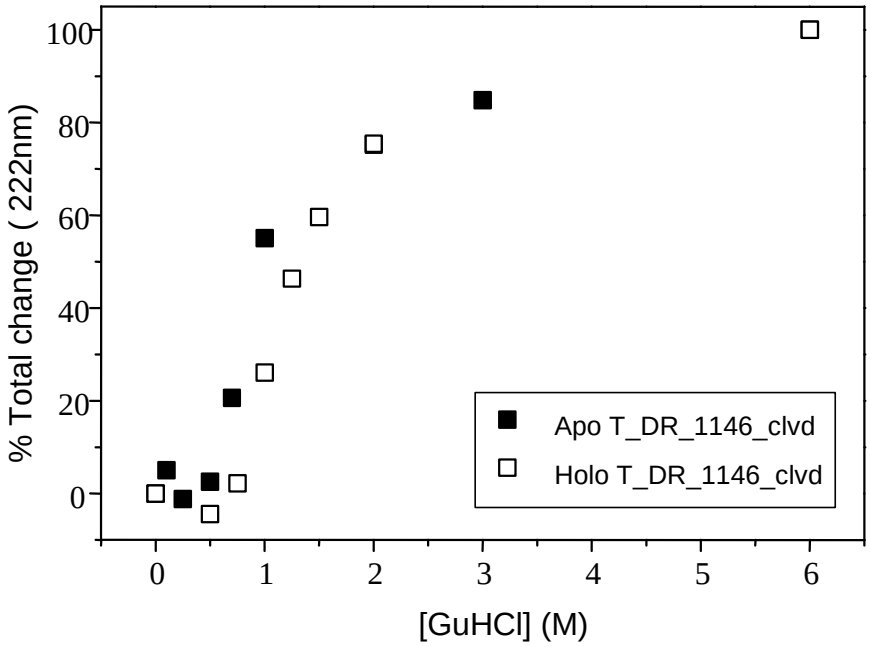


Figure 4.19: Plot of the total percentage change in native ellipticity at 222 nm against GuHCl concentration (1-6 M) for the chemical unfolding of apo and holo T_DR_1146_clvd.

The apo protein is 50 % unfolded at a GuHCl concentration of approximately 0.9 M. However, a higher concentration of approximately 1.4 M is needed to unfold 50 % of the holo protein.

By monitoring the FMN contributions in the visible-near UV region (Figure 4.20) it is evident that the FMN cofactor added to produce the holo protein is subsequently removed on chemical unfolding with GuHCl, with a concentration of 2 M required for complete dissociation. This is shown by reduced peak intensities at 470 nm on addition of increasing concentrations of guanidine hydrochloride. The ease at which the FMN dissociates suggests that the intermolecular interactions between the added cofactor and the protein are non-covalent.

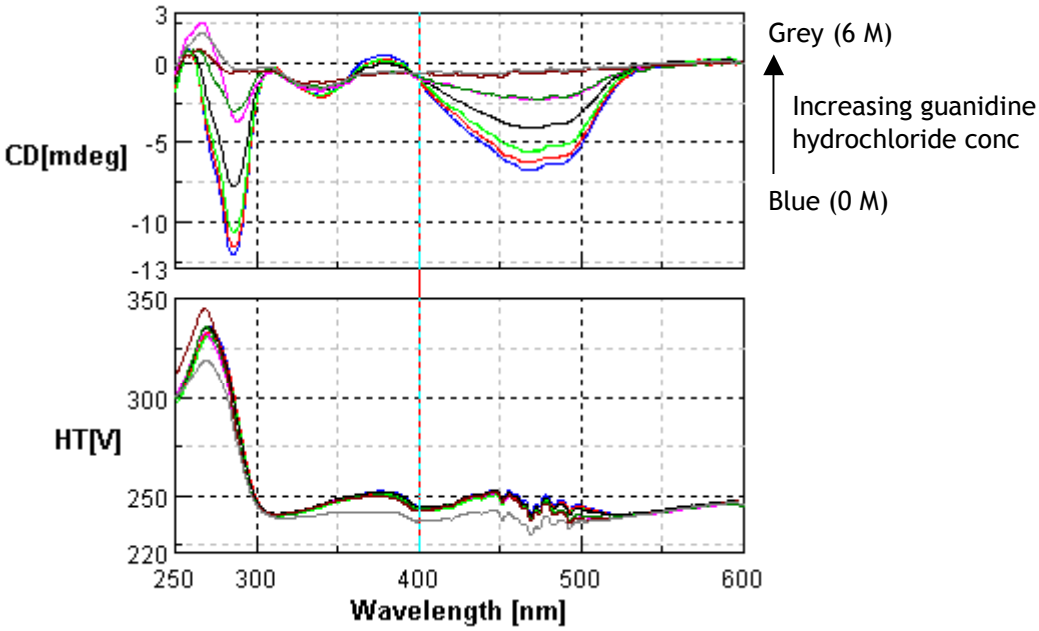


Figure 4.20: Visible-Near UV spectrum for holo T_DR_1146_clvd showing decreased FMN contributions (monitored at 470 nm) on increasing concentrations of GuHCl.

4.5.6 NMR

4.5.7 Experimental aims

The aim of preliminary non-labelled and labelled NMR experiments was to investigate the feasibility of using this technique to determine the 3-D structure of T_DR_1146 in solution. 2-D ^{15}N-^{1}H HSQC experiments were used to establish the amount of FMN needed to saturate cofactor binding sites.

4.5.8 One-dimensional, non-labelled NMR experiments

Well-dispersed methyl and amide regions of 1-D spectra for the apo and holo forms of T_DR_1146_clvd show that the protein is folded and structured both with and without FMN. This is shown in Figure 4.21, which is a spectrum of holo T_DR1146_clvd showing defined peaks in the amide (8.5 ppm) and methyl (0 ppm) spectral regions.

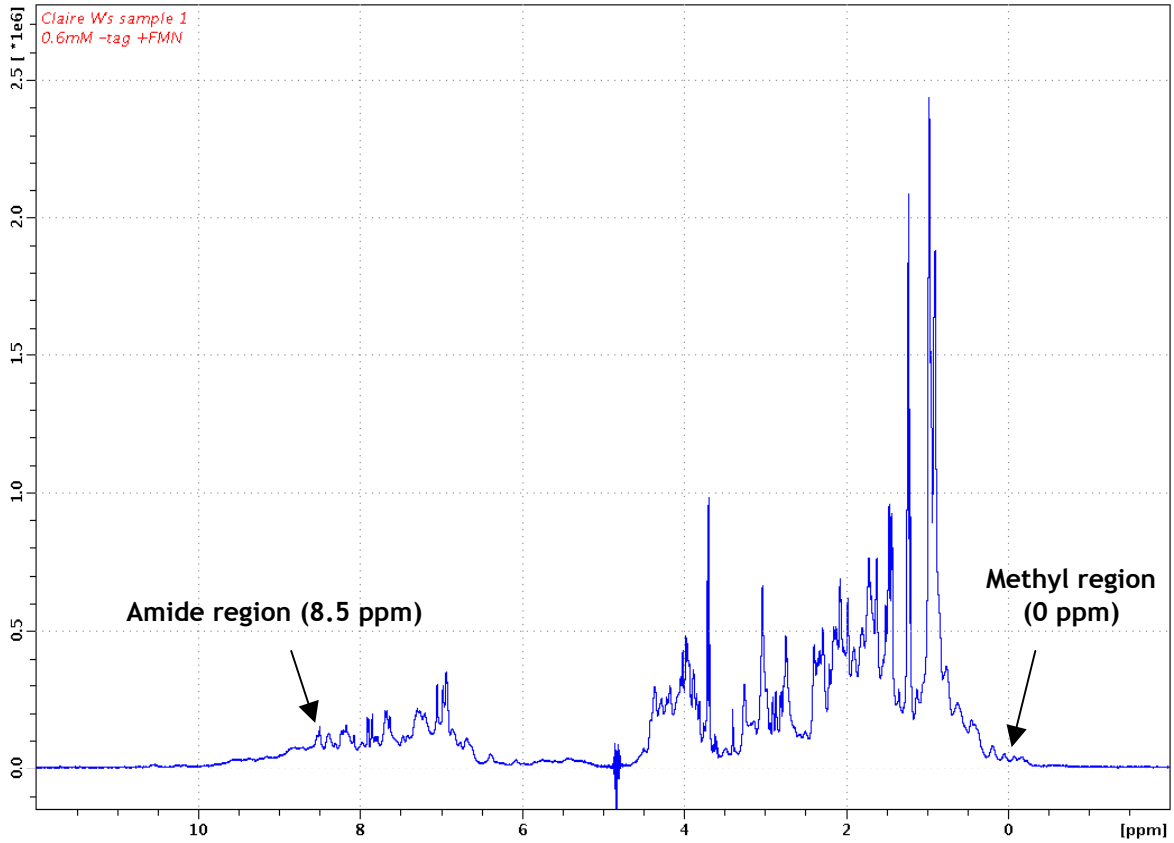


Figure 4.21: 1D NMR spectrum of non-labelled T_DR_1146_clvd with 1µM FMN. Well-dispersed methyl and amide spectral regions suggest T_DR_1146 is folded and structured.

Overall, from 1-D experiments, it was concluded that the addition of FMN has an effect on the structure of T_DR_1146. However, in the absence of the FMN cofactor, the quality of the spectra is increased with sharper peaks in the down field methyl region (0 ppm) (Figure 4.22).

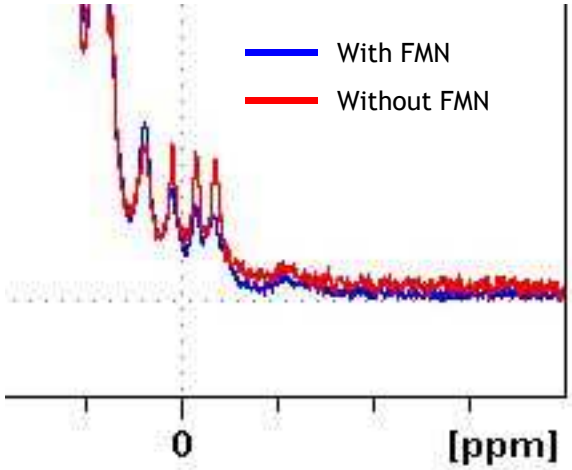


Figure 4.22: Methyl region of 1D NMR spectra for apo (red) and holo (blue) T_DR_1146_clvd shows sharper methyl peaks in the absence of the FMN cofactor.

On comparison of spectra for apo T_DR_1146 and apo T_DR_1146_clvd, it was apparent that spectral quality is also improved by removal of the 6xHis tag; a result which influenced sample preparation for 2-D ^{15}N labelled experiments.

4.5.9 Two-dimensional ^{15}N-^{1}H HSQC experiments

Figure 4.23 (a) to (d) shows the resulting 2-D ^{15}N HSQC spectra for titrations of increasing concentrations of FMN (0 to 2.7 mM) into a 0.9 mM sample of apo T_DR_1146_clvd in order to determine the [protein]:[FMN] ratio needed to fully saturate the protein.

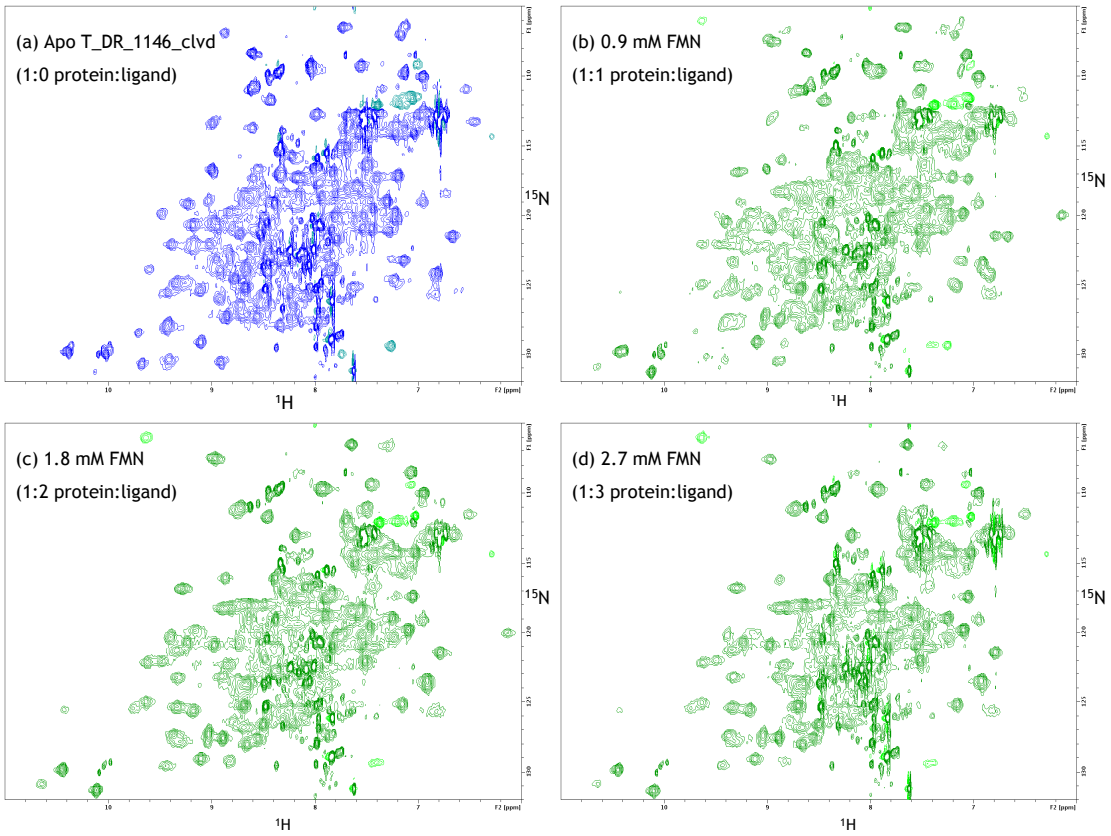


Figure 4.23: 2-D ^{15}N HSQC spectra for titrations of increasing concentrations of FMN (0-2.7 mM) into a 0.9 mM sample of T_DR_1146_clvd: (a) 0.9 mM Apo T_DR_1146_clvd (1:0 protein:ligand ratio), (b) 0.9 mM FMN added to 0.9 mM T_DR_1146_clvd (1:1 protein:ligand ratio), (c) 1.8 mM FMN added to 0.9 mM T_DR_1146_clvd (1:2 protein:ligand ratio) and (d) 2.7 mM FMN added to 0.9 mM T_DR_1146_clvd (1:3 protein:ligand ratio).

Comparison of Figures 4.23 (a) and (b) show that significant chemical shift changes occur on the addition of a 1:1 ratio of FMN to apo T_DR_1146_clvd. Further changes occurred when increasing the protein:FMN concentration ratio to 1:2 (Figure 4.23c) signifying that 1:1 does not fully saturate binding sites. Adding another 0.9 mM aliquot of FMN (1:3 [protein]:[FMN] ratio) produced only slight changes (Figure 4.23d) suggesting that binding sites are virtually saturated at a 1:2 ratio.

This result confirms that a 2:1 excess of FMN:protein is required to ensure that the binding sites of T_DR_1146_clvd are completely saturated with the cofactor.

Based on this result and on data gained from DSC experiments (Section 4.4.3), it was decided that at least a two equimolar excess of FMN would be added to the protein during purification to ensure that all binding sites are saturated and hence that the protein sample is homogenous for crystallisation experiments.

Figure 1.22 is a superimposition of Figures 1.21 (a) and (d) showing the overall pattern of chemical shift changes on addition of 2.7 mM FMN. Such significant peak movements show that global structural changes are induced by FMN binding.

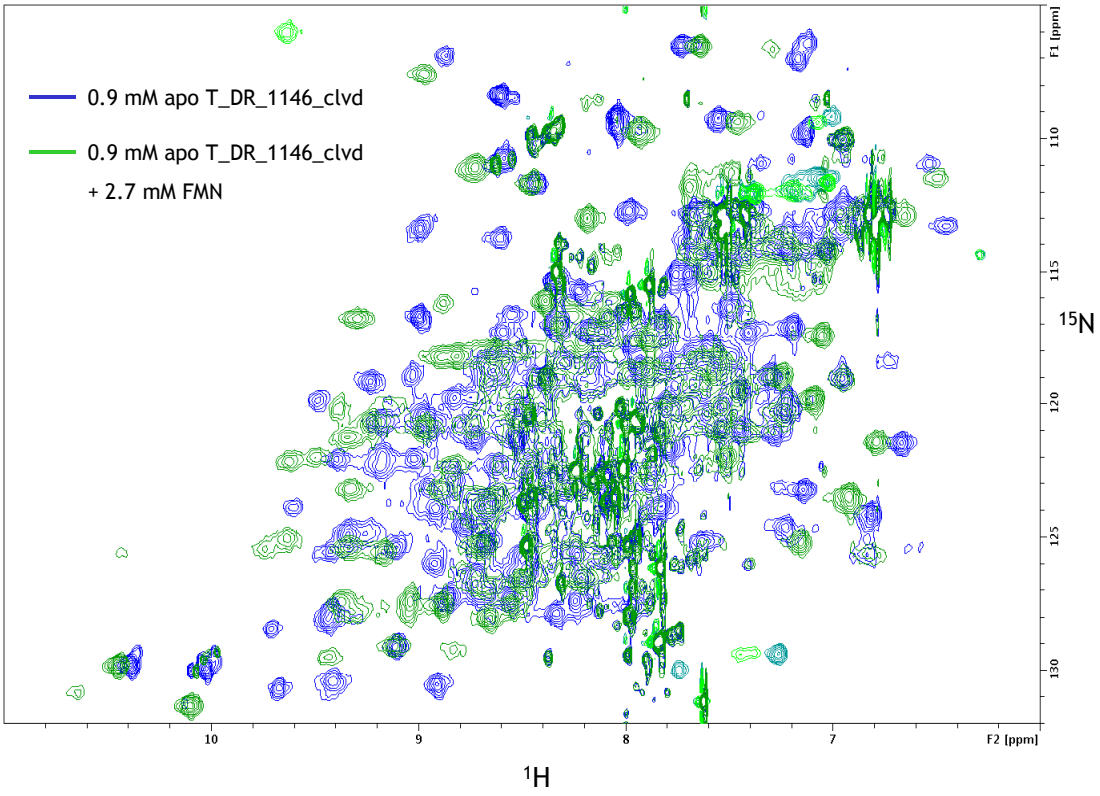


Figure 4.24: Superimposed 2-D ^{15}N HSQC spectra for apo T_DR_1146_clvd without FMN (blue) and apo T_DR_1146_clvd titrated against 2.7 mM FMN (green) showing the pattern of chemical shift changes as a result of FMN binding.

The transition between apo (blue) and holo (green) peaks in Figure 4.24 occurred through the disappearance of the apo peak and reappearance in a different position after addition of FMN. This is suggestive of system in slow chemical exchange on the NMR timescale and is typical for protein-ligand systems with high binding affinity and low dissociation constants (μM or lower) [193].

The overall quality of both non-labelled and ^{15}N-^{1}H HSQC spectra suggests that structure determination by NMR would be feasible for T_DR_1146_clvd.

4.6 Analytical ultracentrifugation (AUC)

The solution behaviour of holo T_DR_1146_clvd was investigated by sedimentation velocity (SV) and sedimentation equilibrium (SE) analytical ultracentrifugation (AUC) as described in Section 2.7.6.

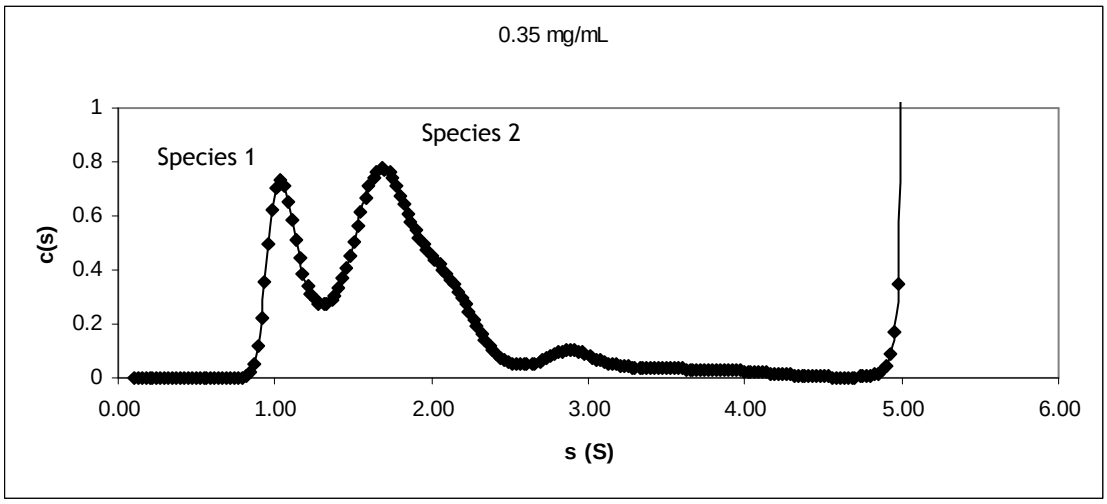
Buffer density and viscosity and the partial specific volume of the protein at 4°C were calculated as 1.00968 g/mL, 0.016197 (Poise) and 0.720 mL/g (0.727 mL/g at 20°C) respectively, using SEDNTERP software [180]. These parameters were utilised for analysis of SV and SE datasets using the SEDFIT [179] and SEDPHAT [181] programmes.

4.6.1 Sedimentation velocity (SV)

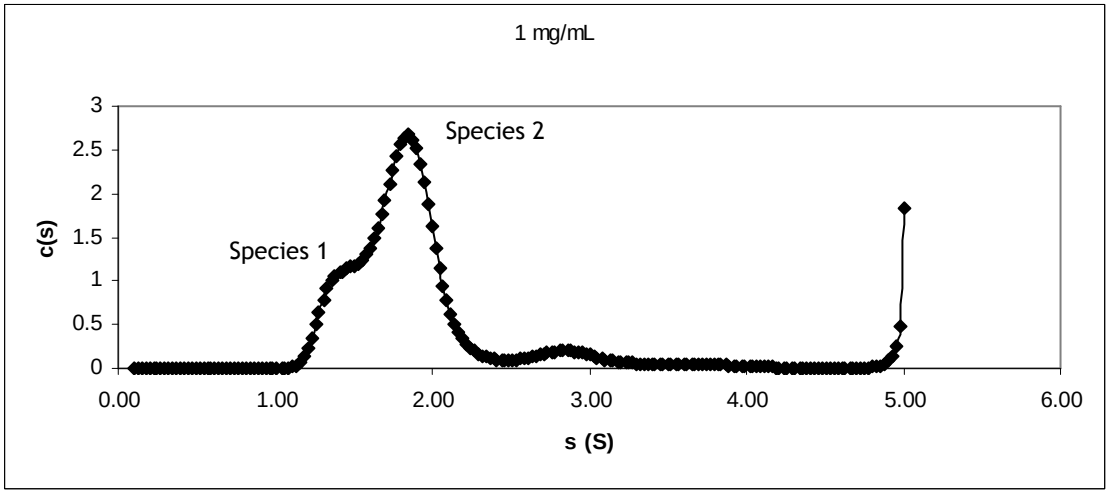
The aim of the SV experiment was to gain insight into the concentration dependent oligomerisation behaviour of holo T_DR_1146_clvd and to determine a sedimentation coefficient for the protein.

Seven concentrations of holo T_DR_1146_clvd (0.35, 1, 2, 3.5, 10, 22 and 35 mg/mL) were analysed in the SV experiment. The resulting size distribution c(s) profiles for SV data analysed by SEDFIT [179] are shown in Figures 4.25 to 4.27.

(a)



(b)



(c)

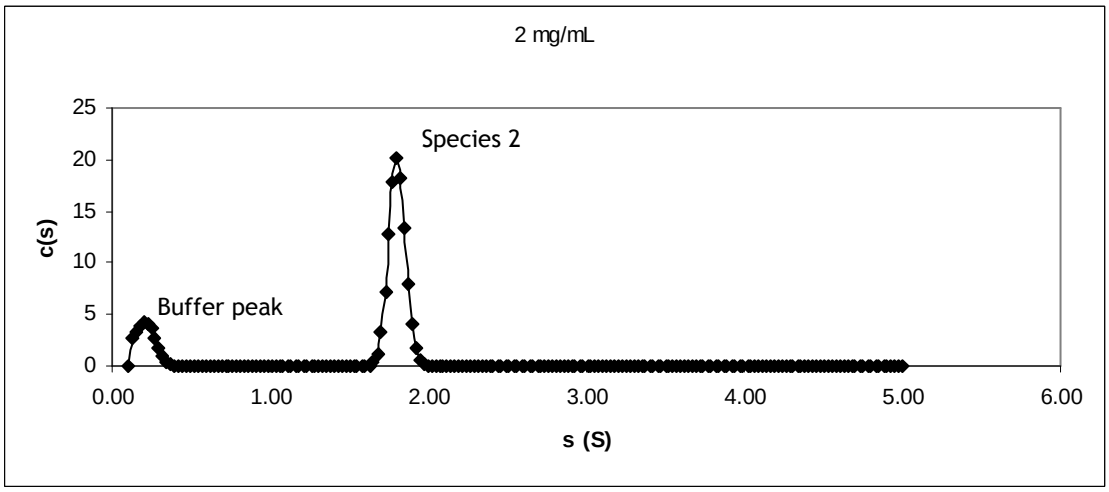
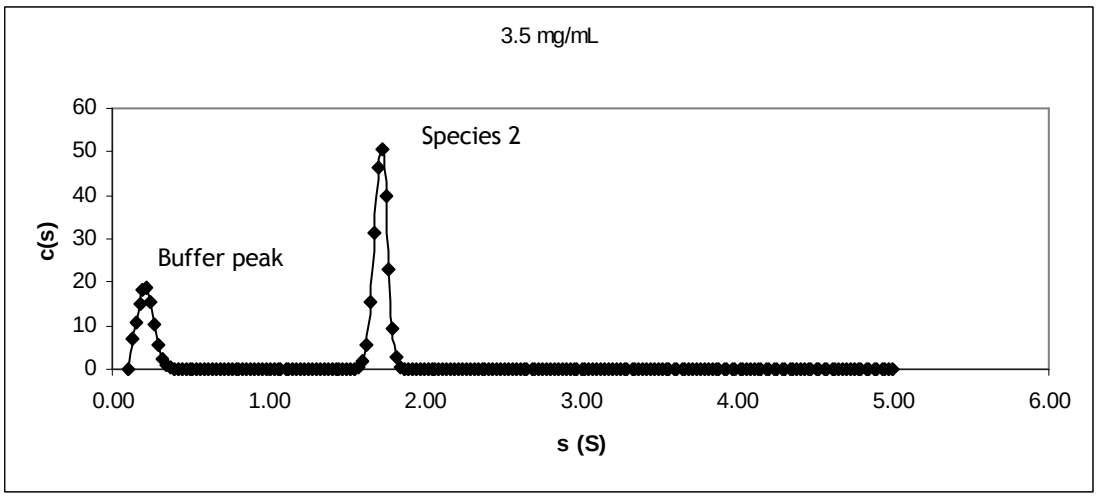
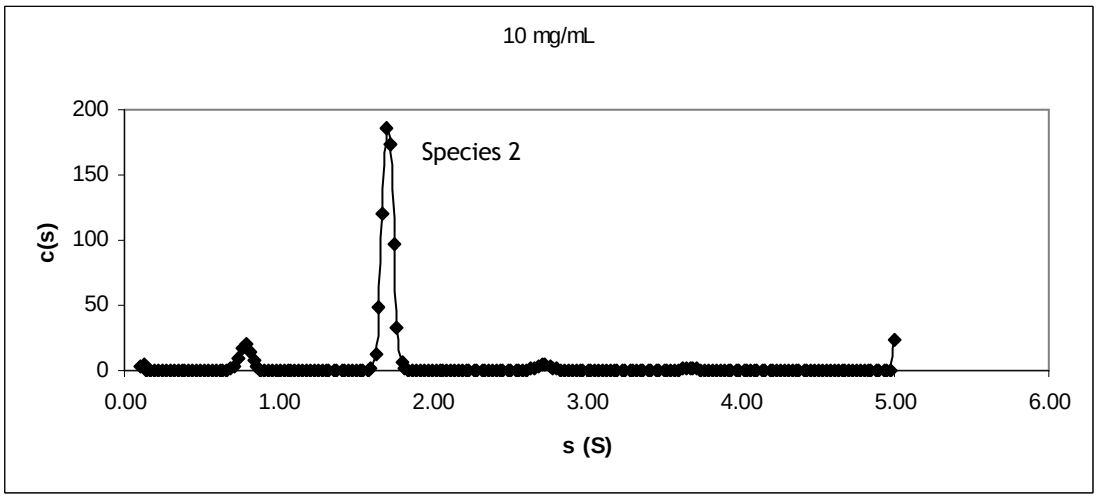


Figure 4.25: Individual size distribution c(s) profiles from SV data analysed with SEDFIT [179] for holo T_DR_1146_clvd at (a) 0.35 mg/mL: showing two main species present in solution (b) 1 mg/mL: shows peaks for Species 1 and 2 merge closer together suggesting self-association (c) 2 mg/mL: two species have merged completely leaving only one main peak. The small peak seen close to zero S is thought to be buffer.

(a)



(b)



(c)

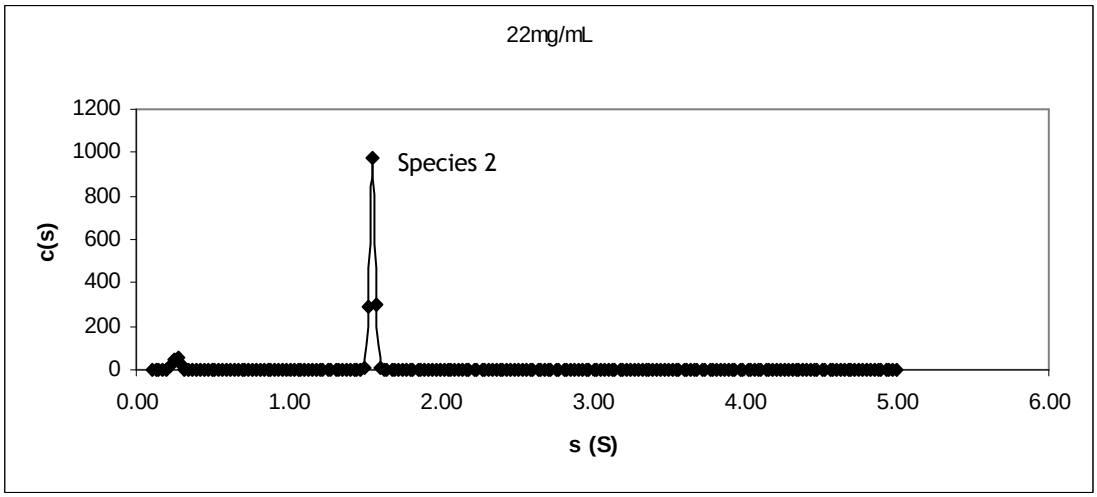


Figure 4.26: Individual size distribution c(s) profiles from SV data analysed with SEDFIT [179] for holo T_DR_1146_clvd at (a) 3.5 mg/mL: shows species 2 begin to migrate back to the left (non-ideal) (b) 10 mg/mL and (c) 22 mg/mL: further migration of species 2 to the left.

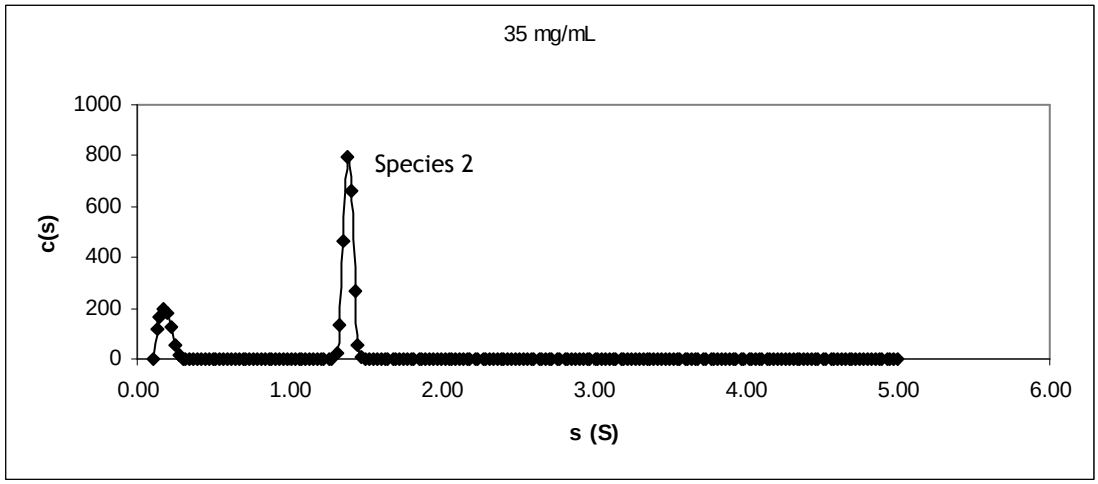


Figure 4.27: Individual size distribution c(s) profile from SV data analysed with SEDFIT [179] for holo T_DR_1146_clvd at 35 mg/mL showing non-ideal behaviour of species 2 (migration to the left).

At a concentration of 0.35 mg/mL, two main species (Species 1 and 2) were observed within the protein solution (Figure 4.25a). At 1 mg/mL, the two peaks corresponding to Species 1 and 2 merged closer together (Figure 4.25b). Such behaviour is suggestive of self-association in solution.

A further increase in protein concentration to 2 mg/mL resulted in complete merging of the two species to form a single peak (Figures 4.25c). At this concentration the protein exists as only species 2 in a single conformation. At concentrations of 3.5 and above, Species 2 is shown to migrate back to the left again (Figures 4.26b and c). It is thought that the reported decrease in sedimentation coefficient (s) of the Species 2 occurred as a result of non-ideal concentration effects caused by overcrowding of protein molecules at the base of the AUC cell.

Figure 4.28 shows an overlay of size distribution c(s) profiles for the seven protein concentrations, which illustrates the overall migration of species 2.

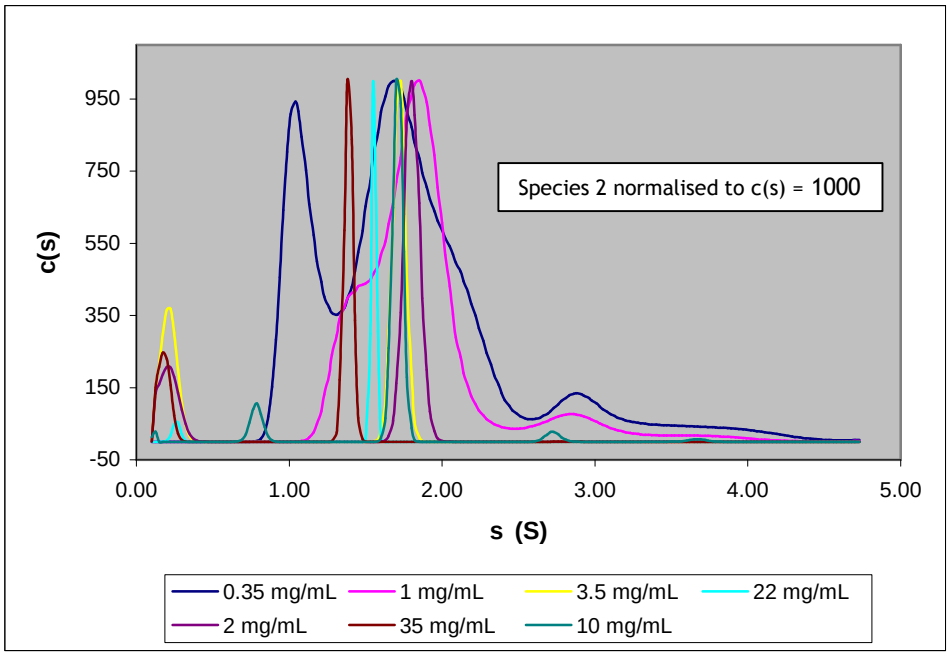


Figure 4.28: Normalised (species 2) comparison of size distribution c(s) profiles for increasing concentrations of holo T_DR_1146_clvd showing migration of species 2.

The apparent sedimentation coefficients ($s_{4,b}$) for species 2 of holo T_DR_1146_clvd at 4˚C in buffer, and subsequent true sedimentation coefficients ($s_{20,w}$), corrected for solvent viscosity and density were calculated as described in Section 2.7.6.1 and are listed in Table 4.6.

[T_DR_1146_clvd] (mg/mL)	Apparent sedimentation coefficient at 4˚C in buffer $s_{4,b}$ (S)	True sedimentation coefficient at 20˚C in water $s_{20,w}$ (S)
0.35	1.70 ± 0.17	2.76 ± 0.28
1	1.81 ± 0.10	2.95 ± 0.18
2	1.81 ± 0.05	2.94 ± 0.09
3.5	1.63 ± 0.05	2.65 ± 0.08
10	1.68 ± 0.02	2.73 ± 0.03
22	1.54 ± 0.03	2.51 ± 0.05
35	1.36 ± 0.03	2.21 ± 0.04

Table 4.6: Apparent sedimentation coefficients determined for species 2. Errors for $s_{20,w}$ values were estimated using the percentage error calculated for $s_{4,b}$.

A plot of $s_{w,\,20}$ against concentration of holo T_DR_1146_clvd (mg/mL) (Figure 4.29) was used to determine ($s^0_{20,w}$) (sedimentation coefficient independent of concentration) value of 2.87 S. The $s^0_{20,w}$ for Species 1 was taken to be equal to that of $s_{20,w}$ at 0.35 mg/mL (1.72 S) as this concentration is close to zero and Species 1 was not resolved at higher concentrations.

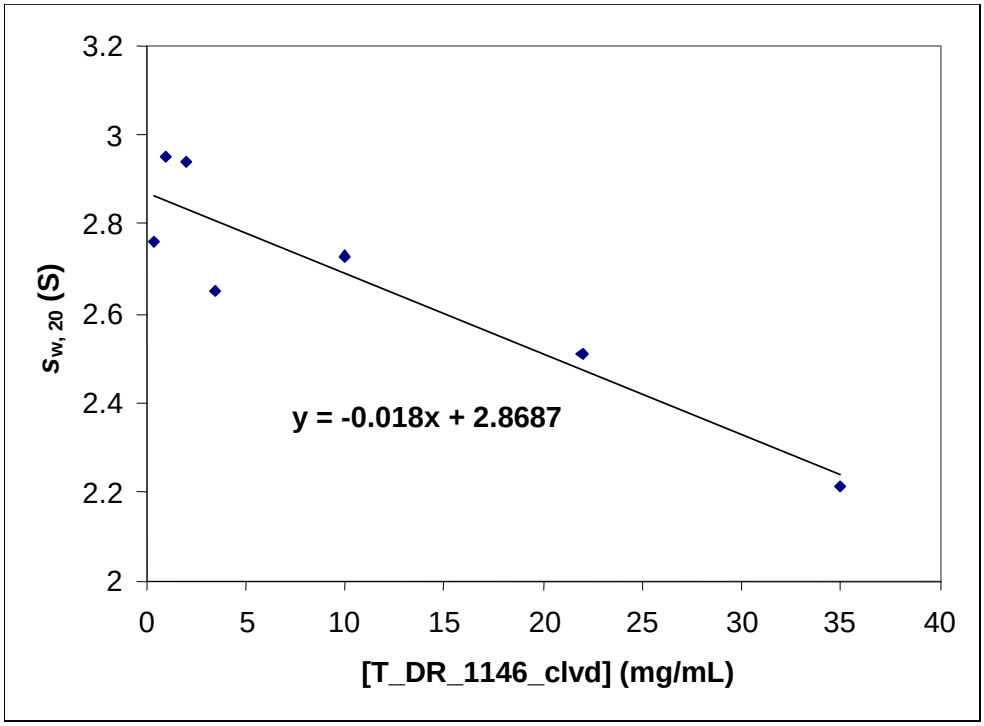


Figure 4.29: Plot of $s_{20,w}$ vs concentration for species 2 of holo T_DR_1146 giving $s^0_{20,w}$ = 2.87 S.

Overall, SV results suggest that holo T_DR_1146_clvd undergoes self-association to form a single species (Species 2) at concentrations above 2 mg/mL. As gel filtration elution profiles suggest that the protein exists as a dimer in solution, it is proposed that the observed change in quaternary structure with increasing concentration is from monomer to dimer, with Species 2 representing holo T_DR_1146_clvd in a dimeric conformation.

The theoretical sedimentation coefficient for dimeric holo T_DR_1146_clvd, if assumed to be a hydrated sphere (s_{hyd}), was calculated as 3.47 S by the method described in Section 2.7.6.3. The $s^0_{20,w}$ experimentally determined for Species 2 (2.87 S) is significantly lower than s_{hyd}, which suggests that the dimer is not a sphere. A lower than expected value of $s^0_{20,w}$ is attributable either to excessive hydration or, more probably, to asymmetry or elongation of the protein. It is therefore reasonable to speculate that dimeric holo T_DR_1146 is slightly elongated in solution.

Comparison of theoretical and experimentally determined sedimentation coefficients for monomeric holo T_DR_1146_clvd, (2.18 S and 1.72 S

respectively) infers a similar arrangement to that reported for the dimer; the monomer is more extended than a sphere of the same molar mass.

To confirm the proposed monomer-dimer self-association of holo T_DR_1146_clvd and to establish the concentration dissociation constant; K_D, an SE experiment was performed.

4.6.2 Sedimentation equilibrium (SE)

A SE experiment was carried out at three rotor speeds (20, 30 and 40 krpm) using the same seven sample concentrations used for SV AUC (0.35, 1, 2, 3.5, 10, 22 and 35 mg/mL). The resulting datasets were analysed using SEDPHAT software [181] as described in Section 2.7.6.2. Unfortunately the quality of data collected decreased over time, resulting in the exclusion of some 20 and 30 krpm datasets, and hence accounts for the absence of several points in graphs generated for data analysis.

4.6.2.1 Molecular weight determination for holo T_DR_1146_clvd

Datasets for each rotor speed were separately fitted with a 'species analysis' model which required values of predicted dimeric mass (36875 Da) and $s^0_{20,w}$ (2.87 S). Individual experiment fits were obtained for each of seven concentrations corresponding to 20, 30, and 40 krpm rotor speeds. The quality of the fit was judged by inspection of generated SE profiles, residuals of the fit and root mean squared deviation (rmsd) values. A typical experimental SE profile, demonstrating a good fit with the species analysis model, is shown in Figure 4.30.

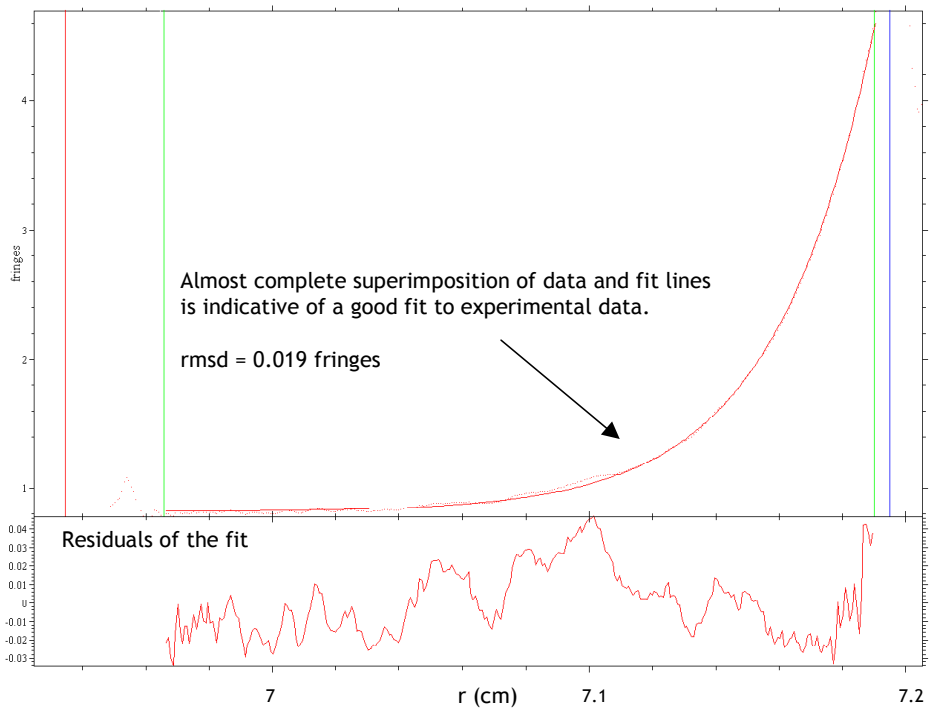


Figure 4.30: Typical SE profile for holo T_DR_1146-clvd at 0.35 mg/mL, 30 krpm, displaying a good fit of the chosen model to experimental data. Characteristics of a good fit include (a) almost complete superimposition of data (red solid curve) and fit (red, broken curve) lines (b) small, well-dispersed residuals and (c) low rmsd values.

The apparent weight-average molecular weights ($M_{w,app}$) determined for the 40, 30 and 20 krpm data are presented alongside rmsd values in Table 4.7.

Rotor speed (krpm)	[Protein] (mg/mL)	$M_{w,app}$ (Da)	Rmsd (fringes)
20	**0.35**	**19389**	**0.28**
	1	40698	0.03
	2	**62831**	**0.24**
	3.5	38456	0.04
	10	36225	0.03
	22	**30907**	**0.94**
	35	30425	0.03
30	0.35	37751	0.02
	1	38701	0.02
	2	**40471**	**0.20**
	3.5	33145	0.16
	10	31095	0.03
	22	**19597**	**1.04**
	35	23976	0.06
40	0.35	34201	0.04
	1	36501	0.03
	2	33509	0.06
	3.5	31568	0.06
	10	28704	0.11
	22	26421	0.14
	35	25688	0.06

Table 4.7: Values of $M_{w,app}$ (with rmsd) determined by fitting to the species analysis model for all datasets using SEDPHAT [181]. The values highlighted in bold were excluded from analysis due to the poor data quality.

$M_{w,app}$ values for well-fitted data were plotted against protein concentration (Figure 4.31) to illustrate the effect of increasing protein concentration.

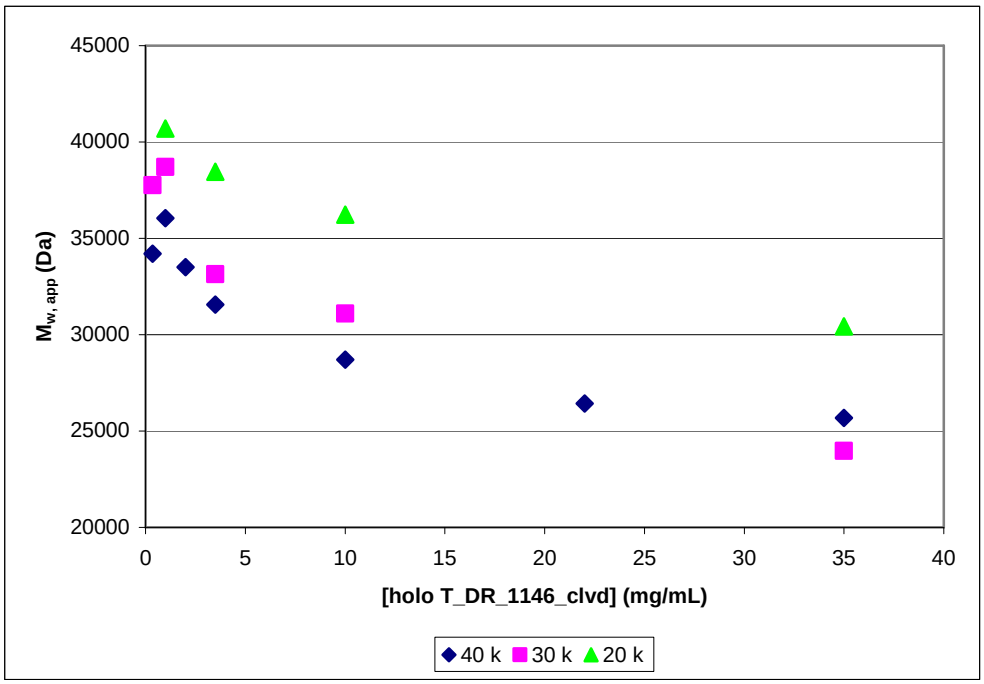


Figure 4.31: Plot of $M_{w,app}$ against concentration of holo T_DR_1146_clvd (mg/mL).

The distribution of points in Figure 4.31 reflects the non-ideal behaviour of the protein at significant concentrations as shown by the reduction in mass on increasing protein concentration for all three-rotor speeds. The anomalous decrease in molecular weight could be attributed to overcrowding of protein molecules towards the base of the AUC cell, which limits the ability of the protein to form steep concentration gradients.

On comparison with predicted dimeric molecular weight of the protein (36,875 Da), the trend depicted in Figure 4.31 shows that the molecular weight of holo T_DR_1146_clvd tends towards that of the dimeric protein at concentrations close to 0 mg/mL.

Individual global fits of all smooth experimental data sets for each rotor speed resulted in the $M_{w,app}$ values tabulated below (Table 4.8). A global fit of smooth data sets for all three speeds gave an overall molecular weight of 31,572 Da. It is thought that this value is lower than the predicted dimeric molecular weight of 36,875 Da due to non-ideal concentration effects.

Rotor speed (krpm)	Global $M_{w,app}$ for each rotor speed (Da)	Global $M_{w,app}$ for all smooth datasets (Da)
20	34663.35 (rmsd = 0.77)	31572.09 rmsd = 0.05
30	30593.07 (rmsd = 0.08)	
40	29073.25 (rmsd = 0.07)	

Table 4.8: Global $M_{w,app}$ values determined for 20, 30 and 40 krpm rotor speeds using SEDPHAT software [181].

4.6.2.2 Investigating different models for self-association

A global fit of all smooth experimental data sets was performed using three different self-association models: monomer-dimer (1-2), monomer-trimer (1-3) and monomer-tetramer (1-4). The model that best globally represented the analysed datasets was the 1-2 model, as shown by almost complete superimposition of data and fit lines and small, well-dispersed residuals (Figure 4.32). In addition, the rmsd value for the 1-2 model, was lower than those determined for the 1-3 and 1-4 models, a characteristic indicative of the best fit.

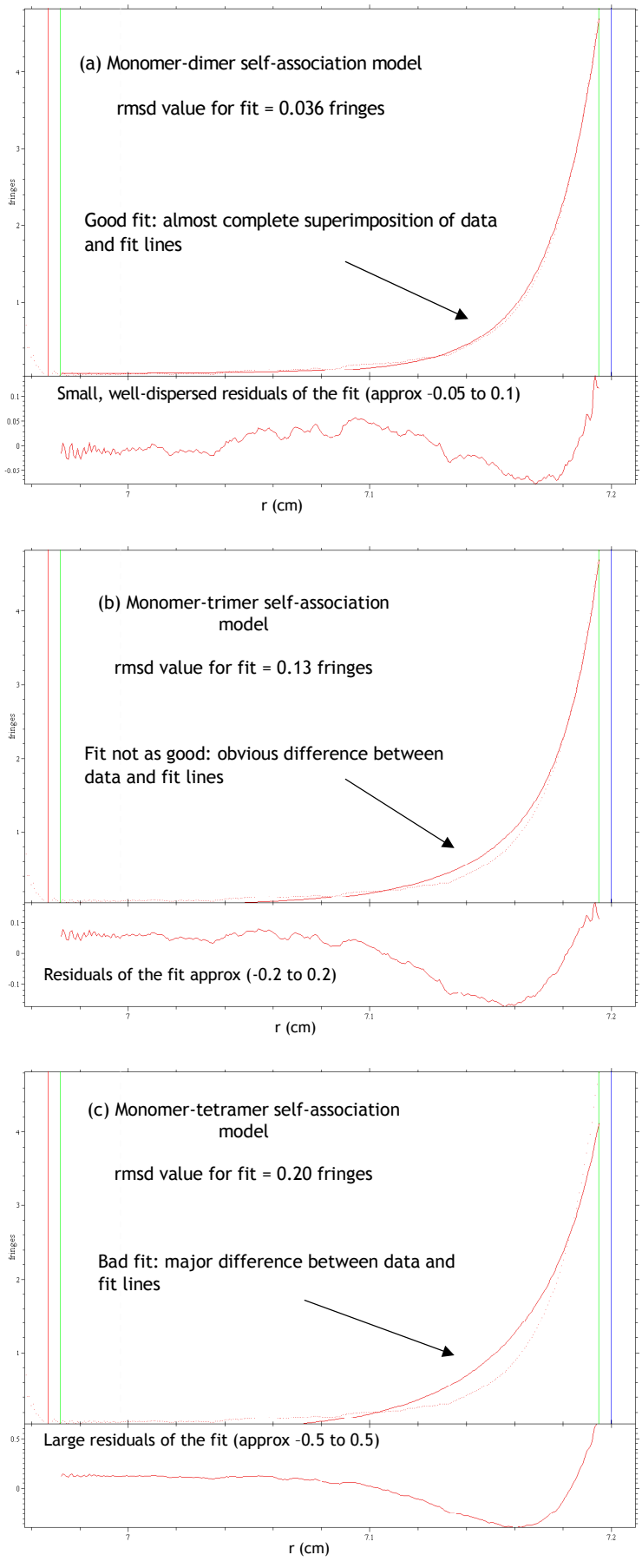


Figure 4.32: Representative SE profiles and residuals of the fits for holo T_DR_1146-clvd (0.35 mg/mL, 40 krpm) for self-association models (a) 1-2 (b) 1-3 and (c) 1-4. Data are best fitted by the 1-2 model confirming that a change from monomer to dimer as concentration increases.

4.6.2.3 Determination of dissociation (K_D^{1-2}) constant for 1-2 self-association

By globally fitting all smooth datasets (concentrations 0.35 to 3.5 mg/mL) with the 1-2 model, a dissociation constant (K_D^{1-2}) of 7.67 μM was obtained. This value is consistent with the results obtained from SV experiments, which showed that self-association was underway at 0.35 mg/mL (19 μM), and hence must be beginning at a concentration below this. The K_D^{1-2} obtained by SE AUC confirms that at concentrations below 7.67 μM, holo T_DR_1146_clvd exists predominantly as a monomer and, above 7.67 μM, mostly as a dimer. From these data, it can be postulated that the structure of holo_T_DR_1146_clvd, derived as a result of high concentration crystallisation experiments, would be dimeric.

5 Crystallisation of degraded T_DR_1146 with FMN

5.1 Research aims

The aim of work described in this Chapter was to gain a stable, homogeneous T_DR_1146 sample for crystallisation experiments. This work involved attempting to saturate protein binding sites with the FMN cofactor and investigating the extent of proteolysis with the objective of isolating a stable protein fragment.

5.2 Purification of T_DR_1146 in the presence of FMN

The following three experiments were performed to investigate the effect of FMN binding on the prevention of proteolysis and the impact on crystallisation of T_DR_1146.

(1) Cell lysis in the presence of FMN

A final concentration of 2 mM of FMN was added to the standard lysis buffer shown in Table 2.8. After cell lysis, the lysate was purified by IMAC and SEC (Sections 2.5.4 and 2.5.7). Following 6xHis tag cleavage, crystallisation experiments were set up. No protein crystals were obtained.

SDS-PAGE analysis of the purified protein showed that although degradation was slowed by the presence of FMN, it was not stopped completely.

Throughout the purification process, the yellow colour attributed to FMN, became paler. This was thought to be due to either the removal of unbound FMN or due to dissociation of bound FMN due to high NaCl concentrations in purification buffers.

It was concluded that partial saturation of binding sites and the occurrence of degradation gave an inhomogeneous sample, which was unfavorable for crystallisation.

(2) Addition of FMN throughout purification

Two equimolar equivalents of FMN were added to T_DR_1146 samples after each purification step to compensate for reversible binding. The last addition was made after concentration for SEC. Any excess FMN was removed by passage through a S200 10/300 HR Global gel filtration column (Section 2.5.7). Subsequent crystallisation experiments were again unsuccessful.

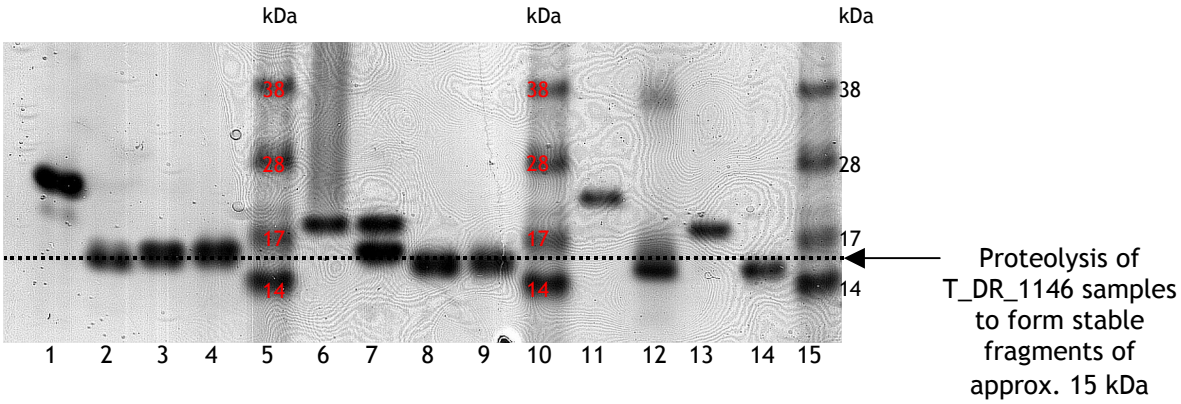
(3) Attempted co-crystallisation with FMN

Purified T_DR1146_clvd was concentrated to 20 mg/mL (0.98 mM) and FMN was added to a final concentration of 2 mM. The mixture was incubated at 4°C for 2 hrs then crystal trials were set up. No credible protein crystals were obtained.

As the protocols followed in experiments (2) and (3) ensured complete saturation of binding sites with FMN, it was thought that sample inhomogeneity due to protein degradation was a major factor in the inability to grow protein crystals.

5.3 Investigating degradation of T_DR_1146

To investigate the extent of protein degradation, concentrated samples of cleaved and non-cleaved T_DR_1146 were stored at 4°C for a maximum time of two months. SDS-PAGE analysis of these degraded samples showed that in general, proteolysis resulted in formation of a stable fragment of approximately 15 kDa (Figure 5.1, lanes 2-4, 8 & 9).



Lane 1: Apo T_DR_1146
Lane 2: Apo T_DR1146 degraded
Lane 3: Apo T_DR1146 degraded
Lane 4: Apo T_DR1146 degraded
} Stored for two months
Lane 5: MW Marker
Lane 6: Apo T_DR_1146_clvd
Lane 7: Apo T_DR_1146_clvd partially degraded
Lane 8: Apo T_DR_1146_clvd degraded
Lane 9: Apo T_DR_1146_clvd degraded
} Stored for two months
Lane 10: MW Marker
Lane 11: Halo T_DR_1146 with tag
Lane 12: Halo T_DR_1146 degraded (stored for five weeks)
Lane 13: Halo T_DR_1146_clvd
Lane 14: Halo T_DR_1146_clvd degraded (stored for weeks)
Lane 15: MW Marker

Figure 5.1: SDS-PAGE analysis of degraded T_DR_1146 samples. Lane 1: non-cleaved apo T_DR_1146 sample before proteolysis. Lanes 2-4: non-cleaved apo T_DR_1146 samples after 2 months showing degradation to a stable fragment. Lane 6: cleaved apo T_DR_1146 sample before proteolysis. Lane 7: partial degradation of cleaved apo T_DR_1146 sample after 2 months. Lanes 8&9: cleaved apo T_DR_1146 sample after 2 months showing degradation to a stable fragment. Lanes 11&12: non-cleaved holo T_DR_1146 before and after proteolysis showing formation of stable fragment after 5 weeks. Lanes 13&14: cleaved holo T_DR_1146 before and after proteolysis showing formation of stable fragment after 5 weeks.

Samples of non-cleaved and cleaved holo T_DR_1146 were stored at 4°C for five weeks. The extent of degradation of theses samples is shown in lanes 12 and 14 respectively, confirming that the addition of FMN in the protein sample does not prevent proteolysis. Comparison of lanes 11 and 12 with 13 and 14 shows that non-cleaved T_DR_1146 degrades to the same level as T_DR_1146_clvd suggesting that the tag is removed by proteolysis without the addition of thrombin. This can also be seen for apo T_DR_1146 (no additional FMN added) by comparing lanes 1-4 (tagged) with 6-9 (cleaved).

The degraded samples shown in lanes 2, 8, 12 and 14 were purified by anion exchange to remove degraded peptides. Crystallisation experiments were then set up however, no protein crystals resulted.

The degraded non-cleaved holo T_DR_1146 sample (lane 12) was analysed by N-terminal sequencing, establishing the first eight residues of the amino acid sequence to be Ser1, His2, Met3, Ser4, Arg5, Asp6, Glu7 and Ala8. Residue 3 (Met3) matches residue 1 of the T_DR_1146 sequence confirming that the 6xHis tag is removed without the addition of thrombin. The residues Ser1 and His2, which were encoded from vector nucleotides, remain intact after N-terminal proteolysis.

ESI-MS analysis of the degraded non-cleaved holo T_DR_1146 sample showed numerous peaks in the 15,700-15,800 Da region with two main peaks at 15,694 and 15,710 Da (Figure 5.2).

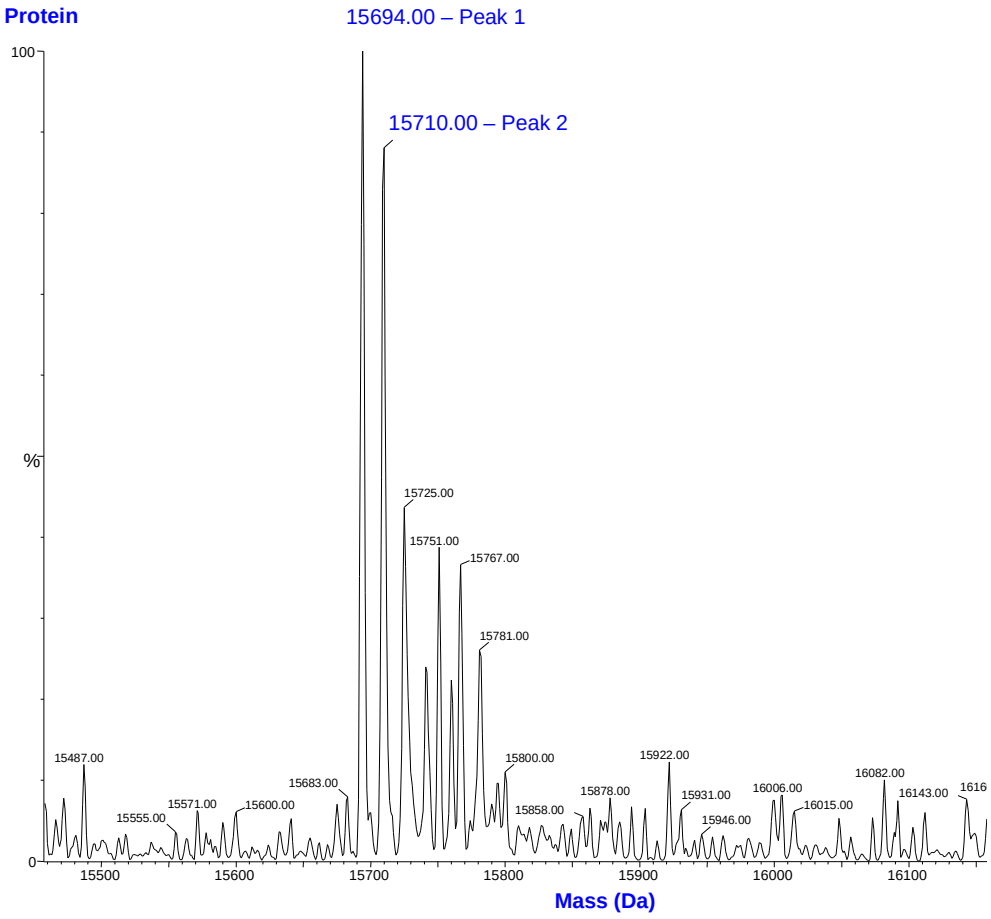


Figure 5.2: ESI-MS analysis of degraded non-cleaved holo DR_1146 showing two main peaks at 15,694 Da (Peak 1) and 15,710 Da (Peak 2).

The reported N-terminal sequencing and ESI-MS results were used in conjunction with the ProtParam tool [171] from the ExPASy proteomics server [172] to predict a possible C-terminal cleavage site for T_DR_1146. A method of subtracting the MWs of individual amino acids from the theoretical MW of DR_1146 (18,380 Da including vector residues Ser1 and His2) was used in order to

match the masses determined by ESI-MS (15,694 and 15,710 Da). Figure 5.3 shows the predicted C-terminal cleavage site, which is located between the residues Phe139 and Ala140.

For these calculations, it was assumed that FMN was removed during ESI-MS as the technique involves use of acetonitrile, which dissociates non-covalent intermolecular forces between the cofactor and protein during the experiment.

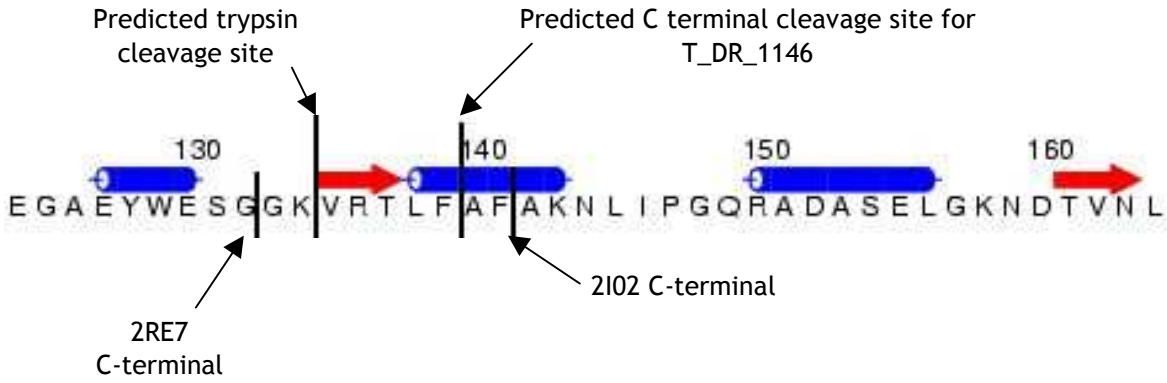


Figure 5.3: Residues 124-164 of the amino acid sequence of T_DR_1146 with the predicted C-terminal cleavage site marked between Phe139 and Ala140. The predicted secondary structure is shown above the amino acid sequence of T_DR_1146 (determined by Network Protein Sequence Analysis[194, 195]). C-terminal sites for PDB structures 2RE7 and 2i02 are also indicated by vertical black lines. The PeptideCutter tool [171] from the ExPASy proteomics server [171, 172] was used to predict a C-terminal trypsin cleavage site between residues K135 and V136. Figure 5.3 was generated using the program Aline [196].

5.4 Analysis of FMN binding in the crystal structures of two T_DR_1146 homologues from the PDB

A search of the RCSB PDB [79, 80] using the predicted amino acid sequence of T_DR_1146 gave the 3-D crystal structures of 10 homologues with between 47 to 21 % sequence identity with the target T_DR_1146 protein. Amino acid sequence alignment (performed using Multalin [197]) of 7 of theses homologues (Figure 5.4) shows complete conservation of four residues (red) and partial conservation of several regions. One of the conserved residues is a methionine that is involved in FMN binding of structure 2I02 (described later in this Section).

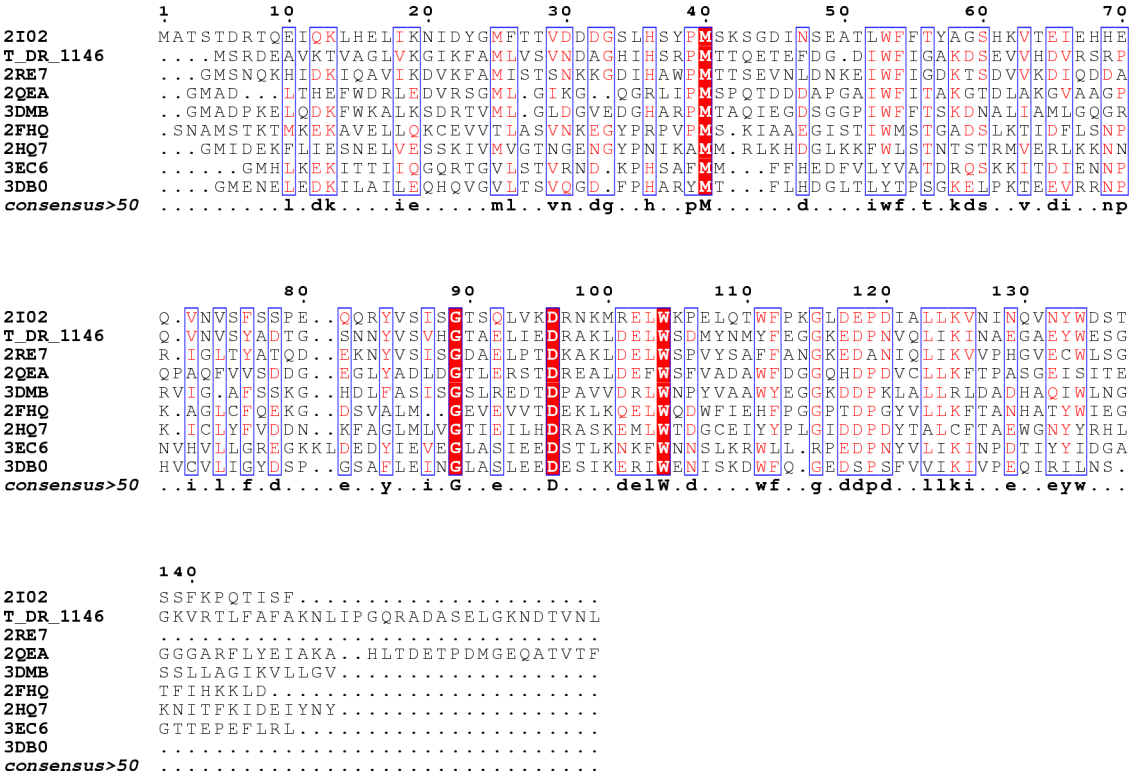


Figure 5.4: Amino acid sequence alignment of T_DR_1146 with homologous PDB structures including 2RE7 and 2IO2 using MultAlin [197]. Figure and secondary structure of 2I02 generated by ESPript 2.2 [198, 199] using data from the PDB.

The structure 2RE7 (PDB ID) is a putative general stress protein (COG3871) from *Psychrobacter arcticus* 273-4 solved at 2.50 Å resolution [200]. The amino acid sequence of 2RE7 shows 47 % identity with that of T_DR_1146 and it was crystallised in the absence of a FMN cofactor (Figure 5.5a).

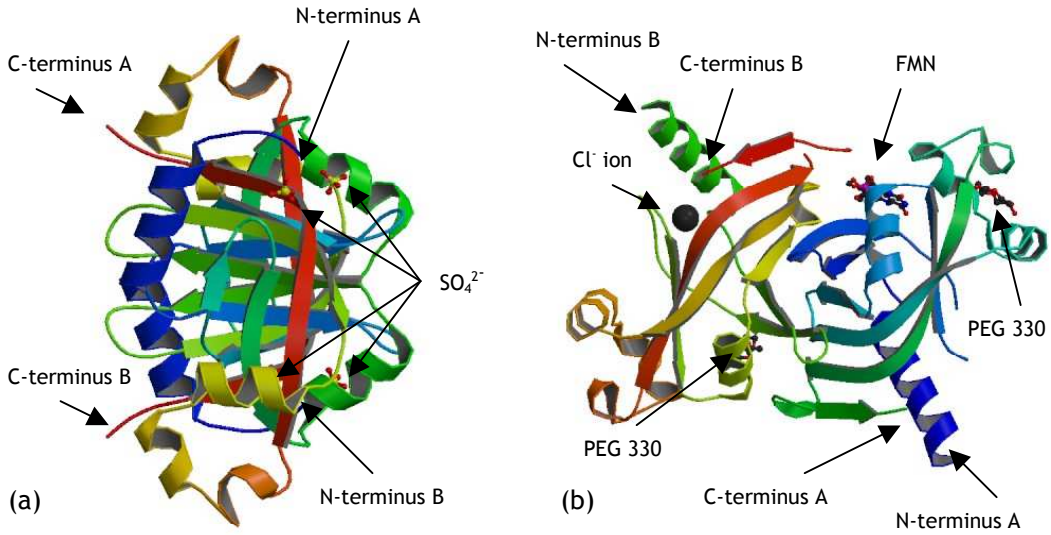


Figure 5.5: 3-D schematic ribbon representation of (a) 2RE7 complexed with four SO_4^{2-} ions [200] and (b) 2IO2 complexed with one molecule of FMN, two molecules of PEG 330 and a Cl^- ion [81] taken from the RCSB PDB [79, 80]. B-strands are represented as arrows, α-helices as coiled ribbons and random coils as thin tubes. Bound prosthetic molecules and ions are represented by ball and stick diagrams. The direction of the polypeptide chains are depicted by the following colour ramps: (a) both chains of 2RE7 go from blue (N-terminus) to red (C-terminus) and (b) Chain A of 2I02 goes from blue (N-terminus A) to green (C-terminus A) and Chain B of 2I02 goes from green (N-terminus B) to red (C-terminus B).

Each peptide chain is 133 amino acids long ending with Gly133. However, the predicted UniProt [201] amino acid sequence was 164 residues in length suggesting that, like T_DR_1146, the protein underwent C-terminal proteolysis before crystallisation. The C-terminal cleavage site (after Gly133) is marked on the equivalent T_DR_1146 sequence in Figure 5.3.

The structure of a putative general stress proteins (COG3871) from *Nostoc punctiforme* PCC 73102 (PDB ID: 2I02), solved at 1.80 Å resolution has 32 % identity with T_DR_1146. In contrast to 2RE7, the dimeric structure of 2I02 was crystallised with one molecule of FMN bond in a hydrophobic pocket positioned at the dimer interface (Figure 5.5b). The Structural Classification of Proteins database (SCOP) [202] classified Chains A and B of 2I02 to have a split-barrel like fold, which is characteristic of the FMN-binding split barrel superfamily. This fold is shared with 47 of the 421 FMN binding proteins entered on the PDB.

In order to predict which T_DR_1146 residues are involved in FMN binding, the interactions of 2I02 with the bound FMN were examined using the programs Coot [203] and Relibase 2.2.1 [204, 205].

The single FMN molecule is involved in extensive hydrophobic and hydrogen bond interactions with 2I02. Most of the interactions involve Chain A however two residues from Chain B bind to the ribityl hydroxyl region of FMN.

The isoalloxazine moiety of the FMN molecule forms hydrophobic contacts (of less than 3.8 Å) with Trp111, Phe55 and Mse40 residues of Chain A. Only an unmodified Methionine residue is conserved in the T_DR_1146 and 2RE7 sequences with Tyr and Ile (T_DR_1146) and Phe and Ile (2RE7) in place of Trp111 and Phe55. In addition to hydrophobic interactions, the isoalloxazine ring and 2I02 are involved in hydrogen-bonded interactions (Figure 5.6). The ring atoms O4 and N3 make direct hydrogen bonds with Ser41(N) and Phe55(O) of Chain A. Neither of theses two amino acids are conserved in the sequences of T_DR_1146 or 2RE7.

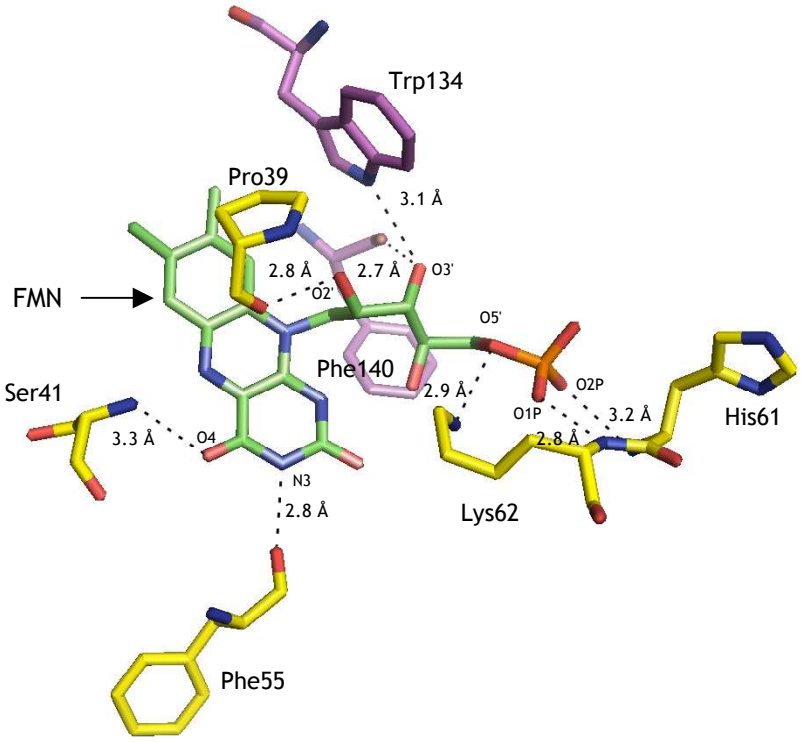


Figure 5.6: Direct hydrogen bond interactions between Chain A (yellow) and Chain B (magenta) of 2I02 and a single FMN molecule. Nitrogen, oxygen and phosphate atoms are coloured blue, red and orange respectively and hydrogen bonds indicated by dotted black lines. Figure generated using PyMOL [206].

The hydroxyl groups of the ribityl chain form three direct hydrogen bonds with Pro39(O) of Chain A, and Trp134(NE1) and Phe140(O) of chain B of which Pro39 (Chain A) and Trp134 (Chain B) are conserved.

His61(N) and Lys62(NZ and N) form direct hydrogen bonds to the prosthetic phosphate moiety of FMN (O2P and O1P), and to the ribityl atom O5'. It can be seen in the sequence alignment (Figure 5.4) that neither His61 nor Lys62 are conserved in the sequences of T_DR_1146 or 2RE7. ITC results suggest that the phosphate moiety of FMN (which is part of a prosthetic group that differs between flavins) plays no part in binding to T_DR_1146. It is therefore possible that the absence of Lys62 and His 61 could have a negative impact on the FMN binding affinity of T_DR_1146. As 2RE7 is also deficient of Lys62 and His61, this may explain why this particular protein was crystallised in the absence of FMN. Only biophysical characterisation of 2I02 and 2RE7 and structure determination of T_DR_1146 will confirm such hypotheses.

A summary of the atoms of FMN and amino acid residues involved in hydrogen bonding is given in Table 5.1.

Amino acid residue	Atoms involved in hydrogen bond	Chain of 2I02	Bond length (Å)	FMN atoms
Phe55	Backbone carboxyl O	A	2.8	Ring N3'
Ser41	Backbone N	A	3.3	Ring O4
Pro39	Backbone carboxyl O	A	2.8	Ribityl O2'
Trp134	Cyclic N in side chain (NE1)	B	3.1	Ribityl O3'
Phe140	Backbone carboxyl O	B	2.7	Ribityl O3'
Lys62	Side chain N (NZ)	A	2.9	Ribityl 05'
Lys62	Backbone N	A	2.8	O1P
His61	Backbone N	A	3.2	O2P

Table 5.1: Summary of the hydrogen bonds formed between 2I02 and FMN

In conclusion, as the 2I02 FMN binding residues Mse40, Pro39 and Trp134 are conserved in the amino acid sequence of T_DR_1146, it is thought that they may be involved in binding of the protein to the isoalloxazine ring system and ribityl chain of the FMN cofactor. It is unknown whether the replacement residues for un-conserved amino acids will interact with FMN in the same way as in 2I02. However, it is thought that the absence of a Lys62 and His61 in T_DR_1146 may

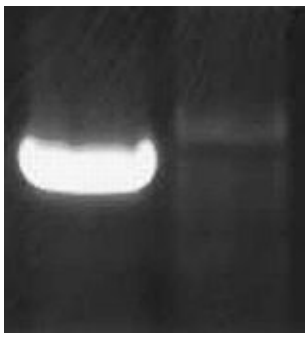
prevent hydrogen bonding to the phosphate moiety and explain predicted redundancy of the prosthetic –X group of flavins in binding to the protein.

5.5 Construct design and cloning of *t_DR_1146* with a C-terminal truncation

To investigate crystallisation of the stable T_DR_1146 fragment, attempts were made to clone a C-terminal truncated form of the *t_dr_1146* gene (*c_t_dr_1146*) into pOPINF. The *c_t_dr_1146* construct was designed to encode a protein of 132 residues ending at Gly132. This was based on the C-terminal cleavage site of the homologous 2RE7 crystal structure, which is thought to have degraded to this point. This C-terminal site was chosen for construct design because of its close proximity to a predicted trypsin cleavage site (determined by PeptideCutter [171, 172]) in a region with no predicted secondary structure (determined by Network Protein Sequence Analysis [194, 195]) (Figure 5.3).

Primers were designed as described in Section 2.2.1 and are listed in Table 2.2. The construct DNA of *t_dr_1146* in pET15b was used as the DNA template in the PCR step and the master mix recipe and temperature cycle shown in Section 2.2.3 and Table 2-3 were used.

An optimum annealing temperature of 58°C resulted in successful amplification of the *c_t_dr_1146* gene fragment (396 bp) shown in Figure 5.7.



Lane 1: Amplified *c_t_dr_1146* (396 bp)

Lane 2: 1 Kb Plus DNA Ladder

Figure 5.7: PCR amplification of *c_t_ dr_1146* gene fragment imaged by agarose gel electrophoresis.

This band was excised and purified as described in Section 2.2.4. The purified PCR product was added to the InFusion reaction mixture as described in Section

2.2.5 and the reaction was immediately quenched with TE buffer and 5 µL of the reaction mixture was used to transform DH5α cells. The absence of colonies after overnight incubation at 37°C proved to be a negative result. The competency of the DH5α cells was confirmed by a positive control.

This work could not be continued due to time constraints.

5.6 Purification and crystallisation of degraded ^{15}N labelled T_DR_1146 sample

A 0.9 mM sample of T_DR_1146, produced for NMR experiments, was incubated at 4°C for three months in a buffer of 50 mM phosphate buffer pH 7.5, 150 mM NaCl, 0.01 % NaN_3 and 2 mM FMN. NMR experiments repeated over the three-month period determined that proteolysis was complete due to no further changes in ^{15}N HSQC spectra.

The sample was purified by SEC using the same buffer but without 2 mM FMN. The resulting A_{280} elution profile (Figure 5.8) showed a single peak (Peak 1) eluted in 15.97 mL of buffer which corresponds to a dimer of 27.9 kDa (including FMN contributions). Later peaks correspond to removal of degradation products and un-bound FMN.

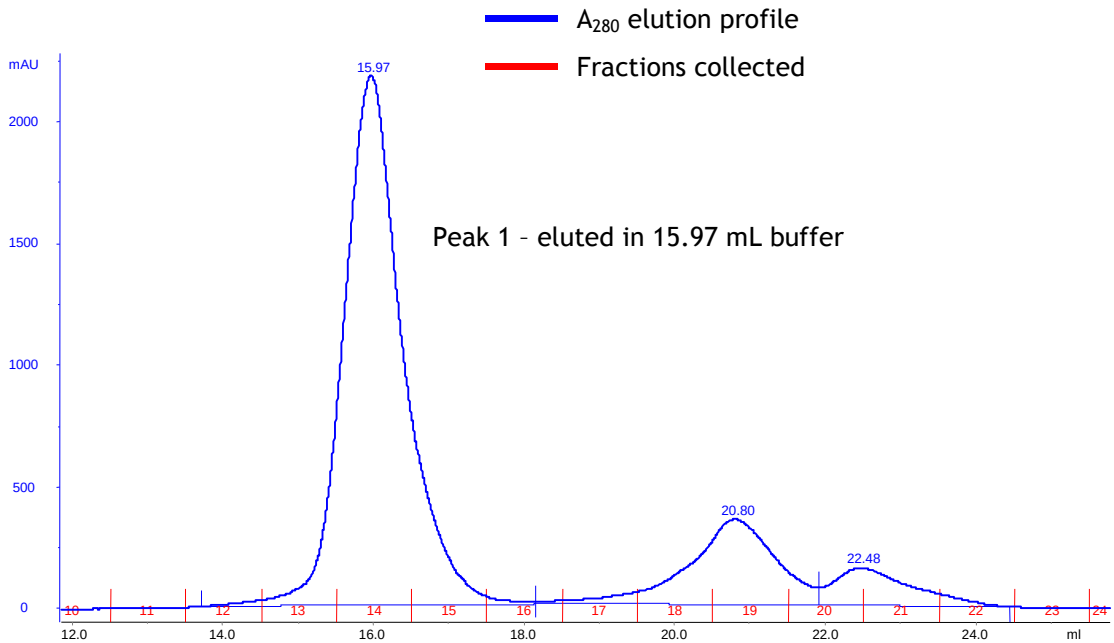


Figure 5.8: A_{280} elution profile for SEC purification of degraded T_DR_1146 shows a single peak eluted with 15.97 mL of buffer corresponding to a dimer.

A high level of purity was confirmed by SDS-PAGE analysis of the degraded T_DR_1146 eluted in Peak 1 (Figure 5.9, lanes 4-6).

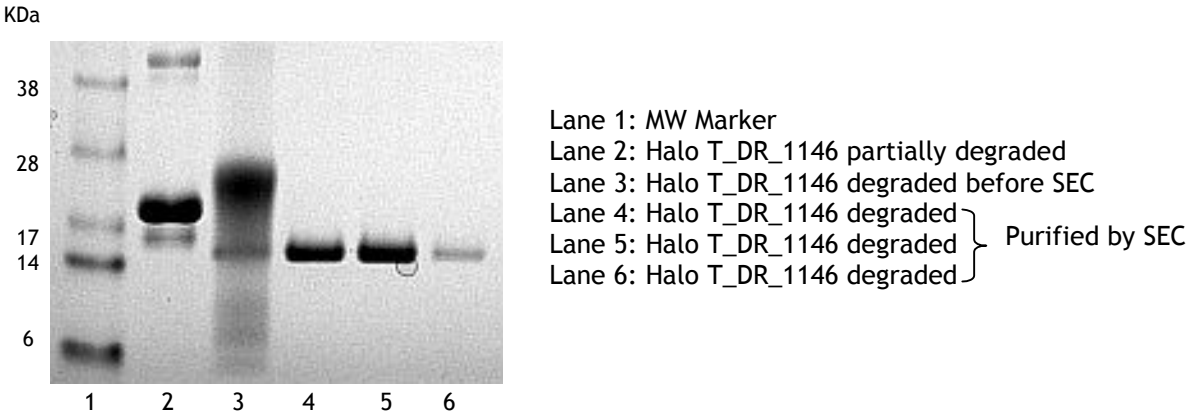


Figure 5.9: SDS-PAGE analysis of degraded T_DR_1146 after purification by SEC.

The purified protein was concentrated to 15 mg/mL and crystal trials were set up following the minimalistic approach described in Section 6.11. This involved the use of PACT 1 & 2 and JCSG+ crystallisation screens with site 1 containing degraded T_DR_1146 (15 mg/mL) and site 2 containing a negative control (gel filtration buffer).

After 3 days, small yellow microcrystals formed in site 1, condition B1 (0.1 M MIB buffer pH 4.0, 25 % w/v PEG 1500) of the PACT screen (Figure 5.10).

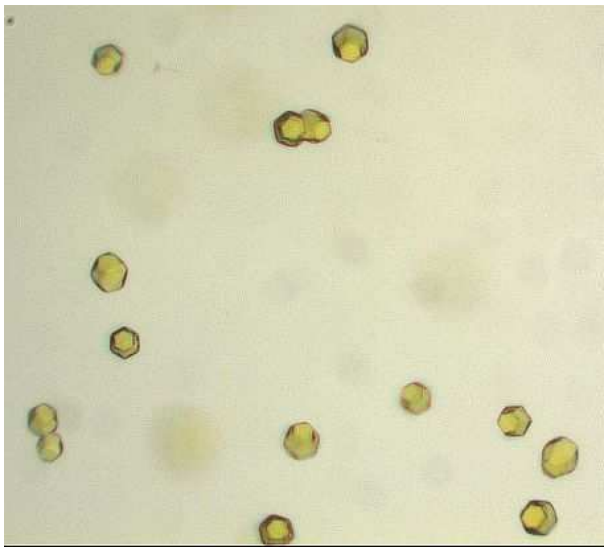


Figure 5.10: Crystals of degraded T_DR_1146 in 0.1 M MIB buffer pH 4.0, 25 % w/v PEG 1500.

The composition of MIB buffer pH 4.0 (Molecular Dimensions) is shown in Table 5.2.

Chemical Name and source	Molar ratio	Formula weight	Mass (g)	Final buffer volume (mL)	Final buffer conc. (M)	Comment
Sodium malonate	2	266.00	4.15	100	1	Adjust to pH 4.0 with 10 M NaOH
Imidazole	3	68.08	2.55			
Boric acid	3	61.83	2.32			

Table 5.2: Composition of MIB buffer pH 4.0 provided by Molecular Dimensions.

The microcrystals displayed no birefringent properties and in most cases were fused together. The observed yellow colour of the crystals was more intense than the pale yellow background drop.

After two weeks of incubation the crystals grew to approximately 50 µm in diameter with more defined edges.

The crystals were mounted into 0.1 mm loops before being immersed in two different cryoprotectants (Table 5.3).

Cryoprotectant	Composition	Comments
A	0.1 M MIB buffer pH 4.0 25 % PEG 1500 25 % Glycerol	MIB buffer taken from 1 M stock made following recipe in Table 5.1
B	0.1 M MIB buffer pH 4.0 40 % PEG 1500	MIB buffer taken from 1 M stock made following recipe in Table 5.1

Table 5.3: Composition of experimental cryoprotectants.

Initial in-house X-ray diffraction experiments resulted in no diffraction. However, on exposure to synchrotron radiation, a diffraction pattern with a resolution limit of 7.3 Å (Figure 5.11) was observed, confirming analysis of protein crystals rather than FMN or salt. The most defined spots were gained using cryoprotectant A, resulting in some lunar patterns (Figure 5.11, zoomed). The presence of rings or merged spots suggests that diffraction resulted from a cluster of single microcrystals rather than a single crystal.

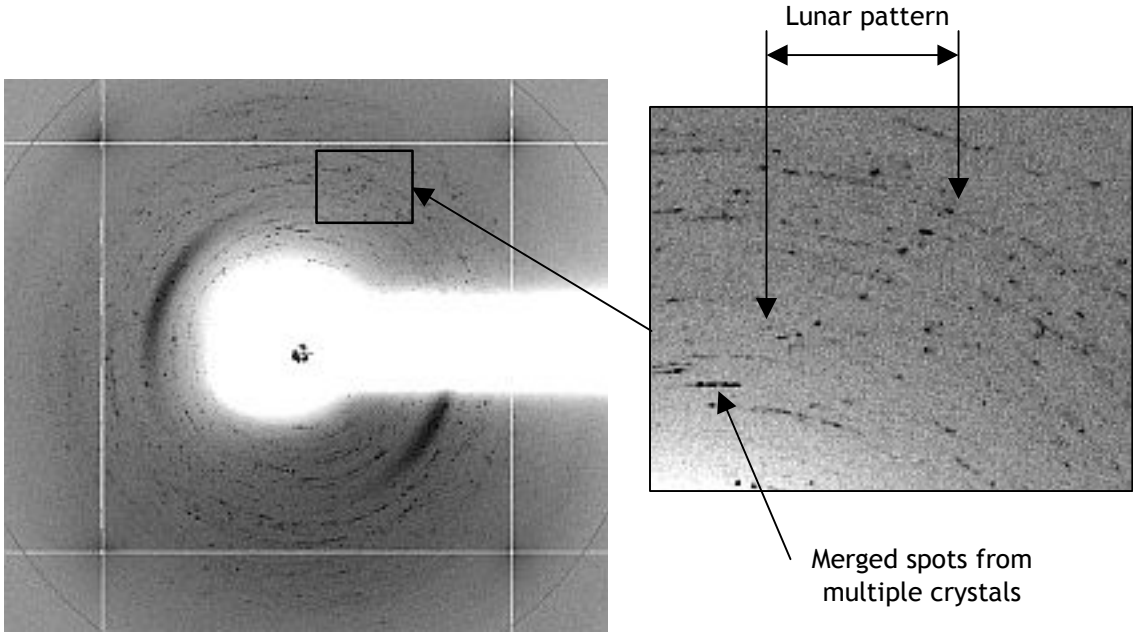


Figure 5.11 - Diffraction pattern for degraded T_DR_1146 crystals. Zoomed region shows lunar patterns as a result of diffraction using cryoprotectant A. The presence of rings or merged spots suggests that diffraction resulted from a cluster of single microcrystals rather than a single crystal.

It was concluded that crystallisation of degraded T_DR_1146 occurred due to sample homogeneity achieved by isolating a stable fragment and saturating binding sites with FMN. As T_DR_1146 shares high sequence homology with the apo protein 2ER7, it is thought it may be possible to crystallise an apo form of degraded T_DR_1146 in the absence of FMN.

6 Expression, purification and crystallisation of SP_1651

6.1 Project aims

The aim of this project was to investigate the expression, purification and crystallisation of SP_1651 from *S. pneumoniae* with the objective of obtaining the 3-D crystal structure. This would allow structural and functional comparisons to be made with a homologue from *D. radiodurans* (DR_2242), for which the structure has been solved by Dr. Dave Hall as part of the ESRF's structural genomics project.

6.2 Construct design

A full length form of the *sp_1651* gene (residues 1 to 519) cloned into a pOPINF vector (Table 2.1) was kindly provided by Dr. Alan Riboldi-Tunnicliffe, University of Glasgow. The cloning work was carried out at the Oxford Protein Production Facility, Oxford, using the In-Fusion™ PCR Reaction system (Clontech).

Throughout this chapter, protein encoded from the full length gene with the N-terminal 6xHis tag still intact will be referred to as SP_1651. The protein with the 6xHis tag cleaved using HRV-3C protease will be referred to as SP_1651_clvd.

6.3 Recombinant expression of *sp_1651* from pOPINF

2 µL plasmid DNA (*sp_1651* in pOPINF) was used to transform B834(DE3) competent cells by the method described in Section 2.2.6. 5 µL of the transformation reaction mixture gave an even spread of single colonies after 20 hrs incubation at 37°C.

Recombinant *sp_1651* was expressed in Overnight Express™ Autoinduction Media (Table 2.5) from the pOPINF vector as described in Section 2.4.6. A 100 x dilution of starter culture was used to initiate cell growth in a 500 mL culture which was incubated at 37°C for 4 hrs, followed by 20 hrs at 25°C at which an

OD_{600} of 7.05 was reached. A cell pellet of 6.18 g was harvested by centrifugation of the 500 mL culture.

6.4 Initial purification trial

The pellet was re-suspended in 35 mL of lysis buffer (Table 2.8) with the addition of 0.2 % Tween 20. The cells were lysed by French press and loaded on to a 1 mL HisTrap HR column as described in Sections 2.5.3 and 2.5.4. IMAC was carried out on the bench, with the lysate incubated on ice, using a peristaltic pump at a flow rate of 1 mL/min. The loaded column was flushed with 12 column volumes (CVs) of IMAC wash buffer followed by 10 CVs of IMAC elution buffer. The composition of buffers used for purification of SP_1651 are shown in Table 2.8.

Successful expression of *sp_1651* is evident in lane 3 of Figure 6.1 which shows a defined band at approximately 21 kDa (ringed). The presence of SP_1651 in this lane, and its absence in lanes 1 and 2, corresponding to the column flow through and the 12 CV wash respectively, shows that the tagged protein had an affinity for the Ni^{2+} matrix making purification by IMAC feasible. The significant number of impurities as seen in lane 3 of Figure 6.1, indicates further optimisation of the IMAC protocol is necessary to obtain pure protein.

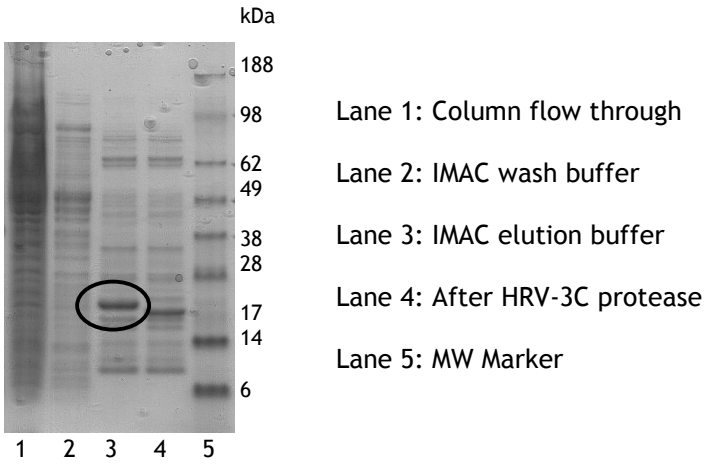


Figure 6.1: SDS-PAGE analysis of fractions eluted from initial IMAC purification.

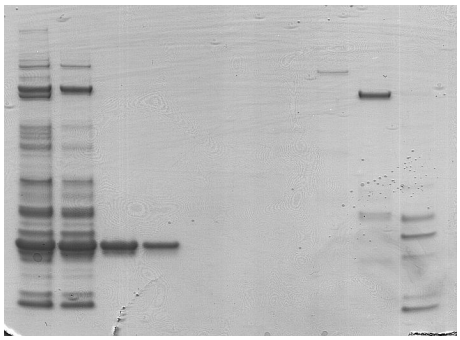
An A_{280} measurement of the elution fraction containing SP_1651 gave a peak maximum at 277.2 nm. A yield of 21.0 mg protein (including impurities) was calculated (Section 2.5.5).

6.5 Initial investigation of 6xHis tag cleavage with HRV-3C protease

Cleavage of the 6xHis tag using HRV-3C protease was carried out as described in Section 2.5.9. The decrease in molecular weight of the SP_1651 band in lane 4 of Figure 6.1 compared to lane 3 confirmed the success of the cleavage reaction.

6.6 Initial investigation into the removal of cleaved 6xHis tag by IMAC

After dialysis to remove imidazole, the cleaved sample was loaded on to a 1 mL HisTrap column and the flow through was collected. The column was flushed with IMAC wash buffer plus increasing concentrations of imidazole in 5 column volume intervals. The single bands in lanes 3 and 4 of Figure 6.2 show that SP_1651 was eluted in the column flow through (lane 3) and with 5 CVs of IMAC wash buffer (lane 4).



Lane 1: After HRV-3C cleavage
Lane 2: Dialysed to remove precipitate
Lane 3: Column flow through
Lane 4: O mM imidazole buffer
Lane 5: 10 mM midazole buffer
Lane 6: 20 mM imidazole buffer (IMAC wash buffer)
Lane 7: 30 mM imidazole buffer
Lane 8: 50 mM imidazole buffer
Lane 9: 100 mM imidazole buffer
Lane 10: 500 mM imidazole buffer (IMAC elution buffer)

Figure 6.2: SDS-PAGE analysis of fractions eluted from IMAC purification of SP_1651_clvd

Lanes 3 and 4 were pooled together and a yield of 6.3 mg protein was calculated. The SP_1651 bands were very pure and thus suitable for crystallisation experiments. The faint band just below the SP_1651 bands in lanes 3 and 4 may either be an impurity or a proteolysis product.

The purified SP_1651 was concentrated to 9 mg/mL and two sparse matrix screens were set up: Hampton 1 and 2, Cryo 1 and 2 and an in house Grid Screen. They were stored and monitored at 20°C in a Rhombix Imaging System.

After six weeks no protein crystals resulted, although crystalline precipitate was observed for several conditions.

6.7 Optimised purification protocol for crystallisation experiments

1 L of cell culture was produced following the protocol outlined in Section 2.4.6. Cell lysis was carried out using the Constant Cell Disruption System (Section 2.5.3). After cell lysis and harvesting an additional Complete EDTA-free Protease Inhibitor tablet was added to the soluble supernatant to avoid possible proteolysis.

IMAC was carried out using a 5 mL HiTrap HR column to accommodate an increased amount of protein. Using an AKTA purifier FPLC, a step gradient including two additional imidazole concentrations (10 and 20% IMAC elution buffer) and an increased number of CVs was employed in an attempt to remove impurities (Table 6.1).

Step number	Number of column volumes (1 CV = 5 mL)	% IMAC wash buffer (20 mM imidazole)	% IMAC elution buffer (500 mM imidazole)	Effective imidazole conc. (mM)
1	12	100	0	20
2	11	90	10	68
3	8	80	20	116
4	8	0	100	500

Table 6.1: Optimised IMAC protocol.

Figure 6.3 is the A_{280} elution profile for the optimised purification of SP_1651. Peaks 1 and 2 (approximately 250 mAU) correspond to protein eluted with the additional 10 and 20 % IMAC elution buffer steps.

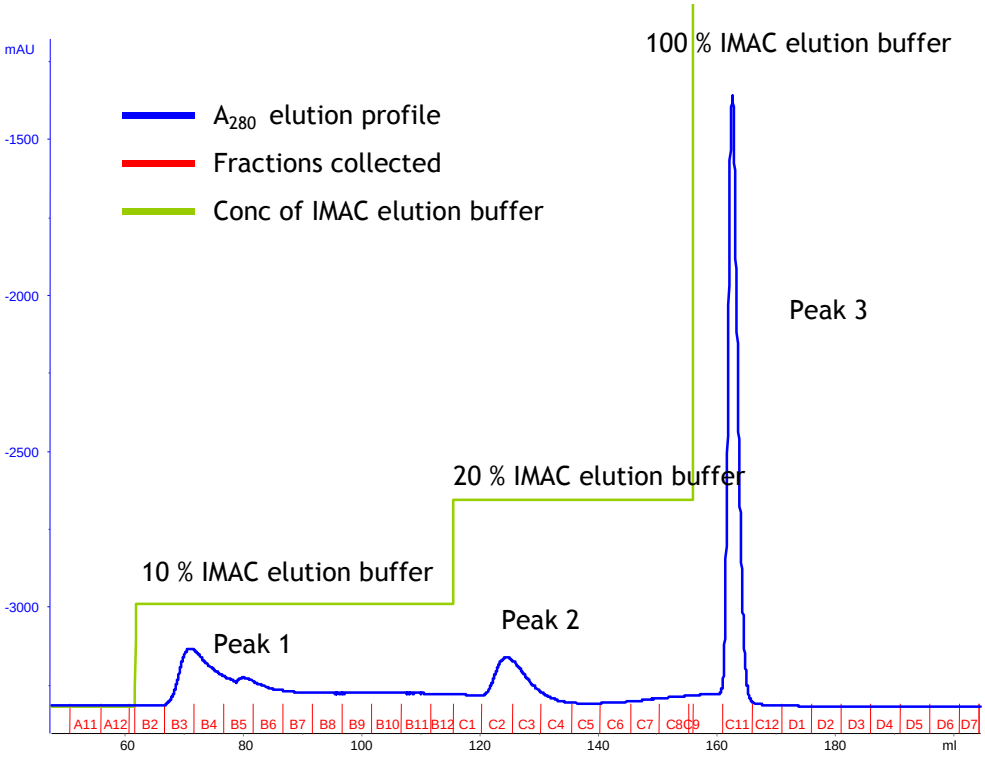


Figure 6.3: A_{280} elution profile for optimised IMAC purification of SP_1651. Peak 3 corresponds to the elution of SP_1651 with 100 % IMAC Elution buffer.

Lanes 4-8 in Figure 6.4 show the impurities eluted with these buffer concentrations. Lane 9 contains a heavy single band corresponding to the protein eluted in peak 3 (Figure 6.3) resulting in an absorbance of 2000 mAU. This single band at approximately 21 kDa was SP_1651 at purity sufficient for crystallisation trials. A yield of 14.4 mg of pure SP_1651 was calculated and then was partitioned onto two aliquots to investigate crystallisation both with and without the 6xHis tag.

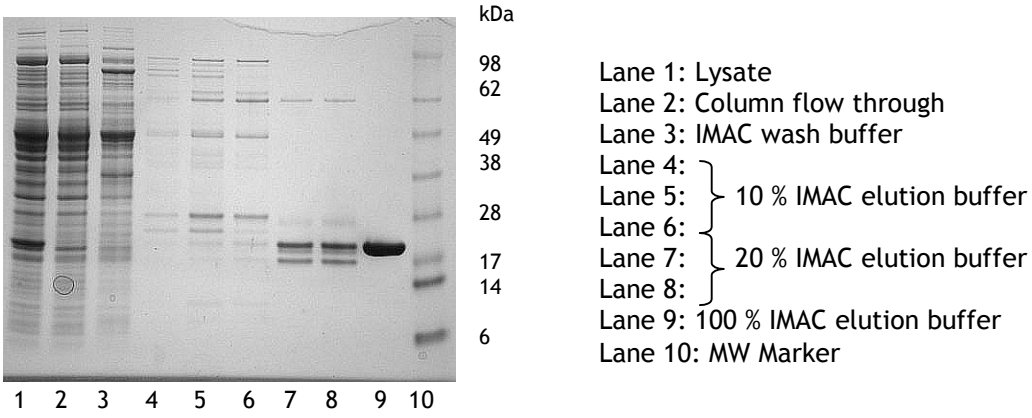


Figure 6.4: SDS-PAGE analysis of optimised IMAC purification of SP_1651.

6.8 Gel filtration of SP_1651 to analyse oligomeric state

Half of the purified SP_1651 prep was dialysed against a gel filtration buffer of 20 mM Tris pH 7.5, 100 mM NaCl as described in Section 2.5.2. It was then concentrated to a volume of 500 µL (approximately 14 mg/mL) and analysed by size exclusion chromatography using a S200 10/300 HR Global gel filtration column as described in Section 2.5.7. The aim of this experiment was to determine the oligomeric state of the protein in solution. Figure 6.5 shows the resulting A_{280} elution profile, which comprises a single peak at an elution volume of 13.28 mL. Comparison with the column calibration profile suggests a molecular weight of 40 kDa, which indicates that SP_1651 exists as a dimer in solution. The eluted protein fractions were pooled and concentrated to 16.3 mg/mL by centrifugation (Section 2.5.6).

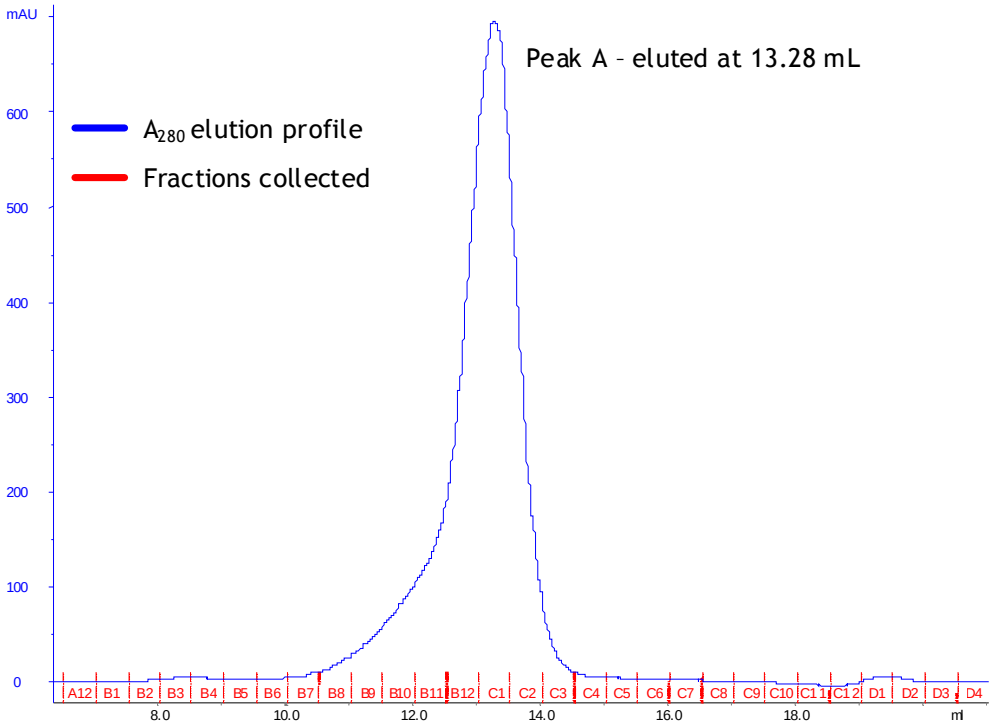


Figure 6.5: A_{280} elution profile for analytical gel filtration experiment to determine oligomeric state of SP_1651 shows a single peak eluted with 13.28 mL of buffer.

6.9 HRV-3C cleavage of 6xHis tag

The second aliquot of the SP_1651 prep was subjected to 6xHis tag removal using HRV-3C protease followed by an IMAC step as described previously. The cleaved protein (SP_1651_clvd) was dialysed into the same buffer as the non-

cleaved protein (20 mM Tris-HCl pH 7.5, 100 mM NaCl) and was concentrated to 25.3 mg/mL by centrifugation (Section 2.5.6).

6.10 Pre-crystallization Test (PCT) for SP_1651 and SP_1651_clvd

A Pre-crystallization Test (PCT) to establish the optimum protein concentration for crystallisation experiments was carried out for SP_1651 and SP_1651_clvd as described in Section 2.6.2. The experiments were observed by light microscope after 24 hr incubation at 20°C. Concentrations of 16.3 mg/mL and 17.0 mg/mL were selected for SP_1651 and SP_165_clvd respectively.

6.11 Crystallisation trials for SP_1651 and SP_1651_clvd

When setting up crystallisation experiments the minimalistic approach designed by Newman *et al.* (2005) [207] was employed. This method was designed primarily for small academic laboratories where cost and experimental monitoring are important considerations. The strategy involved the combination of two 96 condition screens; JCSG+ - a sparse matrix screen developed by the Joint Centre for Structural Genomics, and PACT – a systematic screen which tests pH, anionic and cationic additives. The JCSG+ screen allows initial screening covering a wide range of conditions selected from other sparse matrix screens without redundancy. The PACT screen can be used to gain information about the protein even in the absence of crystal growth [207].

The trays were stored and monitored as outlined in Section 2.6.3, with SP_1651 (16.3 mg/mL) in drop 1 and SP_1651_clvd (17.0 mg/mL) in drop 2 for each condition in both 96 well trays. The remaining protein samples were separated into 60 µL aliquots and stored at –20°C to minimise possible proteolysis.

After 11 days, birefringent crystals were formed in condition A1, site 2 of the PACT screen. This corresponded to the SP_1651_clvd sample at a concentration of 17 mg/mL. The composition of the A1 well solution was 0.1 M SPG Buffer pH 4.0, 25 % w/v PEG 1500. The SPG Buffer pH 4.0 (Molecular Dimensions)

comprised succinic acid, sodium dihydrogen phosphate monohydrate and glycine (Table 6.2).

Chemical Name and source	Molar ratio	Formula weight	Mass (g)	Final buffer volume (mL)	Final buffer conc. (M)	Comment
Succinic acid (Sigma S3674)	2	118.10	1.48	100	1	Adjust to pH 4.0 with 10 M NaOH
Sodium dihydrogen phosphate monohydrate (Fluka 71507)	7	137.99	6.04			
Glycine (Fluka 50046)	7	75.07	3.28			

Table 6.2: Composition of SPG Buffer pH 4.0 provided by Molecular Dimensions.

The crystals had formed as 100 µm clusters (Figure 6.6) and were difficult to separate without breaking.

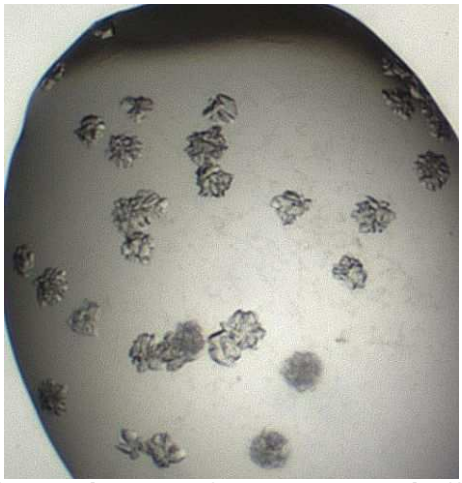


Figure 6.6: SP_1651_clvd crystals grown from 0.1 M SPG buffer pH 4.0, 25 % PEG 1500

No crystals were found in condition A1 for the non-cleaved protein. Phase separation and possible microcrystals were observed, however after 3 months there was no crystal formation suggesting that the presence of the 6xHis tag was unfavourable. No further crystals were found in either of the trays for either protein over a three-month period.

Analysis of the PACT screens for SP_1651 and SP_1651_clvd after two weeks gave an insight into unfavourable crystallisation conditions. Heavy amorphous precipitate appeared in the presence of a combination of acetate buffer and various cations, with 0.2 M $ZnCl_2$ giving particularly heavy precipitation for both SP_1651 and SP_1651_clvd.

6.12 X-ray diffraction experiments on SP_1651_clvd crystals

The PACT crystallisation tray containing the initial SP_1651_clvd crystals was taken to the Diamond Light Source, South Oxfordshire where crystals were mounted into 0.12 mm loops before being immersed in three different cryoprotectants (Table 6.3). It was difficult to loop single crystals as they formed as fragile clusters, which were easily damaged.

Cryoprotectant	Composition	Comments
A	0.1 M SPG buffer pH 4.0 25 % PEG 1500	Taken directly from well
B	0.1 M SPG buffer pH 4.0 25 % PEG 1500 25 % Glycerol	SPG buffer taken from 1 M stock made following recipe in Table 6.2
C	0.1 M SPG buffer pH 4.0 10 % PEG 600	SPG buffer taken from 1 M stock made following recipe in Table 6.2

Table 6.3: Composition of experimental cryoprotectants.

Diffraction confirmed that the crystals were protein and diffraction data with a resolution limit of 3.2 Å (Figure 6.7), were collected using a fractured single crystal immersed in cryoprotectant B (Table 6.3).

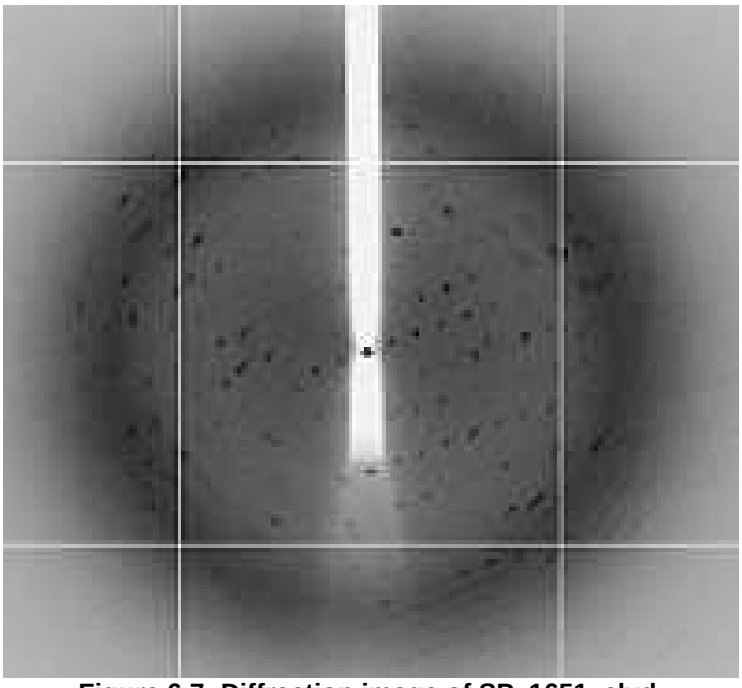


Figure 6.7: Diffraction image of SP_1651_clvd.

Multiple spots in the diffraction pattern exist due to the crystal being split. Although this crystal appears mosaic the diffraction image represents a suitable crystal form for optimisation.

6.13 Attempts to reproduce and optimise SP_1651 crystals

Frozen aliquots of SP_1651_clvd were thawed at room temperature and efforts were made to reproduce/optimise the initial crystallisation. The experiments carried out are listed in Table 6.4. For each experiment the concentrations of protein in drops 1 and 2 were 17.0 and 8.5 mg/mL respectively. A drop size of 1 µL (protein:precipitant ratio of 0.5 µL:0.5 µL) was used. After 6 weeks there was no evidence of crystallisation. It was thought that the protein had been damaged, denatured, unfolded or aggregated by the freeze thaw process making crystallisation unfavourable.

Experiment	Conditions/Screen	Number of conditions	Comments
Reproduction Tray 1	0.1 M SPG buffer pH 4.0 25 % PEG 15000	96 x same condition in each well	SPG buffer made following recipe in Table 6-2.
Reproduction Tray 2	0.1 M SPG buffer pH 4.0 25 % PEG 15000	96 x same condition in each well	SPG buffer made using sodium dihydrogen phosphate dehydrate rather than hygroscopic monohydrate
Reproduction Tray 3	0.1 M SPG buffer pH 4.0 25 % PEG 15000	96 x same condition in each well	SPG buffer purchased from manufacturer
Reproduction Tray 4	PACT tray	96 different conditions	PACT Screen from the same batch which produced initial crystals
Optimisation Tray 1	Varied pH between 3.8-4.5 in 0.1 pH unit increments. Varied % PEG 1500 between 5-30 % in 5 % increments.	48 different conditions	SPG buffer made following recipe in Table 6-2.

Table 6.4: Experiments set up in attempt to reproduce/optimise SP_1651_clvd crystals.

Seeding experiments were performed in an attempt to reproduce the initial crystals however the drop precipitated destroying any remaining crystals. After re-sealing and re-incubation at 20°C for 24 hrs, some crystals reappeared. However, no diffraction was observed despite trying numerous crystals from this batch in the previously successful cryoprotectant.

A fresh batch of SP_1651_clvd was prepared by the same protocol used to obtain the initial crystals. The crystallisation experiments listed in Table 6.4 were repeated in a second attempt to reproduce the initial crystal hit. No crystals grew after continued monitoring of experiments over a period of 3 months.

6.14 JBS Solubility Screen and DLS experiments

A study by Zulauf and D'Arcy (1992) [208] showed that the presence of aggregates in the protein solution may inhibit crystal nucleation or growth. In the past, researchers have tried to overcome such problems by adjusting purification protocols, removing fusion and purification tags and adding reducing agents, ultimately to provide a suitable environment conducive to complex macromolecules. Although these factors are important, buffer composition can also impact on aggregation. Jancarik *et al*. (2004) [173] devised the JBS Solubility Kit which tests a panel of buffers and additives to find the condition in which the recombinant protein is most homogenous. This method uses DLS experiments to assess the monodispersity of the conditions tested. These studies resulted in the crystallisation of 9 out of 14 proteins which had previously been susceptible to aggregation [173].

In an attempt to generate new crystal hits, the JBS Solubility kit was used for SP_1651_clvd as described in Section 2.6.1. The initial 24 well hanging drop experiments resulted in precipitation of the protein in buffers 1-5. This showed that the protein was insoluble in these buffer conditions (pH 3-4.5) therefore they were excluded from the DLS buffer screen. Table 6.5 shows the results gained for Part A: DLS Buffer Screen. The results recorded included the average hydrodynamic radius (Rh) (nm), polydispersity (P) (nm), % polydispersity (% P) and the predicted molecular weight (MW) (kDa) values gained from DLS experiments. These values were analysed in conjunction with the generated regularisation histograms and the range of the 12 individual DLS measurements for each buffer.

Buffer	Composition (100 mM)	R (nm)	P (nm)	% P	% P < 25 %	MW (kDa)
6	Na/K phosphate pH 5.0	2.98	0.51	17.1	Yes	40.9
7	Sodium citrate pH 5.5	2.69	0.57	21.1	Yes	32.0
8	Na/K phosphate pH 6.0	3.25	1.66	51.1	No	50.4
9	Bis-Tris pH 6.0	2.78	0.83	29.9	No	34.5
10	MES pH 6.2	2.98	0.99	33.2	No	41.1
11	ADA pH 6.5	2.96	0.89	30.0	No	40.4
12	Bis-Tris propane pH 6.5	2.61	0.97	37.0	No	31.2
13	Ammonium acetate pH 7.0	2.91	0.97	33.2	No	38.7
14	MOPS pH 7.0	3.03	1.13	37.3	No	42.7
15	Na/K phosphate pH 7.0	2.84	0.83	29.2	No	36.4
16	HEPES pH 7.5	2.73	1.10	40.3	No	33.1
17	Tris pH 7.5	2.96	0.80	27.0	No	40.2
18	EPPS pH 8.0	2.95	1.12	38.0	No	40.0
19	Imidazole pH 8.0	3.00	1.04	34.7	No	41.7
20	Bicine pH 8.5	3.50	1.64	46.9	No	60.8
21	Tris pH 8.5	2.90	0.96	33.1	No	38.5
22	CHES pH 9.0	3.02	1.04	34.4	No	42.2
23	CHES pH 9.5	3.01	1.19	39.5	No	42.3
24	CAPS pH 10.0	2.85	1.01	35.4	No	36.8

Table 6.5: Results of DLS Buffer Screen on SP_1651_clvd.

Jancarik *et al*. (2004) [173] identified the criteria of a suitable crystallisation buffer to be one which gives a radius of less than 5 nm and a percentage polydispersity of less than 25%. Buffers 6 (pH 5.0) and 7 (pH 5.5) gave suitable monomodal size distributions (< 5 nm) and poydispersity percentages below 25 % and thus would be suitable for crystallisation trials. However, it was decided that the best result was gained for buffer 7, which was comprised of 100 mM sodium citrate pH 5.5. It was chosen because of the combination of a high intensity (> 90 %) single peak on the DLS histogram (Figure 6.8), a low hydrodynamic radius of 2.69 nm, and a narrow result distribution shown on the measurement trace (Figure 6.9).

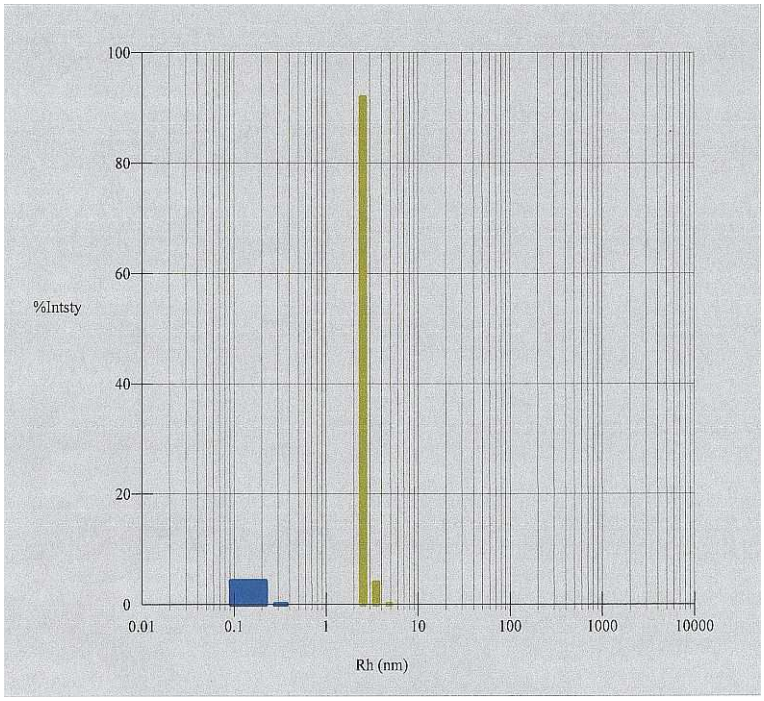


Figure 6.8: Dynamic light scattering histogram of SP_1651_clvd in buffer 7, 100 mM Na citrate pH 5.5. The high intensity single peak (yellow) demonstrates high monodispersity. Figure generated by Dynapro-801 software (Protein Solutions).

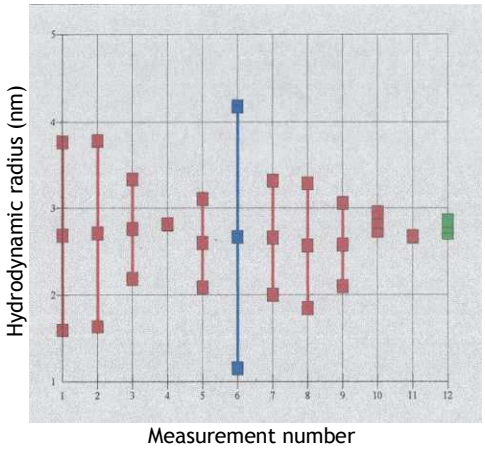


Figure 6.9: Ranges of the 12 measurements taken for SP_1651_clvd in buffer 7, 100 mM Na citrate pH 5.5. Shows a narrow range of results over the 12 measurements taken. Figure generated by Dynapro-801 software (Protein Solutions).

In general, the results gained for SP_1651_clvd were good, with radius values ranging between 2.69 to 3.50 nm (all below 5 nm) and percentage polydispersity values ranging between 17 to 51 %. The worst result was gained for SP_1651_clvd solubilised in buffer 20, which was comprised of 100 mM bicine pH 8.5. The DLS histogram (Figure 6.10) shows evidence of a high molecular weight aggregate which is undesirable for successful crystallisation.

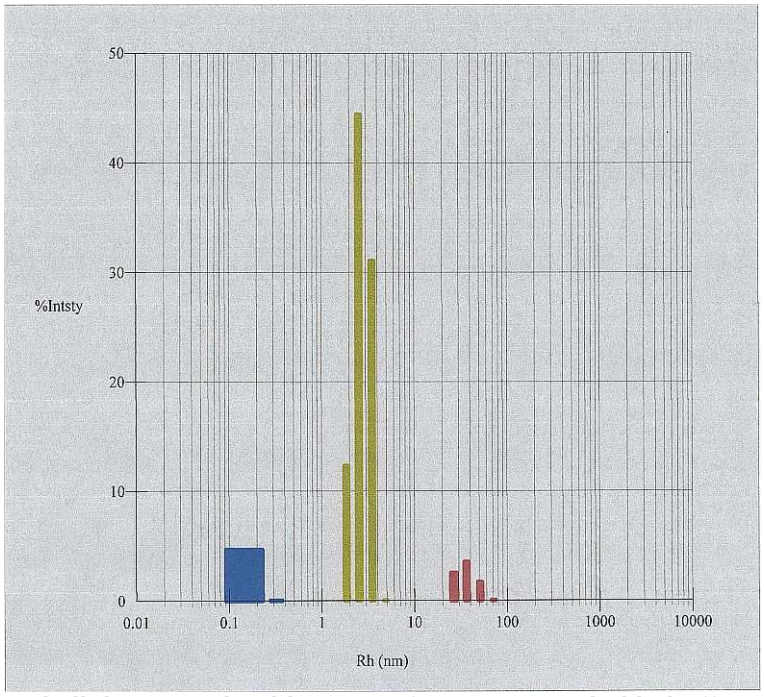


Figure 6.10: Dynamic light scattering histogram for SP_1651_clvd in buffer 20, 100 mM bicine pH 8.5. Figure shows the presence of high molecular aggregates (red) and anomalous low molecular species (blue). Figure generated by Dynapro-801 software (Protein Solutions).

Although the SP_1651_clvd sample in buffer 7 was deemed suitable for crystallisation trials, Part B of the experiment (additive screen) was still carried out to investigate the positive and negative effects of commonly used additives on homogeneity and monodispersity. The experiment was carried out as described in Section 1.6.1 and the results are depicted in Table 6.6.

	Composition	Stock conc.	Final conc.	R (nm)	P (nm)	% P	Compared with initial result	MW (kDa)
1	Sodium chloride	80 mM	20 mM	2.76	0.50	18.3	Slightly better	34.1
2	Sodium chloride	200 mM	50 mM	2.78	0.75	26.8	Same	34.7
3	Sodium chloride	400 mM	100 mM	2.75	0.62	22.5	Same	33.6
4	Glycerol	20%	5%	2.91	0.83	28.6	Slightly worse	38.6
5	Glycerol	40%	10%	3.27	1.26	38.5	Slightly worse	51.4
6	CHAPS	8 mM	2 mM	2.71	0.48	17.8	Same	32.5
7	OG	0.40%	0.10%	2.69	0.51	18.8	Slightly better	31.8
8	OG	4%	1%	2.66	0.60	22.5	Slightly worse	31.0
9	DDM	0.40%	0.10%	3.23	0.88	27.2	Slightly worse	49.8
10	DDM	4%	1%	3.26	0.55	16.8	Slightly worse	51.0
11	BME	40 mM	10 mM	2.69	0.62	23.0	Slightly worse	31.9
12	DTT	4 mM	1 mM	2.62	0.44	16.8	Same	30.0
13	DTT	20 mM	5 mM	2.69	0.36	13.4	Slightly better	31.8
14	TCEP	120 mM	30 mM	-	-	-	-	-

Table 6.6: Results of DLS Additive Screen on SP_1651_clvd in 50 mM Na citrate pH 5.5.

A slight improvement on the initial 100 mM odium citrate pH 5.5 buffer was gained for additive 1 (20 mM NaCl final conc.), additive 7 (0.1 % octyl glucopryanocide final conc.) and additive 13 (5 mM DTT final conc.). In general,

no great change was observed. Additive 14, 120 mM TCEP (stock concentration) resulted in the sample precipitating before the DLS experiment could be performed.

SP_1651_clvd was dialysed into four subsequent buffers: (a) 50 mM Sodium citrate pH 5.5 buffer without any additive, (b) with 20 mM NaCl, (c) with 0.1 % octyl glucopryanocide and (d) with 5 mM DTT. Crystallisation experiments were set up for 20 mg/mL and 10 mg/mL samples of each, using the PACT/JCSG+ strategy described in Section 1.10.

After 3 months of monitoring the experiments using a Rhombix Imaging System (Section 2.6.3), no credible protein crystals were gained.

In some of the histograms generated throughout these experiments, a low molecular weight species of low intensity was present. This anomaly cannot be explained; however it is thought to be concentration dependent and disappears at higher protein concentrations. The average molecular weight calculated for the 24 buffers equalled 33.6 kDa. This correlated with the gel filtration trace (Figure 6.5) corresponding to SP_1651 which suggested that it is dimeric in solution. Although lower than the expected value of 38.8 kDa this may be due to the contribution from the anomalous low molecular weight species.

The JBS Solubility test was also carried out on non-cleaved SP_1651 samples. In comparison with SP_1651_clvd results, none of the 24 buffers gave an improvement in monodispersity. In the majority of cases, large molecular weight aggregates were present resulting in an average molecular weight of 232 kDa, a value almost 7 times that calculated for SP_1651_clvd. The best result obtained for SP_1651 was for buffer 6, sodium/potassium phosphate pH 5.0. Table 6.7 contrasts the difference in radius, polydispersity, % polydispersity and MW values for SP_1651 and SP_1651_clvd samples for this buffer.

Protein sample	R (nm)	P (nm)	% P	% P < 25 %	MW (kDa)
SP_1651	3.26	1.03	31.6	No	51.1
SP_1651_clvd	2.98	0.51	17.1	Yes	40.9

Table 6.7: Comparison of DLS Buffer Screen results for SP_1651 and SP_1651_clvd in buffer 6.

Although both radial measurements are below the 5 nm criteria, the value for SP_1651_clvd is less than half of that gained for the non-cleaved SP_1651 sample.

Figure 6.11 compares the regularisation histograms for both protein samples in buffer 6. For SP_1651_clvd the % polydispersity is below the desired threshold of 25 %, whereas that for the non-cleaved protein is above. For non-cleaved SP_1651 (Figure 6.11b) there is evidence of a high molecular weight species which is absent for SP_1651_clvd (Figure 6.11a). Ultimately, these results confirm that the cleaved protein is less aggregated in solution and therefore crystallisation is more favourable. This could explain why the initial crystals described in Section 6.12 were gained for SP_1651_clvd but not for SP_1651 using the same condition.

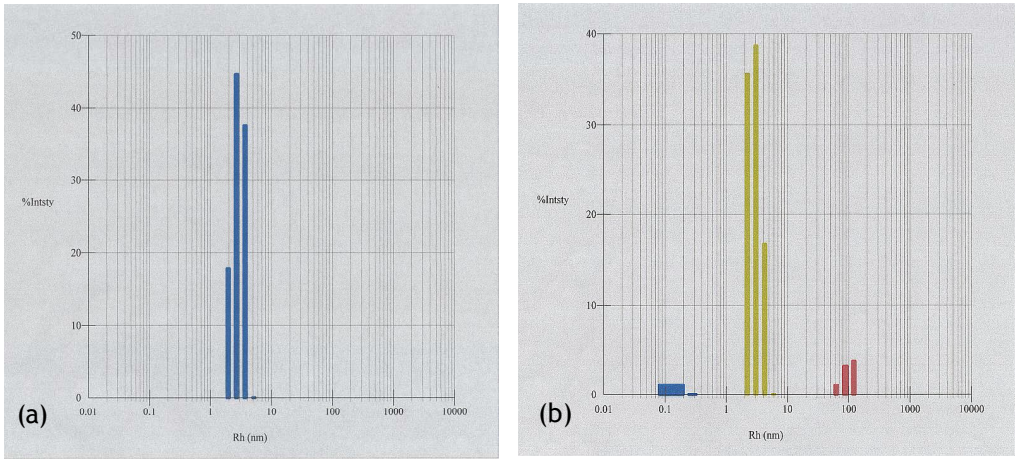


Figure 6.11: (a) DLS histogram of SP_1651_cvld in buffer 6 shows a monodispersed sample (b) DLS histogram of SP_1651 in buffer 6 shows the presence of high molecular aggregates (red) and anomalous low molecular species (blue). Figures generated by Dynapro-801 software (Protein Solutions).

7 Other targets from *D. radiodurans* and *S. pneumoniae*

7.1 DR_0463 from *D. radiodurans*

7.1.1 Project aims

The aim of this project was to investigate the expression and purification of DR_0463 from *D. radiodurans* R1 with the objective of crystallising and characterising the protein.

7.1.2 Construct design

A full length form of the *dr_0463* gene (nucleotide residues 1 to 2937), cloned into a pDEST17 vector (Table 2.1), was kindly provided by Protein'eXpert, Grenoble, France. The nucleotide and putative amino acid sequence for DR_0463 is annotated on the Comprehensive Microbial Resources (CMR) database [6] which can be accessed on the web page of The J. Craig Venter Institute [7]. The gene has been assigned the name *treY* and has been putatively identified as a maltooligosytrehalose enzyme, involved in trehalose biosynthesis [122]. Throughout this chapter the gene and encoded protein will be referred to by the given locus name of DR_0463.

7.1.3 Recombinant expression of dr_0463 from pDEST17

Before beginning expression trials, *dr_0463*-pDEST17 construct DNA was sequenced to confirm the same codon distribution of nucleotides as for the putative sequence annotated on the CMR database.

1 µL plasmid DNA (*dr_0463* in pDEST17) was used to transform B834(DE3), BL21 Star(DE3), Rosetta(DE3)pLysS and BL21(DE3)pLysS competent cells (genotypes given in Table 2.6 of Chapter 2) by the method described in Section 2.2.6. 20 µL of the transformation reaction mixture gave an even spread of single colonies after 20 hrs incubation at 37°C for all four cell lines.

Preliminary expression trials were performed for each cell line using LB media (Table 2.5) as described in Section 2.4.5. A 100 x dilution of starter culture was used to initiate cell growth in each of the four 500 mL cultures which were incubated at 37°C until an OD_{600} of 0.6 was reached. They were then induced with IPTG and incubated for 20 hrs at 22°C.

SDS-PAGE (Section 2.4.8) and western blot analysis (Section 2.4.9) of 0.5 mL cell extracts taken during this incubation period were used to assess the levels of soluble and insoluble expression. Table 7.1 gives a summary of the results gained for the four cell lines tested.

Cell line	Soluble expression	Insoluble expression
B834(DE3)	No	No
BL21 Star(DE3)	No	No
Rosetta(DE3)pLysS	No	Weak
BL21(DE3)pLysS	No	Very strong

Table 7.1: Results of preliminary expression trials for *DR_0463* in four different cell lines as judged by SDS-PAGE and western blot. Strong insoluble expression was observed for BL21(DE3)pLysS. In addition, a small amount of insoluble DR_0463 was produced by Rosetta(DE3)pLysS. No expression was observed for B834(DE3) and BL21 Star(DE3) and no evidence of soluble DR_0463 was found for any of the four cell lines.

The largest quantity of DR_0463 was produced in an insoluble form by BL21(DE3)pLysS competent cells. This is shown by the increasing intensity of a band at approximately 100 kDa in lanes 7-11 of Figure 7.1.

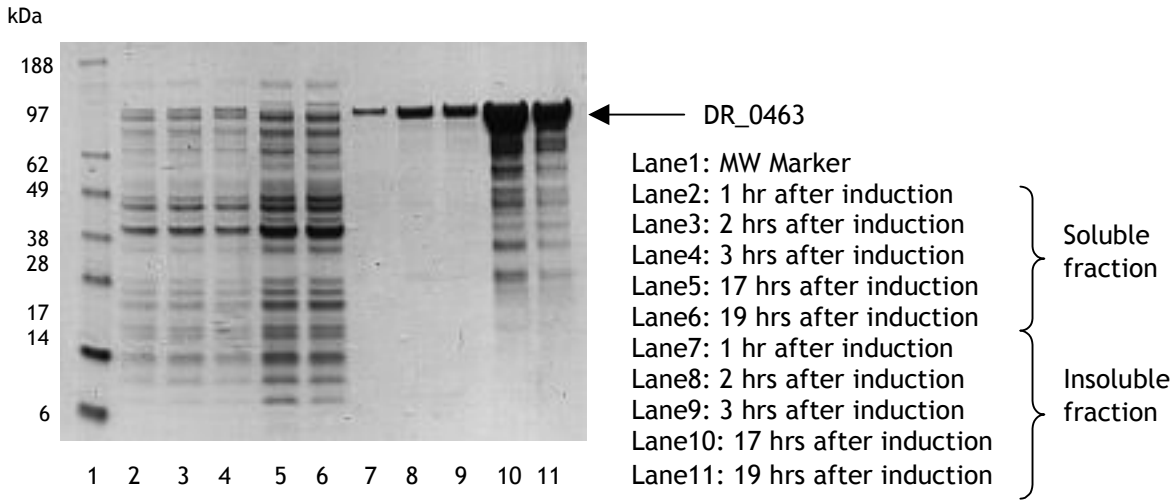


Figure 7.1: SDS-PAGE analysis of soluble (Lanes 2-6) and insoluble (Lanes 7-11) cell extracts (normalised to OD_{600} = 5) taken at time points to assess expression levels of DR_0463. Increased intensity of the band at approximately 100 kDa in lanes 7-11 indicates good expression of DR_0463 in the insoluble cell fraction. Analysis of soluble cell extracts (lanes 2-6) showed several faint bands in the 100 kDa region, however none were recognised by western blot.

A small amount of insoluble DR_0463 was produced by Rosetta(DE3)pLysS but no expression was identified for B834(DE3) and BL21 Star(DE3) cell lines.

No soluble protein was produced by B834(DE3), BL21 Star(DE3) or Rosetta(DE3)pLysS cells, although several possible bands were observed in the 100 kDa region of soluble BL21(DE3)pLysS cell extracts (Figure 7.1, lanes 2-6). These bands were not detected by western blot and did not show any affinity for the HisTrap HR column matrix during initial IMAC purification trials. This suggests that either (a) the bands in lanes 2-6 are not DR_0463 and thus no soluble target protein was expressed or (b) the 6xHis tag is inaccessible for detection and purification.

It was decided to concentrate efforts on solubilisation and purification of insoluble DR_0463 produced by BL21(DE3)pLysS cells. To enable further experiments, protein expression was scaled up to 4 x 500 mL cultures as described in Section 2.4.5.

7.1.4 Solubilisation and IMAC purification of DR_0463

Cell pellet from 2 L of culture was re-suspended in 30 mL of lysis buffer (Table 2-8) and the cells were lysed by French press as described Section 2.5.3. After centrifugation the soluble and insoluble fractions were separated. The insoluble pellet, containing expressed DR_0463, was solubilised by the procedure described in Section 2.5.3, using 20 mL of solubilisation buffer (Table 2.8). After further centrifugation at 5700 rcf for 30 minutes, the solubilised supernatant was loaded on to a 1 mL HisTrap HR column, pre-equilibrated with IMAC wash buffer (Table 2.8) containing 8 M urea. The column was flushed with 10 CV of this buffer to remove any unbound proteins.

IMAC was carried out on an AKTA Purifier FPLC as described in Section 2.5.4. with 8 M urea added to both wash and elution buffers. The program employed is summarised in Table 7.2.

Step number	Number of column volumes (1 CV = 1 mL)	% IMAC Wash Buffer (+8 M urea)	% IMAC Elution Buffer (+8 M urea)	Effective Imidazole conc. (mM)
1	10	100	0	20
2	20	90	10	68
3	12	80	20	116
4	10	0	100	500

Table 7.2: IMAC protocol or purification of solubilised DR_0463.

The eluted fractions were analysed by SDS-PAGE as shown in Figure 7.2.

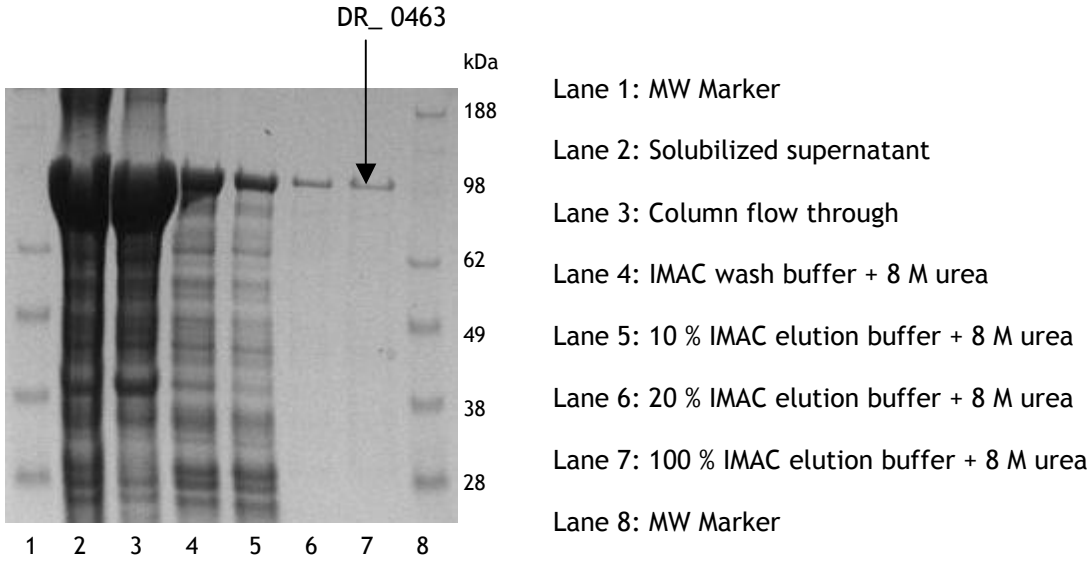


Figure 7.2: SDS-PAGE analysis of IMAC purification of solubilised DR_0463

The large 100 kDa bands in the solubilised supernatant (Lane 2) and the column flow through (lane 3) show similar intensities suggesting that on loading, a large proportion of the target protein passed through the column without binding. Such low affinity for the Ni^{2+} matrix resulted in the elution of DR_0463 with IMAC wash buffer (Lane 4) and 10 % IMAC elution buffer (Lane 5) alongside many impurities. Faint single bands corresponding to target protein eluted with 20 % (lane 6) and 100 % (lane 7) IMAC elution buffer showed insufficient yields for re-folding experiments.

In an attempt to increase binding of DR_0463 to the column resin, the following protocol adaptations were made:

(a) Increasing the potential binding capacity of the column by using a higher volume 5 mL HiTrap column

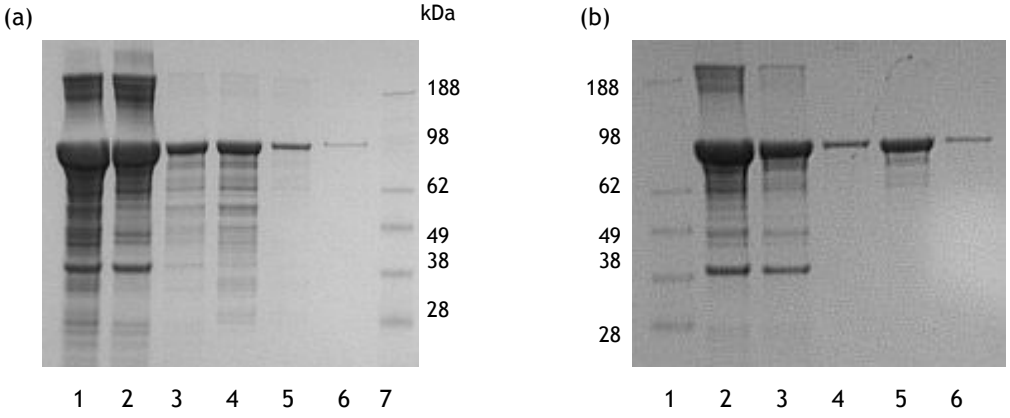
(b) Reducing the imidazole concentration of IMAC wash buffer from 20 to 5 mM in an attempt to prevent the protein passing through the column without binding

(c) Adding 5 % glycerol to the solubilised supernatant before loading onto the column in an attempt to 'loosen' the protein and make the 6xHis tag accessible for binding

(d) Changing denaturation agent from 8 M urea to 6M guanidine hydrochloride

SDS-PAGE analysis of IMAC purification experiments implementing the above changes showed no increase in the amount of DR_0463 binding to the column.

However, it was found that sequentially repeating IMAC purification with the same protein sample resulted in the removal of many unwanted impurities. On the second and third passes (Figures 7.3b and c) through the 5 mL HiTrap column, DR_0463 was purified to a level and yield deemed acceptable for further purification by gel filtration or ion exchange.



Lane 1: Lysate
Lane 2: Column flow through
Lane 3: IMAC wash buffer + 8 M urea
Lane 4: 10 % elution buffer +8 M urea
Lane 5: 20 % elution buffer +8 M urea
Lane 6: 100 % elution buffer +8 M urea
Lane 7: MW Marker

Lane 1: MW Marker
Lane 2: Column flow through
Lane 3: IMAC wash buffer + 8 M urea
Lane 4: 10 % elution buffer + 8 M urea
Lane 5: 20 % elution buffer + 8 M urea
Lane 6: 100 % elution buffer + 8 M urea

Figure 7.3: SDS-PAGE analysis of second (a) and third (b) IMAC purifications of DR_0463. Lanes 3-6 of (a) and Lanes 3-6 of (b) show that repetitive IMAC cycles produced purity levels and yields of DR_0463 acceptable for further purification by ion exchange or gel filtration.

Future work planned for the DR_0463 project involved optimisation of the purification protocol to produce sufficient yields of pure protein for re-folding trials.

7.2 SP_1648 from *S. pneumoniae* and DR_2284 from *D. radiodurans*

7.2.1 Project aims

The aim of this project was to investigate the expression and purification of homologous genes from *S. pneumoniae* (*sp_1648* or *psaB*) and *D. radiodurans* R1 (*dr_2284*). Crystallisation of the encoded proteins would allow structural and functional comparisons to be made.

Work on DR_2284 was carried out at within the Macromolecular Crystallography Group, ESRF, Grenoble, France.

7.2.2 Construct design

A full length form of the *sp_1648* gene (residues 1 to 723) cloned into a pOPINF vector (Table 2.1) was kindly provided by Dr. Alan Riboldi-Tunnicliffe, University of Glasgow. The cloning work was carried out at the Oxford Protein Production Facility, Oxford, using the In-Fusion™ PCR Reaction system (Clontech).

A full length form of the *dr_2284* gene (nucleotide residues 1 to 720), cloned into a pET151/D TOPO vector (Table 2.1), was kindly provided by The Macromolecular Crystallography Group, ESRF, Grenoble, France.

The nucleotide and putative amino acid sequence for both *sp_1648* and *dr_2284* are annotated on the Comprehensive Microbial Resources database [6] which can be accessed on the web page of The J. Craig Venter Institute [7].

7.2.3 Recombinant expression and purification of SP_1648

2 µL plasmid DNA (*sp_16481* in pOPINF) was used to transform B834(DE3) competent cells by the method described in Section 2.2.6. 5 µL of the

transformation reaction mixture gave an even spread of single colonies after 20 hrs incubation at 37°C.

Small-scale expression trials were carried out in Overnight Express™ Autoinduction Media (Table 2.5) as described in Section 2.4.6. A 20 x dilution of starter culture was used to initiate cell growth in two 100 mL cultures which were incubated at 37°C for 4 hrs. One culture was then incubated at 25°C and the other at 22°C for 24 hrs at which OD_{600} values of 7.0 and 6.8 were reached. The respective cell pellets were then harvested by centrifugation.

SDS-PAGE analysis of 0.5 mL samples taken throughout expression trials showed similar results for both cultures incubated at 25 and 22°C. Steady expression in the insoluble cell fraction was observed 6 hrs after the temperature was decreased as shown by the increasing size of the bands at approximately 29 kDa (Figure 7.4a and b). In the soluble cell fractions (Figures 7.4c and d), less intense bands appeared 10 hrs after the temperature decrease, with those incubated at 25°C being more prominent.

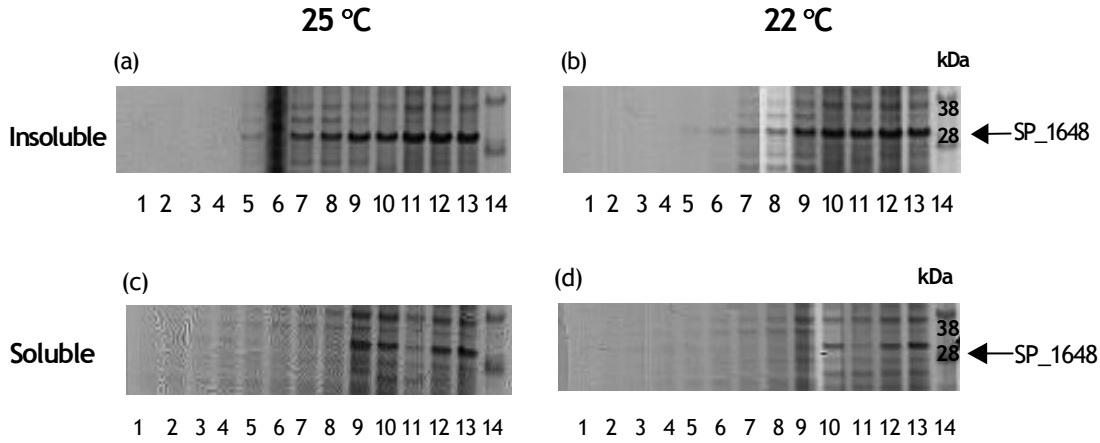


Figure 7.4: SDS-PAGE analysis of insoluble (a and b) and soluble (c and d) cell extracts (normalised to OD_{600} = 5) taken at time points to assess expression levels of SP_1648 at 25 °C (a and c) and 22 °C (b and d). Overall results were similar for 25°C and 22°C samples. Analysis of the insoluble cell fractions (a) and (b) shows a steady increase in intensity of the band at approximately 30 kDa indicating good expression of SP_1648. Analysis of soluble cell extracts (c) and (d) shows some expression after 9-10 hrs, with more prominent bands for the 25°C sample.

Although expression appeared greater in the insoluble cell fraction, it was decided to investigate purification of the soluble protein before attempting solubilisation and re-folding of insoluble SP_1648.

The cell pellets were re-suspended in 30 mL of lysis buffer (Table 2.8) with the addition of 0.2 % Tween 20. The cells were lysed by French press and the soluble lysates were loaded on to separate 1 mL HisTrap HR columns as described in Sections 2.5.3 and 2.5.4. IMAC purification was carried out on the bench, with the lysate incubated on ice, using a peristaltic pump at a flow rate of 1 mL/min. The loaded column was flushed with 12 column volumes (CVs) of IMAC wash buffer (Table 2.8) followed by 10 CVs of IMAC elution buffer (Table 2.8).

SDS-PAGE analysis of the resulting elusions indicated that SP_1648 has no affinity for the column. Lanes 1-2 and 5-6 of Figure 7-5 show similar band intensities (at 29 kDa) in the lysate and column flow through, demonstrating that the target protein had passed through the column without attaching to the matrix. This suggests that the 6xHis tag was unavailable for binding.

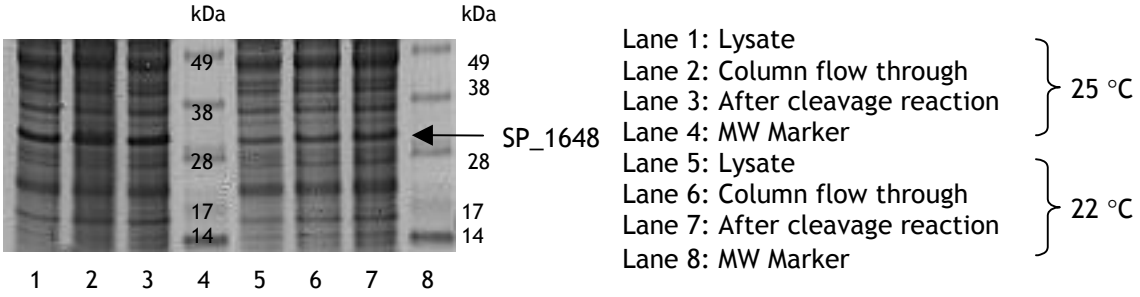


Figure 7.5: SDS-PAGE analysis of attempted IMAC purification and HRV-3C proteolysis of soluble SP_1648 expressed at 25 (lanes 1-3) and 22°C (lanes 5-7). The equal intensity of the 29 kDa band in the cell lysate and column flow through (for both samples) shows that SP_1648 has no affinity for the HisTrap HR column matrix suggesting that the 6xHis tag may be inaccessible. Attempted cleavage of the 6xHis tag using HRV-3C protease was unsuccessful, as shown by no decease in molecular weight (lanes 3 and 7) in comparison to the column flow through (lanes 2 and 6). This supports the supposition that the 6xHis tag may not be available for binding to the column.

To assess the accessibility of the 6xHis tag, the protein samples were subjected to HRV-3C proteolysis as described in Section 2.5.9. For both samples, comparison of SDS-PAGE analysis (Figure 7.5) before and after the cleavage reaction showed no decrease in molecular weight (comparison of lanes 2 and 3 for 25°C sample and Lanes 6 and 7 for 22°C). This suggests that the 6xHis tag

was inaccessible for proteolysis and thus explains lack of affinity for the HisTrap HR column matrix.

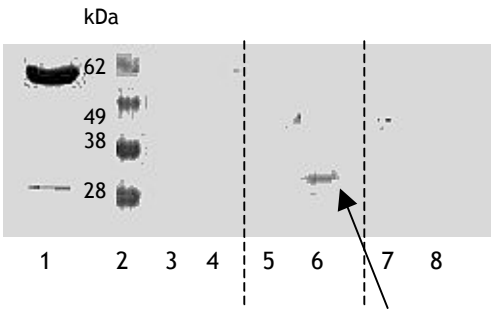
Future work planned for the SP_1648 project included purification of soluble protein by ion exchange and gel filtration in an attempt to overcome the 6xHis tag non-affinity problems encountered with IMAC purification. Alternatively, experiments to solubilise, purify and refold insoluble SP_1648 could be attempted.

7.2.4 Expression of DR_2284

1 µL plasmid DNA (*dr_2284* in pET151/D TOPO) was used to transform BL21 Star(DE3), Rosetta 2(DE3) and BL21-AI competent cells (genotypes given in Table 2.6 of Chapter 2) by the method described in Section 2.2.6. 20 µL of the transformation reaction mixture gave an even spread of single colonies after 20 hrs incubation at 37°C for all three cell lines.

Preliminary expression trials were performed for each cell line using LB media (Table 2.5) as described in Section 2.4.5. The OD_{600} values taken for overnight starter cultures were used to ensure equal cell densities were achieved on inoculation of each 100 mL expression culture. The three cultures were then incubated at 37°C until an OD_{600} of 0.6 was reached upon which the BL21 Star(DE3) and Rosetta2 (DE3) cultures were induced with IPTG and the remaining BL21-AI culture, induced with 0.2 % arabinose. All three cultures were then incubated for a further 20 hrs at 22°C after which cells were harvested by centrifugation. Cell pellets of 2.7 and 3.8 g were gained for Rosetta2 (DE3) and BL21 Star(DE3), with the largest cell growth of 4.3 g obtained for BL21-AI.

A western blot of 0.5 mL cell extracts taken after this incubation period (Figure 7.6) showed no soluble expression of the target protein.



Lane 1: Positive control
Lane 2: MW marker
Lane 3: BL21 Star(DE3) soluble fraction
Lane 4: BL21 Star(DE3) insoluble fraction
Lane 5: BL21-AI soluble fraction
Lane 6: BL21-AI insoluble fraction
Lane 7: Rosetta 2(DE3) soluble fraction
Lane 8: Rosetta 2(DE3) insoluble fraction

DR_2284 from BL21-AI cells

Figure 7.6: Western blot analysis of soluble and insoluble cell extracts taken during expression trials for DR_2284. One faint band appeared in Lane 6, corresponding to insoluble expression by the BL21-AI cell line

This is indicated by the absence of bands in lanes 3, 5 and 7 corresponding to the soluble fractions of BL21 Star(DE3), BL21-AI and Rosetta 2(DE3) cell lines. A faint band of approximately 30 kDa in lane 6 shows a small amount of insoluble target protein expression with the arabinose-induced BL21-AI cells which, as reported earlier, showed the largest cell growth during expression trials.

7.2.5 Future work for DR_2284

As only a very small amount of insoluble DR_2284 was produced, it may be beneficial to continue expression trials with variables including temperature and induction conditions in an attempt to express either soluble protein or a greater yield of insoluble protein. Alternatively, expression of DR_2284 in BL21-AI cells could be scaled up for solubilisation experiments.

8 Discussion

8.1 DR_1146

The *dr_1146* gen from *D. radiodurans* was successfully overexpressed in *E. coli* and purified to a level and yield that was conducive for crystallisation. However, despite significant diligence in the purification protocol initial crystal trials were unsuccessful; probably due to the inhomogeneity of the sample caused by proteolysis of the target protein. Ultimately, it was a combination of rational truncation, complete proteolysis and cofactor saturation that lead to the successful crystallisation of DR_1146 and although of poor quality, some diffraction was achieved.

On a whole, this study reflects the importance and advantage of incorporating information gained from biophysical characterisation into the strategies employed for successful protein crystallisation.

Through the use of biophysical techniques it was established that DR_1146 is a flavoprotein, with a capacity to associate with three flavin cofactors, namely riboflavin, FMN and FAD.

Flavoproteins are ubiquitous proteins; known to indulge in a wide range of biological functions thanks to the extensive redox capacity of the bound flavin molecule [85]. Hence, identification as a flavoprotein, could associate DR_1146 with a number of essential cellular roles including aerobic respiration [85], apoptosis [97, 98] and detoxification [99, 100]. It is likely that whatever system it belongs to, DR_1146 bound to flavin, will undergo either one or two-electron transfers [85]. This could implicate the protein in the production of DNA and protein damaging ROS as a result of irradiation. In the future, further characterisation and elucidation of a 3-D structure may contribute to a greater understanding of how *D. radiodurans* is able to deal with such oxidising conditions in order to maintain viability and prevent cell death.

It was initially thought that DR_1146 could be a flavoprotein based on a combination of information gained from bioinformatics and by the observation

that the target protein was yellow during expression and purification. Although analysis of the extracted yellow prosthetic group by ESI-MS was inconclusive, UV and fluorescence spectra suggested that the yellow colour of DR_1146 was due to a flavin molecule bound to the protein. UV absorbance spectra of a DR_1146 sample showed small absorbance maxima at 375 and 461 nm; a characteristic that is indicative of flavin binding proteins [94]. The presence of the two maxima in this spectral region suggests that the bound flavin cofactor is in an oxidised state on comparison with the spectrum of an oxidised glucose oxidase flavoprotein shown in Figure 1.5 of Section 1 [85, 108]. Fluorescence emission spectra (445 nm excitation) recorded for a sample of DR_1146 displaying a pale yellow colour showed a small fluorescence peak at approximately 520nm, which also suggests the presence of a 'flavin-like' species.

ITC and fluorescence experiments performed to quantify flavin binding, established that DR_1146 exhibits similar binding affinities for all three cofactors, riboflavin, FMN and FAD. K_D values determined were all within the micromolar range (4-11 μM), which, on comparison with other flavin binding proteins suggests moderate binding [187-191]. It is Interesting that DR_1146 is able to bind all three flavins with similar affinities, as many other flavoproteins show specificity for either FMN or FAD [85]. Results showing that DR_1146 binds riboflavin, FMN and FAD within the same micromolar range, suggest that the prosthetic –X group (hydrogen atom for riboflavin, phosphate group for FMN, adenine diphosphate for FAD) differentiates the three flavin structures but has no effect on binding affinity. This finding contrasts the case for the majority of flavoproteins where the prosthetic -X group of FMN or FAD is integral to binding. This also appears to be the case for the DR_1146 homologue, 2I02, a putative general stress protein (GOG3871) from *Nostoc punctiforme* PCC 73102 [81]. On analysis of possible interactions between 2I02 and FMN (using coordinates entered into the PDB) it appears that His61 and Lys62 are integral to FMN binding by forming hydrogen bonds with the oxygen atoms of the 5'-phosphate moiety of the prosthetic group. As neither His61 nor Lys62 are conserved in the putative amino acid sequence of DR_1146, this may explain why there is no increase in binding affinity for FMN and FAD in comparison with riboflavin. Like DR_1146, the apo homologue, 2RE7 (a putative general stress protein (COG3871) from *Psychrobacter arcticus* [200]) is also lacking His 61 and Lys62, which may

explain why this protein was crystallised in the absence of FMN and could support efforts to crystallise an apo form of T_DR_1146.

The binding stoichiometries established by ITC and FS confirm that DR_1146 binds two molecules of FMN per dimer. This suggests that the FMN binding capacity of DR_1146, differs from that of the homologue 2I02. The structure of 2I02 entered on the PDB, indicates that the dimeric protein has only one FMN molecule bound at the dimer interface. Only biophysical characterisation of 2I02 and structure determination of T_DR_1146 will allow firm comparisons to be made. To date, any work carried out on 2I02 and 2RE7 remains unpublished.

To complement observed flavin binding, DSC and CD experiments showed DR_1146 has higher thermal and chemical stability in the presence of an FMN cofactor. An addition of a two-equimolar excess of FMN to a sample of DR_1146, resulted in a change of T_m from 51.2 to 60.2˚C, signifying a marked increase in thermal stability on FMN binding. CD spectra for the chemical unfolding of DR_1146 showed that the apo protein is 50 % unfolded at a GuHCl concentration of 0.9 M whereas a higher concentration of 1.4 M is needed to unfold 50 % of the holo protein. This demonstrated that DR_1146 is chemically more stable when complexed with FMN. By monitoring the FMN contributions in the visible-near UV region of CD spectra, it became apparent that FMN completely dissociates from DR_1146 on addition of approximately 2M GuHCl. As for the majority of known flavoproteins, the observed ease of dissociation suggests that the flavin prosthetic group is non-covalently bound.

Overall changes in near and far CD spectra suggested that DR_1146 undergoes gross tertiary and secondary conformational changes in the presence of FMN. An increase in intensity in the near UV region suggests that the environments of the aromatic amino acids change as a result of FMN binding. Although speculative, it is possible that the aromatics are stacking with the isoalloxazine ring system making the structure of DR_1146 more rigid. Although many of the residues forming hydrophobic contacts in the structure of 2I02 are not conserved in T_DR_1146, their putative replacements are aromatic and hence may play a similar stabilising role.

Data obtained from the far UV spectral region was used to estimate secondary structure compositions. It was concluded that DR_1146 is a mixed alpha-beta protein and that FMN binding causes significant changes in secondary structure. No notable difference could be seen as a result of 6xHis tag removal.

Pronounced chemical shift changes in ^{15}N-^{1}H HSQC NMR spectra on addition of FMN confirms that global structural changes are induced by flavin binding. Such observations, connecting increased structural stability with FMN binding, reinforced efforts to crystallise the holo protein due to the strong relationship between sample stability and crystal growth. Observation of chemical shifts in ^{15}N-^{1}H HSQC NMR spectra on addition of FMN aliquots, also showed that a 2:1 excess of FMN:protein was required to ensure that the binding sites of DR_1146 were completely saturated with the cofactor. It was therefore decided that at least a two-equimolar excess of FMN would be added to the protein during purification to ensure that all binding sites were saturated. It was thought that this would ensure that the protein samples used for crystallisation experiments were homogenous.

In the absence of any known functional assay for DR_1146, CD spectrophotometry and NMR spectroscopy proved to be useful tools in establishing that DR_1146 is a highly folded and structured protein *in vitro*. The overall quality of preliminary non-labelled and ^{15}N- ^{1}H HSQC spectra obtained confirms that structure determination would be feasible by NMR. NMR spectra were a good indicator of quality, purity and homogeneity of DR_1146 preparations and it was perhaps not coincidental that protein crystals were obtained from a sample purified to the high standards required for this technique.

Initial expression trials performed on three different constructs harbouring the *dr_1146* gene, showed that very high yields of soluble target protein were produced in BL21 Star(DE3) competent cells. The resulting 6xHis tagged protein was easy to purify, showing good affinity for Ni^{2+} resin for IMAC. However, homogeneity proved very difficult to attain because of a combination of extensive N- and C-terminal proteolysis (as determined by N-terminal sequencing and mass spectrometry) and reversible binding of flavin cofactors. When all efforts of prevention or retardation of proteolysis failed, it was decided to

investigate the full extent of degradation by storing the protein for three months at 4°C. As discussed earlier, it was in fact a NMR spectrum that indicated that for a sample of DR_1146, degradation had ceased and that the protein was fully saturated with FMN. Subsequent purification by gel filtration isolated a stable dimeric fragment of 27.9 kDa from which small yellow crystals were formed. In total, a loss of approximately 18.6 kDa had occurred in comparison with the original full length protein expressed from pDEST17. Although poor, the resulting diffraction pattern indicated protein crystals rather than FMN or salt, and promising lunar patterns observed showed potential for optimisation.

Progression of this work to attain diffraction quality sufficient for structure determination using crystallography would primarily include reproduction and optimisation of the DR_1146 crystals obtained. To enable these experiments a degraded sample would be prepared under identical conditions to those employed previously. Meanwhile, it would be beneficial to continue attempts to clone an N- and C- terminally truncated gene to encode a protein fragment of similar size for recombinant expression. It is thought that initial efforts to do this failed as a result of a problematic insertion of the gene into the pOPINF vector due to possible deactivation of the essential enzyme needed for the pOPINF InFusion cloning reaction. It may be productive to clone the fragment in to pET-15b by the method used successfully to clone full length and truncated versions of *dr_1146* into pET-15b.

It should be noted that DR_1146 crystals were obtained by following a minimalistic approach designed by Newman *et al.*, (2005) [207] involving use of only two 96-condition crystal screens, PACT and JCSG+. The small, yellow crystals of degraded DR_1146 complexed with FMN were produced by sitting-drop vapour-diffusion, in condition B1 (0.1 M MIB buffer pH 4.0, 25 % w/v PEG 1500) of the PACT screen. Success of this small-scale strategy contrasted previous efforts to crystallise non-homogenous samples using up to nine sets of expensive commercial reagents. This result suggests that it is the quality of the protein preparation that impacts on successful crystallisation rather than the number of conditions screened. However, the consistency, reliability and speed of automating the crystallisation process, whether large, medium or small, is unprecedented. Successful crystallisation described in this study, resulted from

trays compiled by volume dispensing robotics. In addition, crystal growth was identified and monitored using an automated Rhombix system.

Perhaps high-throughput screening would have been better applied to the cloning step of the protein production pipeline. It was bioinformatic software, namely the RONN Disorder Prediction Server [183-185] that guided the successful 30 residue N-terminal truncation of DR_1146. Perhaps, before starting expression trials it would have been beneficial to produce various truncated forms of the *dr_1146* gene, based upon bioinformatic analysis such as disorder prediction and sequence alignment with other successfully crystallised homologues. Parallel expression trials could have then saved time in identifying a stable, homogenous protein fragment for crystallisation.

In terms of construct design, a change in expression vector from pDEST17 (non-cleavable 6xHis tag) to pET-15b, to incorporate a thrombin cleavable 6xHis tag, did not directly affect crystallisation of DR_1146. However, DSC experiments suggested that when in place, the 6xHis tag may obstruct protein refolding and thus may have a negative affect on protein stability. Interestingly, for the final DR_1146 sample that gave diffracting crystals, the 6xHis tag had been degraded away without intervention by thrombin proteolysis. This suggests that removal of the tag would be favourable for crystallisation. Hence for cloning of an N- and C-terminal truncated construct, a vector with a cleavable 6xHis tag would be chosen.

The elution profiles from gel filtration purifications performed on DR_1146, suggested that the protein exists as a dimer in solution. This information was used in conjunction with data obtained from SE AUC experiments to establish that the holo protein undergoes monomer-dimer self-association at a concentration of 7.67 μM (K_D^{1-2}). This value is consistent with the results obtained from SV experiments, which showed that self-association was underway by a concentration of 19 μM. The K_D^{1-2} obtained by SE AUC confirms that at concentrations below 7.67 μM, holo T_DR_1146_clvd exists predominantly as a monomer and, above 7.67 μM, mostly as a dimer. From this data, it can be postulated that if a crystal structure solution for DR_1146 were obtained, as a result of high concentration crystallisation experiments, its quaternary structure would be dimeric.

A global fit of SE data was used to determine an overall predicted molecular weight of 31,572 Da. It is thought that this value is lower than the dimeric molecular weight of 36875 Da (predicted using the ProtParam tool from the ExPASy server [171, 172]) because of two possible reasons; (a) that overcrowding of protein molecules in the experimental cell at high concentration caused non-ideal migration of the dimer species or (b) that the lower than expected molecular weight reflected proteolysis of DR_1146 throughout the duration of the experiment. The most probable scenario would be the former explanation as experiments were conducted within two days to avoid significant proteolysis of DR_1146.

Sedimentation coefficients (at zero concentration) of 1.72 and 2.87 S were established for the respective monomer and dimer conformations of holo DR_1146. This calculation was enabled by analysis of SV data obtained for a range of concentrations between 0.35 and 35 mg/mL. Comparison of experimental values with theoretical sedimentation coefficients (for a hydrated sphere of the same molar mass as the protein) suggest that both monomer and dimer have a slightly elongated shape in solution.

Although this study presents extensive biophysical characterisation of DR_1146, its true biochemical and biological function remains unknown. Significant sharing of amino acid sequence homology suggests DR_1146 may be a pyridoxine 5'-phosphate oxidase protein (PNPOx).

The PNPOx family of enzymes have been shown to catalyse the terminal step in the *de novo* biosynthesis of pyridoxal 5'-phosphate (PLP) in bacteria, by the FMN-dependent oxidation of PNP (pyridoxine 5'- phosphate) or PMP (pyridoxamine 5'-phosphate) [60, 62, 64, 65, 209, 210].

The reaction product, PLP has recently been identified as an effective antioxidant, [72-74] able to quench singlet oxygen radicals [75]. Hence, if DR_1146 was to be identified as a PNPOx enzyme, it could play a role in stress response through scavenging of ROS by its reaction product PLP. Such action may prevent the propagation of cell damaging hydroxyl radicals and hence contribute to the radioresistance of *D. radiodurans*.

This study presents several similarities in the physical and chemical properties of DR_1146 and characterised PNPOx enzymes, beginning with oligomeric state. Gel filtration elution profiles and AUC SV and SE analyses showed that DR_1146 exists as a dimer at concentrations above 7.67 μM. The crystal structures of bacterial, human, mammalian and yeast PNPOx enzymes are also homodimeric with molecular weights (22-29 kDa, monomeric) comparable with those determined for full length DR_1146 constructs [61]. Interestingly, like DR_1146, PNPOx preparations from *E. coli* have been shown to be susceptible to degradation by proteases as indicated by the presence of similar doublet bands on SDS-PAGE gels [62]. Controlled trypsin digest resulted in doublet of approximately 14 kDa that was shown to retain only 1 % PNPOx catalytic activity.

The main analogy between DR_1146 and PNPOx enzymes, highlighted by this study, is that the target protein has a capacity to bind FMN, the integral electron transfer vessel involved in the synthesis of PLP from PNP or PMP. *E. coli* PNPOx has been documented to associate with two molecules of FMN, which are bound in hydrophobic cavities approximately 19 Å apart, along the dimer interface [61, 62]. This correlates with data obtained from ITC experiments performed on DR_1146, which shows 1:1 binding stoichiometry, suggesting that the target protein also binds two molecules of FMN per dimer. However that is where the similarities end.

The amino acid residues forming interactions between *E. coli* PNPOx and FMN occur in sequence motifs highly conserved within the PNPOx family [61]. Such motifs are not conserved in the amino acid sequence of DR_1146 suggesting that the capacity to bind FMN does not signify functional similarities.

Characterised PNPOx enzymes show high specificity for FMN over riboflavin and FAD. Binding affinities for FMN are reported to be within the nanomolar range (K_D = 11-30 nm) [105] whereas those for FAD and riboflavin are not quantified in the literature. FMN has also been shown to be significantly more efficient as a catalytic cofactor when complexed with PNPOx in comparison to the other two flavins [211] suggesting that binding to the single phosphate prosthetic group of FMN may be integral to enzymatic function. In contrast to PNPOx enzymes, DR_1146 displays weaker binding affinities for the three cofactors, and does not show any specificity for FMN over riboflavin and FAD. As the prosthetic

phosphate group of FMN does not appear to be important to binding in holo DR_1146, is most likely that it has no effect on catalysis and hence it is probable that DR_1146 does not share functional similarities with PNPOx.

In addition to FMN binding, *E. coli* PNPOx has also been shown to tightly associate two molecules of PLP. The crystal structure shows one molecule of PLP to be situated within the active site of the enzyme, in close proximity to a FMN molecule, and the other to be stored 11 Å away in a non-catalytic site [68]. PLP has been shown to be an effective inhibitor of the PNPOx catalysed reaction (K_i = 1-8 μM [60, 65]) when in complex with the enzyme. Quantifiably, PLP has a greater affinity for the active site of PNPOx than either PNP or PMP substrates which suggests that product inhibition is a mode of regulation for the reaction [60, 76]. In contrast, none of the residues reported to be involved in PLP binding by *E. coli* PNPOx [68], are conserved in the amino acid sequence of DR_1146 and ITC and FS results suggest that the protein is unable to bind PLP. Of course, it is possible that PLP is already bound to the protein and its presence has remained undetected due to spectral masking by bound FMN, as both cofactors absorb UV light in similar regions [63]. This could be confirmed or disproved by monitoring a decrease in absorbance at 388 nm after addition of 0.1 M NaOH to apo DR_1146 to remove any bound PLP [186]. However, to date, evidence from this study suggests that DR_1146 is unable to bind PLP which appears to be a key reaction controlling substrate; hence it is proposed that DR_1146 does not function as a PNPOx in *D. radiodurans*. This postulation could be either confirmed or dispelled by performing the enzymatic assay used to confirm function for *E. coli* PNPOx as described by Di Salvo *et al*. (1998) [62].

It has been shown that PNPOx from *Arabidopsis thaliana* has tolerance to oxidative stress by comparing growth of the wild type and a deletion mutant in the presence of H_2O_2. The deletion strain showed a distinct growth defect in comparison with the wild type [67]. Independent of whether DR_1146 is shown to be a PNPOx or not, this would be a good way of assessing its putative role in cellular response to oxidative stress.

At this time, the function of DR_1146 is unknown, but is thought that it belongs to a family of PNPOX-like proteins that are able to bind FMN. Some such structures, like 2I02, have been entered on the PBD but details and

characterisation remain unpublished. This study is the first to look at characterisation and crystallisation of this protein or indeed a PNPOx from *D. radiodurans*.

8.2 Other Targets

8.2.1 SP_1651

Expression of the gene *sp_1651* (*psaD*) has previously been shown to encode a thiol peroxidase in *S. pneumoniae* [139]. Its homologue from *D. radiodurans*, DR_2242 is a putative thiol-specific antioxidant protein, belonging to the Thioredoxin-like superfamily [6, 7]. The structure of DR_2242 has been solved by Dr. Dave Hall as part of the ESRF's structural genomics project (unpublished).

A full length form of the *sp_1651* gene was recombinantly expressed from pOPINF in sufficient yields for crystallisation experiments. The resulting protein was purified to homogeneity by IMAC and gel filtration and the corresponding elution profile suggested that SP_1651 is dimeric in solution. After 6xHis tag removal by HRV-3C protease, the minimalistic approach to crystallisation proposed by Newman *et al*. (2005) [207] was followed. After 11 days, birefringent crystal bunches were formed in 0.1 M SPG Buffer pH 4.0, 25 % w/v PEG 1500. Diffraction data with a resolution limit of 3.2 Å was obtained. Although mosaic, the image represented a suitable crystal form for optimisation. Unfortunately, attempts to reproduce SP_1651 crystals failed and hence, structural comparisons with DR_2242 could not be made.

8.2.2 DR_0463, A maltooligosyltrehalose synthase (MTSase)

D. radiodurans encodes a large 108 kDa enzyme known as maltooligosyltrehalose synthase (MTSase) from the gene *treY* (locus name *dr_0463*). This protein, in conjunction with its partner protein maltooligosyltrehalose trehalohyrdolase (MTHase), catalyses the breakdown of maltooligosaccharide (or starch) into trehalose [122]. Trehalose has been shown to protect cell membranes and proteins against oxidative stress [123-128] and thus could facilitate the extraordinary protection and recovery displayed by *D. radiodurans*.

This study describes the expression of *treY* in BL21(DE3)pLysS cells to yield large amounts of DR_0463. Unfortunately the target protein was expressed in inclusion bodies, hence subsequent purification trials required solubilisation in 8 M urea. Once unfolded and extracted from the insoluble cell fraction, DR_0463 was purified by IMAC with buffers containing 8M urea to maintain a denatured state. This step also proved problematic, as the 6xHis tagged MTSase showed weak affinity for the Ni^{2+} column matrix thus allowing non-specific binding of contaminant *E. coli* proteins. In an attempt to increase binding affinity several protocol adaptations were made including, use of a higher volume column, reducing the imidazole concentration of the IMAC wash buffer, adding glycerol to 'loosen' the protein and finally, changing the denaturant to 6 M guanidine hydrochloride. None of these changes increased in the amount of DR_0463 binding to the column.

However, it was found that sequentially repeating IMAC purification up to three times with the same protein sample resulted in the removal of many unwanted impurities. Purity levels and yield deemed acceptable for further purification by gel filtration or ion exchange were obtained.

In order to progress to crystal trials, the purification protocol described for DR_0463 would need to be optimised to produce sufficient yields of pure protein for crystal trials. In order to set up crystallisation experiments, DR_0463 would first have to be re-folded into a structured aqueous state which could be assessed by CD or NMR.

8.2.3 SP_1648 and DR_2284

SP_1648 is an ATP-binding protein that forms part of the Mn-ABC transport system in *S. pneumoniae* [145] and its homologue from *D. radiodurans*, DR_2284 is predicted to share a similar function [6, 7]. Although Mn is known play a pivotal role in preventing oxidative protein damage in *D. radiodurans*, research into the structure of the Mn-ABC transport system remains undocumented.

Expression of the *sp_1648*-pOPINF construct in B834(DE3) cells, using autoinduction media resulted in production of the corresponding protein predominantly in inclusion bodies. On attempting to purifying the small amount

of soluble protein produced by IMAC, it be came apparent that the 6xHis tag was inaccessible for binding to the Ni^{2+} matrix of the HisTrap column. In order to obtain homogenous SP_1648 for crystal trials, alternative methods of purification (ion exchange or gel filtration) would need to be utilised in order to overcome 6xHis tag non-affinity problems. Alternatively, experiments to solubilise, purify and refold insoluble SP_1648 could be attempted.

Small scale trials for the expression of the *dr_2284*-pET151/D TOPO construct in LB media, resulted in the production of only a very small amount of insoluble protein in arabinose-induced BL21-AI cells. Future work would include continuation of expression trials, to investigate variables such as temperature and induction conditions in an attempt to express either soluble protein or a greater yield of insoluble protein.

9 Membrane proteins from *D. radiodurans*

9.1 Working with membrane proteins

To complement information gained from the structures determined for soluble proteins, the ESRF's structural genomics project has been expanded to include investigation into the structure determination of target membrane proteins. The first 6 months of this CASE award was therefore spent at the protein laboratory in Grenoble to gain initial experience in cloning.

Membrane proteins represent about 30 % of the human genome and are prime objectives in the pharmaceutical industry as drug targets. It has been estimated that more than half of all drugs currently on the market are directed against membrane proteins [212]. They also ubiquitous in their function, ranging from proteins involved in complex signalling pathways to basic small molecule transport proteins. However, the number of membrane protein structures determined to date (approximately 100) remains relatively low in comparison to the 30,000 soluble protein structures determined [213-215]. This can be attributed to the many difficulties faced during several stages: in cloning, expression, purification and finally in the crystallisation pipeline. Over the last decade there has been a modicum of success in membrane crystal structure determination, but there is still a shortfall between predicted expansion and reality of actual structures determined [216].

Membrane proteins are notoriously difficult to express by conventional expression routes, for example, the expressed product can be toxic to the *E. coli* host cells due to the preponderance of hydrophobic domains [217-219]. Hence, careful consideration is required when designing plasmid constructs and when selecting systems for recombinant expression. The cell lines C41 and C43, which are mutated versions of the *E. coli* strain BL21, have been shown to overcome much of the toxic effects associated with over-expression of hydrophobic membrane proteins [217] and are therefore a good starting point for expression trials. Alternatively, successful expression has been reported using yeast, insect and mammalian systems [220]. However, these methodologies are more complex

and expensive than the *E. coli* route, which has the advantage of being cheap and easy to use [221].

When extracting and purifying a membrane protein after successful cloning and expression, it is essential to mimic the native lipid environment of the protein by use of a suitable detergent. The detergent functions to maintain the integrity of the hydrophobic domains and prevent the protein from collapsing [222]. Although there is a considerable number of detergents available the task of determining the most suitable for both protein solubilisation and stability can be difficult [223, 224]. The necessity to have detergent present in all purifying and crystallisation steps can also present a greater challenge for nucleation and crystal growth since this substantially increases the number of variables in the crystallisation experiment [213]. While there is a whole range of detergents with a large variety of critical micelle concentrations (CMC) that would suit most requirements it is still very much a trial and error approach to determine which detergent will solubilise and stabilise the extracted membrane protein. One other, not inconsiderable, factor is the expense involved, particularly with detergents with large CMCs.

It is thought that structural genomic projects may be the way forward in the field of membrane protein structural biology [225]. This is due to the availability of high throughput methodologies employed at every step of the journey from clone to structure, which allows cloning, expression, solubilisation and purification experiments to be carried out in parallel [226]. Comparisons can be easily made between a large number and range of targets, allowing selection of potential candidates for crystallisation and structure determination.

A range of membrane proteins from *D. radiodurans* had already been chosen for investigation and two of theses protein targets included in the project were an undecaprenyl diphosphatase (UDP) and a diacylglycerol kinase (DGKA). This study describes the molecular cloning of their corresponding genes *bacA* and *dgkA* to enable recombinant expression trials.

9.2 Diacylglycerol kinase (DGKA)

D. radiodurans is predicted to encode a 13.3 kDa diacylglycerol kinase from the 384 bp gene, *dgkA* [6, 7]. The bioinformatics programs TMpred [227] and SOSUI [228] have been used to predict DGKA will have 3 transmembrane domains once expressed.

While there is no current 3-D structure for *E. coli*-DGKA, NMR assignments have been made [229] and the topology of the protein has been elucidated [230]. It exists as a 40 kDa homotrimer consisting of three 13 kDa subunits each with an active site. In turn each of the subunits has three transmembrane domains. Biochemically, DGKA has been shown to catalyse the conversion of diacylglycerol and MgATP to phosphatidic acid and MgADP. This pathway has served as a valuable model for membrane protein biocatalysis, stability, folding and misfolding studies [231], and thus presents DGKA as a potential drug target.

9.3 Undecaprenyl diphosphatase (UDP)

In *D. radiodurans*, undecaprenyl diphosphatase (locus name DR_0454) is a 29.7 kDa protein predicted to be encoded from the 825 bp gene *bacA* [6, 7]. Using TMpred [227] and SOSUI [228] software, it is predicted that UDP will have 7-8 transmembrane domains once expressed. UDP derived from *E. coli* has been expressed; extracted from the membrane using detergent and purified to near homogeneity in the histidine-tagged form. Although, to date the crystal structure of UDP has not been solved for any organism, the *E. coli* enzyme has been characterised [232, 233]. It has been shown to function as a phosphatase, catalysing the dephosphorylation of undecaprenyl diphosphate to undecaprenyl phosphate, which is an important lipid carrier in the synthesis of peptoglycan and other cell wall components [232, 233].

It has also been shown that the over-expression of the *E. coli bacA* gene encodes a resistance to the polypeptide antibiotic bacitracin [234]. Further investigations into the mechanism employed by UDP to encode such antibiotic resistance could be of great interest from a clinical aspect.

9.4 Materials and methods

Full length forms of the *bacA* and *dgkA* genes were cloned into a pET-28b producing both N- and C-terminally 6xHis tagged constructs. Primers were designed to incorporate BamHI/HindIII (N-terminally tagged) and NcoI/XhoI (C-terminally tagged) restriction enzymes and are depicted in Table 9.1.

Primer	Sequence	Restriction site
Full length *bacA* with N-terminal 6xHis tag	Forward: 5'-GAATTC**GGATCC**GATGGACTGGCTGTA TTCGCTG-3'	*Bam*HI: 5'-GˆGATCC-3'
	Reverse: 5'-GAATTC**AAGCTT**ATCAGGCGAGCGAGC CGT TTTG-3'	*Hind*III: 5'-AˆAGCTT-3'
Full length *bacA* with C-terminal 6xHis tag	Forward: 5'GGAT**CCATGG**ACTGGCTGTATTCGCTG-3'	*Nco*I: 5'-CˆCATGG-3'
	Reverse: 5'-AAGCTT**CTCGAG**GGCGAGCGAGCCGTTT TGC-3	*Xho*I: 5'-CˆTCGAG-3'
Full length *dgkA* with N-terminal 6xHis tag	Forward: 5'-GAATTC**GGATCC**GGTGAACTCGGCG GGCTCGGC-3'	*Bam*HI: 5'-GˆGATCC-3'
	Reverse: 5'-GAATTC**AAGCTT**ACTAGAGCCGACTCA CCCACGC-3'	*Hind*III: 5'-AˆAGCTT-3'
Full length *dgkA* with C-terminal 6xHis tag	Forward: 5'- GGAT**CCATGG**GCGTGAACTCGGCGG GCTCGGC-3'	*Nco*I: 5'-CˆCATGG-3'
	Reverse: 5'-AAGCTT**CTCGAG**GAGCCGACTCACCC ACGC-3'	*Xho*I: 5'-CˆTCGAG-3'

Table 9.1: Oligonucleotide primers for amplification of *bacA* and *dgkA* genes. The restriction sites incorporated into each primer sequence are shown in bold.

The required inserts were amplified from *D. radiodurans* R1 genomic DNA by PCR using the master mix and temperature cycles listed in Section 2.2.3. To establish the optimum concentration of genomic DNA required, several dilutions of the 0.375 mg/mL stock were trialled (1:0, 1:10, 1:20 and 1:40 dilutions of genomic DNA with sterile deionised H_2O). The resulting PCR products were digested with the corresponding restriction enzymes (*Bam*HI and *Hind*III for N-terminally tagged; *Nco*I and *Xho*I C-terminally tagged) and separated by agarose gel electrophoresis (Sections 2.2.8 and 2.2.4). The required genes were extracted, purified and then ligated (Section 2.2.9) into linearised pET-28b (prepared with *Bam*HI and *Hind*III or *Nco*I and *Xho*I restriction enzymes). Positive clones were identified by digestion with the restriction enzyme *Eco*R1 which is unable to cleave plasmid DNA if the insert has been ligated into place effectively.

Positive clones were transformed in DH5-α cells for storage (Section 2.2.6) and into BL21(C41) and BL21(C43) expression strains for preliminary, small scale expression trials using 20 mL LB cultures with IPTG induction (Section 2.2.4). BL21(C41) and BL21(C43) cell lines were selected based on evidence that they function to overcome the toxic effects associated with over-expression of hydrophobic membrane proteins [217] in order to give sufficient yields of protein for characterisation and crystallisation experiments.

9.5 Results and discussion

9.5.1 DGK

Full length forms of the *dgkA* gene were cloned into pET-28b to produce N- and C-terminally Hisx6 tagged constructs for expression trials. Critical to the success of this process was the optimisation of PCR conditions which produced a yield of amplified insert sufficient for ligation into pET-28b. The best amplification was achieved using 1:0 and 1:10 concentrations of genomic DNA; an annealing temperature of 55°C, and by using the additive, Q solution (Purchased from Qiagen, composition not available) to alleviate the possibility of mutation caused by high G-C content (72.4 % for *dgkA* [6, 7]). Figure 9.1 illustrates the positive PCR of *dgkA* inserts (384 bp) which are shown as strong bands between 500 bp and 250 bp marker bands.

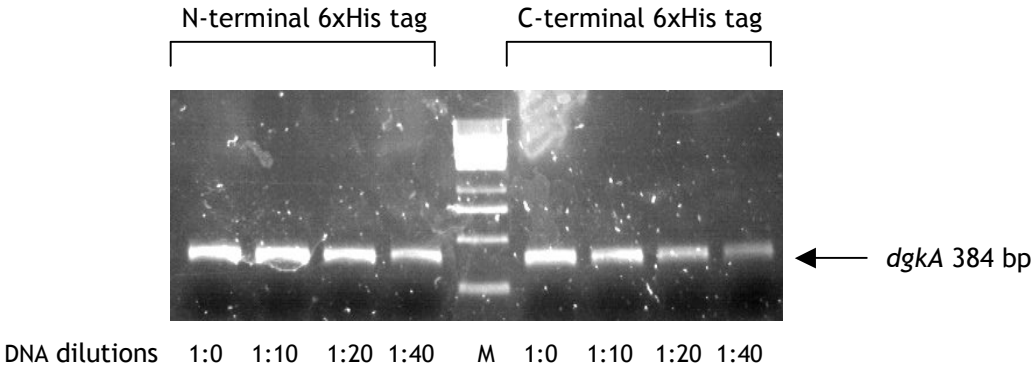


Figure 9.1: Agarose gel showing successful PCR of *dgkA* from 1:0, 1:10, 1:20 and 1:40 dilutions of genomic DNA (stock concentration 0.375 mg/mL). Primers designed to produce N-terminal and C-terminal 6xHis tagged constructs were used for the PCR reaction. PCR products are shown by strong single 384 bp bands between 500 and 250 bp marker bands.

Successful ligation into linearised pET-28b was achieved for almost all of the *dgkA* inserts (both N- and C- terminally tagged) as shown by ineffective digestion with *Eco*R1 (Figure 9.2). Positive clones are depicted by un-cleaved plasmid which appears on an agarose gel as three close bands; nicked (top), open coiled (middle) and supercoiled (bottom) DNA. Several positive clones were selected for sequencing which confirmed the presence of the *dgkA* insert. The single negative clone obtained (marked X in Figure 9.2) was excluded from further studies.

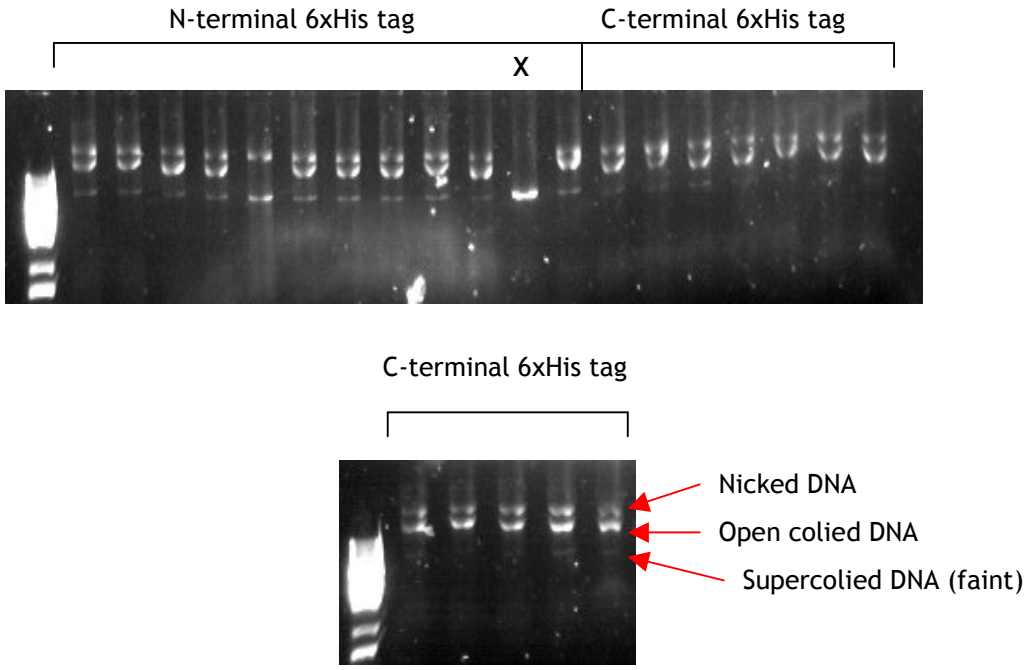


Figure 9.2: Agarose gel imaging *dgkA*-pET-28b plasmids after digestion with *Eco*RI. Positive clones can be identified by three close bands; nicked (top), open coiled (middle) and supercoiled (bottom) DNA. This pattern indicates that plasmid DNA was not cleaved by EcoR1, signifying that the insert has been ligated into place. The single negative clone obtained is marked X and was excluded from further studies.

Research into the expression of *dgkA* from pET-28b was continued at the ESRF, Grenoble by Dr. Rana Roy and Dr. Dave Hall.

9.5.2 UDP

Full length forms of the *bacA* gene were successfully cloned into pET-28b to enable possible expression of N- and C-terminally Hisx6 tagged protein. The best PCR amplification was achieved by using 1:10, 1:20 and 1:40 dilution of genomic DNA and an annealing temperature of 58°C. Positive PCR products are imaged as strong 825 bp fragments in Lanes 2-4 and 7-9 of Figure 9.3. No amplification

occurred from undiluted genomic DNA at 0.375 mg/mL (Figure 9.3, Lanes 1 and 6).

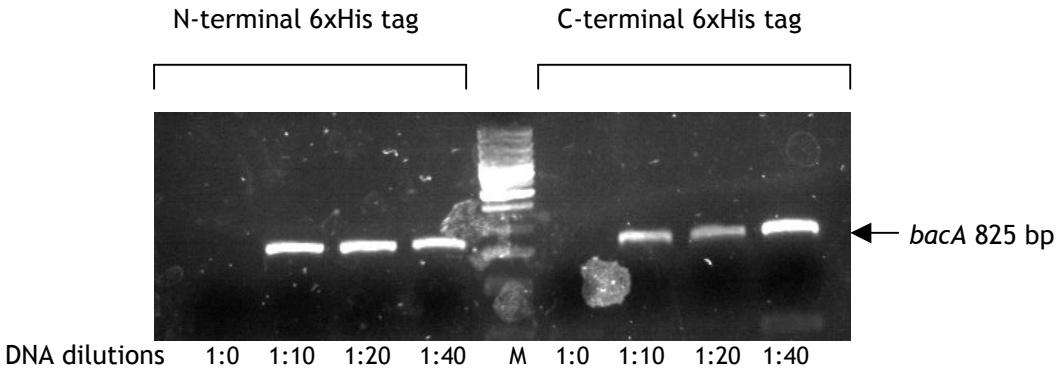


Figure 9.3: Agarose gel showing successful PCR of *bacA* from 1:10, 1:20 and 1:40 dilutions of genomic DNA (0.375 mg/mL). PCR products are shown by strong 825 bp bands between 1000 bp and 500 bp marker bands. The absence of bands in lanes 1 and 6 signify that no amplification occurred from undiluted genomic DNA.

The insert amplified from the 1:40 dilution of genomic DNA was selected for ligation into linearised pET-28b. Successful incorporation of the *bacA* gene into N-terminally tagged pET-28b was confirmed by release of the 825 bp insert from the 5.5 kbp vector on digestion with *Xho*I and *Nco*I restriction enzymes (Figure 9.4).

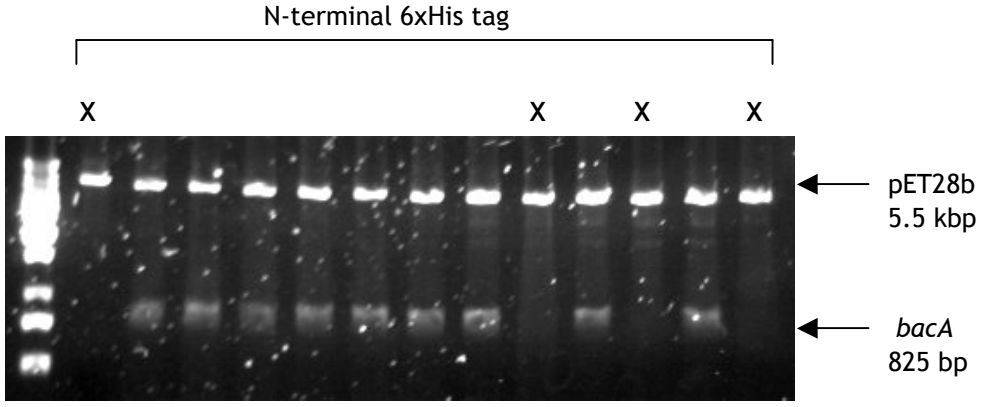


Figure 9.4: Agarose gel showing release of the 825 bp from the 5.5 kbp vector on digestion of *bacA*-pET28b with *Xho*I and *Nco*I restriction enzymes. Release of the insert is indicative of a positive clone. Negative clones (marked by X) are indicated by no evidence of insert release and were excluded from further studies.

clones of *bacA* in pET-28b with residues included to encode an N-terminal 6xHis tag. Positive clones were also obtained for the C-terminal tagged construct as shown by no disruption to plasmid DNA on digestion with EcoR1 (Figure 9.5).

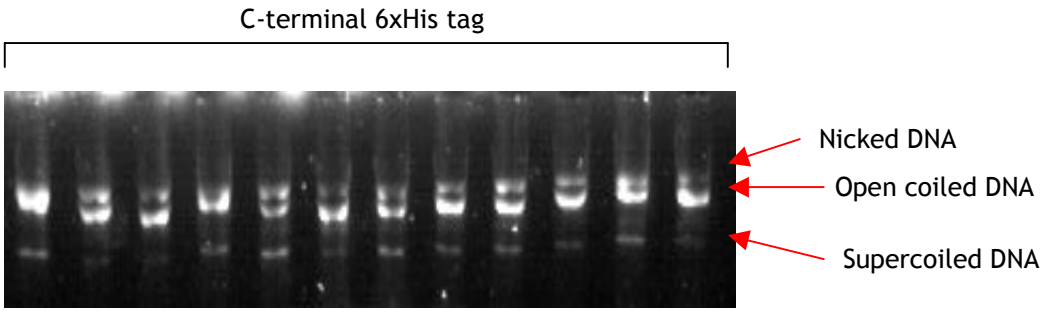


Figure 9.5: Agarose gel imaging *bacA*-pET-28b (C-terminal 6xHis tagged) plasmids after digestion with *Eco*RI. All of the clones are positive as shown by three close bands; nicked (top), open coiled (middle) and supercoiled (bottom) DNA. This pattern indicates that plasmid DNA was not cleaved by EcoR1, signifying that the insert has been ligated into place.

Several positive clones for both N- and C- terminal tagged *bacA* were selected for preliminary small scale expression trials using BL21(C41) and BL21(C43) cell lines. 20 mL LB cultures were incubated until an OD_{600} of 0.6 was reached, and then were induced with 1 mM IPTG. They were then incubated at 16, 20 and 37°C, with shaking at 200 rpm, for 20 hrs. 1 mL samples, taken every hour during cell growth were analysed by SDS-PAGE which showed no trace of *bacA* expression. This negative result was confirmed by western blot.

9.6 Summary and conclusions

The genes *dgkA* and *bacA* are thought to encode the membrane proteins, DGK and UDP in *D. radiodurans*. In order to investigate the enzymes as potential drug targets, the genes have been cloned into pET-28b for recombinant expression in *E. coli* strains. This study describes the successful PCR, restriction enzyme digestion, purification and ligation of the full length target genes into the chosen expression vector. Primers were designed to produce two different constructs for each gene, which, on over-expression would encode target protein with either an N- or a C-terminal 6xHis tag.

Although preliminary expression trials for *bacA* from pET-28b were unsuccessful, there are still avenues to explore in an effort to gain recombinant expression. For example, one could vary the IPTG concentration used for inducing protein production. Furthermore, different induction methods could be employed, such as autoinduction systems or arabinose induction, which is required when using the slow-promoting strain BL21(AI). Alternatively, a change in the constituents of growth media could aid over-expression. As the *bacA* gene consists of many

rare codons which are not recognised by some typical expression strains, cell lines like Rosetta 2(DE3), which can alleviate such codon bias, could be included in expression trials [149, 152, 163, 164].

Ultimately, the cloning of the *D. radiodurans* genes *bacA* and *dgkA* has made an important contribution to the extension of the ESRF's structural genomics project to include the study of membrane proteins. Investigation into the expression of the cloned genes has been continued at the ESRF, by Dr. Rana Roy and Dr. Dave Hall. Access to high-throughput methodologies and the synchrotron radiation source at the ESRF may contribute immensely to the future of this project

Appendix 1 – Sequences of full length DR_1146

DNA sequence

ATGTGTGGTCTGGCCCATTCTCCGGATATGAGCGACAACAAGAACGCCGGACAAAGCAGT
GACAGCCAGGGCAACGGTGCGCCGCAGATGAGCCGTGACGAAGCCGTGAAGACGGTCGCT
GGGCTGGTCAAGGGCATCAAATTCGCCATGCTGGTGTCGGTCAACGACGCGGGCCACATC
CACTCGCGCCCCATGACCACCCAGGAAACCGAGTTCGACGGCGATATCTGGTTTATCGGC
GCCAAGGACAGCGAGGTCGTTCACGATGTCCGCAGCCGCCCCCAGGTCAACGTGAGCTAC
GCCGACACCGGCAGCAACAACTACGTGAGCGTGCACGGCACCGCCGAACTGATCGAAGAT
CGCGCCAAACTCGACGAACTGTGGAGCGACATGTACAACATGTACTTCGAGGGCGGCAAG
GAAGACCCGAATGTCCAGCTCATCAAGATCAACGCCGAGGGCGCCGAATACTGGGAGAGC
GGCGGCAAGGTGCGGACCCTCTTCGCCTTCGCCAAGAACCTGATTCCTGGCCAGCGCGCC
GACGCGAGCGAGCTGGGCAAGAACGACACGGTTAATCTCTAA

Amino acid sequence

```
          10         20         30         40         50         60
MCGLAHSPDM SDNKNAGQSS DSQGNGAPQM SRDEAVKTVA GLVKGIKFAM LVSVNDAGHI

          70         80         90        100        110        120
HSRPMTTQET EFDGDIWFIG AKDSEVVHDV RSRPQVNVSY ADTGSNNYVS VHGTAELIED

         130        140        150        160        170        180
RAKLDELWSD MYNMYFEGGK EDPNVQLIKI NAEGAEYWES GGKVRTLFAF AKNLIPGQRA

         190
DASELGKNDT VNL
```

Appendix 2 - Sequences of truncated DR_1146

DNA sequence

ATGAGCCGTGACGAAGCCGTGAAGACGGTCGCTGGGCTGGTCAAGGGCATCAAATTCGCCAT

GCTGGTGTCGGTCAACGACGCGGGCCACATCCACTCGCGCCCCATGACCACCCAGGAAACCG

AGTTCGACGGCGATATCTGGTTTATCGGCGCCAAGGACAGCGAGGTCGTTCACGATGTCCGC

AGCCGCCCCCAGGTCAACGTGAGCTACGCCGACACCGGCAGCAACAACTACGTGAGCGTGCA

CGGCACCGCCGAACTGATCGAAGATCGCGCCAAACTCGACGAACTGTGGAGCGACATGTACA

ACATGTACTTCGAGGGCGGCAAGGAAGACCCGAATGTCCAGCTCATCAAGATCAACGCCGAG

GGCGCCGAATACTGGGAGAGCGGCGGCAAGGTGCGGACCCTCTTCGCCTTCGCCAAGAACC

TGATTCCTGGCCAGCGCGCCGACGCGAGCGAGCTGGGCAAGAACGACACGGTTAATCTCTAA

Amino acid sequence

```
           11          21          31          41          51          61
M SRDEAVKTVA GLVKGIKFAM LVSVNDAGHI HSRPMTTQET EFDGDIWFIG AKDSEVVHDV

          71          81          91         101         111         121
RSRPQVNVSY ADTGSNNYVS VHGTAELIED RAKLDELWSD MYNMYFEGGK EDPNVQLIKI

         131         141         151         161
NAEGAEYWES GGKVRTLFAF AKNLIPGQRA DASELGKNDT VNL
```

List of References

1. Anderson AW, Nordan HC, Cain RF, Parrish G, Duggan D: **Studies on a Radio-Resistant *Micrococcus* .1. Isolation, Morphology, Cultural Characteristics, and Resistance to Gamma Radiation**. *Food Technology* 1956, **10**(12):575-578.

2. Minton KW: **DNA-Repair in the Extremely Radioresistant Bacterium *Deinococcus radiodurans***. *Molecular Microbiology* 1994, **13**(1):9-15.

3. Thornton JM, Todd AE, Milburn D, Borkakoti N, Orengo CA: **From structure to function: Approaches and limitations**. *Nature Structural Biology* 2000, **7**:991-994.

4. White O: **Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1 (vol 286, pg 1571, 1999)**. *Science* 2004, **303**(5659):766.

5. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL *et al*: **Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1**. *Science* 1999, **286**(5444):1571-1577.

6. **Comprehensive Microbial Resource** [http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi]

7. **J. Craig Venter Institute (JCVI)** [http://www.tigr.org/]

8. Lin JY, Qi R, Aston C, Jing JP, Anantharaman TS, Mishra B, White O, Daly MJ, Minton KW, Venter JC *et al*: **Whole-genome shotgun optical mapping of *Deinococcus radiodurans***. *Science* 1999, **285**(5433):1558-1562.

9. Harsojo, Kitayama S, Matsuyama A: **Genome multiplicity and radiation resistance in *Micrococcus radiodurans***. *J Biochem* 1981, **90**(3):877-880.

10. Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Venkateswaran A, Hess M, Omelchenko MV, Kostandarithes HM, Makarova KS *et al*: **Accumulation of Mn(II) in, *Deinococcus radiodurans* facilitates gamma-radiation resistance**. *Science* 2004, **306**(5698):1025-1028.

11. Masters CI, Murray RGE, Moseley BEB, Minton KW: **DNA Polymorphisms in New Isolates of *Deinococcus-radiopugnans***. *Journal of General Microbiology* 1991, **137**:1459-1469.

12. Counsell TJ, Murray RGE: **Polar Lipid Profiles of the Genus *Deinococcus***. *International Journal of Systematic Bacteriology* 1986, **36**(2):202-206.

13. Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, Bagaley D, Rash BA, Park MJ, Earl AM, Shank NC *et al*: **Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample**. *Appl Environ Microbiol* 2005, **71**(9):5225-5235.

14. Brooks BW, Murray RGE: **Nomenclature for *Micrococcus-radiodurans* and Other Radiation Resistant Cocci *Deinococcaceae* New Family and *Deinococcus* New-Genus Including 5 Species**. *International Journal of Systematic Bacteriology* 1981, **31**(3):339-360.

15. Hirsch P, Gallikowski CA, Siebert J, Peissl K, Kroppenstedt R, Schumann P, Stackebrandt E, Anderson R: ***Deinococcus frigens* sp nov., *Deinococcus saxicola sp* nov., and *Deinococcus marmoris* sp nov., low temperature and draught-tolerating, UV-resistant bacteria from continental Antarctica**. *Systematic and Applied Microbiology* 2004, **27**(6):636-645.

16. Suresh K, Reddy GSN, Sengupta S, Shivaji S: ***Deinococcus indicus* sp nov., an arsenic-resistant bacterium from an aquifer in West Bengal, India**. *International Journal of Systematic and Evolutionary Microbiology* 2004, **54**:457-461.

17. [http://www.usuhs.mil/pat/deinococcus/index_20.htm]

18. Ferreira AC, Nobre MF, Rainey FA, Silva MT, Wait R, Burghardt J, Chung AP, daCosta MS: ***Deinococcus geothermalis* sp. nov. and *Deinococcus murrayi* sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs**. *International Journal of Systematic Bacteriology* 1997, **47**(4):939-947.

19. Cox MM, Battista JR: ***Deinococcus radiodurans* - The consummate survivor**. *Nature Reviews Microbiology* 2005, **3**(11):882-892.

20. Pavlov AK, Kalinin VL, Konstantinov AN, Shelegedin VN, Pavlov AA: **Was earth ever infected by martian biota? Clues from radioresistant bacteria**. *Astrobiology* 2006, **6**(6):911-918.

21. Battista JR, Earl AM, Park MJ: **Why is *Deinococcus radiodurans* so resistant to ionizing radiation?** *Trends in Microbiology* 1999, **7**(9):362-365.

22. Mattimore V, Battista JR: **Radioresistance of *Deinococcus radiodurans*: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation**. *J Bacteriol* 1996, **178**(3):633-637.

23. Krasin F, Hutchinson F: **Repair of DNA Double-Strand Breaks in *Escherichia-Coli*, Which Requires Reca-Function and Presence of a Duplicate Genome**. *Journal of Molecular Biology* 1977, **116**(1):81-98.

24. Imlay JA: **Pathways of oxidative damage**. *Annual Review of Microbiology* 2003, **57**:395-418.

25. Halliwell B, Gutteridge JM: **Free Radicals in Biology and Medicine**, 3rd edn. Oxford: Oxford University Press; 1999.

26. Gerard E, Jolivet E, Prieur D, Forterre P: **DNA protection mechanisms are not involved in the radioresistance of the hyper thermophilic archaea Pyrococcus abyssi and P-furiosus**. *Molecular Genetics and Genomics* 2001, **266**(1):72-78.

27. Burrell AD, Feldschr.P, Dean CJ: **DNA-Membrane Association and Repair of Double Breaks in X-Irradiated *Micrococcus-Radiodurans***. *Biochimica Et Biophysica Acta* 1971, **247**(1):38-&.

28. Moseley BEB, Mattingl.A: **Repair of Irradiated Transforming Deoxyribonucleic Acid in Wild Type and a Radiation-Sensitive Mutant of *Micrococcus-Radiodurans***. *J Bacteriol* 1971, **105**(3):976-&.

29. Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, Petranovic M, Lindner AB, Radman M: **Reassembly of shattered chromosomes in *Deinococcus radiodurans***. *Nature* 2006, **443**(7111):569-573.

30. Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Leapman RD, Lai B, Ravel B, Li SMW, Kemner KM *et al*: **Protein oxidation implicated as the primary determinant of bacterial radioresistance**. *Plos Biology* 2007, **5**(4):769-779.

31. Wang P, Schellhorn HE: **Induction of Resistance to Hydrogen-Peroxide and Radiation in *Deinococcus Radiodurans***. *Can J Microbiol* 1995, **41**(2):170-176.

32. McCord JM, Keele BB, Fridovic.I: **Enzyme-Based Theory of Obligate Anaerobiosis - Physiological Function of Superoxide Dismutase**. *Proceedings of the National Academy of Sciences of the United States of America* 1971, **68**(5):1024-&.

33. Karlin S, Mrazek J: **Predicted highly expressed and putative alien genes of *Deinococcus radiodurans* and implications for resistance to ionizing radiation damage**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(9):5240-5245.

34. Johnson SS, Hebsgaard MB, Christensen TR, Mastepanov M, Nielsen R, Munch K, Brand T, Thomas M, Gilbert P, Zuber MT *et al*: **Ancient bacteria show evidence of DNA repair**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(36):14401-14405.

35. Fredrickson JK, Li SMW, Gaidamakova EK, Matrosova VY, Zhai M, Sulloway HM, Scholten JC, Brown MG, Balkwill DL, Daly MJ: **Protein oxidation: key to bacterial desiccation resistance?** *Isme Journal* 2008, **2**(4):393-403.

36. Stadtman ER: **Oxidation of Free Amino-Acids and Amino-Acid-Residues in Proteins by Radiolysis and by Metal-Catalyzed Reactions**. *Annual Review of Biochemistry* 1993, **62**:797-821.

37. Nystrom T: **Role of oxidative carbonylation in protein quality control and senescence**. *Embo Journal* 2005, **24**(7):1311-1317.

38. Ghosal D, Omelchenko MV, Gaidamakova EK, Matrosova VY, Vasilenko A, Venkateswaran A, Zhai M, Kostandarithes HM, Brim H, Makarova KS *et al*: **How radiation kills cells: Survival of *Deinococcus radiodurans* and *Shewanella oneidensis* under oxidative stress**. *Fems Microbiology Reviews* 2005, **29**(2):361-375.

39. Sweet DM, Moseley BEB: **Resistance of *Micrococcus-Radiodurans* to Killing and Mutation by Agents Which Damage DNA**. *Mutation Research* 1976, **34**(2):175-186.

40. von Sonntag C: **The Chemical basis of radiation biology**. London: Taylor and Francis Group; 1987.

41. Nauser T, Koppenol WH, Gebicki JM: **The kinetics of oxidation of GSH by protein radicals**. *Biochemical Journal* 2005, **392**:693-701.

42. Markillie LM, Varnum SM, Hradecky P, Wong KK: **Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: Radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants**. *J Bacteriol* 1999, **181**(2):666-669.

43. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin EV, Daly MJ: **Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics**. *Microbiology and Molecular Biology Reviews* 2001, **65**(1):44-+.

44. Daly MJ: **Modulating radiation resistance: Insights based on defenses against reactive oxygen species in the radioresistant bacterium *Deinococcus radiodurans***. *Clinics in Laboratory Medicine* 2006, **26**(2):491-+.

45. Harris DR, Tanaka M, Saveliev SV, Jolivet E, Earl AM, Cox MM, Battista JR: **Preserving genome integrity: The DdrA protein of *Deinococcus radiodurans* R1**. *Plos Biology* 2004, **2**(10):1629-1639.

46. McCullough J, Hazen T, Benson S, F B-M, Palmisano AC: **Bioremediation of metals and radionuclotides**. In.: US Dept. of Energy, Office of Biological and Environmental Research, Germantown, MD 1999.

47. Riley RG, Zachara JM, Wobber FJ: **Chemical Contaminants on DOE Lands and Selection of Contaminant Mixtures for Subsurface Science Research**. In.: US Department of Energy, Office of Energy Research, Subsurface Science Program, Washington, DC; 1992.

48. Macilwain C: **Science seeks weapons clean-up role**. *Nature* 1996, **383**(6599):375-&.

49. Tsapin AI, Nealson KH, Meyers T, Cusanovich MA, VanBeuumen J, Crosby LD, Feinberg BA, Zhang C: **Purification and properties of a low-redox-potential tetraheme cytochrome c(3) from *Shewanella putrefaciens***. *J Bacteriol* 1996, **178**(21):6386-6388.

50. Turner JS, Robinson NJ: **Cyanobacterial Metallothioneins - Biochemistry and Molecular-Genetics**. *J Indust Microbiol* 1995, **14**(2):119-125.

51. Voordouw G, Brenner S: **Cloning and Sequencing of the Gene Encoding Cytochrome-C3 from *Desulfovibrio-Vulgaris* (Hildenborough)**. *Eur J Biochem* 1986, **159**(2):347-351.

52. Thornley MJ: **Radiation Resistance among Bacteria**. *J Appl Bacteriol* 1963, **26**(3):334-&.

53. Rugh CL, Senecoff JF, Meagher RB, Merkle SA: **Development of transgenic yellow poplar for mercury phytoremediation**. *Nat Biotechnol* 1998, **16**(10):925-928.

54. Rugh CL, Wilde HD, Stack NM, Thompson DM, Summers AO, Meagher RB: **Mercuric ion reduction and resistance in transgenic Arabidopsis thaliana plants expressing a modified bacterial merA gene**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(8):3182-3187.

55. Lange CC, Wackett LP, Minton KW, Daly MJ: **Engineering a recombinant *Deinococcus radiodurans* for organopollutant degradation in radioactive mixed waste environments**. *Nat Biotechnol* 1998, **16**(10):929-933.

56. Barrineau Pea: **The structure of the *mer* operon**. *Basic Life Science* 1985, **30**:707-718.

57. Brim H, McFarlan SC, Fredrickson JK, Minton KW, Zhai M, Wackett LP, Daly MJ: **Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments**. *Nat Biotechnol* 2000, **18**(1):85-90.

58. **NCBI BLAST (Basic Local Alignment Search Tool)** [http://blast.ncbi.nlm.nih.gov/Blast.cgi]

59. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Research* 1997, **25**:3389-3402.

60. Zhao GS, Winkler ME: **Kinetic Limitation and Cellular Amount of Pyridoxine (Pyridoxamine) 5'-Phosphate Oxidase of *Escherichia-Coli* K-12**. *J Bacteriol* 1995, **177**(4):883-891.

61. Safo MK, Mathews I, Musayev FN, di Salvo ML, Thiel DJ, Abraham DJ, Schirch V: **X-ray structure of *Escherichia coli* pyridoxine 5 '-phosphate oxidase complexed with FMN at 1.8 angstrom resolution**. *Struct Fold Des* 2000, **8**(7):751-762.

62. Di Salvo M, Yang E, Zhao G, Winkler ME, Schirch V: **Expression, purification, and characterization of recombinant *Escherichia coli* pyridoxine 5 '-phosphate oxidase**. *Protein Expres Purif* 1998, **13**(3):349-356.

63. Musayev FN, Di Salvo ML, Ko TP, Schirch V, Safo MK: **Structure and properties of recombinant human pyridoxine 5 '-phosphate oxidase**. *Protein Sci* 2003, **12**(7):1455-1463.

64. Choi JD, Bowerskomro DM, Davis MD, Edmondson DE, Mccormick DB: **Kinetic-Properties of Pyridoxamine (Pyridoxine)-5'-Phosphate Oxidase from Rabbit Liver**. *Journal of Biological Chemistry* 1983, **258**(2):840-845.

65. Choi SY, Churchich JE, Zaiden E, Kwok F: **Brain Pyridoxine-5-Phosphate Oxidase - Modulation of Its Catalytic Activity by Reaction with Pyridoxal 5-Phosphate and Analogs**. *Journal of Biological Chemistry* 1987, **262**(25):12013-12017.

66. Loubbardi A, Marcireau C, Karst F, Guilloton M: **Sterol Uptake Induced by an Impairment of Pyridoxal-Phosphate Synthesis in Saccharomyces-Cerevisiae - Cloning and Sequencing of the Pdx3 Gene Encoding Pyridoxine (Pyridoxamine) Phosphate Oxidase**. *J Bacteriol* 1995, **177**(7):1817-1823.

67. Sang YY, Barbosa JM, Wu HZ, Locy RD, Singh NK: **Identification of a pyridoxine (pyridoxamine) 5 '-phosphate oxidase from Arabidopsis thaliana**. *Febs Lett* 2007, **581**(3):344-348.

68. Safo MK, Musayev FN, di Salvo ML, Schirch V: **X-ray structure of *Escherichia coli* pyridoxine 5 '-phosphate oxidase complexed with pyridoxal 5 '-phosphate at 2.0 angstrom resolution**. *Journal of Molecular Biology* 2001, **310**(4):817-826.

69. Fitzpatrick TB, Amrhein N, Kappes B, Macheroux P, Tews I, Raschle T: **Two independent routes of de novo vitamin B-6 biosynthesis: not that different after all**. *Biochemical Journal* 2007, **407**:1-13.

70. Eliot AC, Kirsch JF: **Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations**. *Annual Review of Biochemistry* 2004, **73**:383-415.

71. Drewke C, Leistner E: **Biosynthesis of vitamin B-6 and structurally related derivatives**. In: *Vitamins and Hormones - Advances in Research and Applications, Vol 61*. vol. 61. San Diego: Academic Press Inc; 2001: 121-155.

72. Ehrenshaft M, Bilski P, Li MY, Chignell CF, Daub ME: **A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis**. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(16):9374-9378.

73. Bilski P, Li MY, Ehrenshaft M, Daub ME, Chignell CF: **Vitamin B-6 (pyridoxine) and its derivatives are efficient singlet oxygen quenchers and potential fungal antioxidants**. *Photochem Photobiol* 2000, **71**(2):129-134.

74. Jain SK, Lim G: **Pyridoxine and pyridoxamine inhibits superoxide radicals and prevents lipid peroxidation, protein glycosylation, and (Na++K+)-ATPase activity reduction in high glucose-treated human erythrocytes**. *Free Radic Biol Med* 2001, **30**(3):232-237.

75. Chen H, Xiong LM: **Pyridoxine is required for post-embryonic root development and tolerance to osmotic and oxidative stresses**. *Plant J* 2005, **44**(3):396-408.

76. Kazarinoff MN, McCormick DB: **Rabbit Liver Pyridoxamine (Pyridoxine) 5'-Phosphate Oxidase - Purification and Properties**. *Journal of Biological Chemistry* 1975, **250**(9):3436-3442.

77. Di Salvo ML, Ko TP, Musayev FN, Raboni S, Schirch V, Safo MK: **Active site structure and stereospecificity of *Escherichia coli* pyridoxine-5 '-phosphate oxidase**. *Journal of Molecular Biology* 2002, **315**(3):385-397.

78. Stryer L, Berg JM, Tymoczko JL: **Biochemistry 5th Edition**. New York: W.H. Freeman and Company; 2002.

79. Berman H, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank**. *Nature Structural Biology* 2003, **10**(12):980-980.

80. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucleic Acids Research* 2000, **28**(1):235-242 [www.pdb.org]

81. (JCSG) JCfSG: **PDB ID: 2I02, Crystal structure of general stress protein of COG3871 (ZP_00108720.1) from *Nostoc punctiforme* PCC 73102 at 1.80 A resolution**. In.; To be published

82. Walsh C: **Flavin Coenzymes - at the Crossroads of Biological Redox Chemistry**. *Accounts Chem Res* 1980, **13**(5):148-155.

83. Muller F: **The Flavin Redox-System and Its Biological Function**. *Topics in Current Chemistry* 1983, **108**:71-107.

84. Fraaije MW, Mattevi A: **Flavoenzymes: diverse catalysts with recurrent features**. *Trends in Biochemical Sciences* 2000, **25**(3):126-132.

85. Massey V: **The chemical and biological versatility of riboflavin**. *Biochem Soc Trans* 2000, **28**:283-296.

86. Bornemann S: **Flavoenzymes that catalyse reactions with no net redox change**. *Nat Prod Rep* 2002, **19**(6):761-772.

87. Joosten V, van Berkel WJH: **Flavoenzymes**. *Curr Opin Chem Biol* 2007, **11**(2):195-202.

88. Hefti MH, Vervoort J, van Berkel WJH: **Deflavination and reconstitution of flavoproteins - Tackling fold and function**. *Eur J Biochem* 2003, **270**(21):4227-4242.

89. Ballou DP, Williams CH, Coon MJ: **Vincent Massey (1926-2002)**. *Trends in Biochemical Sciences* 2002, **27**(641-642).

90. Kao YT, Saxena C, He TF, Guo LJ, Wang LJ, Sancar A, Zhong DP: **Ultrafast dynamics of flavins in five redox states**. *J Am Chem Soc* 2008, **130**(39):13132-13139.

91. Spencer R, Fisher J, Walsh C: **Preparation, Characterization, and Chemical Properties of Flavin Coenzyme Analogs 5-Deazariboflavin, 5-Deazariboflavin 5'-Phosphate, and 5-Deazariboflavin 5'-Diphosphate, 5'-]5'-Adenosine Ester**. *Biochemistry* 1976, **15**(5):1043-1053.

92. Karthikeyan S, Zhou QX, Mseeh F, Grishin NV, Osterman AL, Zhang H: **Crystal structure of human riboflavin kinase reveals a beta barrel fold and a novel active site arch**. *Structure* 2003, **11**(3):265-273.

93. Manstein DJ, Pai EF: **Purification and Characterization of Fad Synthetase from *Brevibacterium-Ammoniagenes***. *Journal of Biological Chemistry* 1986, **261**(34):6169-6173.

94. Muller F, Mayhew SG, Massey V: **Effect of Temperature on Absorption-Spectra of Free and Protein-Bound Flavines**. *Biochemistry* 1973, **12**(23):4654-4662.

95. Palfey BA, Massey V: **Flavin-dependent enzymes. In Comprehensive Biochemical Catalysis**
New York: Academic Press; 1996.

96. Ghisla S, Massey V: **Mechanisms of Flavoprotein-Catalyzed Reactions**. *Eur J Biochem* 1989, **181**(1):1-17.

97. Susin SA, Lorenzo HK, Zamzami N, Marzo I, Snow BE, Brothers GM, Mangion J, Jacotot E, Costantini P, Loeffler M *et al*: **Molecular characterization of mitochondrial apoptosis-inducing factor**. *Nature* 1999, **397**(6718):441-446.

98. Chen YC, Lin-Shiau SY, Lin JK: **Involvement of reactive oxygen species and caspase 3 activation in arsenite-induced apoptosis**. *J Cell Physiol* 1998, **177**(2):324-333.

99. Dagley S: **Lessons from Biodegradation**. *Annual Review of Microbiology* 1987, **41**:1-23.

100. Chen HZ, Hopper SL, Cerniglia CE: **Biochemical and molecular characterization of an azoreductase from *Staphylococcus aureus*, a tetrameric NADPH-dependent flavoprotein**. *Microbiology-(UK)* 2005, **151**:1433-1441.

101. Sancar A: **Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors**. *Chem Rev* 2003, **103**(6):2203-2237.

102. Kao YT, Saxena C, Wang LJ, Sancar A, Zhong DP: **Direct observation of thymine dimer repair in DNA by photolyase**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(45):16128-16132.

103. Mewies M, McIntire WS, Scrutton NS: **Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: The current state of affairs**. *Protein Sci* 1998, **7**(1):7-20.

104. Vermilion JL, Coon MJ: **Identification of the High and Low Potential Flavins of Liver Microsomal Nadph-Cytochrome P-450 Reductase**. *Journal of Biological Chemistry* 1978, **253**(24):8812-8819.

105. Merrill AH, Kasai S, Matsui K, Tsuge H, McCormick DB: **Spectroscopic studies of pyridoxamine(pyridoxine) 5'-phosphate oxidase. Equilibrium**

dissociation constants and spectra for riboflavin 5'-phosphate and analogs. *Biochemistry* 1979, **18**(16):3635-3641.

106. Dym O, Eisenberg D: **Sequence-structure analysis of FAD-containing proteins**. *Protein Sci* 2001, **10**(9):1712-1728.

107. Lee YH, Nadaraia S, Gu D, Becker DF, Tanner JJ: **Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein**. *Nature Structural Biology* 2003, **10**(2):109-114.

108. Massey V, Palmer G: **On Existence of Spectrally Distinct Classes of Flavoprotein Semiquinones . a New Method for Quantitative Production of Flavoprotein Semiquinones**. *Biochemistry* 1966, **5**(10):3181-&.

109. Massey V, Palmer G, Ballou DP: **In Oxidases and Related Systems** Baltimore: University Park Press; 1973.

110. Kemal C, Chan TW, Bruice TC: **Reaction of 3o2 with Dihydroflavins .1. "N3,5-Dimethyl-1,5-Dihydrolumiflavin and 1,5-Dihydroisoalloxazines**. *J Am Chem Soc* 1977, **99**(22):7272-7286.

111. Bruice TC: **Oxygen-Flavin Chemistry**. *Isr J Chem* 1984, **24**(1):54-61.

112. Palfey BA, Ballou DP, Massey ViROSiB: **in Reactive Oxygen Species in Biochemistry** London: Blackie Academic; 1995.

113. Halliwell B. In: *Reactive Oxygen Species in Biochemistry* Edited by Valentine JS, Foote CS, Liebman J, Greenberg A. London: Blackie Academic; 1995: 313–335.

114. Huie RE, Padmaja S: **The Reaction of No with Superoxide**. *Free Radic Res Commun* 1993, **18**(4):195-199.

115. Messner KR, Imlay JA: **Mechanism of superoxide and hydrogen peroxide formation by fumarate reductase, succinate dehydrogenase, and aspartate oxidase**. *Journal of Biological Chemistry* 2002, **277**(45):42563-42571.

116. Heidelberg JF, Paulsen IT, Nelson KE, Gaidos EJ, Nelson WC, Read TD, Eisen JA, Seshadri R, Ward N, Methe B *et al*: **Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis***. *Nat Biotechnol* 2002, **20**(11):1118-1123.

117. Orengo C: **Classification of Protein Folds**. *Curr Opin Struct Biol* 1994, **4**(3):429-440.

118. Holm L, Sander C: **Protein folds and families: sequence and structure alignments**. *Nucleic Acids Research* 1999, **27**(1):244-247.

119. Orengo CA, Todd AE, Thornton JM: **From protein structure to function**. *Curr Opin Struct Biol* 1999, **9**(3):374-382.

120. Maniasetty BA, Turnbull AP, Panjikar S, Bussow K, Chance MR: **Automated technologies and novel techniques to accelerate protein crystallography for structural genomics**. *Proteomics* 2008, **8**(4):612-625.

121. Chayen NE: **Optimization techniques for automation and high throughput**. *Methods Mol Biol* 2007, **363** 175–190.

122. Timmins J, Leiros HKS, Leonard G, Leiros I, McSweeney S: **Crystal structure of maltooligosyltrehalose trehalohydrolase from *Deinococcus radiodurans* in complex with disaccharides**. *Journal of Molecular Biology* 2005, **347**(5):949-963.

123. Eleutherio ECA, Araujo PS, Panek AD: **Protective Role of Trehalose during Heat-Stress in *Saccharomyces-Cerevisiae***. *Cryobiology* 1993, **30**(6):591-596.

124. Paik S-K, Yun H-S, Sohn H-Y, Jin I: **Effect of trehalose accumulation on the intrinsic and acquired thermotolerance in a natural isolate, *Saccharomyces cerevisiae* KNU5377**. *J Microbiol Biotechnol* 2003, **13**:85-89.

125. Park JE, Lee KH, Jahng D: **Effect of trehalose on bioluminescence and viability of freeze-dried bacterial cells**. *J Microbiol Biotechnol* 2002, **12**(2):349-353.

126. Vanlaere A: **Trehalose, Reserve and or Stress Metabolite**. *Fems Microbiology Reviews* 1989, **63**(3):201-210.

127. Potts M: **Desiccation tolerance: a simple process?** *Trends in Microbiology* 2001, **9**(11):553-559.

128. Elbein AD, Pan YT, Pastuszak I, Carroll D: **New insights on trehalose: a multifunctional molecule**. *Glycobiology* 2003, **13**:17R–27R.

129. Richards AB, Krakowka S, Dexter LB, Schmid H, Wolterbeek APM, Waalkens-Berendsen DH, Shigoyuki A, Kurimoto M: **Trehalose: a review of properties, history of use and human tolerance, and results of multiple safety studies**. *Food and Chemical Toxicology* 2002, **40**(7):871-898.

130. Cole JN, Henningham A, Gillen CM, Ramachandran V, Walker MJ: **Human pathogenic streptococcal proteomics and vaccine development**. *Proteom Clin Appl* 2008, **2**(3):387-410.

131. Tseng HJ, McEwan AG, Paton JC, Jennings MP: **Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress**. *Infect Immun* 2002, **70**(3):1635-1639.

132. McAllister LJ, Tseng HJ, Ogunniyi AD, Jennings MP, McEwan AG, Paton JC: **Molecular analysis of the psa permease complex of *Streptococcus pneumoniae***. *Molecular Microbiology* 2004, **53**(3):889-901.

133. Paterson GK, Blue CE, Mitchell TJ: **An operon in *Streptococcus pneumoniae* containing a putative alkylhydroperoxidase D homologue contributes to virulence and the response to oxidative stress**. *Microb Pathog* 2006, **40**(4):152-160.

134. Paton JC: **Novel pneumococcal surface proteins: role in virulence and vaccine potential**. *Trends in Microbiology* 1998, **6**(3):85-87.

135. Paton JC, Andrew PW, Boulnois GJ, Mitchell TJ: **Molecular Analysis of the Pathogenicity of *Streptococcus-Pneumoniae* - the Role of Pneumococcal Proteins**. *Annual Review of Microbiology* 1993, **47**:89-115.

136. Musher DM: **Infections Caused by *Streptococcus-Pneumoniae* - Clinical Spectrum, Pathogenesis, Immunity, and Treatment**. *Clin Infect Dis* 1992, **14**(4):801-809.

137. Tuomanen EI, Austrian R, Masure HR: **Mechanisms of Disease - Pathogenesis of Pneumococcal Infection**. *N Engl J Med* 1995, **332**(19):1280-1284.

138. Janoff EN, Rubins JB: **Invasive pneumococcal disease in the immunocompromised host**. *Microb Drug Resist-Mechan Epidemiol Dis* 1997, **3**(3):215-232.

139. Novak R, Braun JS, Charpentier E, Tuomanen E: **Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa**. *Molecular Microbiology* 1998, **29**(5):1285-1296.

140. **Centers for Disease Control and Prevention. Prevention of pneumococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP)**. In.; Morb Mortal Wkly Rep1997;46(No. RR-8):1–24.

141. Butler JC, Breiman RF, Campbell JF, Lipman HB, Broome CV, Facklam RR: **Pneumococcal Polysaccharide Vaccine Efficacy - an Evaluation of Current Recommendations**. *JAMA-J Am Med Assoc* 1993, **270**(15):1826-1831.

142. Shapiro ED, Berg AT, Austrian R, Schroeder D, Parcells V, Margolis A, Adair RK, Clemens JD: **The Protective Efficacy of Polyvalent Pneumococcal Polysaccharide Vaccine**. *N Engl J Med* 1991, **325**(21):1453-1460.

143. Robinson KA, Baughman W, Rothrock G, Barrett NL, Pass M, Lexau C, Damaske B, Stefonek K, Barnes B, Patterson J *et al*: **Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995-1998 - Opportunities for prevention in the conjugate vaccine era**. *JAMA-J Am Med Assoc* 2001, **285**(13):1729-1735.

144. Dintilhac A, Alloing G, Granadel C, Claverys JP: **Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases**. *Molecular Microbiology* 1997, **25**(4):727-739.

145. Dintilhac A, Claverys JP: **The adc locus, which affects competence for genetic transformation in *Streptococcus pneumoniae*, encodes an ABC transporter with a putative lipoprotein homologous to a family of streptococcal adhesins**. *Res Microbiol* 1997, **148**(2):119-131.

146. Lawrence MC, Pilling PA, Epa VC, Berry AM, Ogunniyi AD, Paton JC: **The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein**. *Structure* 1998, **6**(12):1553-1561.

147. Sambrook J, Fritsch EF, Maniatis T: **Molecular Cloning: A Laboratory Manual**. New York: Cold Spring Harbour Laboratory Press; 1989.

148. Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, Studier FW: **Vectors for Selective Expression of Cloned Dnas by T7 Rna-Polymerase**. *Gene* 1987, **56**(1):125-135.

149. Studier FW, Moffatt BA: **Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes**. *Journal of Molecular Biology* 1986, **189**(1):113-130.

150. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW: **Use of T7 RNA-Ploymerase to Direct Expression of Cloned Genes**. *Meth Enzymol* 1990, **185**:60-89.

151. Shuman S: **Recombination Mediated by Vaccinia Virus-DNA Topoisomerase-I in *Escherichia coli* Is Sequence Specific**. *Proceedings of the National Academy of Sciences of the United States of America* 1991, **88**(22):10104-10108.

152. Grodberg J, Dunn JJ: ***Ompt* Encodes the *Escherichia-coli* Outer-Membrane Protease that Cleaves T7-RNA Polmerase**. *J Bacteriol* 1988, **170**(3):1245-1253

153. Moffatt BA, Studier FW: **T7 Lysozyme Inhibits Transcription by T7 Rna-Polymerase**. *Cell* 1987, **49**(2):221-227.

154. **QIAquick Gel Extraction Kit - Handbooks & Protocols** [http://www1.qiagen.com/literature/handbooks/literature.aspx?id=1000252]

155. Hanahan D: **Studies on Transformation of *Escherichia-Coli* with Plasmids**. *Journal of Molecular Biology* 1983, **166**(4):557-580.

156. **Library Efficiency® DH5α™ Competent Cells (Transformation Protocol)** [http://tools.invitrogen.com/content/sfs/manuals/18263012.pdf]

157. Tartof KD, Hobbs CA: *FOCUS® 9:2* 1987:12.

158. **QIAprep Spin Miniprep Kit - Handbooks & Protocols** [http://www1.qiagen.com/literature/handbooks/literature.aspx?id=1000248]

159. Grunberg-Manago M: **Messenger RNA stability and its role in control of gene expression in bacteria and phages**. *Annual Review of Genetics* 1999, **33**:193-227.

160. Miyada CG, Stoltzfus L, Wilcox G: **Regulation of the araC Gene of *Escherichia coli*: Catabolite Repression, Autoregulation, and Effect on araBAD Expression**. *Proc Natl Acad Sci USA* 1984, **81**:4120-4124.

161. Lee N: **Molecular Aspects of ara Regulation**. New York: Cold Spring Harbor Laboratory Press; 1980.

162. Lee N, Francklyn C, Hamilton EP: **Arabinose-Induced Binding of AraC Protein to aral2 Activates the araBAD Operon Promoter**. *Proc Natl Acad Sci USA* 1987, **84**:8814-8818.

163. Novy R, Drott D, Yaegaer K, Mierendorf R: *in**Nova**tions* 2001, **12**:1-3.

164. Baca AM, Hol WG: **Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in *Escherichia coli***. *Int J Parasitol* 2000, **30**:113-118.

165. Wood WB: **Host Specificity of DNA Produced by *Escherichia Coli* - Bacterial Mutations Affecting Restriction and Modification of DNA**. *Journal of Molecular Biology* 1966, **16**(1):118-133.

166. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP: **Structure of a Fibronectin Type-Iii Domain from Tenascin Phased by Mad Analysis of the Selenomethionyl Protein**. *Science* 1992, **258**(5084):987-991.

167. **NuPAGE® Technical Guide: General information and protocols for using the NuPAGE® electrophoresis system (Version E)** [http://resources.invitrogen.com/content/sfs/manuals/nupage_tech_man.pdf]

168. Laemmli UK: **Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4**. *Nature* 1970, **227**(5259):680-&.

169. **B-PER® II Bacterial Protein Extraction Reagent Instructions** [http://www.piercenet.com/files/0783lea-4.pdf]

170. **SuperSignal® West HisProbe™ Kit Instructions** [http://www.piercenet.com/files/0711as4.pdf]

171. Gasteiger E, et al.: **Protein Identification and Analysis Tools on the ExPASy Server**. *The Proteomic Protocols Handbook* 2005:571-607.

172. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: the proteomics server for in-depth protein knowledge and analysis**. *Nucleic Acids Research* 2003, **31**(13):3784-3788.

173. Jancarik J, Pufan R, Hong C, Kim SH, Kim R: **Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins**. *Acta Crystallogr D* 2004, **60**:1670-1673.

174. **PCT™ - Pre-Crystallization Test User Guide** [http://www.hamptonresearch.com/assets/products/attachments/0000000010-0000000164.pdf]

175. Cooper A, Johnson CM: **Microscopy, Optical Spectroscopy, and Macroscopic Techniques**, vol. 22. Totowa, NJ: Humana Press; 1994.

176. Wiseman T, Williston S, Brandts JF, Lin LN: **Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter**. *Analytical Biochemistry* 1989, **179**(1):131-137.

177. Whitmore L, Wallace BA: **DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data**. *Nucleic Acids Research* 2004, **32**:W668-W673.

178. Sreerama N, Woody RW: **Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set**. *Anal Biochem* 2000, **287**(2):252-260.

179. Schuck P: **Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling**. *Biophysical Journal* 2000, **78**(3):1606-1619.

180. Laue TM, Shah BD, Ridgeway TM, Pelletier SL: **Computer-aided interpretation of analytical sedimentation data for proteins. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science**. Cambridge, UK: Royal Society for Chemistry; 1992.

181. Schuck P: **On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation**. *Anal Biochem* 2003, **320**:104-124.

182. Hwang TL, Shaka AJ: **Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients**. *J Magn Reson Ser A* 1995, **112**(2):275-279.

183. Yang ZR, Thomson R, McNeil P, Esnouf RM: **RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins**. *Bioinformatics* 2005, **21**(16):3369-3376.

184. Thomson R, Esnouf R: **Prediction of natively disordered regions in proteins using a bio-basis function neural network**. *Lect Notes Comput Sc* 2004, **3177**:108-116.

185. Thomson R, Hodgman TC, Yang ZR, Doyle AK: **Characterizing proteolytic cleavage site activity using bio-basis function neural networks**. *Bioinformatics* 2003, **19**(14):1741-1747.

186. Yang ES, Schirch V: **Tight binding of pyridoxal 5 '-phosphate to recombinant *Escherichia coli* pyridoxine 5 '-phosphate oxidase**. *Archives of Biochemistry and Biophysics* 2000, **377**(1):109-114.

187. Tang CK, Jeffers CE, Nichols JC, Tu SC: **Flavin specificity and subunit interaction of Vibrio fischeri general NAD(P)H-flavin oxidoreductase FRG/FRase I**. *Archives of Biochemistry and Biophysics* 2001, **392**(1):110-116.

188. Grandori R, Khalifah P, Boice JA, Fairman R, Giovanielli K, Carey J: **Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins**. *Journal of Biological Chemistry* 1998, **273**(33):20960-20966.

189. Caldeira J, Palma PN, Regalla M, Lampreia J, Calvete J, Schafer W, Legall J, Moura I, Moura JJG: **Primary Sequence, Oxidation-Reduction Potentials and Tertiary-Structure Prediction of *Desulfovibrio-***

***Desulfuricans* Atcc-27774 Flavodoxin**. *Eur J Biochem* 1994, **220**(3):987-995.

190. Meissner B, Schleicher E, Weber S, Essen LO: **The dodecin from Thermus thermophilus, a bifunctional cofactor storage protein**. *Journal of Biological Chemistry* 2007, **282**(45):33142-33154.

191. Duurkens RH, Tol MB, Geertsma ER, Permentier HP, Slotboom DJ: **Flavin binding to the high affinity riboflavin transporter RibU**. *Journal of Biological Chemistry* 2007, **282**(14):10380-10386.

192. Woody RW: **Contributions of Tryptophan Side-Chains to the Far-Ultraviolet Circular-Dichroism of Proteins**. *Eur Biophys J Biophy* 1994, **23**(4):253-262.

193. Fielding L: **NMR methods for the determination of protein-ligand dissociation constants**. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2007, **51**(4):219-242.

194. Combet C, Blanchet C, Geourjon C, Deleage G: **NPS@: Network Protein Sequence Analysis**. *Trends in Biochemical Sciences* 2000, **25**(3):147-150.

195. Garnier J, Gibrat JF, Robson B: **GOR method for predicting protein secondary structure from amino acid sequence**. *Computer Methods for Macromolecular Sequence Analysis* 1996, **266**:540-553.

196. Bond CS: **Aline. University of Western Australia**. Unpublished work.

197. Corpet F: **Multiple Sequence Alignment with Hierarchical-Clustering**. *Nucleic Acids Research* 1988, **16**(22):10881-10890.

198. Gouet P, Courcelle E, Stuart DI, Metoz F: **ESPript: analysis of multiple sequence alignments in PostScript**. *Bioinformatics* 1999, **15**(4):305-308.

199. Gouet P, Robert X, Courcelle E: **ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins**. *Nucleic Acids Research* 2003, **31**(13):3320-3323.

200. (JCSG) JCfSG: **PDB ID: 2RE7, Crystal structure of General Stress Protein COG3871 (YP_263493.1) from *Psychrobacter arcticus* 273-4 at 2.50 A resolution**. In.; To be published.

201. Bairoch A, Bougueleret L, Altairac S, Amendolia V, Auchincloss A, Puy GA, Axelsen K, Baratin D, Blatter MC, Boeckmann B *et al*: **The Universal Protein Resource (UniProt)**. *Nucleic Acids Research* 2008, **36**:D190-D195.

202. Murzin AG, Brenner SE, Hubbard T, Chothia C: **Scop - a Structural Classification of Proteins Database for the Investigation of Sequences and Structures**. *Journal of Molecular Biology* 1995, **247**(4):536-540.

203. Emsley P, Cowtan K: **Coot: Model-Building Tools for Molecular Graphics**. *Acta Crystallographica Section D - Biological Crystallography* 2004, **60**:2126-2132.

204. Hendlich M, Rippmann F, Barnickel G, Hemm K, Aberer K: **RELIBase - An object-oriented comprehensive receptor-ligand database.** *Abstracts of Papers of the American Chemical Society* 1996, **211**:49-Cinf.

205. Boyd S: **Relibase**. *Chemistry World* 2007, **4**(11):67-68.

206. DeLano WL, Lam JW: **PyMOL: A communications tool for computational models**. *Abstracts of Papers of the American Chemical Society* 2005, **230**:U1371-U1372.

207. Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, Sussman JL, Stuart DI, Perrakis A: **Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG plus strategy**. *Acta Crystallogr D* 2005, **61**:1426-1431.

208. Zulauf M, Darcy A: **Light-Scattering of Proteins as a Criterion for Crystallization**. *Journal of Crystal Growth* 1992, **122**(1-4):102-106.

209. Churchich JE: **Brain Pyridoxine-5-Phosphate Oxidase - a Dimeric Enzyme Containing One Fmn Site**. *Eur J Biochem* 1984, **138**(2):327-332.

210. Lam HM, Winkler ME: **Characterization of the Complex Pdxh-Tyrs Operon of *Escherichia-Coli*-K-12 and Pleiotropic Phenotypes Caused by Pdxh Insertion Mutations**. *J Bacteriol* 1992, **174**(19):6033-6045.

211. Wada H, Snell EE: **The Enzymatic Oxidation of Pyridoxine and Pyridoxamine Phosphates**. *J Biol Chem* 1961, **236**(7):2089-2095.

212. Klabunde T, Hessler G: **Drug design strategies for targeting G-protein-coupled receptors**. *Chembiochem* 2002, **3**(10):929-944.

213. Lundstrom K: **Structural genomics for membrane proteins**. *Cell Mol Life Sci* 2006, **63**(22):2597-2607.

214. Ostermeier C, Michel H: **Crystallization of membrane proteins**. *Curr Opin Struct Biol* 1997, **7**(5):697-701.

215. Hunte C, Michel H, Carola H, Gebhard Von J, Hermann S: **Membrane Protein Crystallization**. In: *Membrane Protein Purification and Crystallization (Second Edition)*. San Diego: Academic Press; 2003: 143-160.

216. Elofsson A, von Heijne G: **Membrane protein structure: Prediction versus reality**. *Annual Review of Biochemistry* 2007, **76**:125-140.

217. Miroux B, Walker JE: **Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels**. *Journal of Molecular Biology* 1996, **260**(3):289-298.

218. Dumon-Seignovert L, Cariot G, Vuillard L: **The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3)**. *Protein Expres Purif* 2004, **37**(1):203-206.

219. Wiener MC: **A pedestrian guide to membrane protein crystallization**. *Methods* 2004, **34**(3):364-372.

220. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE *et al*: **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* 2000, **289**(5480):739-745.

221. Lavallie ER, McCoy JM: **Gene Fusion Expression Systems *Escherichia-Coli***. *Curr Opin Biotechnol* 1995, **6**(5):501-506.

222. Prive GG: **Detergents for the stabilization and crystallization of membrane proteins**. *Methods* 2007, **41**(4):388-397.

223. Byrne B, Jormakka M: **Solubilization and purification of membrane proteins**. Boca Raton: CRC; 2006.

224. Keyes MH, Gray DN, Kreh KE, Sanders CR: **Solubilizing detergents for membrane proteins**. San Diego: IUL; 2003.

225. Todd AE, Marsden RL, Thornton JM, Orengo CA: **Progress of structural genomics initiatives: An analysis of solved target structures**. *Journal of Molecular Biology* 2005, **348**(5):1235-1260.

226. Lundstrom K: **Structural genomics on membrane proteins: Mini review**. *Combinatorial Chemistry & High Throughput Screening* 2004, **7**(5):431-439.

227. Hofmann K, Stoffel W: **TMbase - A database of membrane spanning proteins segments**. *Biol Chem Hoppe-Seyler* 1993:374,166.

228. Hirokawa T, Boon-Chieng S, Mitaku S: **SOSUI: classification and secondary structure prediction system for membrane proteins**. *Bioinformatics* 1998, **14**(4):378-379.

229. Oxenoid K, Kim HJ, Jacob J, Sonnichsen FD, Sanders CR: **NMR Assignments for a Helical 40 kDa Membrane Protein**. *J Am Chem Soc* 2004, **126**(16):5048-5049.

230. Smith RL, O'Toole JF, Maguire ME, Sanders CR, 2nd: **Membrane topology of *Escherichia coli* diacylglycerol kinase**. *J Bacteriol* 1994, **176**(17):5459-5465.

231. Lau FW, Chen X, Bowie JU: **Active Sites of Diacylglycerol Kinase from *Escherichia coli* Are Shared between Subunits**. *Biochemistry* 1999, **38**(17):5521-5527.

232. Ghachi ME, Bouhss A, Blanot D, Mengin-Lecreulx D: **The bacA Gene of *Escherichia coli* Encodes an Undecaprenyl Pyrophosphate Phosphatase Activity**. *J Biol Chem* 2004, **279**(29):30106-30113.

233. Ghachi ME, Derbise A, Bouhss A, Mengin-Lecreulx D: **Identification of Multiple Genes Encoding Membrane Proteins with Undecaprenyl Pyrophosphate Phosphatase (UppP) Activity in *Escherichia coli***. *J Biol Chem* 2005, **280**(19):18689-18695.

234. Cain BD, Norton PJ, Eubanks W, Nick HS, Allen CM: **Amplification of the bacA gene confers bacitracin resistance to *Escherichia coli***. *J Bacteriol* 1993, **175**(12):3784-3789.