Imperial College London

BBSRC 20 Years of Pioneering Great British Bioscience

Structural and Functional Studies of the Als1 Adhesin from *Candida albicans*

Lisa Hale

Abstract

The dimorphic fungus *Candida albicans*, a harmless commensal in approximately 60% of healthy adults, is also an opportunistic pathogen in humans. Immuno-compromised individuals are mainly at risk, for example those under immunosuppressive therapies, HIV/AIDS patients, the elderly and early neonates. Diseases caused include invasive candidiasis, which has a 40-60% mortality rate. The majority of invasive candidiasis is caused by *C. albicans* forming biofilms on medical devices (e.g. catheters, heart valves) from which yeast cells disperse and enter the blood stream.

The adherence properties of *C. albicans* are crucial for colonisation and biofilm formation. An important adhesive factor in the early stages of colonisation is Als1 (Agglutinin-like Sequence protein 1). Another crucial virulence factor is hyphal-expressed Als3, which contributes to adhesion and invasion of host cells. These proteins belong to the Als family of glycoproteins (Als1-7 and Als9), all of which share similar domain organisation: an N-terminal adhesion region comprised of two immunoglobulin-like domains, plus central and C-terminal regions, which are highly glycosylated.

Previously, the N-terminus of Als9 had been shown to bind the flexible C-termini of host cell proteins as ligands. Due to their relevance in *C. albicans* adhesion, the N-terminal domains of Als1 and Als3 are ideal templates for ligand binding studies and the development of antifungal compounds. Shortened constructs Als1 and Als3 with improved solubility and structural properties (relative to the full length adhesins) were employed in this project. Here I describe the X-ray crystallographic structure of sNT-Als1 with a peptide ligand, the backbone resonance assignments by Nuclear Magnetic Resonance (NMR) of the ligand-free protein and ligand-binding experiments using Differential Scanning Fluorimetry (DSF). Comparison of the structure and ligand binding behavior of sNT-Als1 with other Als proteins will guide the design of therapeutic molecules that block the adhesive properties of these adhesins, and thus the formation of disease-causing biofilms by *C. albicans*.

Table of Contents

Abstract **1**

Declaration of Originality **13**

List of Abbreviations **14**

1. Introduction **17**

1.1. *Candida spp*. - Pathogenic Fungi **17**

1.1.1. Who is at Risk from Candidal Diseases? 17

1.1.2. Invasive Candidiasis 18

1.1.3. Oropharyngeal Candidiasis and Denture Stomatitis 19

1.1.4. Vulvovaginal Candidiasis and Balanitis 20

1.2. Pathogenesis of *Candida albicans*: Adhesion, Invasion & Cell Damage **21**

1.2.1. Adhesion 22

1.2.2. Invasion 22

1.2.3. Cell Damage 23

1.3. Pathogenesis Tricks **24**

1.3.1. Polymorphism 24

1.3.1.1. Yeast vs. Hypae 25

1.3.2. Enzyme Secretion and Nutrient Acquisition 25

1.3.2.1. Enzyme Secretion 25

1.3.2.2. Iron Acquisition 26

1.3.3. Biofilm Formation 26

1.3.3.1. Stages of Biofilm Formation 26

1.3.3.2. Mixed-Species Biofilms – Collaborating Pathogens 27

1.3.4. The Adhesive *C. albicans* Cell Wall 29

1.4. Medical Treatment – Past, Present and Future **31**

1.4.1. Current Treatments 31

1.4.2. Drug Resistance – Candida Fights Back 31

1.4.3. The Future of *Candida spp*. Treatment 32

1.5. The Als Glycoproteins **33**

1.5.1. The ALS Gene Family 33

1.5.2. The Adhesive Als Proteins 38

1.5.2.1. Sticky Studies on Als Protein Adhesiveness 38

1.5.2.2. Localisation of Als proteins in C. albicans 40

1.5.2.3. Diseases, Disease Models and ALS Gene Expression 42

1.5.2.4. Biofilms and Als Adhesins 43

1.5.2.5. Ferritin Adhesion with Als3 and Iron Acquisition 44

1.5.2.6. Induced Endocytosis 44

1.5.3. Medical Significance 45

1.6. Structural and Functional Studies of Als Proteins and Their Ligand-Binding Behaviour **46**

1.6.1. Initial Structural Studies and Bioinformatics Predictions 46

1.6.2. Previous Peptide Ligand Binding Specificity Studies 46

1.6.3. Structural and Biophysical Studies of Als Adhesins 47

1.7. Project Aims and Objectives **53**

2. Structural Biology Theory **54**

2.1. X-Ray Crystallography **54**

2.1.1. Protein Crystallization 54

2.1.2. Mathematical 'Artwork; – Crystals Up Close 56

2.1.2.1. The Unit Cell 56

2.1.2.2. The Seven Crystal Systems and 14 Bravais Lattices 58

2.1.2.3. The Miller Indices 59
2.1.3. X-ray Diffraction of Crystals 61
2.1.3.1. X-rays and X-ray Sources 61
2.1.3.2. The X-ray Diffraction Experiment 63
2.1.3.3. Bragg's Law 63
2.1.3.3. Reciprocal Space, Bragg's Law and the Unit Cell 64
2.1.3.4. The Diffraction Pattern 65
2.1.3.5. From Reciprocal Space to Real Space 66
2.1.4. Data Processing 67
2.1.4.1. Indexing and Integration of the Diffraction Pattern Reflections 67
2.1.4.2. Scaling and Truncation 67
2.1.4.3. Asymmetric Unit and Solvent Content Estimation with the Matthews Coefficient 68
2.1.5. Solving the Phase Problem with Molecular Replacement 68
2.1.5.1. Patterson Function Methods 68
2.1.5.1. Molecular Replacement 69
2.1.6. Model Refinement 71
2.1.6.1. Model Bias and R_{free} 72
2.1.6.2. Least Squares and Maximum Likelihood Refinement 72
2.1.6.3. Rigid Body Refinement 72
2.1.6.4. Non-Crystallographic Symmetry (NCS) Refinement 73
2.1.6.5. TLS Group Refinement 73
2.2. Nuclear Magnetic Resonance (NMR) Spectroscopy 74
2.2.1. The NMR Spectrometer 74
2.2.2. NMR Fundamentals 75
2.2.2.1. Nuclear Spin and Larmor Precession 75
2.2.2.2. The Boltzmann Distribution 76
2.2.2.3. Pulse NMR to the NMR Spectrum 77
2.2.2.4. The Chemical Shift 78
2.2.2.5. The Nuclear Overhauser Effect - NOE 79
2.2.3. Protein NMR Experiments 80
2.2.3.1. The NMR Sample 80
2.2.3.2. 1D ^{1}H NMR Experiments 80
2.2.3.3. 2D NMR Experiments and TROSY 81
2.2.2.4. Protein Assignment of Unlabelled Proteins with 2D NMR Spectroscopy 82
2.2.2.5. Protein Assignment of Isotopically Labelled Proteins with 2D and 3D NMR Spectroscopic Techniques 84

3. Materials and Methods 88
3.1. Cloning 88
3.1.1. Cloning of sNT-Als1 and sNT-Als3 into pET32 Plasmids 88
3.1.2. Cloning of sNT-Als1 and sNT-Als3 into pRSETa Plasmids 88
3.1.2.1. Primer Design 88
3.1.2.2. Polymerase Chain Reaction (PCR) 89
3.1.2.3. DNA Agarose Gels 89
3.1.2.4. DNA Extraction and Purification 89
3.1.2.5. Digestion of pRSETa Plasmid 90
3.1.2.6. DNA Ligation 90
3.1.2.7. The pRSETa Plasmid 90
3.1.2.8. Transformation of Plasmid 91
3.1.2.9. Plasmid Amplification and Purification 91
3.1.2.10. DNA Sequence Analysis 91
3.1.2. Cloning of sNT-Als3 into pET46 Ek/LIC Plasmid 92
3.1.2.1. Primer Design 92
3.1.2.2. DNA Ligation 92

3.1.2.3. The pET46 Plasmid and Amplification 93
3.2. Expression of sNT-Als1 and sNT-Als3 Proteins 93
3.2.1. Shuffle Cells 93
3.2.2. Protein Expression in LB Media 93
3.2.3. Protein Expression in ^{15}N-Minimal Media 94
3.2.4. Protein Expression in ^{15}N-^{13}C-D_2O Minimal Media 94
3.3. Protein Purification 94
3.3.1. Cell Harvesting and Lysis 94
3.3.2. Nickel-IMAC Purification 94
3.3.3. Factor Xa Protease Digestion 94
3.3.4. Size Exclusion Chromatography 95
3.3.5. SDS-PAGE 95
3.3.6. Protein Concentration Determination via UV Spectroscopy 95
3.4. NMR Spectroscopy 96
3.4.1. NMR Sample Preparation 96
3.4.2. Data Acquisition 96
3.5. X-ray Crystallography 97
3.5.1. Pre-Crystallisation Screens 97
3.5.2. Crystallisation Screens 97
3.6. Differential Scanning Fluorimetry 97

4. Production of sNT-Als Constructs for Crystallographic and NMR Spectroscopic Studies. 98
4.1. Production of sNT-Als1 for Crystallography 98
4.1.1. Expression Systems and Constructs for NT-Als1 98
4.1.2. sNT-Als1 Sample Check with One-Dimensional ^{1}H NMR Spectroscopy 100
4.2. sNT-Als1 Crystallography 102
4.2.1. Initial Crystallisation Screens of sNT-Als1 102
4.2.2. Co-Crystallisation of sNT-Als1 with Hepta-Threonine Ligand 102
4.2.3. The Addition of Deuterium Oxide to the Co-Crystallisation of sNT-Als1 with Hepta-Threonine Peptide 103
4.2.4. sNT-Als1 Crystal Optimisation 104
4.3. sNT-Als1 Expression and Purification for Backbone Assignment and HSQC-Titrations using NMR Spectroscopy 106
4.3.1. ^{15}N-Labelled and Triple-Labelled sNT-Als1 Expression and Purification 106
4.3.2. Optimisation of Sample Conditions 107
4.4. sNT-Als3 Expression and Purification for Backbone Assignment Studies with NMR Spectroscopy 111
4.4.1. ^{15}N-Labelled sNT-Als3 Expression and Purification 111
4.4.2. Triple-Labelled sNT-Als3 Expression and Purification 114
4.4.3. Further sNT-Als3 Expression and Purification Experiments 116

5. Structural Determination of sNT-lAls1 with a Hepta-Threonine Peptide 119
5.1. From Data to Structure 119
5.1.1. X-Ray Diffraction Data Collection 119
5.1.2. Data Processing 120
5.1.3. Structural Determination 122
5.1.4. Structural Refinement and Model Building 122
5.1.5. Model Validation 124
5.2. The Crystal Structure of SNT-Als1 with Hepta-T Peptide 136
5.2.1. Overall Structure and Topology of sNT-Als1 136
5.2.2. Crystal Packing 139
5.2.3. Crystal Contacts 140
5.2.4. The Peptide Binding Cavity (PBC) 143

5.2.5. Interactions of sNT-Als1 PBC with Hepta-T 144
5.2.5.1. Residues that Interact Directly with the Hepta-T Peptide 144
5.2.5.2. Hydrophobic and Water Molecule Network Interactions of sNT-Als1 PBC Residues 146
5.3. The PBC Residue Differences of sNT-Als1 with sNT-Als3 149
5.4. PBC Residue Differences of sNT-Als1 and NT-Als9.2 156
5.5. Overall Summary of PBC Comparisons between sNT-Als1, sNT-Als3 and NT-ALS9.2 160

6. sNT-Als1 Backbone Assignment with NMR Spectroscopy 163
6.1 – Background 163
6.2 – The $^{2}H^{15}N^{13}C$ Triple Labelled sNT-Als1 Sample. 164
6.3 – Backbone Assignment of sNT-Als1 with NMR Spectroscopy 166
6.3.1. NMR Backbone Assignment Experiments 166
6.3.2 Backbone Assignment Data 167
6.4.2. Manual Assignments 168
6.4.3. Assignment Results 170

7. Ligand Binding Studies of sNT-Als1 174
7.1. Differential Scanning Fluorimetry 174
7.2. Peptide Ligand Screening using DSF 175
7.3. The DSF Peptide Ligand-Binding Study in a Structural Context 182
7.4. Small Drug-Like Molecules Ligand-Binding Screens 186
7.5. Summary 189

8. Discussion 190
8.1. Differences between the Adhesion PBCs of Als1, Als3 and Als9.2. 190
8.2. A functional relationship between Als proteins and secreted aspartyl proteinases (SAPs)? 190
8.3. Recognition of cytoplasmic C-termini by Als proteins 191
8.4. Future studies and Drug Design 191

9. References 193

10. Acknowledgement 216

11. Appendix 217

List of Figures

1. Introduction 17
Figure 1.1: *C. albicans* epidemiology. 17
Figure 1.2: Invasive candidiasis risk factors. 19
Figure 1.3: *Lactobacillus spp.* communities protect vaginal mucosal epithelial cells from *C. albicans* colonization. 20
Figure 1.4: Pathogenesis of *C. albicans* occurs in three stages, adherence, invasion and cell damage. 21
Figure 1.5: Induced endocytosis. 22
Figure 1.6: The different morphologies of *C. albicans*, including yeast cells, pseudohyphae and hyphae 24
Figure 1.7: The different stages of *C. albicans* biofilm formation. 27
Figure 1.8: The process of mixed-species biofilm formation of *C. albicans* and *S. gordonii.* 28
Figure 1.9: Cartoon representation of the *C. albicans* cell wall 29
Figure 1.10: The general domain organisation and function of the Als proteins. 34
Figure 1.11: The ALS family of eight agglutinin-like sequence genes. 36
Figure 1.12: Phylogenetic analysis of the ALS gene family. 37
Figure 1.13: Localisation of Als1, Als3 and Als4 proteins on the *C. albicans* cell surface. 41
Figure 1.14: Crystal structures of NT-Als9.2 and sNT-Als3 with peptide ligands. 47
Figure 1.15: The invariant K59 in NT-Als9.2 and sNT-Als3 Crystal Structures. 48
Figure 1.16: The PBC water molecules networks in both NT-Als9.2/Fib-γ-CT and sNT-Als3/Hepta-T Crystal structures 49
Figure 1.17: The NT-Als9.2/Fib-γ-CT Interaction Up Close. 50
Figure 1.18: The sNT-Als3/Hepta-T Interaction Up Close. 50
Figure 1.19: The Interaction of NT-Als9.2 G299W Mutant with its own Flexible C-terminus. 52

2. Structural Biology Theory 54
Figure 2.1: Vapour diffusion crystallisation techniques hanging drop and sitting drop. 54
Figure 2.2: The zone-diagram relationship between protein solubility and precipitant concentration. 55
Figure 2.3: The unit cell - a building block of the crystal lattice. 56
Figure 2.4: The dimensions of a unit cell 57
Figure 2.5: The Miller indices. 59
Figure 2.6: 2D representations of Miller Index Planes. 60
Figure 2.7: 3D representation of the Miller index (111). 60
Figure 2.8: 3D representation of an example of a negative Miller index (0-10). 61
Figure 2.9: X-rays 62
Figure 2.10: Experimental set-up for the X-ray diffraction of crystals. 63
Figure 2.11: Diagram representing Bragg's Law. 64
Figure 2.12: The Ewald Sphere 65
Figure 2.13: Example of a Patterson Map of a Molecule. 69
Figure 2.14: Eulerian Angles of Rotation used for the Rotation of the Search Model. 70
Figure 2.15: The Interior of an NMR Spectrometer. 74
Figure 2.16: Spin ½ Particles, energy levels and Larmor Precession. 75
Figure 2.17: Energy difference between the two spin states. 77
Figure 2.18: Phase coherence, net magnetization and the FID 78
Figure 2.19: A 1D ^{1}H protein spectrum. 81
Figure 2.20: The overall layout of a 2D-COSY spectrum 82

Figure 2.21: The overall layout of a 2D-NOESY spectrum. 83
Figure 2.22: The Sequential Assignment of Residues using COSY and NOESY Combinations. 83
Figure 2.23: Example of a 2D-TOCSY spectrum. 84
Figure 2.24: Example of a ^{1}H-^{15}N HSQC spectrum. 85
Figure 2.24: Magnetization transfer between nuclei in backbone assignment experiments. 87

3. Materials and Methods 88
Figure 3.1: Diagram of the pRSETa plasmid. 91
Figure 3.2: Diagram of the pET46 Ek/LIC plasmid. 93

4. Production of sNT-Als Constructs for Crystallographic and NMR Spectroscopic Studies. 98
Figure 4.1: SDS-PAGE gel of samples from the induced expression and Nickel IMAC purification of the sNT-Als1-thioredoxin fusion protein, plus the Factor Xa cleavage of sNT-Als1 from thioredoxin. 98
Figure 4.2: SDS-PAGE gels and a size exclusion chromatogram. 99
Figure 4.3: (1D) ^{1}H-NMR Spectrum of sNT-Als1. 100
Figure 4.4: 1D ^{1}H spectral comparisons of ring-current shifted methyl peaks between NT-Als1 and sNT-Als1. 101
Figure 4.5: ^{1}H-^{15}N-HSQC-TROSY spectrum of sNT-Als1. 101
Figure 4.6: A comparison of the sNT-Als1 negative control melting curve with the melting curve of the Hepta-T peptide trial. 102
Figure 4.7: Crystallisation plates of sNT-Als1 protein or in complex with Hepta-T peptide. 103
Figure 4.8: Comparison of crystals from optimising crystallisation conditions in ICL11-F11. 104
Figure 4.9: Comparison of the crystals produced from optimisation plate screens around ICL11-F11 conditions. 105
Figure 4.10: SDS-PAGE gels of labelled sNT-Als1 expression and purification. 106
Figure 4.11: SDS-PAGE Gels of triple labelled sNT-Als1 expression and purification.. 107
Figure 4.12. Comparisons of 1D ^{1}H-NMR spectra for sample buffer optimization. 108
Figure 4.13: Comparison of 1D 1H-NMR spectra of sNT-Als1 for pH optimization. 109
Figure 4.14: Comparison of 1D ^{1}H-NMR spectra of sNT-Als1 before and after incubation with 4M urea. 110
Figure 4.15: SDS-PAGE Gel of sNT-Als3 expression and purification using a pET46 plasmid. 111
Figure 4.16: SDS-PAGE gels for expression and purification of sNT-Als3-TRX using pET32 plasmid. 112
Figure 4.17: Checking the sNT-Als3 construct with ^{1}H-NMR spectroscopy. 113
Figure 4.18: Checking the sNT-Als3 construct with ^{1}H-^{15}N-TROSY-HSQC-NMR spectroscopy. 113
Figure 4.19: SDS-PAGE gels of sNT-Als3-TRX expression and purification tests from D_2O-minimal media. 114
Figure 4.20: sNT-Als3 and Sample Contamination Problems. 115
Figure 4.21: SDS-PAGE gels of size exclusion chromatography optimisation trials. 116
Figure 4.22: SDS-PAGE gels from sNT-Als3 low temperature expression and purification. 117

5. Structural Determination of sNT-Als1 with a Hepta-Threonine Peptide 119
Figure 5.1. Diffraction image from an sNT-Als1/Hepta-T crystal. 120
Figure 5.2: An example of an $2F_{obs}$-F_{calc} (Electron Density Map) of the sNT-Als1/Hepta-T crystal structure and Model Building. 123
Figure 5.3: Main Chain Parameters (PROCHECK). 125

Figure 5.4: Side Chain Parameters (PROCHECK)..... 127
Figure 5.5: Ramachandran Plots from MOLPROBITY Validation Tool..... 129
Figure 5.6: Main chain Ramachandran Plots from PROCHECK (Lasowski et al 1993) analysis of the sNT-Als1/Hepta-T Crystal structure..... 130
Figure 5.7: 2Fo-Fc Map Comparison with the PDB validation server check. 131
Figure 5.8: The alignment of sNT-/Als1 Chains A and C, and Hepta-T Chains B and D, shown as a Cα Trace..... 132
Figure 5.9: Charts for Average B-factor values calculated for each residue of sNT-Als1 molecules Chain A and Chain C.....133
Figure 5.10: Locations of residues with high average B-factors on cartoon representations of sNT-Als1 molecules Chain A and Chain C..... 134
Figure 5.11: The Average B-factors for each residue of Hepta-T Peptide Chains B and D, alongside Stick Representations of the Hepta-T Peptide Chains, coloured according to a B-factor scale. 135
Figure 5.12: Simplified representation of the crystal structure of an sNT-Als1 molecule (Chain A) bound to the Hepta-T peptide ligand (Chain B), from two stereo views..... 136
Figure 5.13: Individually coloured secondary structure elements in an sNT-Als1 cartoon diagram from two stereo views and the labelling of secondary structural elements within a topology diagram..... 137
Figure 5.14: The primary sequence of sNT-Als1 alongside the labelled secondary structure elements..... 138
Figure 5.15: The Overall Surface Electrostatic Profile of sNT-Als1 (Chain A) from Two Stereo Views. 138
Figure 5.16: The Asymmetric Unit of the sNT-Als1/Hepta-T Crystal Structure. 139
Figure 5.17: Crystal Packing Arrangement of sNT-Als1/Hepta-T asymmetric units within the Orthorhombic Unit Cell from Two Stereo Views..... 140
Figure 5.18: Ligplot+ and PyMOL Representations of Interactions between sNT-Als1 molecules Chain and Chain C that pinpoint potential crystallography contacts. 141
Figure 5.19: Hydrogen-bond crystal contact between N250 of sNT-Als1 Chain A and T1 of the Hepta-T Peptide in the sNT-Als1 Chain C PBC, predicted with Ligplot+ and Depicted in PyMOL..... 142
Figure 5.20: Cartoon and Surface Representations sNT-Als1 PBC with the Hepta-T peptide. 143
Figure 5.21: Electrostatic profile of the sNT-Als1 PBC without Hepta-T and Representations of the virtual 'grid' used to estimate PBC volume..... 144
Figure 5.22: Cartoon and Stick Figure of the sNT-Als1 PBC residues that make direct Hydrogen-Bond Contacts with the Hepta-T Ligand. 145
Figure 5.23: Residues of the sNT-Als1 PBC that interact with the Hepta-T peptide..... 146
Figure 5.24: Cartoon and Stick Figure of sNT-Als1 PBC residues that interact with the Hepta-T peptide through water-molecule-mediated H-bonds..... 147
Figure 5.25: Alignment of sNT-Als1 molecule Chain A with the sNT-Als3 structure from two stereo views..... 149
Figure 5.26: The sNT-Als1 PBC secondary structure elements and residues that form direct H-bonds with the Hepta-T aligned with those of the sNT-Als3 PBC..... 151
Figure 5.27: A detailed view of sNT-Als3 PBC residues that interact with the Hepta-T peptide. 152
Figure 5.28: A detailed view of sNT-Als1 PBC residues that interact with the Hepta-T peptide. 153
Figure 5.29: A1-B1 loop differences between sNT-Als1 and sNT-Als3..... 154
Figure 5.30: Comparison of the Surface representations of A1-B1 loop residues at positions 20 and 22 in sNT-Als1 and sNT-Als3..... 155
Figure 5.31: Comparison of the Surface Representations of the A2 Strand residue at amino acid position 168 in sNT-Als1 and sNT-Als3..... 155

Figure 5.32: Alignment of sNT-Als1 molecule Chain A with the NT-Als9.2 structure from two stereo views. 156
Figure 5.33: The sNT-Als1 PBC secondary structure elements and residues that form direct H-bonds with the Hepta-T ligand aligned with those of the NT-Als9.2 residues that interact with fibrinogen-gamma and a Cartoon and Stick representation of the A1-B1 Loop of NT-Als9.2. 158
Figure 5.34: A detailed view of NT-Als9.2 PBC residues that contact the fibrinogen-γ peptide. 159
Figure 5.35: Volume Estimations of the PBCs of sNT-Als1, sNT-Als3 and NT-Als9.2… 160
Figure 5.36: Average B-factors of Peptide Ligand Residue Positions Whilst Bound to the PBC's of sNT-Als1, sNT-Als3 and NT-Als9.2, Alongside a B-factor Scale. 161
Figure 5.37: Comparison of the Surface Electrostatic Charges between the PBCs of snT-Als1, sNT-Als3 and NT-Als9.2. 162

6. sNT-Als1 Backbone Assignment with NMR Spectroscopy 163
Figure 6.1: ^{1}H-^{15}N-TROSY-HSQC titration of NT-Als1 with Hepta-T peptide ligand and how the chemical shift changes of the assigned peaks correspond to residues on the NT-Als1 protein structure. 164
Figure 6.2: ^{1}H 1D spectrum of partially deuterated ^{13}C-^{15}N-sNT-Als1 165
Figure 6.3: A comparison of the 1D ^{1}H one-dimensional spectra of unlabelled sNT-Als1 sample with that of the triple-labelled sNT-Als1. 166
Figure 6.4: A typical strip plot of TROSY-HNCACB (i-1 & i) and TROSY-HN(CO)CACB (i-1) data 167
Figure 6.5: The $^{1}H^{15}N$-NOESY spectrum of sNT-Als1. 169
Figure 6.6: An overlay of the ^{1}H-^{15}N TROSY-HSQC spectrum of deuterated sNT-Als1 with that of non-deuterated NT-Als1. 170
Figure 6.7: ^{1}H-^{15}N-TROSY-HSQC of sNT-Als1 with approximately 55% of assigned peaks. 171
Figure 6.8: The result of the backbone assignment experiments as displayed on the sNT-Als1 topology diagram. 172
Figure 6.9: The result of the backbone assignment experiments as displayed on the sNT-Als1 cartoon diagram. 173

7. Ligand Binding Studies of sNT-Als1 174
Figure 7.1: Plotting fluorescence intensity against temperature yields a sigmoidal curve that represents the melting profile of a protein. 175
Figure 7.2: A comparison of the sNT-Als1 negative control melting curve (N = 2) with the melting curve of the BSA-CT peptide trial (N = 2). 179
Figure 7.3: A comparison of the sNT-Als1 negative control melting curve (N = 2) with the melting curve of the Bα1-CAS-CT peptide trial (N = 3). 180
Figure 7.4: Theoretical representation of the BSA-CT in the sNT-Als1 PBC. 182
Figure 7.5: Theoretical representation of a potential steric clash between the aspartate residue at peptide position 1 of SspB-CT peptide in the sNT-Als1 PBC. 183
Figure 7.6: Theoretical representation of a potential steric clash between the glutamate residue at peptide position 2 of SspB-CT peptide in the sNT-Als1 PBC. 184
Figure 7.7: Theoretical representation of a potential steric clash between the lysine residue at peptide position 3 of SspB-CT peptide in the sNT-Als1 PBC. 184
Figure 7.8: Theoretical representation of a potential steric clash between the arginine residue at peptide position 4 of SspB-CT peptide in the sNT-Als1 PBC. 185
Figure 7.9: A comparison of the sNT-Als1 negative control melting curve (N = 3) with the melting curve of the 3,5-SAL drug molecule trial (N = 1). 188

List of Tables

1. Introduction 17
Table 1.1: Percentage identity comparisons between each of the full-length Als proteins. 35
Table 1.2: Percentage identity comparisons between the amino acid sequences of the N-terminal adhesion domains of all the Als proteins. 35
Table 1.3: Percentage identity comparisons between the amino acid sequences of the central and C-terminal domains of all the Als proteins. 36
Table 1.4: The detection of individual ALS genes from clinical samples or *C. albicans* disease models. 42
Table 1.5: The key residues of NT-Als9.2 and sNT-Als3 that are predicted to contribute to their interactions with their respective peptide ligands via either direct hydrogen bonds or via water molecules 51

2. Structural Biology Theory 54
Table 2.1: The 7 Crystal Systems and 14 Bravais lattices. 58

3. Materials and Methods 88
Table 3.1: Forward and Reverse Primers. 88
Table 3.2: Volumes and concentrations of PCR reagents used to amplify sNT-Als1 and sNT-Als3 DNA sequences. 89
Table 3.3: Times and temperatures of PCR reaction steps performed on the thermocycler. 89
Table 3.4: Reagents and concentrations for the double digestion of pRSETa plasmid with Restriction Enzymes BamHI and HindIII. 90
Table 3.5: Primers designed for the amplification of DNA insert coding for sNT-Als3 for cloning into pET46 Ek/LIC plasmid. 92
Table 3.6: The reagents for T4 DNA polymerase treatment of the PCR product DNA insert of sNT-Als3. 92
Table 3.7: The plates used for the large scale sNT-Als1 crystallisation screen. 97

5. Structural Determination of sNT-lAls1 with a Hepta-Threonine Peptide 119
Table 5.1: Summary of Overall Data Collection, Processing and Scaling Statistics for sNT-Als1/Hepta-T crystal. 121
Table 5.2: Summary of Matthews Coefficient values for the sNT-Als1/Hepta-T Crystal. 121
Table 5.3: Molecular Replacement Statistics for sNT-Als1/Hepta-T Crystal (Z-scores) 122
Table 5.4: Refinement Statistics for Final Refinement of sNT-Als1/Hepta-T Model 124
Table 5.5: Main Chain Parameters (PROCHECK) Statistics for sNT-Als1/Hepta-T Structural Model 126
Table 5.6: Side Chain Parameters (PROCHECK) Statistics for sNT-Als1/Hepta-T Structural Model 128
Table 5.7: MOLPROBITY Ramachandran Plot Statistics for sNT-Als1/Hepta-T Crystal 129
Table 5.8: MOLPROBITY Ramachandran Plot Statistics for sNT-Als1/Hepta-T Crystal 130
Table 5.9: The number of atoms and average B-factors for each chain, as calculated using the program Baverage (CCP4). 133
Table 5.10: PBC Residues of sNT-Als1 and sNT-Als3 that have been identified by Ligplot+ to interact with the Hepta-T ligand 150

Table 5.11: PBC Residues of sNT-Als1 and NT-Als9.2 that have been identified to interact with either the Hepta-T ligand, or Fib-γ C-terminal peptide, respectively. 157
Table 5.12: Estimated volumes of sNT-Als1, sNT-Als3 and NT-Als9.2 PBCs … 160

7. Ligand Binding Studies of sNT-Als1 174
Table 7.1: Peptides used in a DSF screen to test the probability of their binding to sNT-Als1……..178
Table 7.2: Summary of drug molecules used in a DSF screen to test the probability of their interaction with sNT-Als1. 187
Table 7.3: A repeated DSF trial of 3,5-SAL with sNT-Als1. 188

List of Equations

2. Structural Biology Theory 54
Equation 2.1: The Energy Emitted by X-rays. 61
Equation 2.2: The Bragg Equation. 63
Equation 2.3: The Structure Factor Equation. 66
Equation 2.4: The Fourier Transform Equation to Derive Electron Density from Amplitude Information from the Bragg Reflections and Calculated Phases. 66
Equation 2.5: The Fourier Transform Equation Re-Arranged to Show the Relationship Between Phase Information and the Structure Factor. 66
Equation 2.6: The Equation to Derive R_{sym}. 67
Equation 2.7: The Matthews Coefficient Equation. 68
Equation 2.8: The Patterson Function Equation. 69
Equation 2.9: The Rotation Function Equation. 70
Equation 2.10: The Translation Function Equation. 71
Equation 2.11: The Correlation Coefficient Equation. 71
Equation 2.12: The Correlation Coefficient Equation. 71
Equation 2.13: The Equation to Derive R_{free}. 72
Equation 2.14: The Least Squares Equation. 72
Equation 2.15: The Nuclear Angular Frequency Equation. 75
Equation 2.16: The Larmor Frequency Equation. 76
Equation 2.17: The Energy Difference Equation. 76
Equation 2.18: The Boltzmann Distribution Equation. 77

7. Ligand Binding Studies of sNT-Als1 174
Equation 7.1: The Equation For the Fitting of the Sigmoidal Melting Curve of a Protein According to the Relationship Between Fluorescence Intensity and Temperature. 174
Equation 7.2: The Standard Deviation Equation for each T_m and ΔT_m Value. 177

Declaration of Originality

I hereby declare that the results presented in this thesis are from my own work except where specifically indicated in the text. Information derived from others has been referenced.

Signed…….

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

List of Abbreviations

ALS – Agglutinin Like Sequence

BSA – Bovine serum albumin

ClfA – Clumping factor A

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

DTT – Dithiothreitol

GPI – Glycophosphatidylinositol

HUVEC – Human umbilical chord vein endothelial cells

HSQC – Heteronuclear signle quantum coherence

Hwp1 – Hyphal wall protein 1

Ig – Immunoglobulin

INT – Integrin like protein

IPTG – Isopropyl-β-D-thio-galactopyranoside

LB – Lysogeny broth

LIC – Ligation independent cloning

MSCRAMM – Microbial surface components recognizing adhesive matrix molecules

MWCO – Molecular weight cut off

NMR – Nuclear magnetic resonance

NT-Als1 – N-terminal domain of Als1

NT-Als3 – N-terminal domain of Als3

NT-Als9.2. – N-terminal domain of Als9.2

NOE – Nuclear Overhauser effect

OD – Optical density

PCR – Polymerase chain reaction

PEG – Polyethylene glycol

Hepta-T – Hepta-Threonine

ppm – parts per million

PR – Pseudoresidue

PHYRE – Protein Homology/analogy Recognition Engine

RF – Radio frequency

RHE – Reconstituted human endothelial cells

RMSD – Root mean square deviation

SAP – Secreted aspartyl protease

SDS PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

VVC – vulvovaginal candidiasis

WT – Wild-type

1. Introduction

1.1. *Candida spp.* - Pathogenic Fungi

1.1.1. Who is at Risk from Candidal Diseases?

Fungi from the genus *Candida* are responsible for an extensive range of diseases, with *Candida albicans* being the most prominent (Fig 1.1). As a commensal, *C. albicans* colonises 30-60% of healthy adults and can adapt to different biological environments such as the human skin and gastrointestinal, respiratory, vaginal and oral membranes (Karkowska-Kuleta et al 2009, Kumamoto et al 2011).

Candida spp. can cause superficial infections such as oropharyngeal candidiasis (OPC), vulvovaginal candidiasis (VVC) (also known as thrush) and a nail plate infection known as *Candida* onchomycosis (Shirwaiker et al 2009, Williams et al 2013). *Candida spp.* can also infiltrate the human bloodstream via contaminated hospital equipment such as intravenous catheters, shunts and pacemakers and cause the life-threatening infection called disseminated (or invasive) candidiasis, also known as candidemia (Douglas 2003, Lim et al 2012, Delaloye & Calandra 2014).

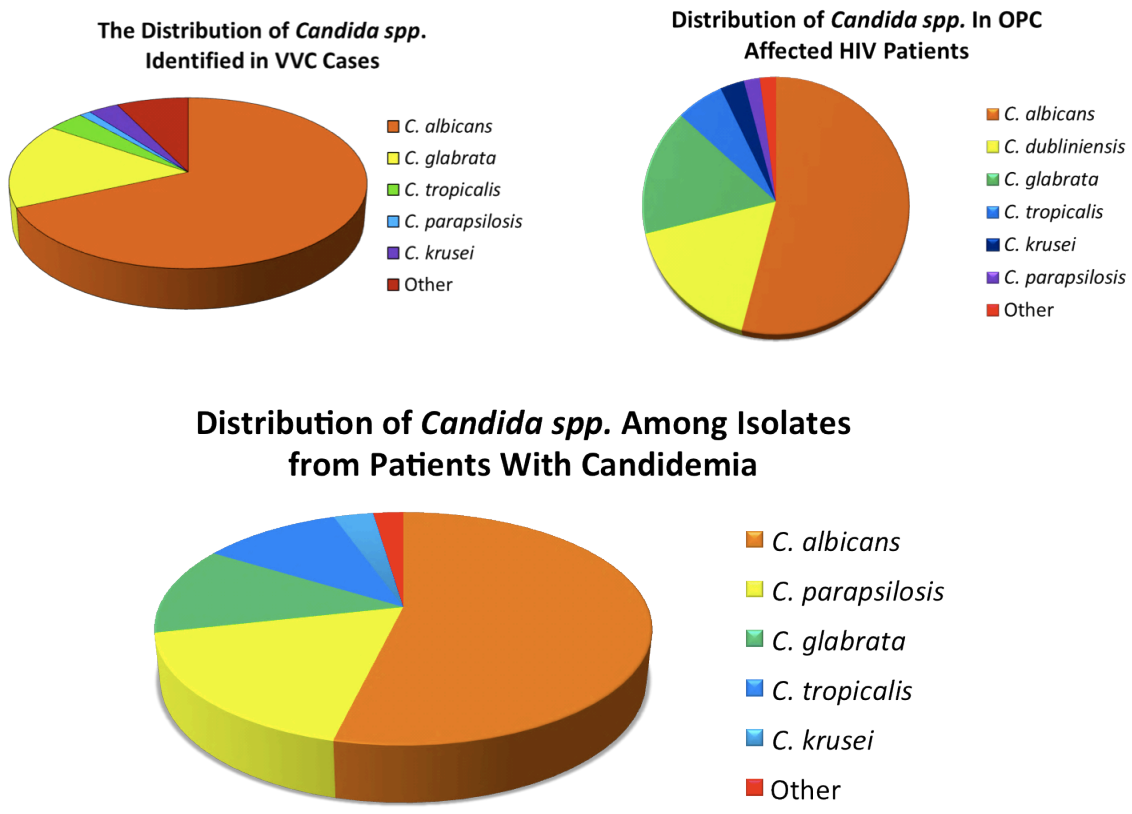


Figure 1.1: *C. albicans* epidemiology.
C. albicans is the most frequent cause of diseases caused by Candida spp., as shown by results from A) a worldwide study of the distribution of *Candida spp*. associated with VVC infections in women (Achkar & Fries 2010), B) a clinical study of *Candida spp*. associated with OPC infections of HIV patients (Thompson et al 2010) and C) a review of the distribution of *Candida spp*. in candidaemic blood isolates from various studies between 1996 to 2009 (Falagas et al 2010).

As opportunistic pathogens, *Candida spp.* take advantage of disruptions to the immune system, microbial flora and skin/mucosal barriers (Yan et al 2013). Many at risk are patients taking immunosuppressive drugs (for instance, following surgery) or suffering from conditions such as acquired immunodeficiency syndrome (AIDS), lymphoma and neutropenia (Hazen 1995). They also

affect diabetes mellitus patients, people with gastro-intestinal (GI) tract inflammation and those taking long-term antibiotic therapy, corticosteroids and chemotherapy (Hazen 1995, Kumamoto 2011, Delaloye & Calandra 2014).

1.1.2. Invasive Candidiasis

Invasive candidiasis has 40-75% mortality rate in immuno-compromised individuals, and affects more than 90,000 people per year in the USA alone. There has been a dramatic increase in patients affected by candidemia over the past 11 years, and disconcertingly alongside a rise in *Candida spp*. resistant to current antifungal drugs (Pfaller 2012). *C. albicans* is the most frequent causal species of candidemia worldwide (Fig 1.1), although there has been a rise in non-*C.albicans* species in recent years (Falagas et al 2010, Delaloye & Calandra 2014).

Risk factors for contracting invasive candidiasis are summarised in Figure 1.2. Whilst *C. albicans* adhesion and colonization of skin and mucosal layers of patients can predict invasive candidiasis, approximately 65-90% of cases of diagnosed candidemia, are linked to the use of central venous catheters (Pfaller & Diekema 2007). Utilizing the blood stream, *Candida* cells can reach and colonize vital organs including the brain, kidneys, lungs and liver (Grubb et al 2008). If it reaches the heart fatal endocarditis can occur (Baddley et al 2008).

Although a rare disease, *Candida* endocarditis can be caused from invasive candidiasis, but also from biofilms on pacemakers and on the damaged vascular endothelium of native heart valves in patients with heart disease. Disseminated Candida cells can colonize thrombi in the blood vessels of those with atherosclerosis. Candida endocarditis is often fatal, with a reported mortality rate of 80% (Douglas 2003, Baddley et al 2008).

Approximately 5-9% of patients with invasive candidiasis can be affected by *Candida spp.* colonization of meninges in the central nervous system, causing *Candida* meningitis. This condition is also seen in almost a quarter of neonates with low birth weight and results in a mortality rate of approximately 70% (Fernandez et al 2000, Faix & Chapman 2003, Ancalle et al 2010). *C. albicans* can infect the brain from the bloodstream via passing through the blood-brain barrier or by contamination of equipment in craniotomy operations (Jong et al 2001).

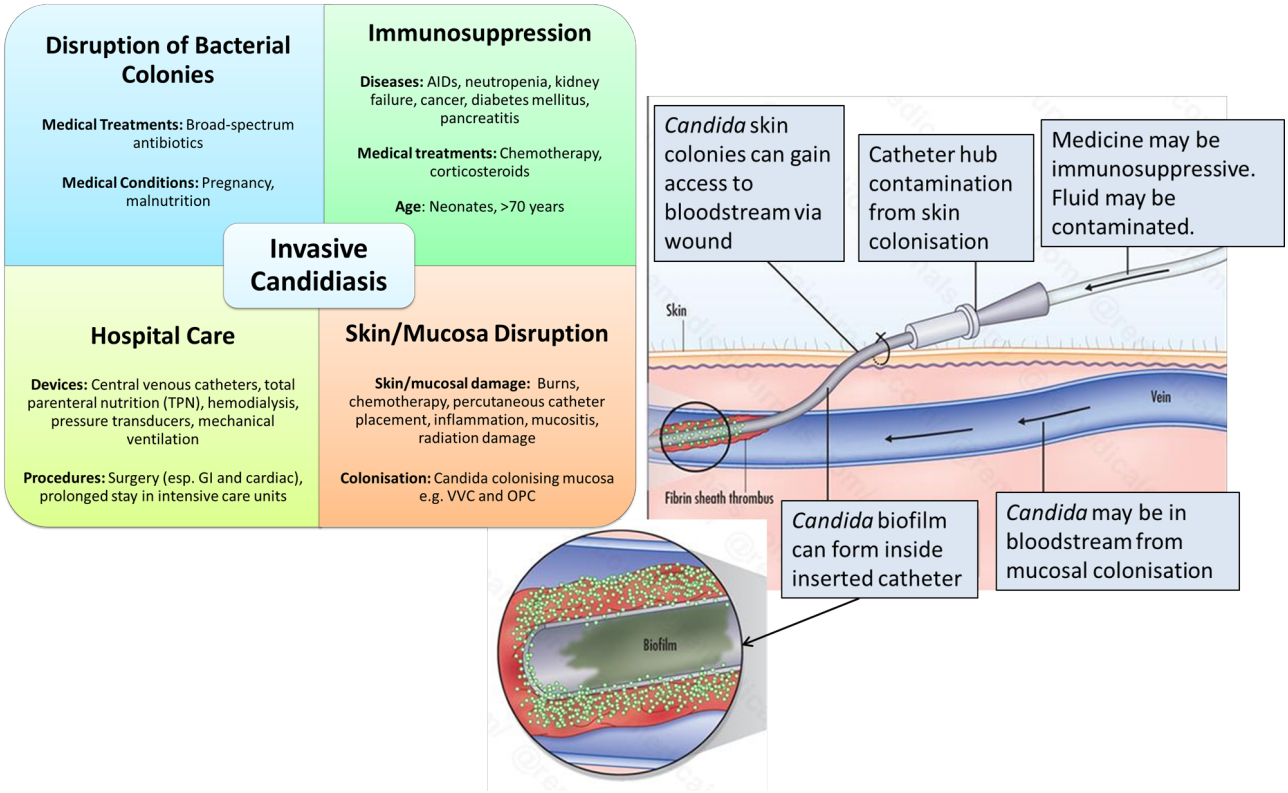


Figure 1.2: Invasive candidiasis risk factors.
A table summarizing risk factors for the development of invasive candidiasis alongside a diagram of how contaminated catheters plus Candidal skin/mucosal candidiasis can lead to disseminated candidiasis (Pelroth et al 2007, Pfaller & Diekema 2007, Lim et al 2012, Delaloye & Calandra 2014. Catheter diagram is a modified version of image from James et al 2011).

Invasive candidiasis can also result from *C. albicans* colonizing in the gastrointestinal tract - the main colonization site in the human body (Kumamoto et al 2011). Since *C. albicans* has no known environmental reservoir, it is almost always associated with the GI tracts of mammals (Harriott & Noverr 2011). *C. albicans* cells are normally in commensal homeostasis with the host but should the intestinal mucosal layer be damaged, *C. albicans* can reach and invade the blood stream, leading to candidemia. As has been demonstrated in mice models, the use of antibiotics and also inflammation of the GI tract can increase the likelihood of *C. albicans* colonization and invasion by disrupting commensal bacterial colonies. The mucosal colonization of *C. albicans* has also been shown to stimulate the host production of cytokines IL-17 and IL-23, resulting in GI tract inflammation and subsequent bloodstream invasion (Kumamoto et al 2011). It is a possibility that therapies aimed at decreasing the colonization of the gut by *Candida* species may cut the number of people affected by candidemia (Nucci and Anaissie 2001, Kumamoto et al 2011).

1.1.3. Oropharyngeal Candidiasis and Denture Stomatitis

In the oral cavity, *Candida* species (predominantly *C. albicans*) can be found in approximately 60% of the human population (Mochon et al 2010. It can become pathogenic when certain diseases result in the ineffective defense of CD4+ T-lymphocytes and proinflammatory cytokines towards *Candida* colonization (for example in HIV patients), and oropharyngeal candidiasis (OPC) can result (Farah et al 2002, Fidel Jr 2011). OPC is seen in 50-90% of HIV patients and is often one of the earliest signs of an underlying HIV infection (Rabeneck et al 1993).

C. albicans biofilms can contaminate dentures and tongue piercings and can mix with other infectious micro-organisms, for example *Streptococcus gordonii* (Morales & Hogan 2010, Zadik et al 2010). These mixed-species biofilms can cause denture stomatitis, an inflammatory infection that can be life-threatening for disabled elderly patients, and that affects 50-75% of otherwise healthy

denture wearers. Furthermore, denture stomatitis is the most prevalent type of OPC (Morales & Hogan 2010, Harriott & Noverr 2011).

1.1.4. Vulvovaginal Candidiasis and Balanitis

In 20-30% of women, *Candida spp.* (mostly *C. albicans*) is present in the vaginal cavity as a harmless commensal, but certain risk factors such as diabetes mellitus, antibiotic use, contraception, pregnancy and hormone therapy, can result in vulvovaginal candidiasis (VVC) (Sobel 2007). Interestingly, the increase in the glycogen content in vaginal mucosal cells resulting from the effects of oestrogen hormones during pregnancy, hormone therapy and contraceptive use, and also from elevated glucose levels during diabetes mellitus, provide an environment with an enriched source of carbon ideal for *C. albicans* colonisation (Sobel 2007). The resulting VVC affects approximately 75% of all women and 5-8% of women will experience four recurrences per year (Fidel 2004).

Within the vaginal cavity, the bacteria *Lactobacillus spp*. competes for adhesion sites on vulvovaginal mucosal cells and so helps to protect this environment from *C. albicans* colonization (Fig 1.3). This may explain why broad-spectrum antibiotics are a risk-factor for developing VVC – as they disrupt these protective *Lactobacillus* communities and lead to *C. albicans* colonization (Boris & Barbes 2000, Morales & Hogan 2010).

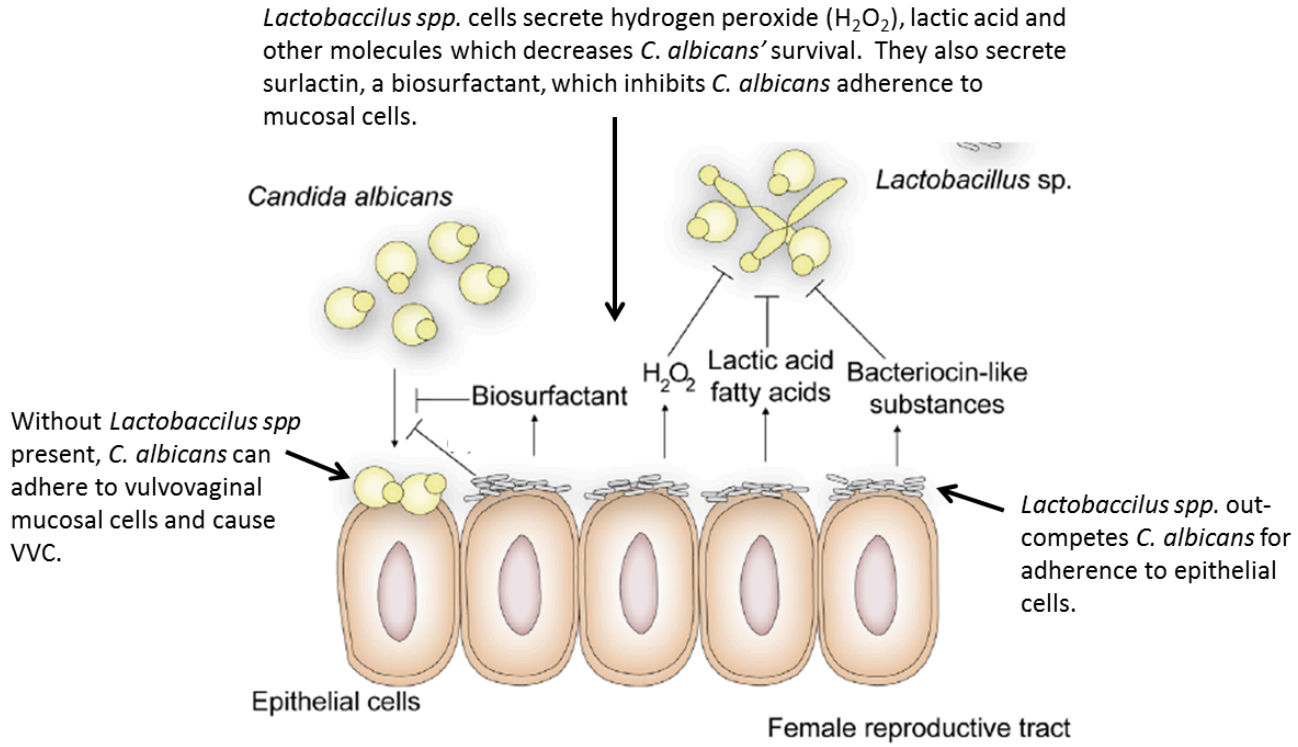


Figure 1.3: *Lactobacillus spp.* communities protect vaginal mucosal epithelial cells from *C. albicans* colonization.
One risk factor for VVC is broad-spectrum antibiotic use, which can deplete *Lactobacillus spp.* populations and thus provide an opportunity for *C. albicans* to establish a colony that can result in VVC (Boris & Barbes 2000, Modified figure from Morales & Hogan 2010).

Candida spp. can also infect the glans penis and cause an inflammatory condition called Candidal balanitis. This affects 14-18% of men and can be acquired from sexual intercourse. However, similarly to VVC, it is also associated with immunosuppression, antibiotic use and diabetes mellitus (Boris & Barbes 2000, Achkar & Fries 2010).

1.2. Pathogenesis of *Candida albicans*: Adhesion, Invasion & Cell Damage

The process by which *C. albicans* starts as a commensal colony to then infecting vital internal organs of humans is quite complex. However, it can be simplified by identifying the three major stages involved in the pathogenesis of *C. albicans*; adhesion, invasion and cell damage, summarized and explained in Figure 1.4 (Zhu & Filler 2010, Naglik et al 2011).

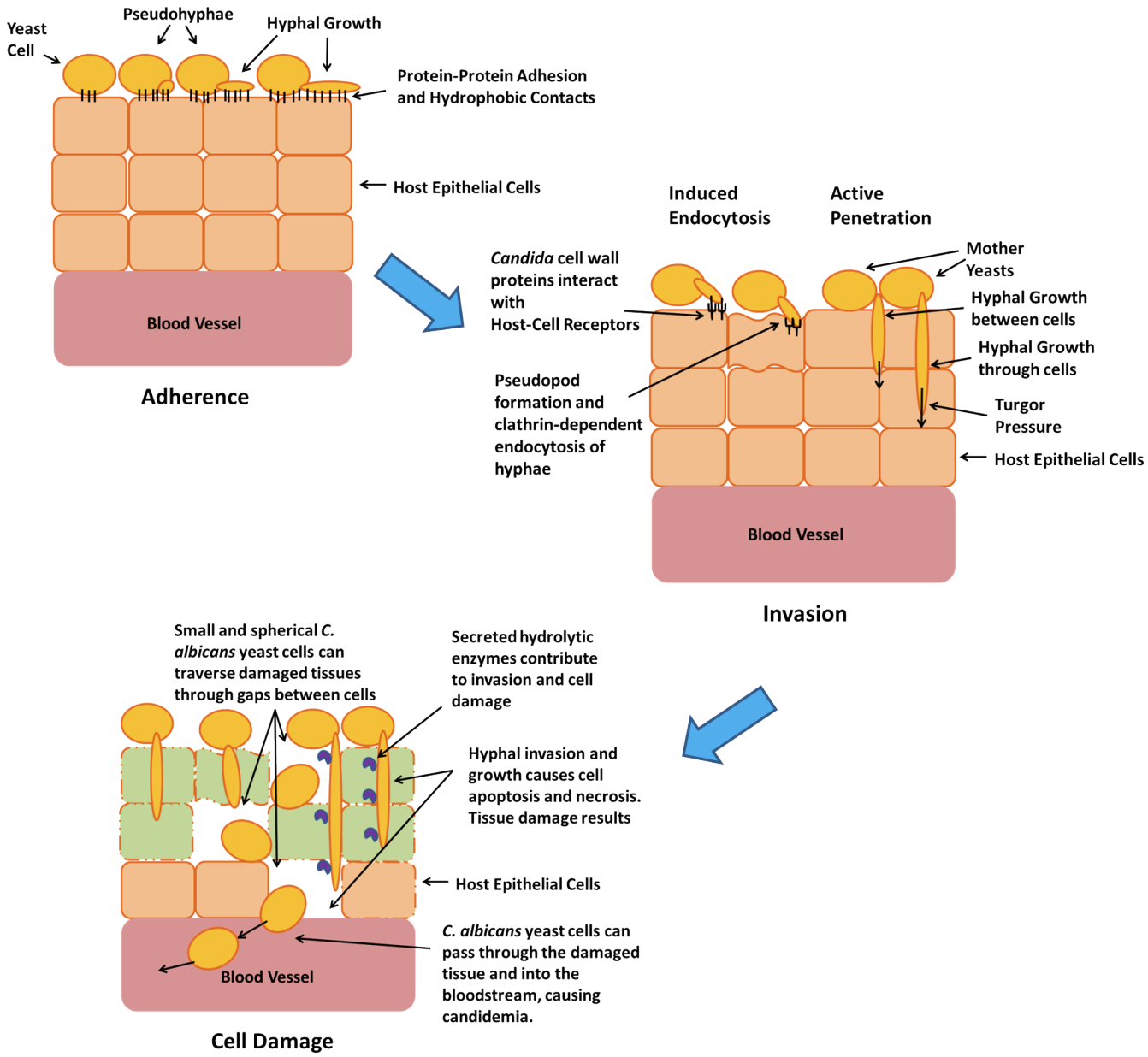


Figure 1.4: **Pathogenesis of *C. albicans* occurs in three stages, adherence, invasion and cell damage.** Once C. albicans has adhered to host cell membranes (e.g. of gut, oral or vaginal mucosal cells) via both hydrophobic interactions and protein-protein contacts, Invasion can occur. C. albicans invades host cells via two processes, a passive process called induced endocytosis, whereby it interacts with host cell endocytotic proteins, and an aggressive process called active penetration, whereby turgor pressure of growing hyphae force their way through host cells. This invasion leads to host cell apoptosis, which damages host tissues enough that yeast cells can traverse from the initial colony to the blood stream, resulting in invasive candidiasis (Zhu & Filler 2010 and Naglik et al 2011).

1.2.1. Adhesion

C. albicans cells are adept at adhering tenaciously and tightly to a variety of different surfaces – human skin, endothelial and epithelial mucosal host tissues and also the surfaces of medical equipment and dentures. Its aptitude for adhesion is an important virulence factor and contributes greatly to laying the foundations for infections (Zhu & Filler 2010).

C. albicans' cells adhere to host cell surfaces via various interactions (Fig 1.4). Adhesins on the *C. albicans* cell surface interact with extracellular matrix proteins laminin, fibronectin, collagen, entactin, vitronectin and tenascin, as well as proteins such as cadherins, integrins and serum components (Chaffin 2008, Williams et al 2013). The level of adhesiveness to host surfaces is directly related to the pathogenicity of the *Candida* species. This highlights why *C. albicans*, the most adherent of *Candida spp.*, is the most pathogenic (Calderone and Braun 1991).

Once *C. albicans* adheres to a surface, the cells are induced to form hyphae. The elongated hyphal morphology contributes to greater adhesion to host cells via up-regulation of 'stickier' hyphal-specific adhesins. The emergence of hyphae is theorized to be the crucial point at which *C. albicans* becomes pathogenic as they not only increase adhesion, but also promote invasion (Jacobsen et al 2012).

1.2.2. Invasion

After adherence occurs, the growth of *C. albicans* hyphal extensions subsequently leads to invasion. Mutant *C. albicans* cells with a reduced ability to invade host cells in vitro have impaired virulence in oropharyngeal and disseminated candidiasis models in mice, emphasizing the importance of the invasion step in pathogenesis. Invasion occurs via two distinct mechanisms; induced endocytosis and active penetration (Fig 1.4) (Zhu & Filler 2010).

Induced endocytosis involves the interaction of *C. albicans* hyphal outer cell wall invasin proteins with host endothelial and epithelial receptor ligands such as E-cadherin, N-cadherin, EGF receptor and HER2 receptor, which triggers clathrin-dependent endocytosis pathways in the host cells, resulting in formation of pseudopods that then engulf *C. albicans* cells (see Fig 1.5) (Filler et al 1995, Phan et al 2007, Zhu et al 2012). This is a passive process requiring hyphae, not yeast cells, but does not require hyphal cells to be alive (Wächtler et al 2012). It depends entirely on the host cells endocytotic machinery of host cells and can be inhibited by actin microfilament polymerization inhibitors such as cytochalasin D (Wächtler et al 2012). In fact, if the expression of clathrin, dynamin or contactin is abolished from endothelial and epithelial cells, this blocks the endocytosis of *C. albicans* (Moreno-Ruiz et al 2009).

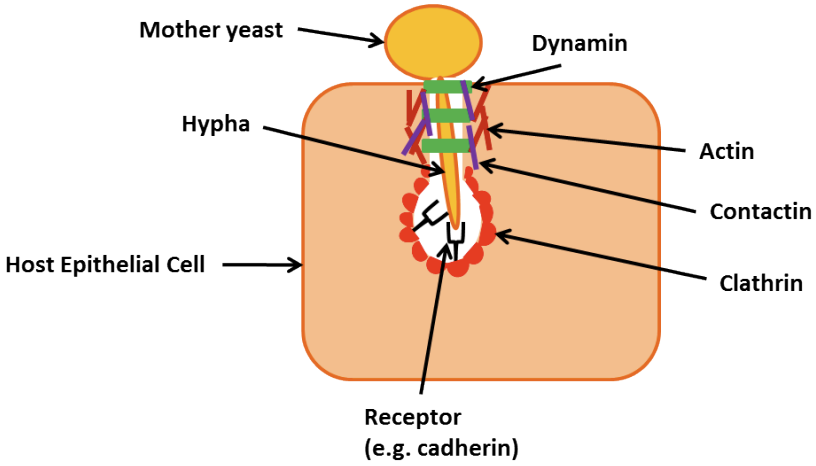


Figure 1.5: Induced endocytosis.
This process involves proteins on the hyphal cell wall interacting with host cell ligands, and inducing pseudopod formation requiring the host cell proteins clathrin, contactin, dynamin and actin. The hypha is then engulfed into the host cell (Moreno-Ruiz et al 2009).

Whilst induced endocytosis is important for *C. albicans* invasion in early stages of infection, active penetration occurs later and represents the majority of *C. albicans'* invasive behavior (Dalle et al 2009, Wächtler et al 2012). Active penetration involves the growth of hyphae, which then drive their way through host cells via turgor pressure (Fig 1.4). This seems to be the predominant mechanism of invasion seen in gastrointestinal *Candida* infections (Dalle et al 2010). The hyphal germ tubes of *C. albicans* also secrete hydrolytic enzymes, which facilitate the invasion process. For instance, the cleavage of E-cadherin proteins via these enzymes disrupts tight junctions between epithelial cells, facilitating invasion (Frank & Hostetter 2007, Dalle et al 2010).

1.2.3. Cell Damage

Both invasion mechanisms lead to damage of the host epithelial, endothelial and mucosal cells, and access to the bloodstream for *C. albicans* yeast cells (Filler et al 1995, Wächtler et al 2011). Host cells die via apoptosis and necrosis, paving the way for *C. albicans* yeast cells to travel in the bloodstream (Naglik et al 2011). These cells can adhere to and invade the endothelial cell lining of blood vessels in order to reach tissues and organs in disseminated candidiasis (Zhu & Filler 2010).

1.3. Pathogenesis Tricks

1.3.1. Polymorphism

C. albicans is a polymorphic organism - the cells can switch between unicellular, ovoid-shaped yeast (or Blastoconidia) and long, filamentous hyphal forms (Soll 1992). It also appears as intermediate pseudohyphae, which are elongated ellipsoid-shaped cells (Mitchell 1998, Berman & Sudbery 2002) (Fig 1.6). Environmental cues such as pH, temperature and nutrient availability influence *C. albicans* cellular morphology and Figure 1.6 provides a summary of the different morphologies and their environmental influences (Heinz 2000, Berman & Sudbery 2002, Sudbery et al 2004). Quorum-sensing, or cell-cell communication via particular molecules, also has a significant influence - for example, the isoprenoid farnesol can inhibit hyphal formation, whilst tyrosol was found to promote it (Hornby et al 2001, Chen et al 2004).

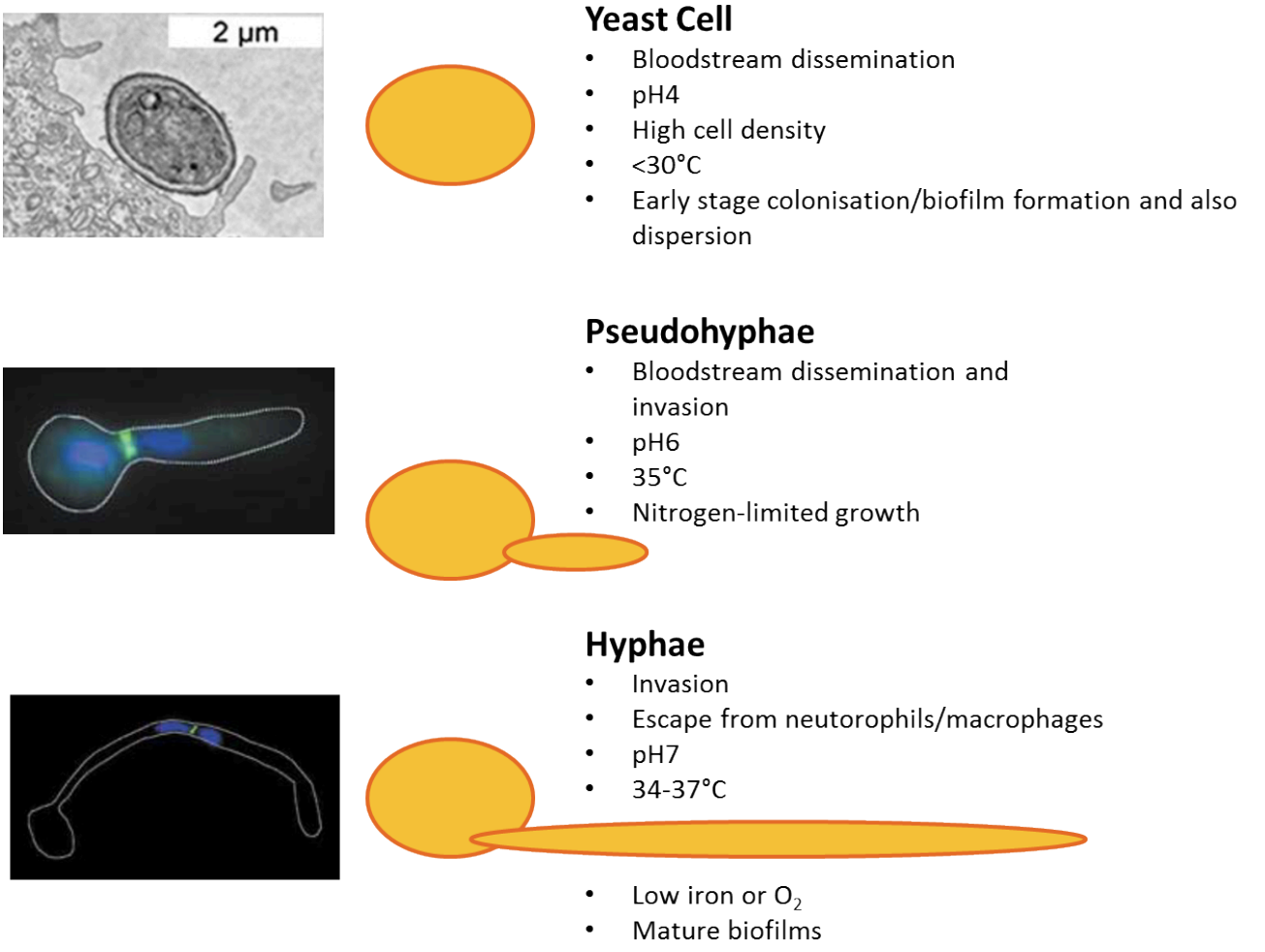


Figure 1.6: The different morphologies of *C. albicans*, including yeast cells, pseudohyphae and hyphae. Plus, the stages of infection they are important for and the conditions these morphologies can be induced (*C. albicans* hyphae and pseudohyphage images from Berman & Sudbery 2002 and Sudbery et al 2004, *C. albicans* yeast image from Naglik et al 2011).

This ability to switch forms is a highly advantageous trait for its commensal and pathogenic, and has been regarded as a major virulence factor (Jacobsen et al 2012). This process is seen uniquely in *C. albicans* and also *C. dubliniensis*. The switch from *C. albicans* yeast to hyphal cells has been regarded as the stage when *C. albicans* transitions from commensal to pathogen due to the prominent role of hyphae in invasion; however both morphological forms are important (Berman & Sudbery 2002, Jacobsen et al 2012).

1.3.1.1. Yeast vs. Hypae

The hyphal form of *C. albicans* is more adherent and also more specialised for invasion in comparison to yeast cells (Phan et al 2000, Jacobsen et al 2012). It has been reported to have a higher propensity for both induced endocytosis and active penetration due to expression of invasive protein factors (e.g. Als3) and turgor pressure exerted from growing germ tubes (Phan et al 2007). They are also important for evading the immune system, as they are more resistant to death from neutrophils and phagocytic attack due to their cellular shape being more difficult to engulf. Interestingly, *C. albicans* yeast cells are induced to form hyphae once they are phagocytised, which kill the immune cell by growing outwards (Fradin et al 2005). The nutrient restrictions within immune cells results a starvation response, which can also trigger hyphal formation (Lorenz et al 2004).

However, the yeast form is more suited to bloodstream travel during dissemination and also endothelial cell binding (Jacobsen et al 2012). In an experiment using a continuous flow system to mimic the bloodstream environment, *C. albicans* yeast cells were shown to adhere to endothelial cells in greater numbers than hyphal or pseudohyphal cells (Grubb et al 2009). Their small size enables them to be more adept than hyphal or pseudohyphal cells at escaping from the bloodstream in order to invade and colonise host organs (Jacobsen et al 2012).

It is generally accepted that both forms are important for adherence and virulence, and especially the formation of, and dispersal from biofilms (Jacobsen et al 2012). It has been shown that *C. albicans* mutants 'trapped' in either hyphal or yeast forms result in less adherence, virulence and damage to host cells (Saville et al 2003). Both morphologies are seen to contribute to disseminated candidiasis; when *C. albicans* yeast cells infiltrate the bloodstream, they travel to and colonise vital organs. Once they adhere to blood vessel endothelial cells, hyphal growth enables them to invade host tissues rapidly. Although hyphal growth is rarely seen in *C. albicans* colonies in the spleen or liver, yeast cells in the blood develop germ tubes early, which are induced by serum factors (Chandra et al 2001, Fradin et al 2005, Jacobsen et al 2012).

1.3.2. Enzyme Secretion and Nutrient Acquisition

1.3.2.1. Enzyme Secretion

C. albicans secretes hydrolytic enzymes, which can damage host tissues during invasion and also, release nutrients into the environment for uptake by the pathogen. For example, phospholipases break down plasma membrane phospholipids and lipases hydrolyse triacylgycerols (Naglik et al 2011). A family of secreted aspartyl proteases (SAPs) hydrolyse host proteins such as ECM components collagen, laminin, fibronectin, and also immune system factors such as proinflammatory cytokines and complement proteins (Naglik et al 2003). Their major contribution to virulence is such that inhibiting their action or removing SAP genes can decrease *C. albicans* pathogenicity (Borg-Von Zepelin et al 1999, Naglik et al 2003, Dos Santos et al 2006).

1.3.2.2. Iron Acquisition

C. albicans has evolved strategies of acquiring host iron, as it is important for many of its intracellular processes including respiration, oxygen transport and DNA synthesis (Almeida et al 2009). It can take host iron from haemoglobin by latching onto red blood cells and then lysing them, finally acquiring haemoglobin using Rbt5 and Rbt51 receptors and other factors (Weissman & Kornitzer 2004).

In hyphal form, *C. albicans* can also adhere to ferritin via Als3 (discussed later in section 1.5.2.5). *C. albicans* subsequently uses a plasma membrane located reductive pathway in order to make use of acquired iron from ferritin and other sources (Almeida et al 2008). In fact, the concentration of iron in host cells seems to positively correlate with the amount of cell damage caused by *C. albicans* as its invasion into host cells is decreased if host cells have a minimal iron content (Almeida et al 2008, Almeida et al 2009).

1.3.3. Biofilm Formation

C. albicans cells can evade the immune system, and become resistant to antifungal drug action, by aggregating together and/or to a surface in single or mixed-species biofilms, and subsequently secreting a protective extracellular matrix (Douglas 2003, Tornu & Van Dijck 2012). *Candida* cells in these biofilms have greater immune resistance than freely floating 'planktonic' cells (Donlan & Costerton 2002). These biofilms can form on mucosal surfaces, teeth, medical tools (e.g. catheters, implants and dentures), and also on tongue piercings (Zadik et al 2010). *C. albicans* cells in biofilms will have different gene expression patterns and phenotypic characteristics from planktonic cells, for example, SAP secretion and certain adhesins are up-regulated (Douglas 2003, Mendes et al 2007).

1.3.3.1. Stages of Biofilm Formation

C. albicans forms biofilms in four stages (Fig 1.7), as observed *in vitro* in laboratory based biofilm studies, and also *in vivo* in animal catheter and denture models (Chandra et al 2001, Andes et al 2004, Ramage et al 2004, Nobile et al 2012, Fox & Nobile 2012). Once *C. albicans* yeast cells have adhered to a surface (mucosal or non-biological) and the cells have multiplied profusely across it and reached optimal density, hyphae sprout at the top of the biofilm. The transcription factor Bcr1 is important for biofilm formation in C. albicans as it up-regulates several adhesion genes including HYR1, HWP1, ALS1 and ALS3 (Nobile et al 2005, Nobile et al 2006, Dwivedi et al 2011, Finkel & Mitchell 2011).

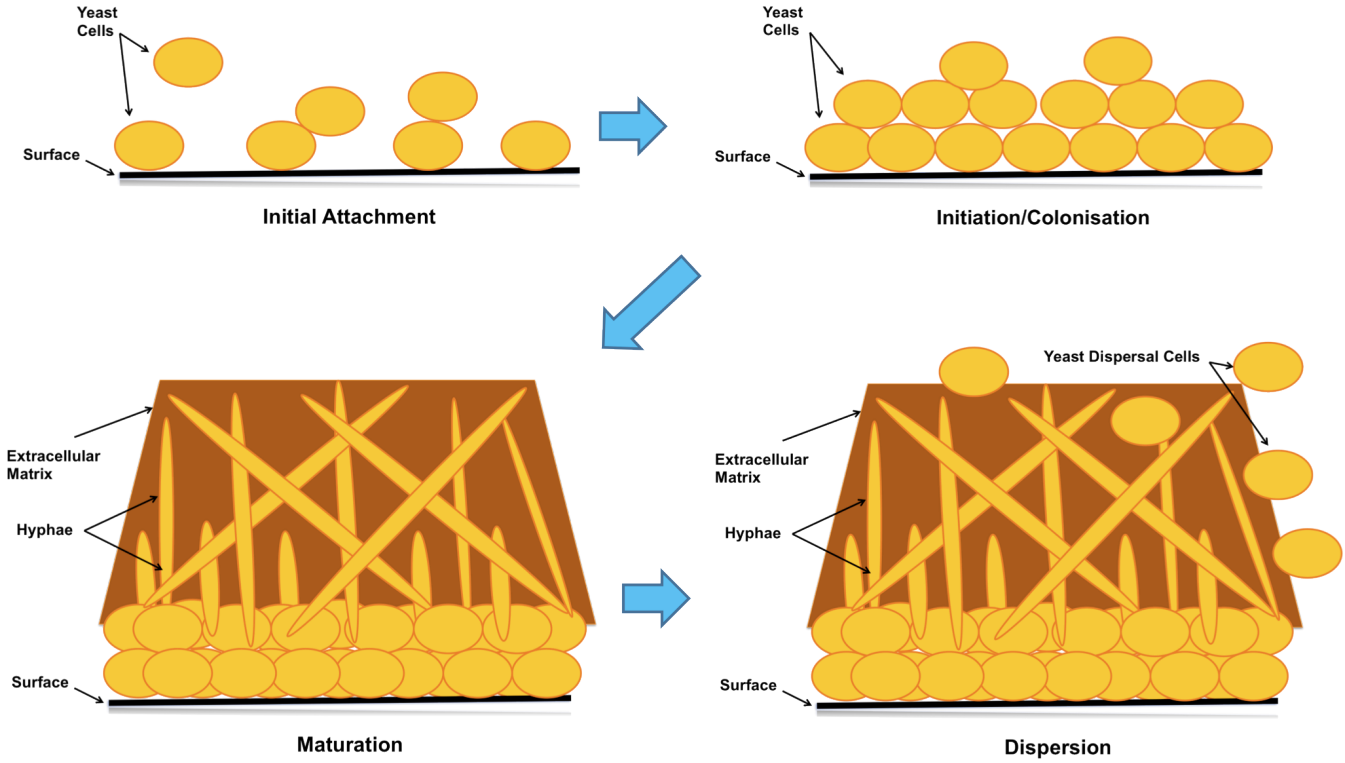


Figure 1.7: The different stages of *C. albicans* biofilm formation.
1) Initial Attachment phase – yeast cells adhere to a surface. 2) Initiation/Colonisation phase – yeast cells proliferate to form a colony and basal biofilm foundation. 3) Maturation phase – Hyphae and pseudohyphae grow and extracellular matrix components are secreted, forming a fully mature biofilm encased in a matrix. 4) Dispersion phase – yeast cells disperse from the established biofilm in order to form colonies elsewhere (Baille & Douglas 1999, Chandra 2001, Douglas 2003, Uppuluri et al 2010 and Fox & Nobile 2012).

The mature biofilms of *C. albicans* are comprised mainly of hyphal-phase cells, as yeast-phase growth becomes repressed (Douglas 2003). The cells secrete a polysaccharide rich extracellular matrix which encloses and protects the cells, and which can be difficult to penetrate by current anti-fungal drugs (Chandra 2001, Tornu & Van Dijck 2012).

When the biofilm has reached maturity and/or, there is a lack of nutrients, yeast cells will disperse (Finkel & Mitchell 2011). If they are in contact with the bloodstream, yeast cells are then able to travel to vital organs and establish new biofilm colonies as seen in candidemia/disseminated candidiasis (Mayer & Hube 2013). This dispersion stage from biofilms is a direct factor in virulence, as seen during a mouse model of disseminated candidiasis (Uppuluri et al 2010).

1.3.3.2. Mixed-Species Biofilms – Collaborating Pathogens

C. albicans can co-operate with other organisms when forming biofilms. In terms of mixing with bacteria, almost half of hospital-acquired *C. albicans* bloodstream infections also involve other pathogens, most commonly *Staphylococcus aureus*, *S. epidermidis* and *Entercoccus spp.* (Klotz et al 2007, Harriott & Noverr 2011). In a catheter biofilm model, *C. albicans* can promote the growth of *S. epidermidis* and increase its resistance to the drug vancomycin (Adam et al 2002). In return, the matrix secreted by *S. epidermidis* can increase antifungal resistance of *C. albicans* to fluconazole (Adams et al 2002).

S. aureus is unable to form a single-species biofilm, but it can form extensive biofilms in combination with *C. albicans.* *C. albicans* hyphae can serve as an effective scaffold for *S. aureus* cells to stick to and form microcolonies on, leading to in increased resistance of *S. aureus* to vancomycin (Harriott & Noverr 2009, Harriott & Noverr 2010, Peters et al 2012).

During denture stomatitis, *C. albicans* and *Streptococcus gordonii* interact (Fig 1.8). *S. gordonii*, which often forms dental plaque on tooth surfaces, promotes *C. albicans* hyphal growth and

colonization by other oral pathogens (Bamford et al 2009). *S. gordonii* is a commonly found microorganism in the oral cavity, and it forms colonies on host mucosal cells and also dentures, and can form mixed-species communities with *C. albicans* in vivo and in vitro (Holmes et al 1995). *S. gordonii* prefers to mix with hyphal germ tubes, but can also associate with yeast and pseudohyphal cells. These microorganisms have a complementary relationship in that *S. gordonii* encourages *C. albicans*' hyphal growth by interfering with the suppression of *C. albicans*' hyphal growth via the quorum sensing molecule farnesol. As farnesol suppresses the Ras1/Efg1 biochemical pathway responsible for hyphal growth, the disruption of farnesol signalling leads to an increase in *C. albicans* filamentation. The authors speculate that *S. gordonii* may prevent farnesol from interacting with its receptors, or may induce the *C. albicans* Ras/Efg1 hyphal-growth pathway regardless of the influence of the farnesol signal (Bamford et al 2009, Morales & Hogan 2010, Silverman et al 2011).

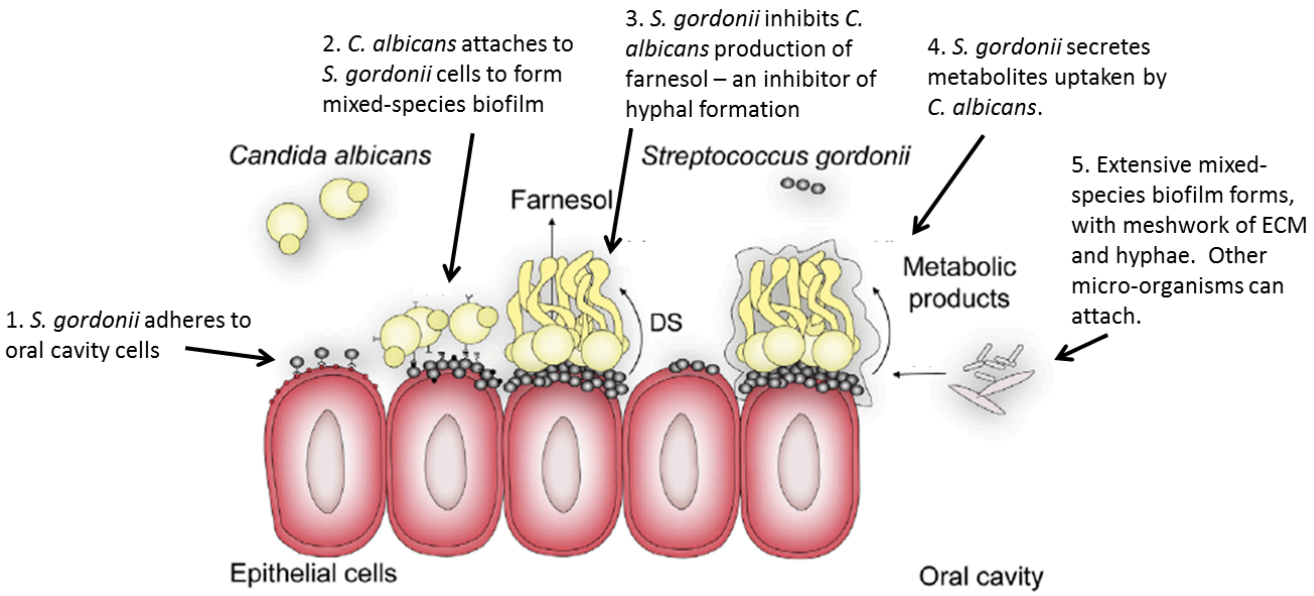


Figure 1.8: The process of mixed-species biofilm formation of *C. albicans* and *S. gordonii*.
This mixed-species biofilm, which can lead to denture stomatitis, initially begins with *S. gordonii* colonisation of oral cavity cells, leading to attachment of *C. albicans*. *S. gordonii* promotes the hyphal formation by inhibiting hyphal-repression quorum-sensing factor farnesol, so encouraging *C. albicans* adhesion and invasion of host cells. Plus, metabolites secreted by *S. gordonii* are uptaken by *C. albicans*, further benefiting candidal colonisation. In return, *S. gordonii* benefits from attachment to *C. albicans* hyphal germ tubes and becoming embedded within the protective biofilm ECM that is secreted by *C. albicans* (Modified image from Morales & Hogan 2010).

C. albicans however can have antagonistic relationships with other micro-organisms. For example, in the vaginal cavity, *Lactobacillus spp.* has antifungal activity and can block *C. albicans* hyphal growth (Boris & Barbes 2000). *Salmonella typhimurium*, *Bacillus amyloliquefaciens* and *E. coli* also secrete factors that kill *C. albicans* hyphae and reduce their colonisation of the host (Kim & Mylonakis 2011, Bandara et al 2013a, Rautela et al 2014). *P. aeruginosa* can also secrete molecules (e.g. phenazine derivatives and lipopolysaccharides) that suppress hyphal formation, and furthermore, is in competition with *C. albicans* for host colonisation sites, even though they can form mixed-species biofilms on intravenous catheters (Holcombe et al 2010, Morales et al 2010, Harriott & Noverr 2011, Purschke et al 2012). Interestingly, *P. aeruginosa* can exploit *C. albicans'* killed hyphal filaments and establish colonies on them, effectively using the germ tubes as a scaffold on which to build a biofilm. Indeed, pre-infection of lung tissue by *C. albicans* can increase occurrence of *P. aeruginosa*-associated pneumonia, yet *P. aeruginosa* colonisation in patients can reduce the likelihood of acquiring pulmonary candidiasis (Hogan & Kolter 2004, Holcome et al 2010, Harriott & Noverr 2011).

1.3.4. The Adhesive *C. albicans* Cell Wall

Electron microscopy studies revealed that the cell wall of *C. albicans* is highly specialised, with layers of differently functioning components (see Figure 1.9). The innermost layer is comprised of chains of N-acetylglucosamine that form chitin (1-3% of the cell wall) and a meshwork of polysaccharides. The outermost layer is comprised mainly of GPI-anchored proteins, mannoproteins and hydrophobic fibrils (30-40% of the wall) (Kapteyn et al 2000). Glucose polymers in the form of a flexible, extensive network of β-1,3 and β-1,6 glucans make up 50-60% of the cell wall's weight (Chaffin 2008). The tightly packed matrix of hydrogen bonds throughout the chitin and glucan chains, results in a strong, rigid and highly insoluble cell wall, which protects the microorganism from the host immune system and environmental stresses such as osmotic pressure (Gow & Hube 2012).

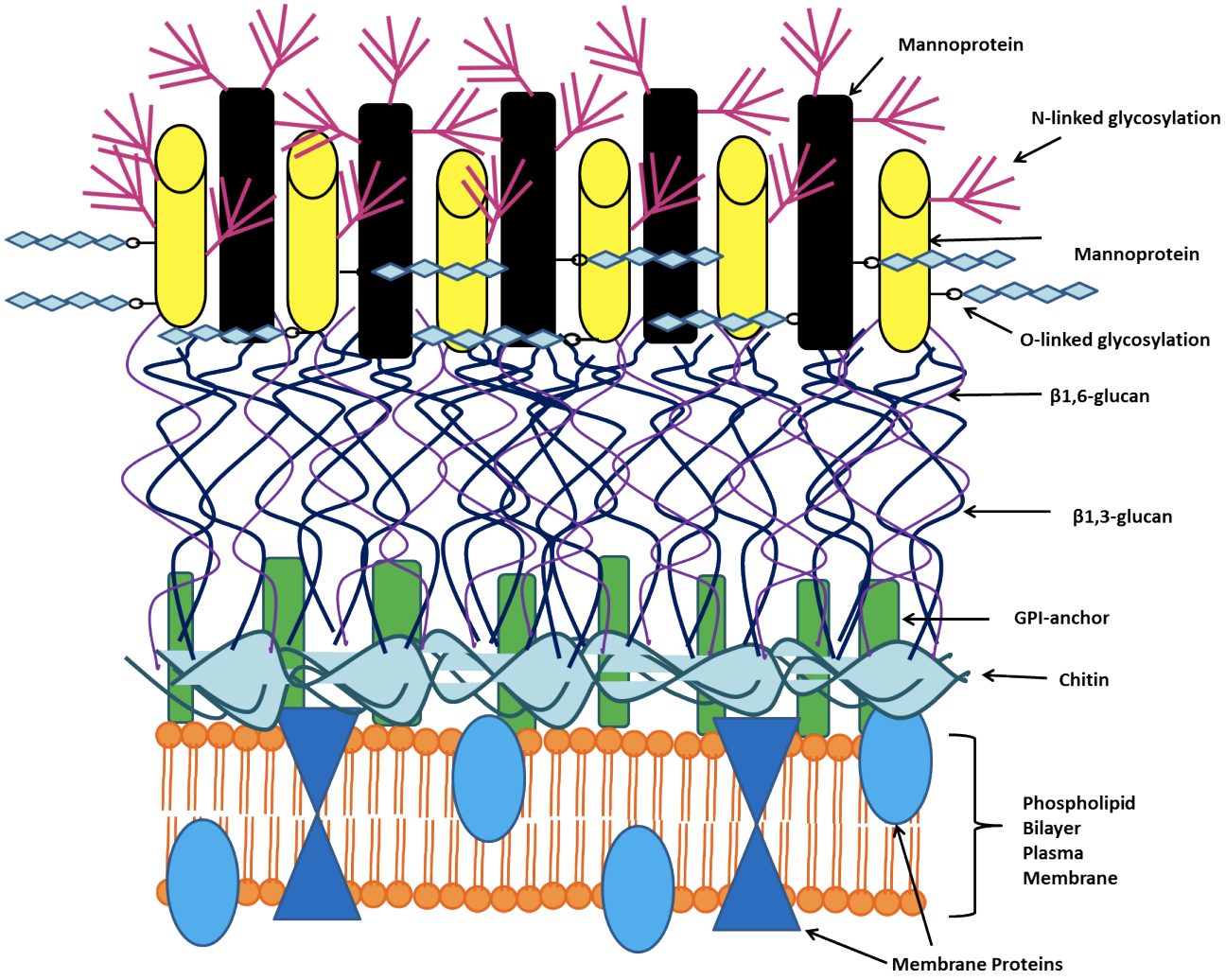


Figure 1.9: Cartoon representation of the *C. albicans* cell wall.
Layers of chitin, β1,3-glucans and β1,6-glucans and O- and N-glycosylated mannoproteins (indicated by arrowed labels) make *the C. albicans* cell wall a tough and adhesive barrier (Grubb et al 2008 and Luo et al 2013).

The cell wall is highly adapted for adherence to host cell surfaces. Its extensive collection of adhesins expressed on the cell surface enable it to 'stick' to extracellular matrix components and host epithelial and endothelial cells. This armoury of cell wall factors changes depending on whether *C. albicans* is in yeast or hyphal form (Chaffin 2008, De Groot et al 2013).

Exclusive to hyphae are the adhesins (Hyphal Wall Proteins) Hwp1 and Hwp2, which can covalently cross-link to host cell transglutaminases via glutamine residues on its N-terminal domain (Staab 1999, Sunstrom et al 2002). Hwp1 has been reported to have a prominent role in biofilm formation (Nobile et al 2006b). Members of the IFF-HYR (Individual Protein File Family F-Hyphally

Regulated) protein family, especially Hyr1, are also associated with hyphae. This family has 12 members of GPI-anchored proteins that have not all been fully characterised, although Hyr1 is noted to be glycosylated and Iff4 reported to have a role in adherence (Bailey et al 1996, Fu 2008, Zhu & Filler 2010, Klis et al 2011, Richard & Plaine 2012, De Groot et al 2013).

Expressed on both yeast and hyphae is Enhanced Adherence to Polystyrene (Eap1) protein, which contains conserved cysteines that form disulphide bridge attachments between proteins as well as possessing amyloidogenic 42 amino acid repeats. It is important for biofilm formation, adhesion to host cells and other materials and also for adhering to *S. gordonii* cells (Li & Palecek 2003, Li et al 2007, Nobbs 2010, Ramsook et al 2010). Pra1 (pH-regulated antigen-1) is also expressed on both yeast and hyphae, and has been reported to bind both fibrinogen and an integrin receptor (Soloviev et al 2007). Likewise, the protein Int1 (Integrin-like protein 1) is involved in adhesion and has been shown to facilitate *C. albicans* binding to platelets and endothelial cells, and may also have a role in hyphal formation (Gale et al 1997, Gale et al 1998, Asleson et al 2001).

The most prominent of adhesins expressed during *C. albicans* infections are the Als protein family, including 8 proteins with an ability to adhere to many different host cell membrane proteins, ECM components, medical device surfaces and also to proteins on *S.gordonii* cells in the formation of mixed-species biofilms (Hoyer et al 2008). Particular Als proteins are expressed in yeast phase (e.g. Als1), and some are expressed on hyphae (e.g. Als3) (Hoyer et al 1998, Hoyer et al 1999, Hoyer 2001, Hoyer et al 2008, Silverman et al 2010, Klis et al 2011). These proteins are reviewed more extensively in section 1.5.

1.4. Medical Treatment – Past, Present and Future

1.4.1. Current Treatments

Once a *Candida* infection has been diagnosed, it is crucial to start treatment as soon as possible – delays can significantly increase the risk of morbidity and mortality. For high-risk patients, prophylaxis is administered, which is the provision of antifungal drugs even if they do not present clinical symptoms of *Candida* infection (Nucci et al 1998, Blot et al 2002, Mikulska et al 2012).

The most common prophylactic drug is fluconazole and this has reduced the number of invasive candidiasis occurrences by 50% (Cruciani et al 2005, Shorr et al 2005). Fluconazole and other azole-based drugs act by inhibiting the cytochrome P450 enzyme 14α-demethylase in *C. albicans*, blocking the ability of this enzyme to catalyse the synthesis of ergosterol from lanosterol. As ergosterol is a crucial plasma membrane component, important for regulating membrane permeability, these fungistatic azole drugs lead to disrupted membrane structure in *C. albicans* (Ghannoum & Rice 1999). However, such drugs can result in toxic effects to the patient, and resistance to azoles is ever increasing in *Candida spp.* (Kanafani & Perfect 2008, Milkulska et al 2012).

Polyene drugs like amphotericn B and nystatin, interact with *Candida spp*. membrane sterols (including ergosterol), which then triggers the formation of pores - the walls of which pores are formed by eight polyene molecules attached to sterol molecules. This compromises the membrane's control of permeability, and leads to the killing of the *Candida spp*. Cell (Ghannoum & Rice 1999). However, these drugs can lead to kidney toxicity, and it has been reported that nystatin can cause allergies in patients (Cruciani et al 2005, Pfaller 2012, Mikulska et al 2012).

When patients present severe invasive candidiasis, including organ colonization, echinocandin drugs are often administered, for example, caspofungin (Andes et al 2012, Mikulska et al 2012). Echinocandins inhibit the enzyme β-1,3-glucan synthetase, so blocking the production of β1,3-glucans necessary for maintaining cell wall strength and integrity in Candida spp., resulting in cells that are susceptible to osmotic pressure and their death by lysis (Ghannoum & Rice 1999).

These drugs disrupt the synthesis of cell wall component beta-1-3,D-glucan, and thus reduces *Candida spp.* cells' resistance to osmotic and mechanical pressure, resulting in death by lysis (Ghannoum & Rice 1999). However, although originally reported to have low toxicity (Mikulska et al 2012), recent research has indicated possible cardiac and embryo toxicity associated with echinocandin drugs (Grover 2010, Stover & Farley, 2014).

1.4.2. Drug Resistance – Candida Fights Back

There has been a significant increase in azole-resistant strains of *C. albicans*, especially those that infect HIV patients. Genes encoding drug ATP-binding cassette efflux pumps, CDR1 and CDR2, plus multi-drug resistance gene MDR1 are up-regulated in *C. albicans* strains resistant to azole drugs (White 1997, Kanafani & Perfect 2008, Nett et al 2009). Furthermore, some *Candida* species have evolved mutations in a gene encoding a component of the β-1,3-D-glucan synthase complex that the echinocandin drugs target (Balashov et al 2006, Walker et al 2010), and echinocandin resistance is also increasing. Less resistance has been reported for polyene drug action against *Candida* strains, but some strains have developed resistance by being able to cope with low ergosterol synthesis due to building cell membranes with other types of sterols (Dick et al 1980, Kanafani & Perfect 2008).

As well as these evolving mechanisms to thwart current anti-*Candida* therapies, there is also the constant problem of impenetrable biofilms. The resistance of *C. albicans* cells to azole and polyene drugs increases with biofilm development and maturation (Chandra et al 2001, Vediyappan et al 2010). *C. albicans* cells are not only surrounded by a protective extracellular-matrix but also *C. albicans* cells in biofilms tend to increase their expression drug efflux pumps (Ramage et al 2002,

Mukherjee & Chandra 2004). Although echinocandin drugs have been reported to be effective at disrupting biofilms, caspofungin has been recently found to trigger up-regulation of cell wall adhesion and biofilm-forming proteins Als1 and Als3 and increase aggregation of *C. albicans* cells (Vediyappan et al 2010, Gregori et al 2011). This aggregative response, leading to biofilm formation, is likely to be a mechanism utilised by *C. albicans* to resist the effects of caspofungin (Mikulska et al 2012, Kanafani & Perfect 2008, El-Kirat-Chatel 2013).

As with many pathogens, there is an arms race between the development of drugs and the evolution of microbial resistance to them, and *C. albicans* is no exception. New antifungal drugs and treatment strategies need to be developed.

1.4.3. The Future of *Candida spp*. Treatment

Fortunately, new methods of treatment for *C. albicans* infections are currently being researched and tested. Drug molecules are being developed that are aimed at different targets, including berberrubine derivatives that target chitin synthase enzymes, sordarin derivatives that block ribosomal subunits and interfere with protein synthesis, and bicyclic peptidomimetic inhibitors of SAP proteins are currently being developed (Kamai et al 2005, Park et al 2006, Calugi et al 2011). The yeast-to-hyphae transition is being targeted too, as honey flavonoid extract has been shown to inhibit this process, and riccardin D from Chinese liverwort has been found to disrupt expression of hyphal-associated genes in *C. albicans* cells (Candiracci et al 2011, Li et al 2012).

As well as small molecule drugs, nanoparticles and monoclonal antibodies are other treatment options. For example, nanoparticles that emit nitric oxide may be effective at treating burns infected by *C. albicans* colonies, and silver nanoparticles that damage *C. albicans* cells via increasing reactive oxygen species levels show promise (Hwang et al 2012, Macherla et al 2012). The monoclonal antibodies (MAbs) that are being developed block vital processes in *C. albicans* survival - for example, the MAb C7 interrupts *C. albicans*' acquisition of iron, and also adhesion to host-cells via interacting with Als3 (Brena et al 2007, Brena et al 2011).

Some drugs already used for treating other ailments have been tested for their effectiveness against *C. albicans*. For example, diclofenac, a non-steroidal anti-inflammatory drug, as been shown to be effective against *C. albicans* biofilms when used in combination with caspofungin and has been shown to target the yeast-to-hyphal transition (Ghalehnoo et al 2010, Bink et al 2012). Verapamil, a current angina drug has similar effects (Yu et al 2013). Similarly, the drug amiodarone, traditionally used for treating arrhythmia, has been shown to have anti-fungal affects and aid the effects of fluconazole on *C. albicans* cells (Gamarra et al 2010). Interestingly, non-steroidal anti-inflammatory drugs including aspirin, ibuprofen and celecoxib were tested on their ability to reduce biofilm development, with aspirin reported to significantly inhibit biofilm formation (Alem & Douglas 2004).

An effective treatment strategy of *C. albicans* infections should be to prevent *C. albicans* from adhering to host cells or medical devices in the first place, and thus halt the production of *C. albicans* biofilms or establishment of hyphae-producing colonies. Drug-like molecules cis- and trans-2-dodecenoic acid from *Burkholderia cenocepacia* may be effective at halting biofilm formation by interfering with the expression of biofilm adhesins Als3 and Hwp1 (Zhang et al 2011). A similar effect was seen with the molecules purpurin and tetrandine (Tsang et al 2012, Zhao et al 2013).

Adhesive proteins themselves on the cell wall surface of *C. albicans*, particularly those of the Als adhesion glycoprotein superfamily, have been added to the repertoire of possible drug targets in this study. Drug design using structural biology studies of adhesion proteins combined with drug-screening assays could isolate therapeutic compounds with the potential to be incorporated into medical devices, or used in combination with other drugs in a prophylactic mix in order to prevent biofilm formation.

1.5. The Als Glycoproteins

During the late 1990s, the Agglutinin-like sequence (Als) family was identified by Hoyer et al and colleagues. Since then, an extensive body of work has aimed to characterise the functions and structures of the Als proteins and the expression patterns of the ALS gene family. As prominent adhesins of the *C. albicans* cell wall, they are interesting drug targets and it is important to study them in-depth (Hoyer et al 1996, Gaur & Klotz 1997, Salgado et al 2011).

1.5.1. The ALS Gene Family

The Agglutinin-Like Sequence (ALS) gene family codes for eight adhesion glycoproteins (Als1, Als2, Als3, Als5, Als6, Als7, Als9.2 and 9.2) with similar protein domain organisation and functions. At their N-terminus there is an 18 amino-acid signal sequence for cell wall targeting (which is eventually cleaved) and a non-glycosylated N-terminal 300-amino acid adhesion domain (NT) with two immunoglobulin-like components N1 and N2 held together by four disulphide bonds (Hoyer et al 2008). This adhesion domain has a binding pocket specialised for interacting with the C-termini of protein and peptide ligands (discussed more in depth in section 1.7.2) (Salgado et al 2011). At the end of this domain there is an amyloid-forming region (AFR) of approximately 12 amino acids that is important for the aggregative properties of *C. albicans* (Lipke et al 2012, Lin et al 2014). Following the AFR, there is β-sheet structured threonine-rich (TR) region of approximately 104 amino acids (Fig 1.10).

At their central, Als proteins have varying numbers of 36 amino acid tandem repeats – the exact number can vary between different *C. albicans* strains and also between alleles in the same strain, but their sequences are always rich in serine and threonine with many O- and N-glycosylation sites (Hoyer et al 1995, Hoyer et al 2008, Alsteens et al 2013) (Fig 1.10).

Finally, their C-terminal domain (CT) contains numerous serine, threonine and asparagine residues that attaches to the cell wall via a GPI-anchor remnant. Als7 is unique in that in its C-terminal domain, it contains 137-147 amino acid tandem repeats that have within them a five-amino acid tandem repeat motif of Val-Ala-Ser-Glu-Ser, hence these repeats are nicknamed the 'VASES repeats' (Hoyer & Hecht 2000, Hoyer 2001).

An ALS8 gene was proposed, but this was found to correspond to the same locus as ALS3 so was removed from the 'gene family'. The genes ALS1, ALS2, ALS4, ALS5 and ALS9 are found on chromosome 6 and ALS3 on chromosome R. They are roughly between 3-7kb and interestingly, none of the ALS genes have been found to have pseudogenes (Hoyer et al 2008).

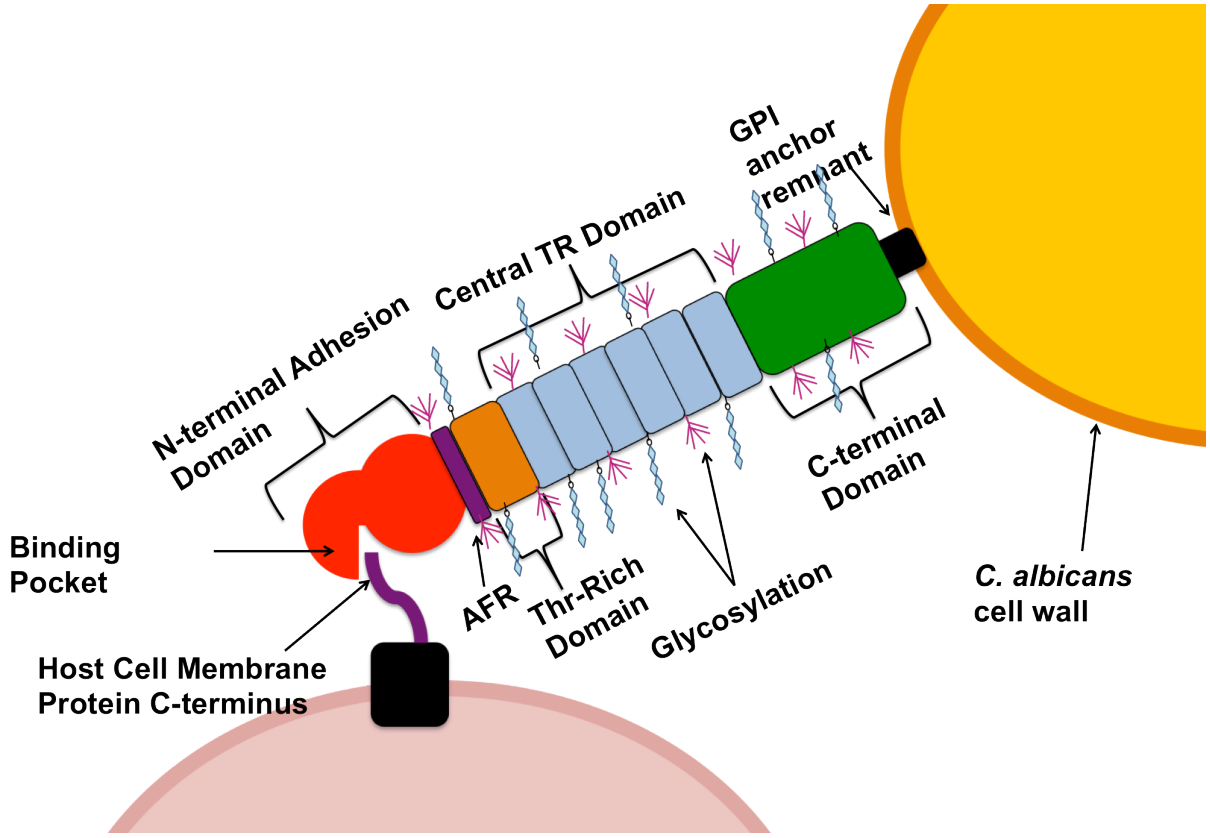


Figure 1.10: The general domain organisation and function of the Als proteins.
The N-terminal adhesion domain contains a binding pocket for interacting with host cell membrane proteins (discussed more in detail in section 1.6.3). Following the adhesion domain is a glycosylated AFR, plus glycosylated threonine-rich, central and C-terminal domains (Hoyer et al 2008, Salgado et al 2011 and Lin et al 2014).

Each gene in the family has a similar structure, with an approximately 1300bp 5' domain that is 40-80% identical, a central domain comprised of differing numbers of 108bp tandem repeats, and another variable domain at the 3' (Hoyer 2001) (Fig 1.11). Most of the variation between in the genes occurs in the 3' region, yet ALS2 and ALS4 share approximately 95% sequence identity and ALS5 and ALS6 share 93% in this segment (Hoyer et al 2008). ALS7 has the longest 3' domain and a very different sequence structure due to the VASEs repeats, sharing 60% nucleotide sequence identity with the rest of the genes (Hoyer & Hecht 2000). There is 11% sequence variation in the 5' domain of both ALS9 alleles, and in the 3' domain of ALS9.2, there are two Variable Block regions absent in ALS9.1 (Zhao et al 2003, Hoyer et al 2008).

The percentage identities at the genetic level of the ALS genes reflect those of the Als proteins in terms of primary sequences. Percentage identity comparisons between each of the full-length Als protein amino acid sequences are listed in Table 1.1. High percentage identities of approximately 70-80% are seen for Als1, Als3 and Als4 amino acid sequences. However, full-length Als4 and Als2 have a high percentage identity of 88.13%. Als5 and Als6 also have a high percentage identity of approximately 82%, and the percentage identity of Als9.1 with Als9.2 is approximately 95%. Als7 is the least similar to any other Als protein in terms of percentage identity (Zhao et al 2003, Hoyer et al 2008).

Table 1.1: Percentage identity comparisons between each of the full-length Als proteins.
Percentage identities (%) were computed using Clustal Omega (Sievers et al 2011) from Als primary sequences (Hoyer et al 2008) obtained from the Candida Genome Database.

	Als1	Als2	Als3	Als4	Als5	Als6	Als7	Als9.1	Als9.2
Als1	100.00	64.35	76.68	75.23	55.58	47.44	35.32	59.04	58.42
Als2	64.35	100.00	68.06	88.13	44.78	39.89	29.63	53.19	53.35
Als3	76.68	68.06	100.00	74.16	55.02	47.92	36.72	58.72	59.00
Als4	75.23	88.13	74.16	100.00	45.91	43.64	32.82	51.60	51.38
Als5	55.58	44.78	55.02	45.91	100.00	82.16	47.75	46.91	46.05
Als6	47.44	39.89	47.92	43.64	82.16	100.00	53.68	45.27	44.18
Als7	35.32	29.63	36.72	32.82	47.75	53.68	100.00	34.22	33.98
Als9.1	59.04	53.19	58.72	51.60	46.91	45.27	34.22	100.00	95.20
Als9.2	58.42	58.35	59.00	51.38	46.05	44.18	33.98	95.20	100.00

When comparing the amino acid sequences of the N-terminal adhesion domains of the Als proteins (between amino acids 18-315), percentage identities between approximately 40-85% are seen (see Table 1.2). A high percentage identity is seen between the N-terminal adhesion domains of Als1 and Als3 (approximately 82%), and between Als1 and Als5 N-termini (approximately 82%). The N-terminal domains of Als9.1 and Als9.2 have approximately 84% identity. Interestingly, the high percentage identity between the full-length sequences of Als2 and Als4, and also between Als5 and Als6, decreases to approximately 60% when comparing N-terminal adhesion domain sequences.

Table 1.2: Percentage identity comparisons between the amino acid sequences of the N-terminal adhesion domains of all the Als proteins.
Percentage identities (%) were computed using Clustal Omega (Sievers et al 2011) from Als primary sequences (Hoyer et al 2008) obtained from the Candida Genome Database.

	Als1	Als2	Als3	Als4	Als5	Als6	Als7	Als9.1	Als9.2
Als1	100.00	74.25	82.67	62.33	82.00	64.43	45.33	69.90	69.23
Als2	74.25	100.00	74.25	60.54	73.91	61.62	44.15	63.55	65.22
Als3	82.67	74.25	100.00	62.00	78.00	63.76	46.67	70.23	69.23
Als4	62.33	60.54	62.00	100.00	58.33	57.32	42.00	63.88	63.21
Als5	82.00	73.91	78.00	58.33	100.00	60.40	42.33	66.22	66.56
Als6	64.43	61.62	63.76	57.38	60.40	100.00	49.33	62.63	61.95
Als7	45.33	44.15	46.67	42.00	42.33	49.33	100.00	45.15	45.15
Als9.1	69.90	63.55	70.23	63.88	66.22	62.63	45.15	100.00	84.33
Als9.2	69.23	65.22	69.23	63.21	66.56	61.95	45.15	84.33	100.00

However, high percentage identities are seen between Als2 and Als4, and also between Als5 and Als6 when comparing the primary sequences of the central and C-terminal domains (from amino acids 315 to their respective C-termini). Percentage identities of approximately 92% are seen when comparing both these regions in Als2 and Als4, and are approximately 90% when comparing these in Als5 and Als6. Apart from these proteins, and also Als9.1 and Als9.2 which have an approximately 97% percentage identity, Als1 and Als3 which have approximately 74% and Als2 and Als3 which have approximately 67% identity, when comparing their central and C-terminal domains, the other Als proteins share relatively low percentage identities (between 30-63%) (see Table 1.3).

Table 1.3: Percentage identity comparisons between the amino acid sequences of the central and C-terminal domains of all the Als proteins.
Percentage identities (%) were computed using Clustal Omega (Sievers et al 2011) from Als primary sequences (Hoyer et al 2008) obtained from the Candida Genome Database.

	Als1	Als2	Als3	Als4	Als5	Als6	Als7	Als9.1	Als9.2
Als1	100.00	61.85	74.48	59.00	44.73	35.55	31.17	45.68	46.41
Als2	61.85	100.00	67.44	92.90	35.17	34.33	28.46	54.36	54.10
Als3	74.48	67.44	100.00	63.26	44.15	34.09	33.33	46.94	48.07
Als4	59.00	92.90	63.26	100.00	33.00	32.43	26.01	48.75	48.61
Als5	44.73	35.17	44.15	33.00	100.00	90.23	49.48	31.54	33.59
Als6	35.55	34.33	34.09	32.43	90.23	100.00	54.56	30.98	32.57
Als7	31.17	28.46	33.33	26.01	49.48	54.56	100.00	25.99	27.63
Als9.1	45.68	54.36	46.94	48.75	31.54	30.98	25.99	100.00	97.07
Als9.2	46.41	54.10	48.07	48.61	33.59	32.57	27.63	97.07	100.00

Importantly, ALS gene sizes vary between alleles and *C. albicans* strains due to the differing quantities of 108bp tandem repeats in the central domain (Fig 1.11). One particular strain (SC53214) has been observed to have two different sized alleles of ALS1, and similar diversity for ALS5 and ALS9 for example (Hoyer et al 1995, Hoyer et al 1998b). The most common ALS1 allele has 16 blocks of the tandem repeat sequence (Lott et al 1999), ALS3 has either 9 or 12 copies and ALS7 alleles tended to have 14-17 copies of the tandem repeat (Zhang et al 2003, Oh et al 2005, Hoyer et al 2008). ALS5 and ALS6 are shorter genes, with ALS5 having 5 copies and ALS6 having 4. The tandem repeats may serve as a mechanism for mixing-up protein characteristics and producing allelic variation (Hoyer et al 2001).

Sequence polymorphisms also produce variation in these genes, and have been most commonly seen in the 5' domain of ALS5 (Hoyer et al 2001, Hoyer & Hecht 2001, Hoyer et al 2008). The ALS genes are also reported to contain CUG codons, which are translated into serine 95-97% of cases, and into leucine 3-5% (Santos et al 1999, Gomes et al 2007).

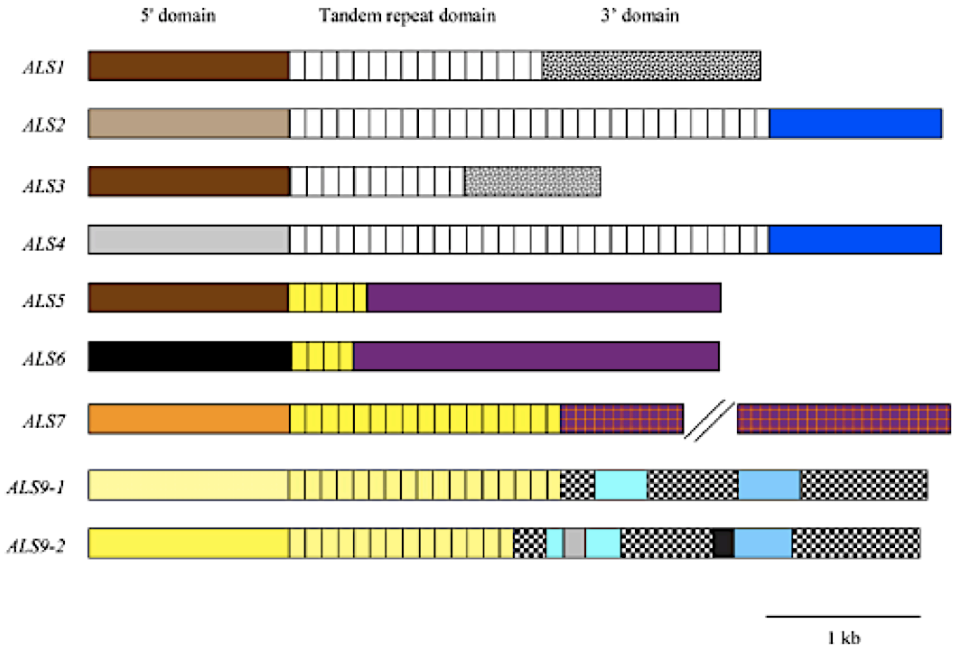


Figure 1.11: The ALS family of eight agglutinin-like sequence genes.
The family includes ALS1, ALS2, ALS3, ALS4, ALS5, ALS6, ALS7, ALS9.1 and Als9.2. An ALS8 was proposed, but was found to be on the same genetic locus as ALS3, so is now not included in the gene family. All genes have a highly similar 5' domain and a central domain of varying tandem repeats. The 3' domains of ALS7, ALS9.1 and ALS9.2 differ from the rest due to the VASES repeats in ALS7 and the variable regions of ALS9.1 and ALS9.2 (Figure from Hoyer et al 2008).

According to phylogenetic analysis, ALS7 seems to be the oldest of the gene family, despite its differences in comparison to the other genes (See Figure 1.12). ALS6 is the second oldest in terms of evolution. The 'youngest' genes are ALS1, ALS3 and ALS5, with Als3 the oldest among the three (Hoyer et al 2001, Zhao et al 2003).

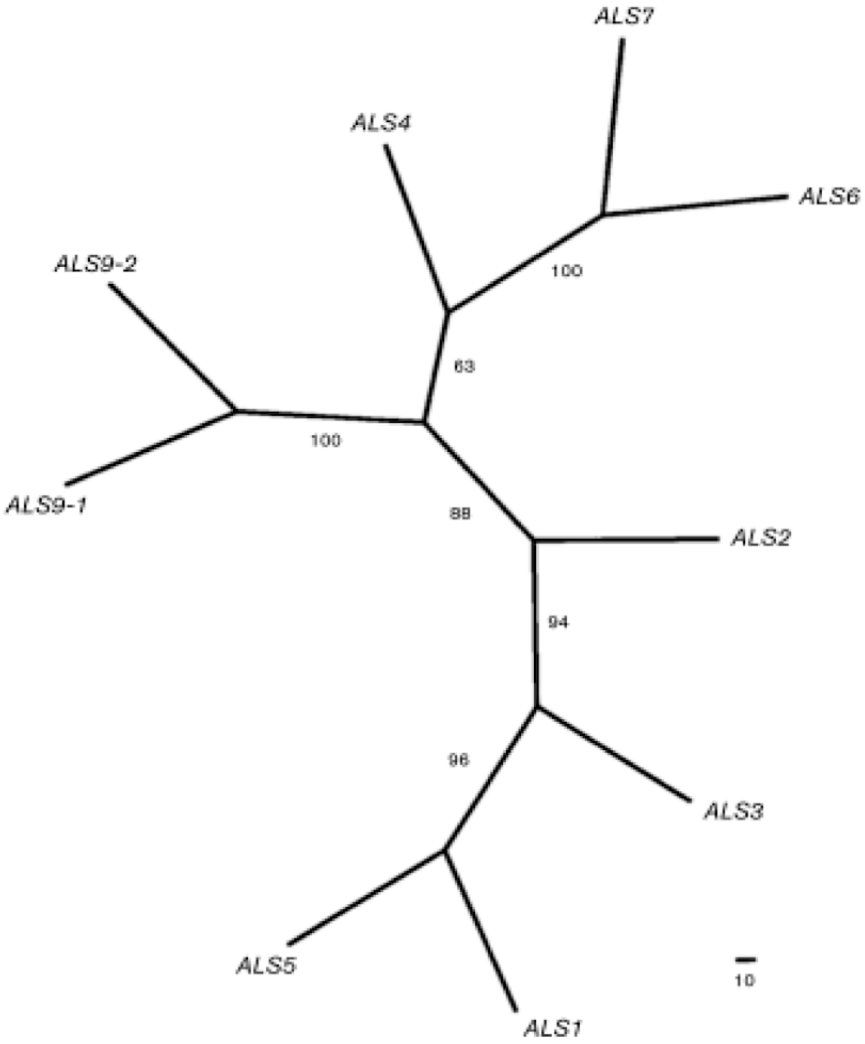


Figure 1.12: Phylogenetic analysis of the ALS gene family.
The oldest gene is ALS7, with ALS6 being the second oldest in terms of evolution. The youngest genes include ALS3, ALS5 and ALS1, with ALS3 being the oldest of the three and ALS1 being the youngest (figure from Zhao et al 2003).

Most of the Als protein research has been conducted in strains 1177, 1161 and SC5314 - the latter of which there is now a complete genome sequence (Jones et al 2004, Hoyer et al 2009, Goshorn et al 1992, Gillum et al 1984). Some strains lack particular ALS genes, for instance B311 and B792 are missing ALS5 (Hoyer & Hecht 2001). In strain WO-1, ALS1 expression seems to be different in comparison to other strains, for example its expression was not identified when the strain was cultured in Als1-inducing conditions, and it was theorized that the regulation of ALS1 expression may be different from other strains (Hoyer et al 1998). However, it was recently discovered that in WO-1, a new Als protein, Als51, has been formed which seems to be a genetic fusion of the Als5 N-terminal domain plus the Als1 tandem repeat region (Zhao et al 2011).

The ALS and SAP protein families may be functionally related. Similar pathways regulate the expression of their proteins and phylogenetic analysis revealed that they appear together on the same chromosomes in *C. albicans*, *C. dubliniensis* and *C. tropicalis*. Plus, a higher number of ALS and SAP genes seems to correlate with the pathogenicity of a particular *Candida* species, with *C. albicans* possessing the most genes from either family. In an intraperitoneal mouse model, their expressional relationship has been documented in vivo and some SAP genes are highly expressed alongside ALS

proteins in *C. albicans* samples from infected patients (Nas et al 2008). However, the ALS gene family has arisen more recently than the SAP family, and it is likely that they have evolved separately (Hube et al 1994, Hoyer 2001, Hoyer et al 2001).

ALS gene orthologues, found through cross-hybridizing experiments, appear in the most pathogenic *Candida spp.* including *C. dubliniensis, C. parapsilosis*, *C. lusitaniae*, *C. guilliermondii* and *C. tropicalis* (Hoyer 2001, Jackson et al 2001). Although the α-agglutinin mating protein has been considered a distant ortholog of the Als proteins, ALS orthologues have not been recorded in *S. cerevisiae* (Hoyer et al 1995). Similarly, ALS orthologues have not been seen in *C. glabrata*, making it and *S. cerevisiae* useful tools in studying the ALS gene functions (Hoyer et al 2001).

In *C. parapsilosis* and *C. tropicalis* genomes, the ALS genes display the greatest sequence diversity, with approximately 50-85% nucleotide sequence identity to *C. albicans* ALS genes (Hoyer et al 2001). In *C. tropicalis* there are 13 ALS genes. ALS5 and ALS3 do not feature in the *C. dubliniensis* genome, but there seems to be a duplicate of the ALS2 gene (Butler et al 2009, Jackson et al 2009, Groot et al 2013).

1.5.2. The Adhesive Als Proteins

1.5.2.1. Sticky Studies on Als Protein Adhesiveness

Als1 was the first of the protein family to be discovered and characterised and was of interest due to the similarity in sequence and domain structure to the *S. cerevisiae* cell-surface adhesion glycoprotein, α-agglutinin, and its corresponding gene, AGα1 (Hoyer et al 1995). However, whilst the N- and C-termini of Als1 have high similarity to α-agglutinin, the latter protein does not have the central tandem repeat domain (Hoyer et al 1995). The protein α-agglutinin in *S. cerevisiae* functions in mating, but since the *C. albicans* mating process is different from *S. cerevisiae* and does not involve Als1, it was theorized that the adhesive properties of Als1 function in host-pathogen interactions instead (Hoyer 2001, Magee et al 2002).

S. cerevisiae cells expressing Als1 were able to bind human epithelial and endothelial cells at a much greater propensity than non-adherent control cells (Fu et al 1998). *C. albicans* homozygous null mutants for the ALS1 gene had a pronounced reduction of adherence to the mice tongues in a murine OPC model, and also to endothelial cells in vitro (Kamai et al 2002, Fu et al 2002). Plus, the over-expression of Als1 in *C. albicans* cells dramatically increased their adherence to host cells, peptide ligands and also to themselves in a process called flocculation (Fu et al 2002, Klotz et al 2004).

Als5 (originally named ALA1) was identified as an adhesin that could bind to buccal epithelial cells and to magnetic beads coated in ECM proteins including laminin, fibronectin, bovine-serum albumin, casein and type IV collagen and was noted to be similar to Als1 (Gaur & Klotz 1997, Gaur et al 2002). *S. cerevisiae* cells and *C. albicans* cells expressing Als5 were reported to adhere to ECM proteins and then induce cell-cell aggregation. This adherence was present even when the cells were vortexed, subjected to high salt concentrations and pH ranges or were in the presence of detergents, but could be disrupted by foramide, which abolishes hydrogen-bonding (Gaur et al 1999, Klotz et al 2004).

The deletion of the ALS3 gene in *C. albicans* resulted in cells that had a pronounced reduction of adhesion to human umbilical vein endothelial cells and buccal epithelial cells (Zhao et al 2004, Murciano et al 2012). A similar reduction was seen in *C. albicans* ALS1 deletion mutants, but less so in comparison to ALS3 mutants. Deletions of these genes also resulted in less adhesion and cell damage in an OPC model formed from buccal reconstituted human epithelial cells (Zhao et al 2004). Plus, targeting of Als3 with antibodies reduces *C. albicans* adhesion to epithelial and endothelial cells (Brena et al 2007, Laforce-Nesbitt et al 2008, Beucher et al 2009, Coleman et al 2009).

Interestingly, *S. cerevisiae* expressing each of the different Als proteins had different adherence specificities in a screen including gelatin, laminin, fibronectin, epithelial and endothelial cells.

S. cerevisiae expressing Als6 was only found to bind gelatin, with Als9 could only bind minimally to laminin, with Als7 could bind to none of the ligands tested, but when expressing Als1, Als3 or Als5 could adhere to all ligands tested. The adherence to gelatin was greatest when *S. cerevisiae* expressed Als1, and to epithelial cells when expressing Als3 (Sheppard et al 2004).

S. cerevisiae cells expressing the Als1 N-terminus (amino acids 1-414) subjected to a glycan-array screen and surface plasmon resonance experiments, were shown to bind ligands containing terminal fucose, and also fucose on its own with milimolar affinity. Free fucose was also able to partially inhibit the Als1 N-terminus interaction with laminin and fibronectin, and this hinted at Als1 having a possible lectin function. However, these experiments did not determine exactly where fucose bound to Als1 as the Als1 N-terminal construct used was glycosylated, included the threonine-rich domain and also the AFR, all of which could have been responsible for the Als1-fucose interaction (Rauceo et al 2006, Donohue et al 2011).

The identification of the N-terminus as the domain responsible for specific ligand-binding interactions was first noted when chimeras of particular Als N-termini were combined with different Als C-termini and expressed in *S. cerevisiae* and the ligand-binding specificities followed the particular N-terminus (Sheppard et al 2004). Plus, deletion mutants within the N-terminus of Als1 disrupted the contribution of this protein to adhesion when expressed in *S. cerevisiae* (Loza et al 2004). Moreover, targeting Als3 with an antibody specific for binding to its N-terminus inhibited the adhesion of *C. albicans* to buccal epithelial cells. Similarly, adherence of a *C. albicans* ALS3 null mutant to epithelial cells was rescued upon expression of the Als3 N-terminus fused to α-agglutinin C-terminus (Zhao et al 2006).

The AFR in between the NT domain and TR domain has been reported recently to contribute primarily towards the adhesion of neighbouring Als proteins on one cell, and to other Als proteins on different cells. Current research states that the AFR seems to be more important for cell-cell aggregation rather than for specific host cell ligand adhesion (Lin et al 2014). Previous research has regarded the AFR as important for both cell-cell aggregation and host-cell ligand adhesion, and been described as a 'molecular zipper' that strengthens early adhesion via the NT domain (Rauceo et al 2006, Otoo et al 2008, Ramsook et al 2010, Lipke et al 2012).

There seems to be functional redundancy in the adhesiveness of the different Als proteins, yet some Als proteins have shown to contribute more to adhesion than others, or have a greater propensity for interacting with ligands. For example, a *C. albicans* strain with an Als3 null mutant exhibits less adherence to endothelial and buccal epithelial cells, but yet adheres normally to fibronectin, possibly because other Als proteins are present. Intriguingly, cells with Als5, Als6 and Als7 deletions show an increase in adhesiveness (Zhao et al 2004, Murciano et al 2012). Als1, Als3 and Als5 adhere to similar substrates when expressed in *S. cerevisiae* and are therefore expected to have similar ligand binding specificities (Klotz et al 2004, Sheppard et al 2004). However, some researchers put forward that the differences in binding behaviour are due to variations in the expression of particular ALS genes (Nobile et al 2008).

It is also important to note that many of the studies on the contribution of individual Als proteins to adhesion have often produced conflicting results due to differences in cell types studied. For example the expression of a particular Als protein in *S. cerevisiae* will have a different impact on cellular adhesion in comparison to the same Als protein expressed on *C. albicans*, amidst a cell wall populated with many other adhesins. To alleviate this problem, the adhesive properties of Als3 wild-type (WT) and loss-of-function mutants, both expressed in *C. albicans* at native levels, have been recently compared, the results of which are discussed in section 1.6.3 (Lin et al 2014).

1.5.2.2. Localisation of Als proteins in *C. albicans*

RT-PCR and flow cytometry studies utilising *C. albicans* cells with ALS gene promoters engineered to induce GFP expression have been used to study ALS expression in both yeast and hyphal cells. ALS1 is highly expressed when *C. albicans* cells are transferred between different types of media (e.g. RPMI and YPD) and prior to hyphal formation, whilst ALS3 expression occurs during the development of hyphae (Zhao et al 2004, Green et al 2005a). The transcription of ALS1, ALS2 and ALS3 is much more prominent compared to ALS5, ALS6, ALS7 and ALS9 genes, with ALS2 also being highly transcribed in hyphal cells (Green et al 2005a). Intriguingly, the expression of both Als1 and Als3 proteins is dependent on the zinc-finger transcription and biofilm inducing factor Bcr1 (Braun & Johnson 200, Nobile et al 2012).

Als1 is up-regulated in yeast-phase cells, and immunofluorescence studies showed that their distribution is even along the yeast surface (Fig 1.13) but when hyphae start to grow, Als1 proteins are localised at the base of hyphal filaments (Coleman et al 2010). The reported decreased rate of hyphal formation in *C. albicans* ALS1 null mutants suggested that Als1 might have a role in hyphal formation as well as its adhesive function (Fu et al 2002). However, this effect seems to differ depending on the media environment of the *C. albicans* cells as both ALS1 and ALS3 null mutants were capable of hyphal formation in RPMI medium, but ALS1 null mutant cells had a reduction in hyphal formation in Lee's medium (Zhao et al 2004). Interestingly, in comparison to *C. albicans* grown in vitro, fluorescent immunolabelling with anti-Als1 monoclonal antibodies revealed that Als1 could be detected on hyphal germ tubes in *C. albicans* samples from a mouse model of invasive candidiasis, not just on yeast cells or at the yeast-hyphal junction (Coleman et al 2010).

The expression of Als3 was initially reported to be restricted to pseudohyphal cells and hyphal germ tubes (Hoyer et al 1998a), and this was confirmed by studies involving GFP-reporter studies similar for ALS1, in which fluorescence was detected on *C. albicans* hyphae (Fig 1.13). Other studies using Fourier-transform MS quantification, atomic force microscopy, and also Coleman et al's stunning anti-Als3 immunolabelling methods also confirmed hyphal localisation (Fig 1.13) (Zhao et al 2006, Coleman et al 2010, Heilmann et al 2011, Beaussart et al 2012).

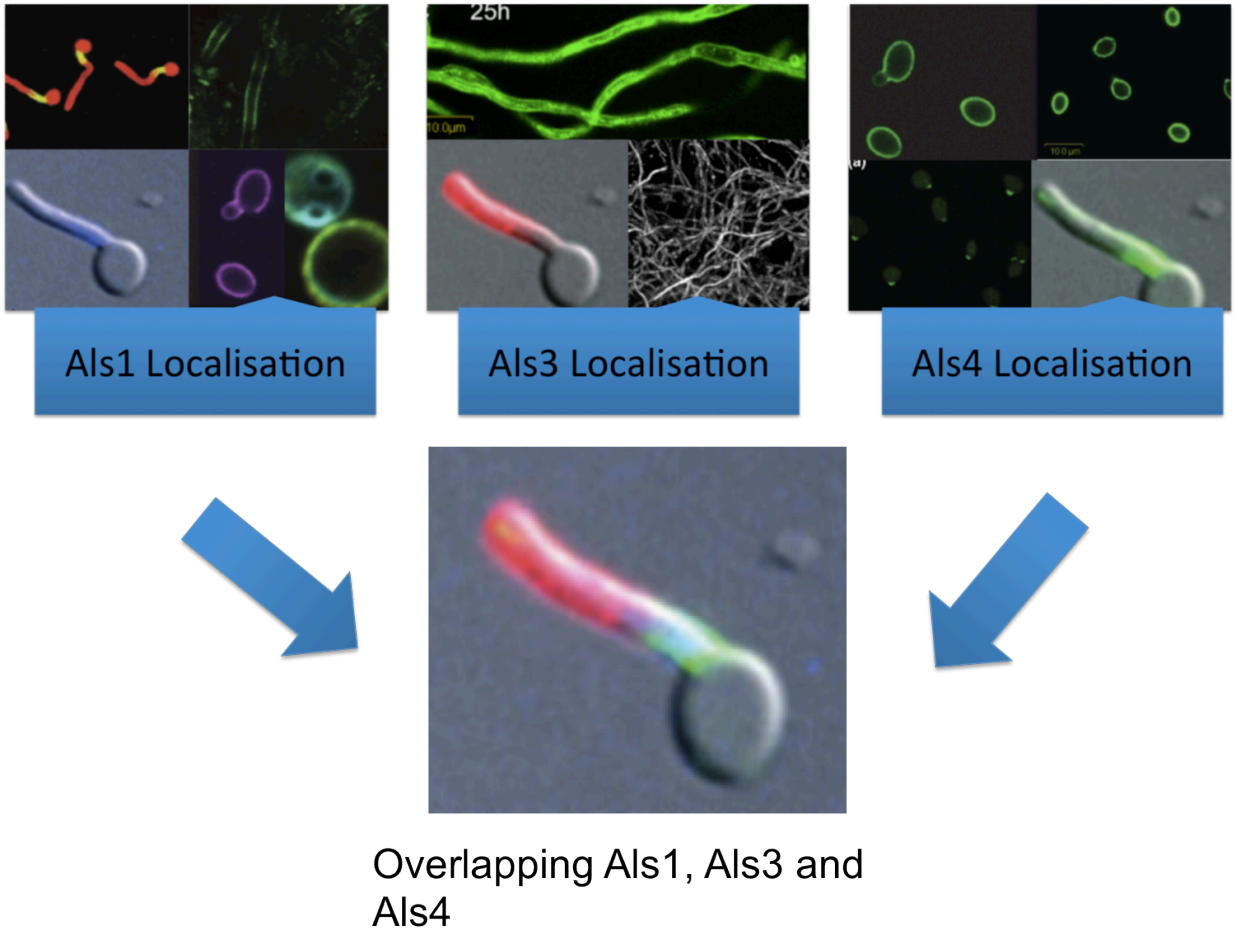


Figure 1.13: Localisation of Als1, Als3 and Als4 proteins on the *C. albicans* cell surface.
Als1 and Als4 are primarily localised to the yeast cell surface or at the base of the yeast-hyphae bud. Als3 is localised on hyphal germ tubes. The localization of these three proteins overlaps (Images from Fu et al 2002, Zhao et al 2006, Coleman et al 2009, Coleman et al 2010 and Coleman et al 2012).

Als4 expression seems to depend on the growth stage of *C. albicans*, specifically during early stages of colonisation (Hoyer et al 1998b). In a recent immunolabelling study, anti-Als4 antibodies showed that Als4 is expressed on the yeast cell surface and at the yeast-hyphae junction, similarly to Als1 (Fig 1.13). However, its expression is temperature dependent – it appearing on the cell surface when cells were grown at 30°C rather than 37°C. In mouse models of invasive candidiasis, Als4 could also be detected on hyphal germ tubes, but much less so than Als1 (Coleman et al 2012). This same elegant study showed that the expression of Als1, Als3 and Als4 on the cell surface is not restricted to discrete areas, but rather their protein localisation merges (Coleman et al 2012).

1.5.2.3. Diseases, Disease Models and ALS Gene Expression

The large number of studies of ALS expression in *C. albicans* disease models, or from cell samples taken from afflicted patients underlines their regular occurrence in disease and highlights them as important virulence factors. However, when observing Table 1.4, some ALS proteins are detected at a higher frequency than others in all of the studies. It appears that different stages of disease and also infection environments may produce variation in the expression of particular genes. For example, Green et al suggest that timescale of infection impacts on ALS gene expression patterns, with ALS1, ALS2, ALS3 and ALS4 being detected at early stages of infection, and ALS5, ALS6 and ALS9 at later stages (Green et al 2006). Other researchers put forward that frequently expressed genes (e.g. ALS1, ALS3 and ALS9) may be more important for *C. albicans* virulence, whilst other genes (e.g. ALS4 and ALS5) may function more during commensalism of the fungus (Cheng et al 2005).

Table 1.4: The detection of individual ALS genes from clinical samples or *C. albicans* disease models. ✔ – Gene expression detected. Low – Low expression of gene detected. ND – Not Detected.

DISEASE STUDY	ALS1	ALS2	ALS3	ALS4	ALS5	ALS6	ALS7	ALS9
Serum samples from human candidemia patients. Microarrays (Mochon et al 2010).	✔	ND	✔	ND	✔	ND	ND	✔
Samples from human HIV patients with OPC. RT-PCR (Green et al 2006).	✔	✔	✔	✔	✔	✔	ND	✔
Samples from human female VVC patients in Mexico. RT-PCR (Monroy-Perez et al 2012).	✔	✔	✔	✔	✔	✔	✔	✔
Samples from human VVC patients in USA. RT-PCR (Cheng et al 2005).	✔	✔	✔	Low	Low	Low	Low	✔
Reconstituted human buccal epithelial model. RT-PCR (Green et al 2004).	✔	✔	✔	✔	✔	Low	Low	✔
Reconstituted human vaginal epithelial model. RT-PCR (Cheng et al 2005).	✔	✔	✔	Low	Low	Low	Low	✔
Mouse Invasive candidiasis model, GFP-reporters & immunohistochemistry. (Green et al 2005b).	✔	✔	✔	✔	ND	ND	ND	✔
Mouse VVC model. RT-PCR (Cheng et al 2005).	✔	✔	✔	Low	Low	Low	Low	✔

Research on the contribution of particular ALS genes on *C. albicans* virulence has shown that mice infected with *C. albicans* null mutant strains for ALS1 and ALS3 in murine OPC and invasive candidiasis models, a reduction in mortality rate is seen (Kamai et al 2002, Tsai et al 2011). However, in two separate murine models of invasive candidiasis, it was reported that *C. albicans* ALS3 null mutant (yeast or hyphal) cells were just as virulent as the WT strain (Cleary et al 2011, Fu et al 2013). However, in one murine model of invasive candidiasis, the expression of ALS3 in *C. glabrata* (which does not have ALS orthologues) leads to greater fungal colonisation of the mouse brain and kidneys in comparison to a control *C. glabrata* strain (Fu et al 2013). It is clear that ALS proteins are important for the *C. albicans* infection process, but it seems apparent that *C. albicans* may counteract loss of particular ALS proteins by expressing others.

1.5.2.4. Biofilms and Als Adhesins

The role of Als proteins in the formation of biofilm was hinted at in early studies of *S. cerevisiae* and *C. albicans* cells expressing Als5 and aggregating following their adherence to ECM components (Gaur et al 1999). Als1 was shown to be important in flocculation – *C. albicans* cell-cell adherence (Fu et al 2002). The adherence of Als5 protein to ECM components was reported to result in an increase in *C. albicans* cell-surface hydrophobicity, leading to cell-cell aggregation (Rauceo et al 2004). *S. cerevisiae* expressing Als1 was demonstrated to aggregate and bind to peptide ligands in sticky clumps of cells (Klotz et al 2004).

In an in vitro microfermentor *C. albicans* biofilm model, the over-expression of ALS1 was detected (Garcia-Sanchez et al 2004), and in RT-PCR and qPCR studies of *C. albicans* biofilm grown on dentures and on silicone elastomer, a material regularly used to make hospital equipment showed significant up-regulation of ALS1 in biofilm associated cells as opposed to 'free' cells (O'Connor et al 2005, Nailis et al 2006, Nett et al 2010).

In similar studies of biofilm models, Als3 has also been highlighted as highly expressed and an important adhesin during biofilm formation (Murillo et al 2005, Uppuluri et al 2009). *C. albicans* ALS3 null mutants presented biofilms with a weaker structure and 50% reduction in biomass in comparison to the wild-type, which reverted to normal biomass when the ALS gene was re-introduced (Nobile & Mitchell 2005, Zhao et al 2006). However in an in vivo venous catheter biofilm model, it was reported that a *C. albicans* ALS3 null mutant could still form a biofilms (Nobile et al 2006a).

It has been suggested that the function of Als1 and Als3 (and Hwp1) proteins are interchangeable but still significant for biofilm formation, and also that Als proteins may elicit cell-cell aggregation by interacting with Hwp1 proteins on other *C. albicans* cells (Liu & Filler 2011). Biofilms formed from *C. albicans* null mutants of ALS1 and ALS3 have a weaker structure, and an ALS1 null mutant formed a highly disordered biofilm in comparison to the wild-type strain. ALS5, ALS6, ALS7 and ALS9 can partially rescue the effects of ALS1 and ALS3 deletion, yet ALS2 and ALS4 are unable to. Plus, a *C. albicans* null mutant of the BCR1 transcription factor that is unsuccessful at forming biofilm can be salvaged by re-introduction of either ALS1 or ALS3 genes (Braun & Johnson 2000, Nobile et al 2006a, Nobile et al 2008). Plus, both ALS1 and ALS3 are up-regulated when *C. albicans* cells come into contact with epithelial cells (Rosenbach et al 2010, Wächtler et al 2011). Finally, in a comprehensive RT-PCR study of various different biofilms, including those grown on silicone elastomer, on polyurethane in a reactor, on a reconstituted human epithelial model and on a rat subcutaneous catheter model, up-regulation of ALS1, ALS2, ALS3, ALS4 and ALS5 was confirmed (Nailis et al 2010).

There is some discussion regarding different stages of biofilm formation, and ALS gene expression. In a venous catheter biofilm model, ALS1 and ALS2 were expressed early in initial biofilm formation (Nett et al 2009). In a silicone biofilm model subjected to continuous flow conditions (in order to mimic a bloodstream environment), ALS1 and ALS3 expression was detected during biofilm, but suggested that ALS1 and ALS3 are important for initial stages of biofilm formation, and that ALS1 is down-regulated at the biofilm maturation stage (Nailis et al 2009). Similar results were reported in microarray studies of denture and catheter biofilm models (Yeater et al 2007). With regards to dispersion, one study of gene transcription patterns in a silicone elastomer biofilm model reported that ALS1 expression was down-regulated during detachment of *C. albicans* cells from biofilms (Sellam et al 2009). However, a recent study reported up-regulation of both Als1 and Als3 in dispersal cells (Bujdáková et al 2012).

Although studies contrast with regards to individual Als protein function in biofilms, it is confirmed that Als proteins are important for biofilm formation and have medical relevance. Furthermore, it has been demonstrated that most of the Als proteins (especially Als1 and Als3) bind to non-biological surfaces such as polypropylene, borosilicate glass and poly-vinyl chloride – materials often used to make medical devices such as stents and catheters (Aoki et al 2012). Plus, the Als proteins may be contributing to the increase of *C. albicans* resistance to current antifungal drugs via

their influence on biofilm formation. This aggregation may protect *C. albicans* cells from antifungal drug damage, as Als1 expression is also induced when cells are exposed to a range of antifungal drugs including ketoconazole and amphotericin B (Dhamgaye et al 2012).

As mentioned previously, another factor that can result in *C. albicans'* increased resistance to antifungals is the formation of mixed-species biofilms. *C. albicans* cells can aggregate with each other and host cells, and have also been shown to co-aggregate with other micro-organisms, especially *S. gordonii*. This particular mixed-species biofilm has implications in denture stomatitis, and mutation studies highlighted the interaction between *C. albicans* Als3 with *S. gordonii* SspB as being important for adhesion between the two species (Holmes et al 1995, Klotz et al 2007, Bamford et al 2009, Nobbs et al 2010, Silverman et al 2010). *C. albicans* can also co-aggregate with *S. aureus*, and it has recently been discovered using mutation studies, that Als3 and to a lesser extent Als1, are implicated in the adherence of *S. aureus* to *C. albicans* (Peters et al 2012).

1.5.2.5. Ferritin Adhesion with Als3 and Iron Acquisition

Als3 expressed on *C. albicans* hyphae was identified as important for the fungal acquisition of iron via binding to host ferritin. A *C. albicans* ALS3 null mutant had a significantly decreased ability to bind ferritin in comparison to wild-type cells. Plus, *S. cerevisiae* cells that normally do not bind ferritin, was found to adhere to it when expressing Als3, but not Als1 or Als5 (Almeida et al 2008). Hyphae seem to be essential for this iron acquisition from ferritin, yet a separate study showed that in an environment with reduced iron concentration in a VVC model resulted in *C. albicans* yeast cells as well as hyphal cells expressing Als3 proteins in a possible 'iron starvation' response (Sosinkska et al 2008, Almeida et al 2009).

1.5.2.6. Induced Endocytosis

The involvement of Als proteins in host-cell invasion was first documented when *S. cerevisiae* expressing Als1, Als3 and Als5 was able to induce endocytic uptake by endothelial cells, with Als3 inducing this at a higher rate than the others (Sheppard et al 2004). Als3 has been reported on multiple occasions to be the most prominent Als protein at inducing endocytosis and facilitating *C. albicans* invasion, as latex beads coated with the Als3 N-terminus were shown to induce endocytosis in oral epithelial cells and human umbilical vein endothelial cells by interacting with E-cadherin on epithelial cells and N-cadherin in endothelial cells (Phan et al 2005, Phan et al 2007). Additionally, *C. albicans* hyphae with deleted ALS3 that had also been killed (so they can only invade epithelial cells by induced endocytosis, not active penetration) were reported to have approximately 65% reduced invasion in comparison to wild-type *C. albicans* hyphae, emphasising importance of this protein in induced endocytosis. However, as invasion was not completely abrogated, other factors apart from Als3 (for example other Als proteins), may contribute to induced endocytosis (Wächtler et al 2011, Wächtler et al 2012).

A clathrin-mediated pathway carries out this process, as Als3-coated beads were endocytosed into epithelial cells only when clathrin was expressed. Plus, cell-imaging and immunofluorescence studies were used to follow the induced endocytotic process involving N- and E-cadherins, clathrin, dynamin, actin filaments, septin 7 and contactin (summarized in section 1.2.2) are also involved (Phan et al 2013). This mode of invasion by induced endocytosis was noted as similar to that employed by *Listeria monocytogenes* (Moreno-Ruiz 2009).

Apart from the cadherins as Als3 invasive endocytosis receptors, the epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor (HER2) cooperatively interact with Als3 and can trigger induced endocytosis of *C. albicans* cells. Interestingly, this process can be blocked by kinase inhibitors, which may have therapeutic potential (Zhu & Filler 2010, Zhu et al 2012).

Research on causes of Candidal brain infections in the context of candidemia focussed on the mechanism of *C. albicans* traversal into brain microvascular endothelial cells. The cell-surface heat-shock protein gp96, specific to brain endothelial cells, was elucidated as another receptor for Als3 mediated induced endocytosis. *S. cerevisiae* expressing Als3 was shown to adhere to gp96 in comparison to wild-type, and *C. albicans* ALS3 null mutants had a significantly reduced ability to induce endocytosis in brain endothelial cells. However, despite this reduction, and alike to other studies on induced endocytosis, some invasion was present, hinting at other factors (perhaps other Als proteins) or processes (active penetration) that may be important for *C. albicans* brain invasion (Liu et al 2011).

The rate of active penetration caused by *C. albicans* hyphae into cytochalasin D-treated epithelial cells was reduced if the hyphae were ALS3 null mutants in comparison to WT. As cytochalasin D is a toxin that disrupts the endocytic pathway in epithelial cells, this suggests that Als3 may stabilise hyphal contact, which could promote active penetration (Wächtler et al 2012).

1.5.3. Medical Significance

It is no surprise that Als proteins are interesting with regards to future medical treatment of *C. albicans* infections. The recombinant Als1 N-terminus was shown to be useful as a vaccine for immunocompetent and immunocompromised mice with invasive candidiasis, and significantly improved their survival – showing a 40-50% survival rate in those vaccinated in comparison to those injected with a control (Ibrahim et al 2005, Spellberg et al 2005). This Als1 vaccine has also been shown to be effective against other *Candida spp.*, including *C. glabrata*, *C. krusei*, *C. tropicalis* and *C. parapsilosis* (Ibrahim et al 2006).

A vaccine developed from recombinant Als3 N-terminus has also been shown to protect against invasive candidiasis to the same extent as the Als1 N-terminal vaccine, and seems better at protecting mice against VVC and OPC (Spellberg et al 2006). Interestingly, the Als3 N-terminal vaccine was also able to protect mice infected with methicillin-resistant *S. aureus*, possibly due to the homology of Als3 N-terminus with *S. aureus* clumping factor A (ClfA) protein (Spellberg et al 2008b). These studies with Als1 and Als3 vaccines also showed that the immune response against them was mainly mediated by T-cells (Ibrahim et al 2005, Spellberg et al 2006, Spellberg et al 20308a, Lin et al 2009).

Currently, the Als3-based anti-*Candida* vaccine, NDV-3, is being developed for medical use, and has recently been introduced to humans in a Phase I clinical trial with some success and no adverse effects. Success in Phase II clinical trials against VVC has also been recently reported (Spellberg et al 2008, Edwards Jr 2012, Cassone 2013, Ibrahim et al 2014).

The monoclonal antibody C7 may be used to treat *C. albicans* infections in the future and has been found to adhere to Als3 and block the adhesion of *C. albicans* to buccal epithelial cells, Als3-mediated iron acquisition and also hyphal formation (Brena et al 2007, Brena et al 2011). Similar blocking of adhesion is seen for another monoclonal antibody, 3D9.3 (Beucher et al 2009).

1.6. Structural and Functional Studies of Als Proteins and Their Ligand-Binding Behaviour

1.6.1. Initial Structural Studies and Bioinformatics Predictions

Initial clues regarding the structure of the adhesive N-termini of Als proteins came from noting that the Als1 N-terminus was similar in nucleotide and amino acid sequence to the N-terminus of *S. cerevisiae* mating adhesion α-agglutinin (Sag1), which was predicted to contain three IgV-like motifs based on circular dichroism, molecular modelling and peptide mapping studies (Wojciechowicz et al 1993, Hoyer et al 1995). Als5 was the subject of the first structural study of an Als protein, in which circular dichroism predicted that its N-terminal adhesion domain featured anti-parallel β-sheets, which they hypothesized to be characteristic of an immunoglobulin-like domain (Hoyer & Hecht 2001).

1.6.2. Previous Peptide Ligand Binding Specificity Studies

An early study of Als protein ligand binding behaviour revealed that *S. cerevisiae* cells expressing Als5 adhered to beads coated with ECM proteins and also to peptides comprised of 4-10 amino acids of alanine, serine or threonine, with 4 residues being the minimum-sized peptide capable of adherence. However, if lysine was present at the C-termini of peptides, no adherence was observed. Furthermore, single amino acids on beads did not promote adherence (Gaur et al 1999, Gaur et al 2002).

The differences in ligand-binding specificity between individual Als proteins was highlighted in a study expressing different Als proteins in *S.cerevisiae*, with some proteins binding certain substrates at a higher propensity than others, for example Als1 adherence to gelatin was greater in comparison to Als3. This study also showed that despite the high sequence identity and similarity between the N-termini of each Als protein, regions of sequence variation were perhaps responsible for the different ligand-binding behaviours (Sheppard et al 2004).

Research into the peptide-binding behaviour of Als proteins gave further clues about different ligand binding specificities. The binding of *S. cerevisiae* cells expressing Als1 or Als5 to particular heptamers on beads was tested, with the finding that there was some redundancy yet some differences in peptide sequences adhered to by either Als1 or Als5. It was theorized that peptide ligands with a structural motif ‘τφ+’ – representing a turn-like residue, a bulky hydrophobic residue and positively charged Lysine/Arginine – were more able to interact with Als1 and Als5 (Klotz et al 2004).

1.6.3. Structural and Biophysical Studies of Als Adhesins

Recent crystallographic and NMR structural information showed that the N-terminus of Als9-2, Als3 and Als1 are comprised of two immunoglobulin-like domains. The 33kDa N-terminus of Als9-2 (NT-Als9.2, amino acids 1-314, not including signal sequence) in ligand free form and in complex with a peptide from the C-terminus of human fibrinogen-γ (Fib-γ-CT) have been solved using X-ray crystallography to a resolution of 2 and 1.5Å, respectively (Salgado et al 2011a, Salgado et al 2011b) (Fig 1.14). Yan et al obtained the backbone and side chain methyl (Ile, Leu, Val) assignments of the 33.5kDa Als1 N-terminus using NMR spectroscopy (NT-Als1, amino acids 1-315) (Yan et al 2010). Finally, recent structural studies of both the N-terminus of Als3 (NT-Als3, amino acids 1-315) and a shortened N-terminal construct of Als3 (sNT-Als3, amino acids 1-302) with and without a threonine heptapeptide (Hepta-T) added valuable knowledge on the structures of Als protein N-termini and their ligand binding specificity (Lin et al 2014) (Fig 1.14).

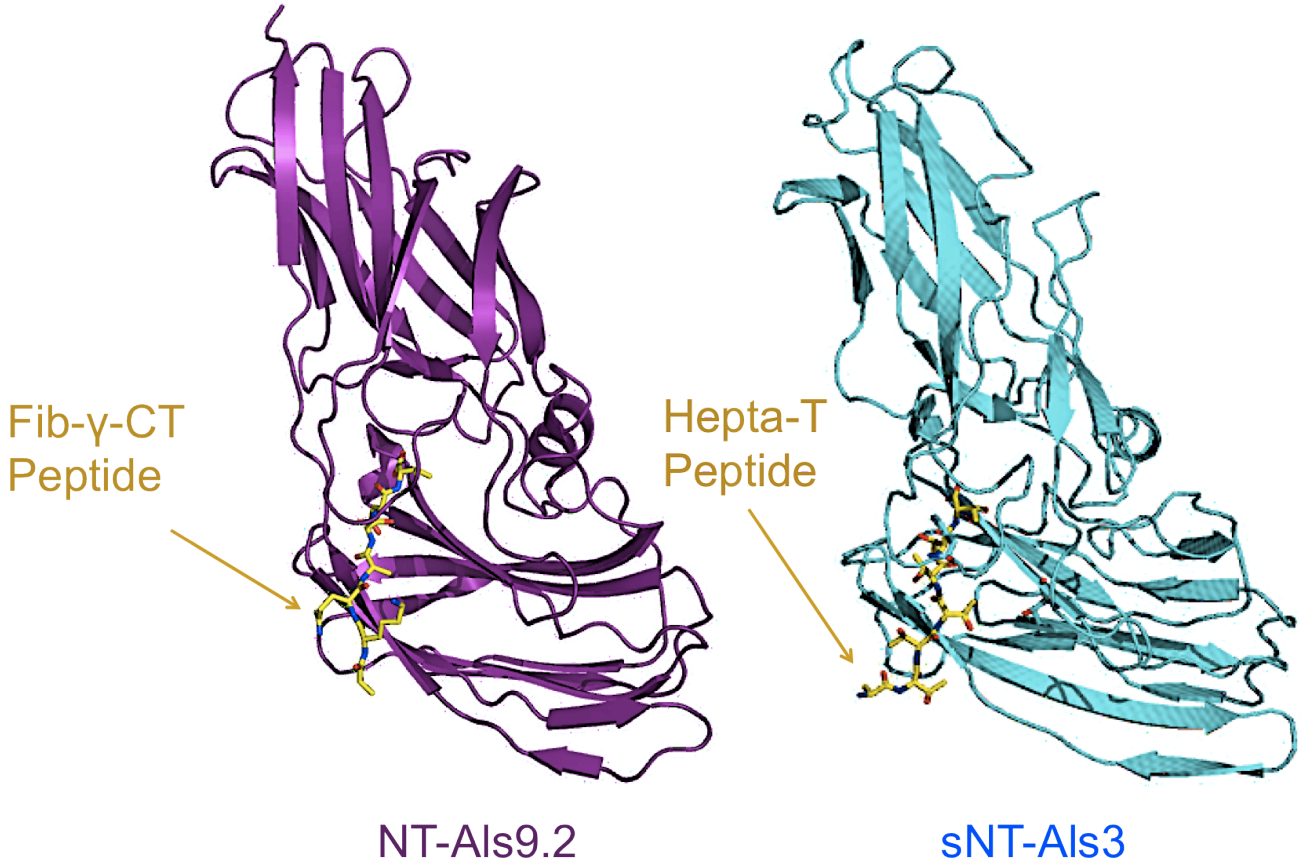


Figure 1.14: Crystal structures of NT-Als9.2 and sNT-Als3 with peptide ligands.
These figures show the overall structure of one NT-Als9.2 adhesion domain (amino acids 1-314, coloured violet) with a bound peptide from the C-terminus of human fibrinogen-γ (Fib-γ-CT, coloured yellow), and the overall structure of one sNT-Als3 adhesion domain (amino acids 1-302, coloured cyan) with a polythreonine heptapeptide (Hepta-T, coloured yellow). Images made using PyMOL, and with structures from PDB codes 2Y7L (NT-Als9.2/Fibγ-CT) and 4LEB (sNT-Als3/Hepta-T) (Salgado et al 2011 and Lin et al 2014).

These studies established that the N-termini of the Als proteins have a peptide binding cavity (PBC) that enables them to interact 'promiscuously' with the C-terminal carboxylates of many peptide ligands. A conserved K59 residue at the innermost end of the cavity, on the C1* strand, is mainly responsible for these ligand-binding interactions, with its amine-side chain forming a salt bridge with the C-terminal carboxylate of the peptide ligand (Salgado et al 2011) (See Figure 1.15).

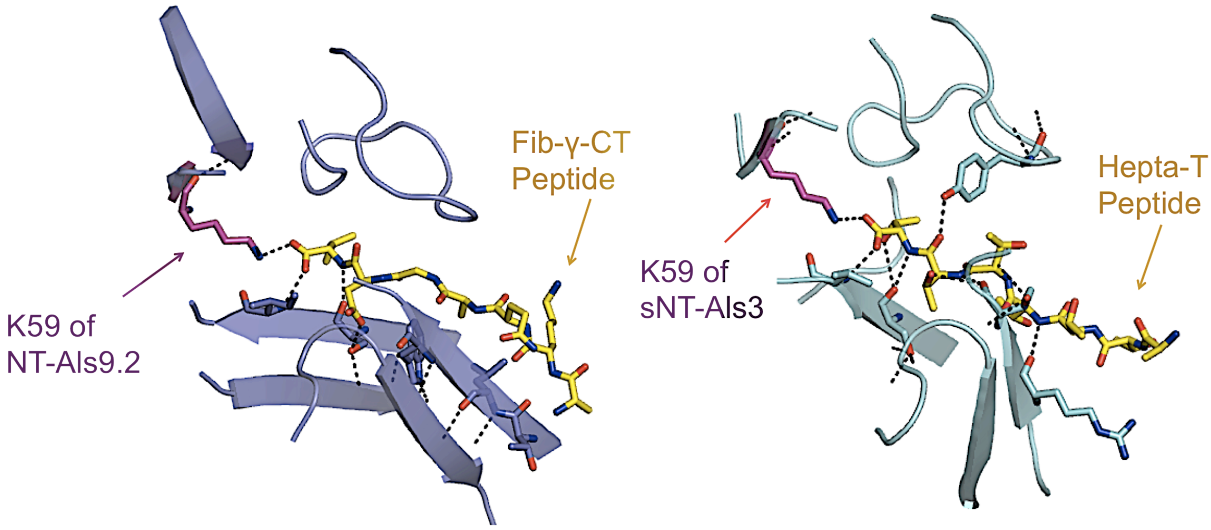


Figure 1.15: The invariant K59 in NT-Als9.2 and sNT-Als3 Crystal Structures.
The invariant K59 (highlighted in magenta), conserved in all Als proteins, has been shown to be important for the salt-bridge interaction of NT-Als9.2 (violet) with its fibrinogen-γ C-terminal peptide ligand (yellow), and for the interaction of sNT-Als3 (cyan) with its Hepta-T ligand (yellow). Images made using PyMOL, and with structures from PDB codes 2Y7L (NT-Als9.2/Fibγ-CT) and 4LEB (sNT-Als3/Hepta-T) (Salgado et al 2011 and Lin et al 2014).

The function of K59 was confirmed by studies demonstrating that peptides with an amidated C-terminus blocked their interaction with Als1 and Als9.2 (Salgado et al 2011). In addition, ^{1}H-^{15}N TROSY-HSQC NMR experiments of NT-Als1 titrated with Hepta-T (N-TTTTTTT-C) peptide showed chemical shift changes for a peak assigned as K59. This residue may also have a structural role, as mutation to either methionine or valine leads to aberrant folding in this domain (Salgado et al 2011). These NMR titration experiments also suggested that the dissociation constant (K_D) of the Als protein-peptide interaction was approximately less than 10μM (Salgado et al 2011). However, ITC experiments on the interaction of Als1 and Als3 with Hepta-T are at present being carried out in order to determine specific K_D values (Lin et al 2014).

There are gaps between residues of the PBC and the bound peptide, in which water molecules can reside and contribute to a network of hydrogen-bond interactions (Fig 1.16). This means that there is some flexibility regarding the shape and size of the peptides bound with regards to the side chains that interact with this water molecule network (Salgado et al 2011, Lin et al 2014). This contribution of water molecules to binding versatility is akin to that seen in the crystal structures of the *Salmonella typhimurium* adhesion protein OppA in complex with a variety of peptides (Tame et al 1994), and also in the molecular modelling of the hydration network shown in the crystal structure of *S. aureus* adhesion protein CNA bound to a collagen-like peptide (Vigagliano et al 2011).

Within the binding pockets of OppA and CNA as well as the Als protein PBC, water molecules can plug the gaps between ligand and binding cavity side-chains that are spaced too far away in Angstrom length for binding to occur. By acting as a bridge between such side-chains, and also those that could interact unfavourably, they can accept or donate hydrogen bonds, or negate unfavourable charge clashes. In this way, the hydration of protein binding pockets can compensate for weak specificity of interactions (Tame et al 1996). However, despite such binding flexibility, the Als PBC is restricted in its capacity to bind a peptide of only six amino acids in length (Salgado et al 2011, Lin et al 2014).

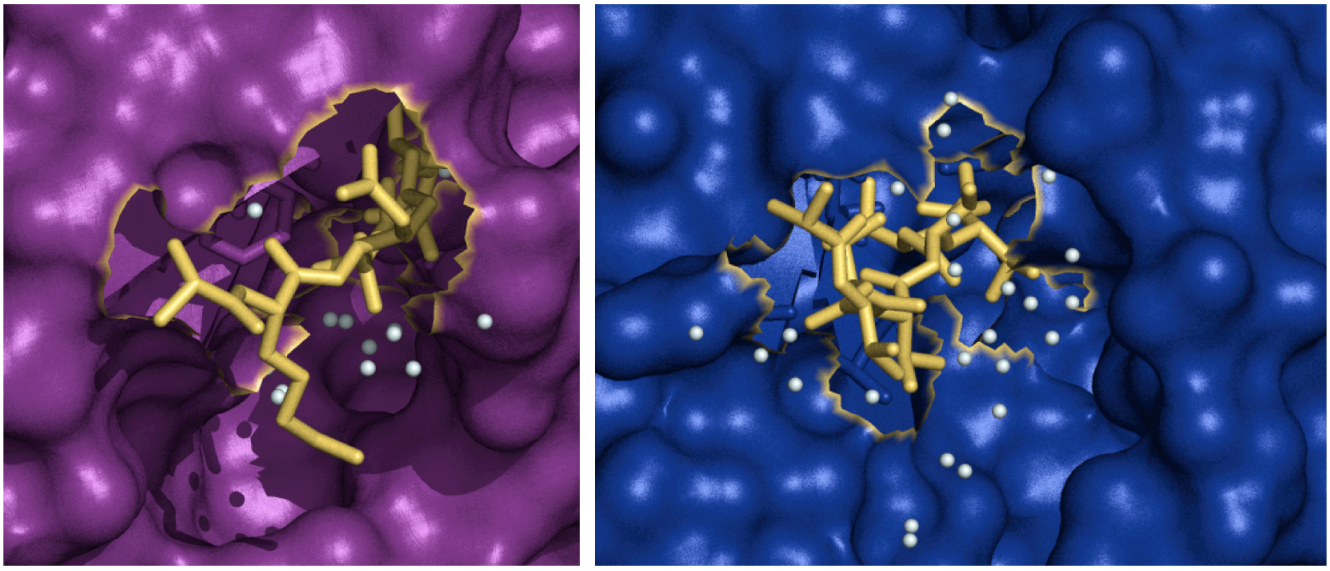


Figure 1.16: The PBC water molecules networks in both NT-Als9.2/Fib-γ-CT and sNT-Als3/Hepta-T Crystal structures.
Water molecules are depicted with the pale cyan circles, and peptide ligands are shown as yellow stick diagrams. The water molecule network bridges gaps between the peptide ligand and the residues of the PBCs in both NT-Als9.2 (violet) and sNT-Als3 (blue). Images made using PyMOL, and with structures from PDB codes 2Y7L (NT-Als9.2/Fibγ-CT) and 4LEB (sNT-Als3/Hepta-T) (Salgado et al 2011 and Lin et al 2014).

Contributing to the rest of the PBC structure is a loop between β-strands A and B in the N1 domain, named the A1-B1 loop, which acts as the 'cap' of the cavity, whilst the G2b and A2 strands are the 'walls' (Salgado et al 2011). As well as K59, key residues that were predicted to contribute to the interaction of Als proteins via direct hydrogen-bond interactions with their respective peptide ligand are listed in Table 1.5. Such hydrogen-bonds are either direct contacts between Als PBC residues and the peptide ligand or those mediated by water molecules. These key residues are also depicted in Figure 1.17 for Als9.2 and Figure 1.18 for Als3 (Salgado et al 2011, Lin et al 2014). Specific atomic details regarding the interactions are discussed in Chapter 5.

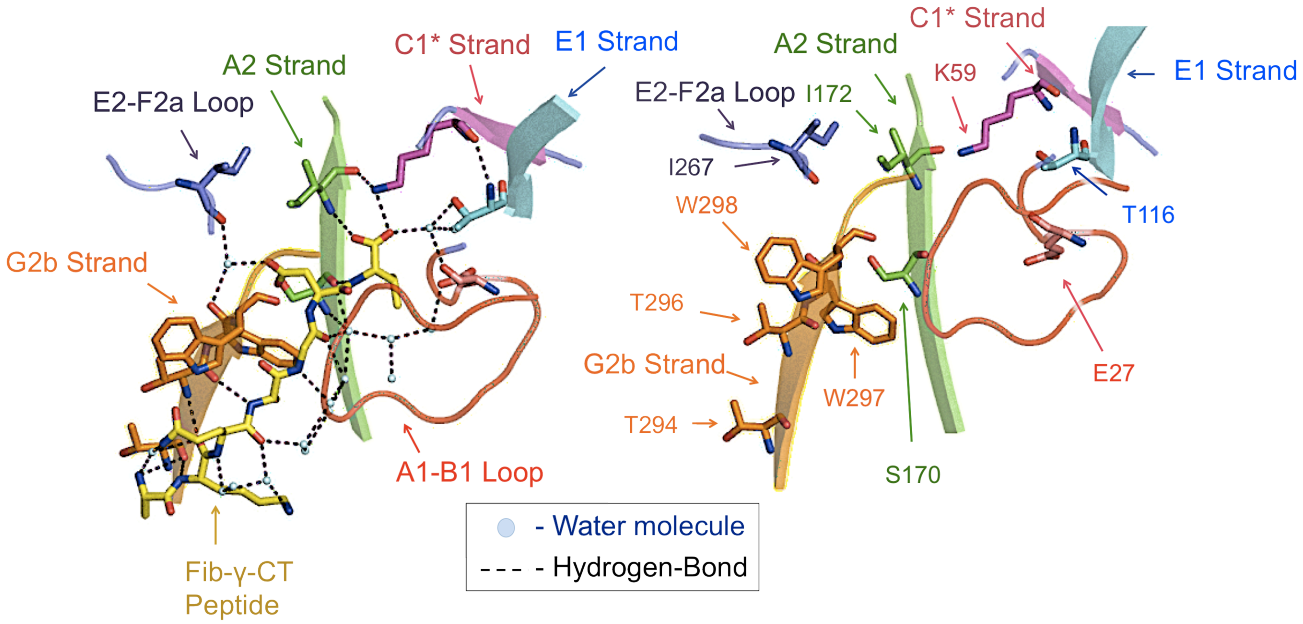


Figure 1.17: The NT-Als9.2/Fib-γ-CT Interaction Up Close.
On the left, the complexity of the NT-Als9.2/Fib-γ-CT is depicted, with labelled secondary structure elements of which the PBC is comprised. The pale cyan spheres are water molecules. Dashed lines show the hydrogen bonds occurring between the atoms of the NT-Als9.2 PBC amino acids with the Fib-γ-CT peptide and/or with water molecules. On the right, the specific amino acids involved in these hydrogen bonding interactions, are labelled. Figure made using PyMOL, using the structure of PDB code 2Y7L (Salgado et al 2011).

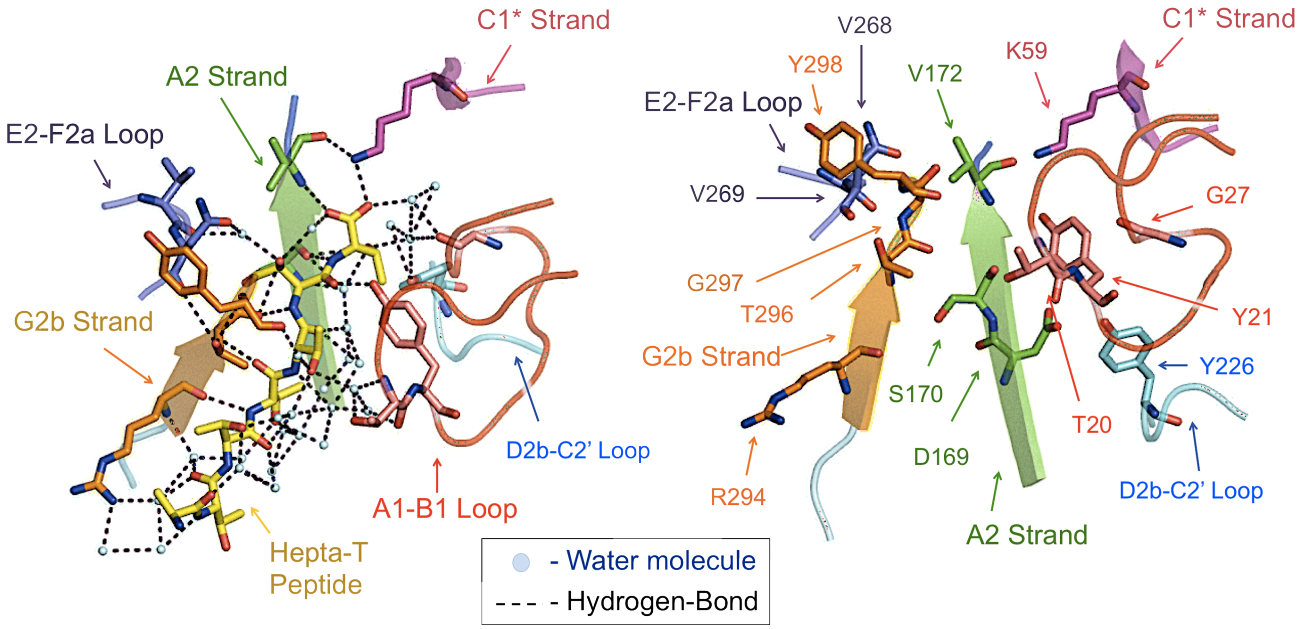


Figure 1.18: The sNT-Als3/Hepta-T Interaction Up Close.
On the left, the entire interaction of sNT-Als3 with Hepta-T is depicted, with labelled secondary structure elements of which the PBC is comprised. The pale cyan spheres are water molecules. Dashed lines show the hydrogen bonds occurring between the atoms of the sNT-Als3 PBC amino acids with the Hepta-T peptide and/or with water molecules. On the right, the specific amino acids involved in these hydrogen bonding interactions, are labelled. Figure made using PyMOL, using the structure of PDB code 4LEB (Lin et al 2014).

Table 1.5: The key residues of NT-Als9.2 and sNT-Als3 that are predicted to contribute to their interactions with their respective peptide ligands via either direct hydrogen bonds or via water molecules. These residues were pinpointed using the program Ligplot+ (Lasowski and Swindells 2011), using structural models from PDB codes 2Y7L (NT-Als9.2/Fibγ-CT) and 4LEB (sNT-Als3/Hepta-T) (Salgado et al 2011 and Lin et al 2014). The boxes coloured grey indicate that a protein-peptide interaction has not been predicted for that particular amino acid. However, since the peptide ligands used in the structural ligand-binding studies of NT-Als9.2 and and sNT-Als3 were different, and also since the protein-peptide interaction is likely to be dynamic, it is not entirely certain that the amino acids with grey boxes do not contribute to Als protein-peptide binding.

Protein Location	Residue Number	Amino acid in Als3 and Bond Type		Amino acid in Als9.2 and Bond Type	
A1-B1 Loop	20	T	H-Bond via Water	E	
	21	Y	Direct H-Bond	Y	
	27	G	H-Bond via Water	E	H-Bond via Water
C1* Strand	59	K	Salt-bridge	K	Salt-bridge
E1 Strand	116	A		T	H-Bond via Water
A2 Strand	169	D	H-Bond via Water	S	
	170	S	Direct H-Bond plus H-Bond via Water	S	Direct H-Bond plus H-Bond via Water
	172	V	Direct H-Bond	I	Direct H-Bond
D2b-C2' Loop	226	Y	H-Bond via Water	P	
E2-F2a Loop	267	N		I	H-Bond via Water
	268	V	H-Bond via Water	D	
	269	N	H-Bond via Water	K	
G2b Strand	294	R	Direct H-Bond plus H-Bond via Water	T	Direct H-Bond
	296	T	Direct H-Bonds	T	Direct H-Bonds
	297	G	Hydrophobic Interaction	W	H-Bond via Water
	298	Y	H-Bond via Water	W	Direct H-Bonds

Additional residues predicted using Ligplot+, to stabilize the PBC-peptide interaction in NT-Als9.2 via hydrophobic interactions (and not via hydrogen-bonds) include V19, E20, Y21, V22, P29 of the A1-B1 loop, R171 of the A2 strand and L295, L179 of the B2 strand, G299 and Y301 of the G2b strand (Lasowski & Swindells 2011, Salgado et al 2011). For sNT-Als3, residues that were predicted to contribute hydrophobic interactions (not via hydrogen-bonds) to the PBC-peptide contact include A19, N22, G27 and T28 of the A1-B1 loop, T61 of the C1* strand, A116 of the E1 strand, R171 of the A2 strand, and W295 and G297 (Lasowski & Swindells 2011, Lin et al 2014).

A crystal structure of NT-Als9.2 shows that in the absence of peptides in solution, the flexible C-terminus of the protein can be recognised by the PBC as a ligand by an adjacent NT-Als9.2 domain in the crystal lattice (Salgado et al 2011). In an attempt to stop this effect, G299 was mutated to W299, in order to decrease the flexibility of the C-terminal strand. Unexpectedly, this led to NT-Als9.2 molecules capable of recognising their own C-terminus as ligand (See Figure 1.19). This intra-molecular interaction was also confirmed with NMR spectroscopic studies of NT-Als1, in which NOEs revealed that C-terminal V315 and A316 were in close proximity to residues in the PBC, whereas other parts of this C-terminal section (residues 305-310) are highly flexible in solution (Yan et al 2010, Salgado et al 2011).

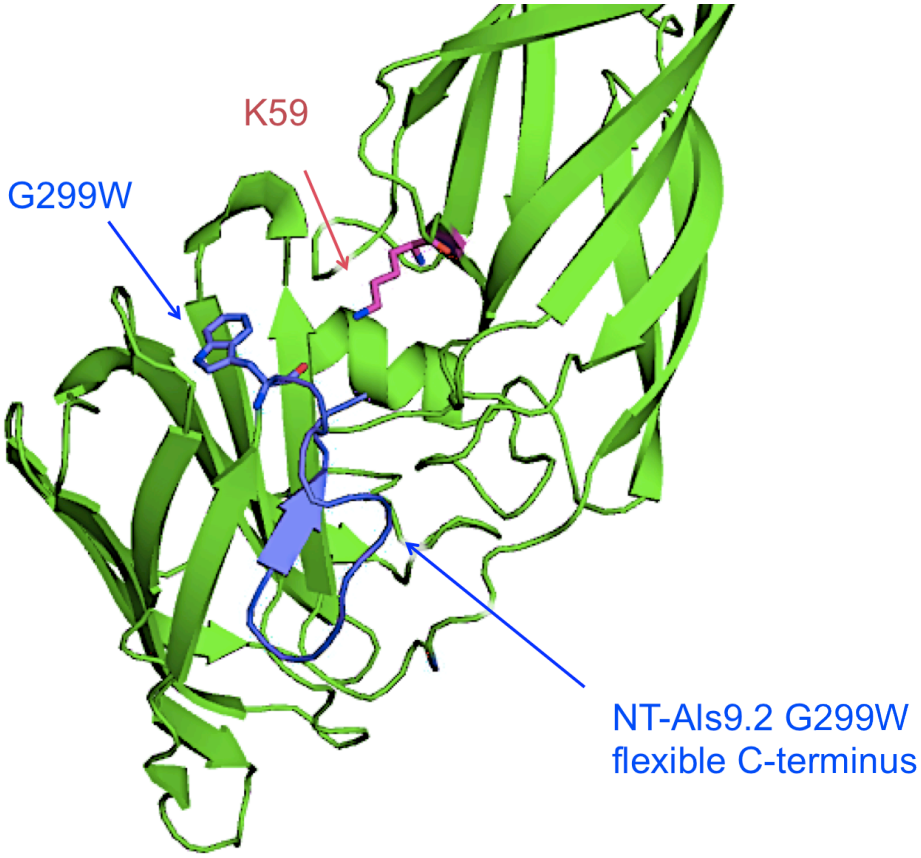


Figure 1.19: The Interaction of NT-Als9.2 G299W Mutant with its own Flexible C-terminus.
Residue G299 was mutated to a tryptophan in order to decrease the flexibility of the C-terminus (amino acids 299-314, shown in magenta). However, this resulted in the K59 of the NT-Als9.2 PBC interacting with its own C-terminus (Salgado et al 2011, PDB code (2YLH).

Structural studies on NT-Als3 have been carried out, but a shorter construct was produced, referred to as sNT-Als3 (amino acids 1-302), which lacks both the flexible C-terminal region and AFR (Lin et al 2014).

Additional structural and functional studies of Als proteins focussed on Als3, due to the relevance of this protein in the adhesive and invasive properties of *C. albicans*. Based on the structural details of NT-Als9.2, loss-of-function mutants were created in the NT-Als3 PBC and AFR. A PBC triple mutant (K59M, A116V, Y301F, or NT-Als3-pbc), replaces the hydrophilic contacts at the end of the cavity necessary for ligand binding and leaves an empty, inactive PBC. Interestingly, this mutant also stabilises the 'ligand bound' conformation of the protein in absence of ligands, meaning that the C-terminal region containing the AFR is attached to the N1 domain as an additional β-strand (Lin et al 2014).

Residues 311-315 in Als3 (-IVIVA-) contain an amyloidogenic region predicted by the TANGO algorithm (Fernandez-Escamilla et al 2004). The AFR double mutant (I311S, I313S, or NT-Als3-afr) abolishes the amyloidogenic potential of this region and can be used to monitor its involvement to the aggregative/adhesive properties of *C. albicans*. Bidimensional NMR spectra were used to confirm that these mutations do not affect the overall structure and folding of the adhesion domain. Subsequently, these mutations were integrated into full-length Als3 and expressed in *C. albicans* ALS3-null strains. Relative to the *C. albicans* WT, the Als3-pbc strain had reduced binding capability to buccal epithelial, human umbilical vein endothelial and FaDu cells. Importantly, these levels of adhesion are identical to those observed for the ALS3-null strain, indicating that mutation of the PBC is responsible for the contribution of Als3 to host cell adhesion (Lin et al 2014).

The Als3-afr strain led to an increase in *C. albicans* adhesion to FaDu cells and human umbilical vein endothelial cells in comparison to the ALS3 WT control strain. This was an unexpected result, given the proposed ability of the AFR to induce host cell adhesion of *C. albicans* cells (Lipke et

al 2012). To explain this effect, the authors hypothesise that removing amyloidogenic aggregates promoted by the AFR may increase access of the PBC to ligands on the host cell surface (Lin et al 2014). Also observed was a reduction in the formation of *C. albicans* 'aggregates' after initial binding to these substrates, indicating that the AFR is important for *Candida-Candida* contacts, but not specific interactions with host cells (Lin et al 2014).

Further evidence for the involvement of the PBC in specific ligand binding was provided by a 'gatekeeper' (gk) mutant. Serine-170 is located at the entrance of the PBC and mutation to a larger side chain should impair access to ligand into the end of this cavity. ITC and bidimensional NMR spectra show that, unlike the WT protein, the purified S170Y mutant (NT-Als3-gk) is unable to bind model peptides such as Hepta-T. This result reinforces the importance of the PBC in specific ligand interactions and that its function is independent from the AFR (Lin et al 2014). The *C. albicans* strain Als3-gk shows host cell binding properties that closely resemble the functions of the Als3-pbc strain (Lin et al 2014).

1.7. Project Aims and Objectives

The primary aim of this work was to solve the X-ray crystallographic structure of the Als1 N-terminus in ligand-free form and in complex with ligands. As previous attempts to crystallise the full-length NT-Als1 domain were unsuccessful, my efforts focussed on a short NT-Als1 construct (sNT-Als1) lacking thirteen flexible C-terminal amino acids in this domain containing the conserved AFR. With this structure, we expect to gain an accurate understanding of the ligand-binding specificity differences between the Als1, Als3 and Als9.2 N-termini. This structure will also facilitate the formation of sNT-Als1 complexes with a variety of ligands, including small chemical fragments that can be used in the design of molecules that inhibit the function of Als proteins.

Another aim of this work was to complete the backbone resonance assignment of sNT-Als1 and sNT-Als3 using NMR spectroscopy. Although the previous assignment of NT-Als1 provided valuable information regarding its interactions with peptides, the flexible C-terminus included in the construct made it difficult to delineate the specific surface involved in the formation of the binding site. In combination with crystallographic data, resonance assignments will be used for future ^{1}H-^{15}N-TROSY-HSQC NMR chemical shift perturbation analyses for the identification of the binding site for peptides and/or small chemical fragments in solution.

A final aim was to perform high-throughput ligand-binding screens of C-terminal peptides from various biologically interesting ligands, and also potential drug-molecules, in order to test for their interaction with Als1. This was to elucidate interesting binding partners for sNT-Als1 and any potential lead compounds for further drug design.

2. Structural Biology Theory

2.1. X-Ray Crystallography

X-ray crystallography analyses the diffraction patterns produced after directing an X-ray beam at a macromolecular crystal. By interpreting the intensities, phases and positions of diffraction spots, it is possible to create a three-dimensional model of the atomic structure of the macromolecule via electron density maps built from the diffraction data (Rhodes 2000). As the majority of structures deposited in the PDB have been solved using X-ray crystallography, it is still referred to as the 'gold standard' of structural determination. However, a prerequisite of this technique is to produce a crystal of your macromolecule, which in this project is a globular protein.

2.1.1. Protein Crystallization

A purified protein at a high concentration (in mg/ml scale) and in the correct conditions should separate from its solution and assemble itself into a highly ordered lattice that is the protein crystal. The crystallisation process is affected by protein concentration, protein purity, pH, temperature, solvent, precipitant, ions, ligands, reagents and additives used, and also the technique used to produce crystals. Frequently, vapour-diffusion methods are used to grow protein crystals – these include hanging- and sitting-drop, and involve mixing a protein solution in a 1:1 ratio with crystallisation solution in a sealed container, in which a reservoir of the crystallisation solution with a higher concentration of precipitants acts to absorb water from the protein-solution droplet (Fig 2.1) (Rupp 2009). Crystal plates are normally incubated at either 4°C or 22°C in an environment away from vibrations.

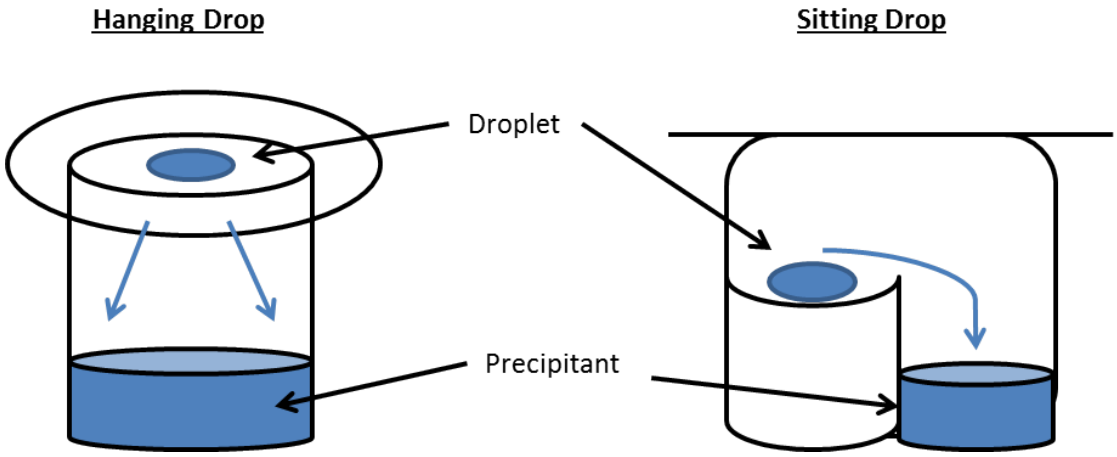


Figure 2.1: Vapour diffusion crystallisation techniques hanging drop and sitting drop.
These techniques involve a droplet of concentrated protein in solution is mixed in a 1:1 ratio with crystallisation and separated from a reservoir of crystallisation solution in a sealed container. The diffusion of water molecules from the droplet occurs due to the higher concentration of precipitant in the reservoir in comparison to the droplet, resulting in super-saturated proteins solution which may be able to form crystals (Rupp 2009).

The protein solution becomes 'super-saturated', and if the protein aggregates to form nucleation site, then the protein has reached the nucleation zone, or labile phase. Crystal growth will proceed in favourable, or 'metastable' conditions (Fig 2.2) (Rupp 2009). Ideally crystals of approximately 0.1-0.3mm in length and diameter will grow, with clear and defined geometric shapes to be useful for X-ray diffraction.

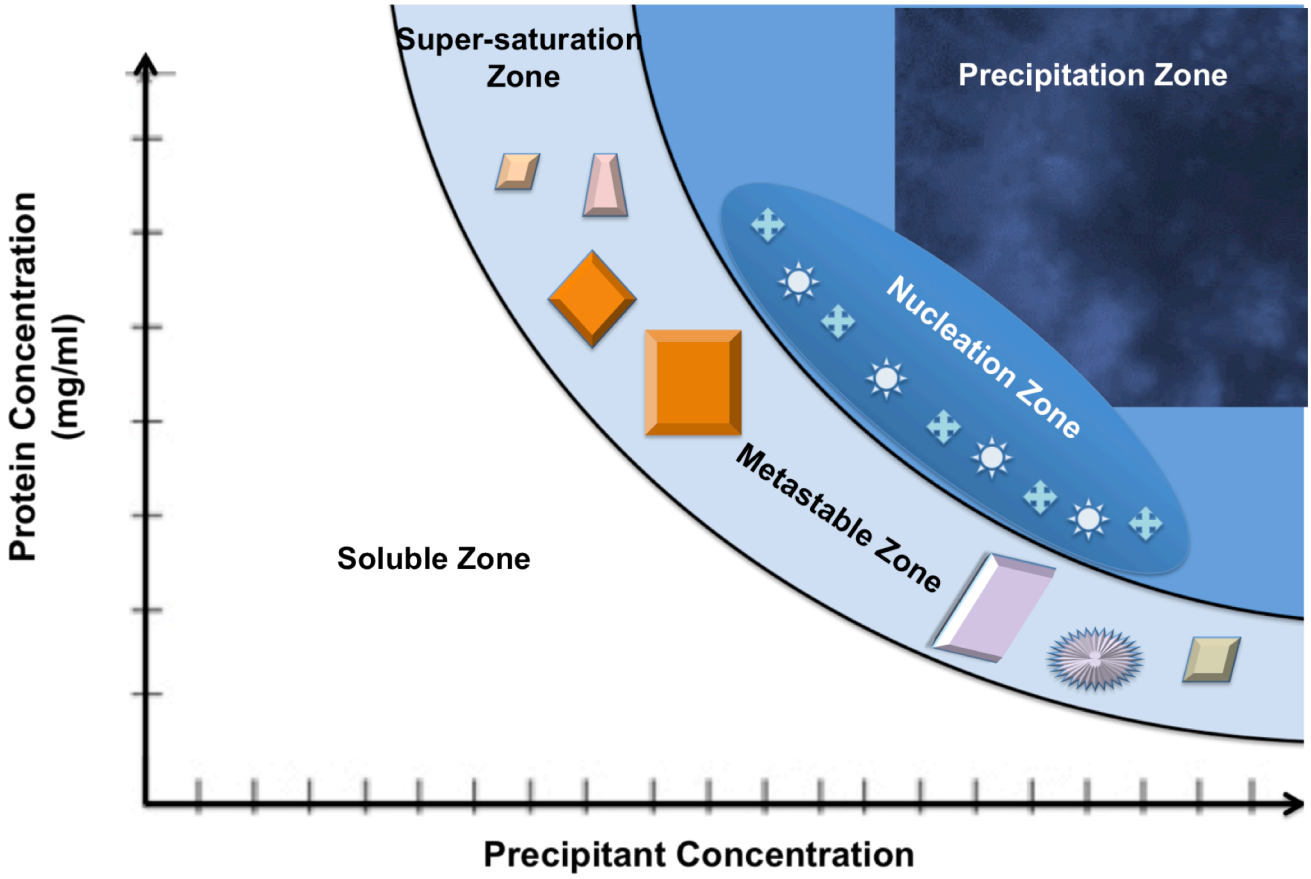


Figure 2.2: The zone-diagram relationship between protein solubility and precipitant concentration.
The relationship between protein solubility and precipitant concentration can be illustrated using a zone diagram. At high protein and precipitant concentrations, protein may form non-diffractable precipitate, as conditions favouring nucleation have not been reached. However, by achieving the appropriate ratio of protein and precipitant concentration, and usually including the use of additives or pH modification, nucleation and metastable zones can be generated in which crystals can form (Rupp 2009).

2.1.2. Mathematical 'Artwork; – Crystals Up Close

2.1.2.1. The Unit Cell

The molecular 'building-blocks' of crystals are in an ordered, regular arrangement termed a 'crystal lattice'. 'Lattice points' in the crystal have identical geometrical surroundings, and when these are grouped into a 'box', they can form the repeating-unit of the crystal known as the 'unit cell' (Fig 2.3). Each unit cell contains exactly the same contents in terms of translational symmetry, and is the simplest 'building block' of the crystal lattice, in that it could create the entire crystal using symmetry operations. Zooming in further, a unit cell can usually be divided by two or more asymmetric units. An asymmetric unit is the smallest repetitive unit from which the crystal is constructed. Asymmetric units in each unit cell are symmetrically related, in that they are identical if reflected, rotated or inverted (Rhodes 2000).

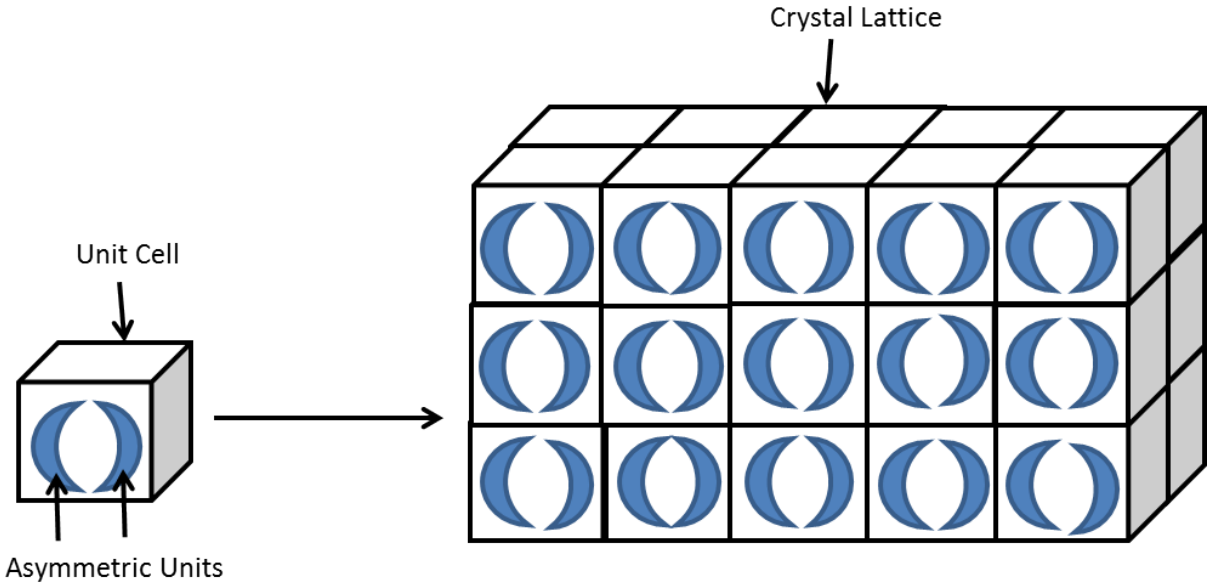


Figure 2.3: **The unit cell - a building block of the crystal lattice.**
The unit cell itself contains the smallest quantity of components, which can specify the assembly of the entire crystal via symmetry operations. Within the unit cell, two or more asymmetric units can be found, in that one asymmetric unit is related to the other via symmetry, for example reflection, rotation or inversion (Blow 2005).

The unit cell can be defined by the parameters of axes, angles and faces (Fig 2.4). It has axes a, b and c; the a-axis is horizontal, the c-axis is vertical and perpendicular to these is the b-axis. In between these axes, there are the angles α, β and γ; the α angle is between the b and c axes, the β angle is between 'a' and 'c' axes and γ is between 'a' and 'b' axes. In terms of faces, the A faces are perpendicular to a-axis, the B faces perpendicular to the b-axis and the C faces perpendicular to the c-axis. The lengths of the unit cell axes can be related to a three-dimensional Cartesian plot where 'a' corresponds to 'x', 'b' corresponds to 'y' and 'c' to 'z' (Rhodes 2000).

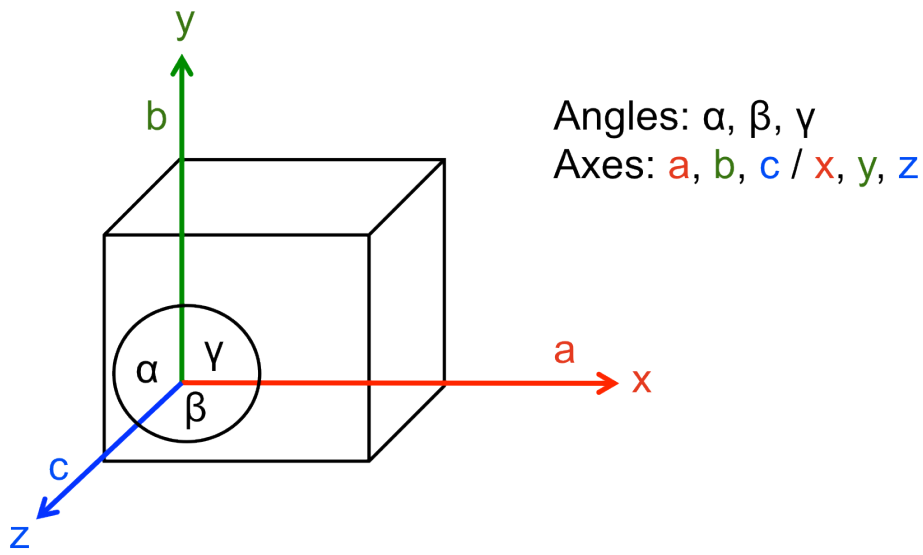
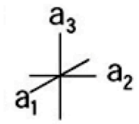
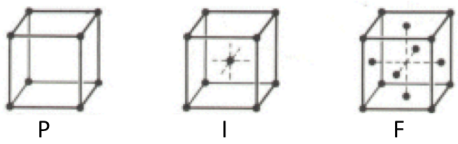
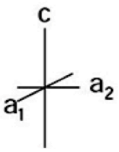
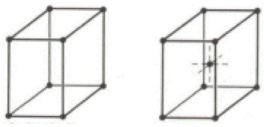
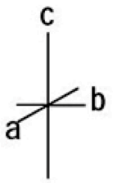
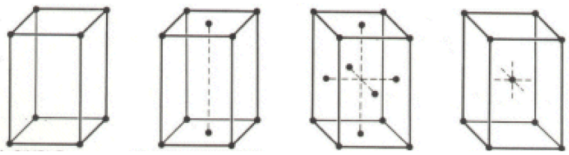
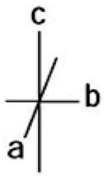
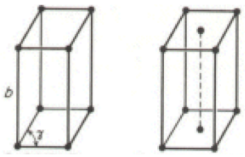
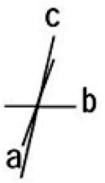
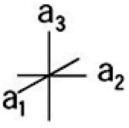
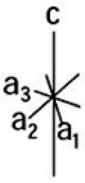
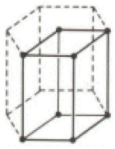


Figure 2.4: The dimensions of a unit cell.
A unit cell can be terms in terms of angles α, β and γ (x, y and z) and the axes a, b and c. The angles and axes will determine the shape of the unit cell and also the shape of the overall crystal formed (Blow 2005).

2.1.2.2. The Seven Crystal Systems and 14 Bravais Lattices

The unit cell shapes are limited depending on the sizes of the angles and axes, with there being seven different types of unit cell. These seven types are known as the crystal systems; cubic, tetragonal, orthorhombic, hexagonal, trigonal, monoclinic and triclinic.. For instance cubic unit cells contain all 90° angles and axes of equal length, tetragonal unit cells resemble a rectangular shape (See Table 2.1) (Rupp 2009).

Table 2.1: **The 7 Crystal Systems and 14 Bravais lattices**.
These crystal systems and Bravais lattices are defined by angles, size of their axes and the positions of their lattice points (Rupp 2009).

Crystal System	Unit Cell Axis Lengths and Angles	Unit Cell Geometry and Bravais Lattices
Cubic (Isometric)	α, β, γ = 90° a = b = c ($a_1 = a_2 = a_3$) 	
Tetragonal	α, β, γ = 90 a = b ≠ c ($a_1 = a_2 \neq c$) 	
Orthorhombic	α, β, γ = 90° a ≠ b ≠ c 	
Monoclinic	α, Y = 90° β ≠ 90° a ≠ b ≠ c 	
Triclinic	α, β, γ ≠ 90° a ≠ b ≠ c 	
Trigonal (Rhombohedral)	α, β, γ ≠ 90° a = b = c ($a_1 = a_2 = a_3$) 	
Hexagonal	α, β = 90° γ = 120° a = b ≠ c ($a_1 = a_2 = a_3 \neq c$) 	

There are also different types of crystal lattices in terms of lattice points. Primitive lattices (P) are described in all crystal systems and have their lattice points only at the unit cell corners. Body-centred (I) lattices are found in cubic, tetragonal, orthorhombic and monoclinic systems, and are similar to primitive in that their points are also at unit cell corners, but also they have one in the unit cell centre. Face-centred (F) lattice types have again lattice points at their corners, but one point in the centre of every unit cell face, seen in cubic and orthorhombic systems. There are also base-centred lattice types (C) that are only found in monoclinic systems, with lattice points at unit cell corners and in the centre of both C-faces. As well as the seven crystal systems, these lattice point types form the 14 Bravais lattices (see Table 2.1) (Rupp 2009).

2.1.2.3. The Miller Indices

The faces of the unit cells form planes in the crystal lattice, as well as other planes resulting from grouping of other lattice points. These ordered planes are capable of diffracting X-ray beams, and specific sets of such planes can be defined by the Miller (or lattice) indices – h, k and l. The total of planes in the 'a' (x) orientation of the unit cell are defined by the index 'h'. Likewise, the total of planes in the 'b' (y) and 'c' (z) directions are determined by the indices 'k' and 'l', respectively. If a Miller index is '0', then the plane it describes will be parallel to its axis, and larger numbers will describe planes that are more perpendicular to their axes (Fig 2.5) (Rhodes 2000).

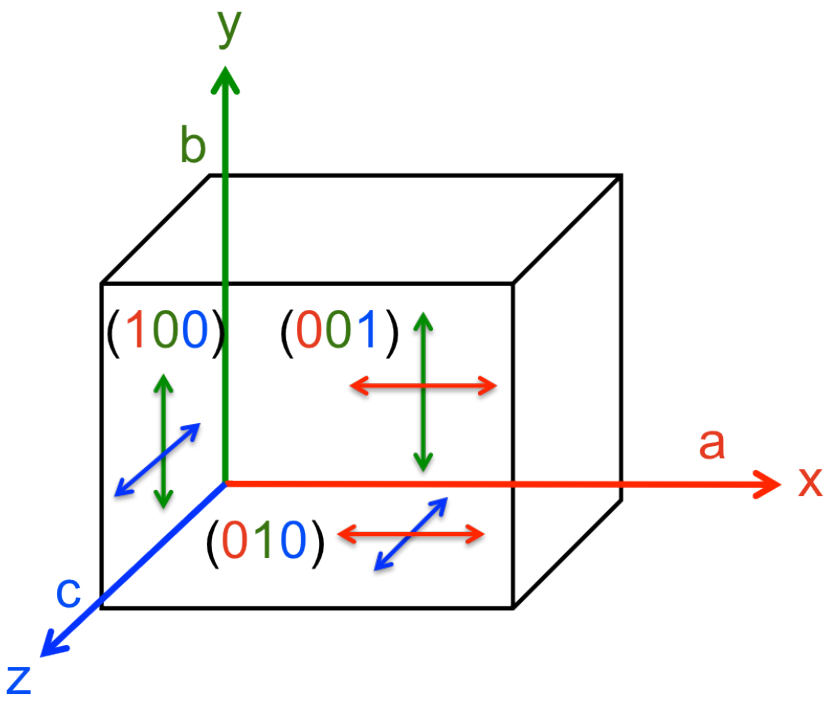


Figure 2.5: **The Miller indices**.
These indices indicate the planes of the unit cell they refer to and other parallel planes within the unit cell, which face a certain direction depending on their axes. If the indexes are denoted (abc) (and coloured accordingly), the zeros refer to the axes that the planes are parallel to. For instance, the index (001) refers to a plane parallel to the ab face, as well as planes which have axes parallel to the a and b (x and y) directions (Rhodes 2000).

Other planes apart from those of the unit cell faces can diffract X-rays. Planes can also intersect planes, for example those that intersect the (100) and (010) planes that are perpendicular to the ab axes of the unit cell, can be defined as (210) planes (Fig 2.5) and other planes that lie between diagonal x and y lattice points can have indexes of (230) (Fig 2.5). Furthermore, many planes are not parallel or perpendicular to the a, b and c (x, y and z) axes of the unit cell and form slices through the interior of the unit cell, for example those with the Miller indices (111) (Fig 2.6). Some planes that are identical in terms of axes, but may run in opposite directions can be represented with negative index values, for example (0 -1 0) is the same but opposite to (010) (Fig 2.7) (Rhodes 2000).

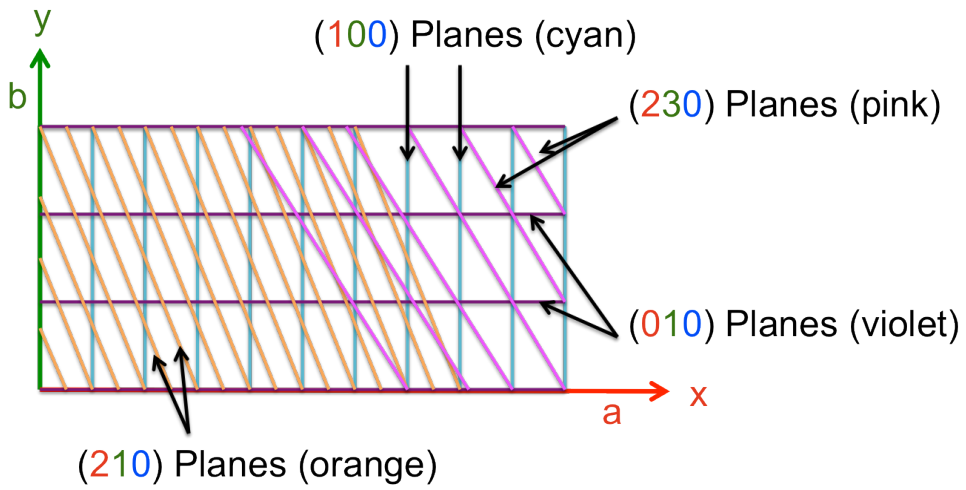


Figure 2.6: 2D representations of Miller Index Planes.
Planes of Miller indices (100) and (010), which run perpendicular and parallel to the a/x axis, respectively, can have intersecting (210) and (230) planes depending on lattice points present in the unit cell (Rhodes 2000).

.

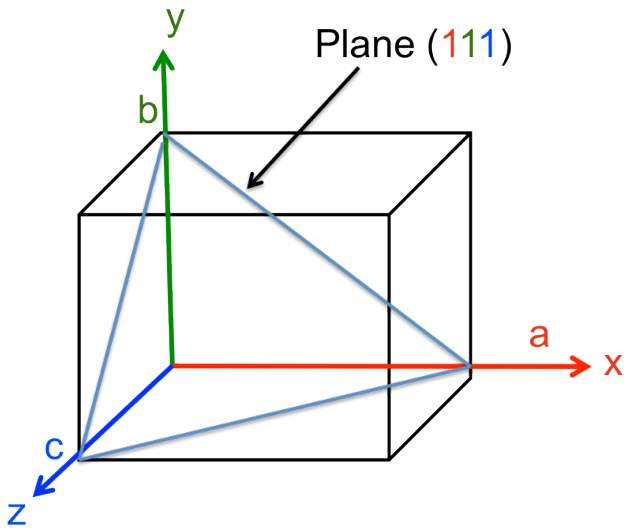


Figure 2.7: 3D representation of the Miller index (111).
This corresponds to a lattice plane that is not parallel or perpendicular to axes of the unit cell, and forms 'slices' through the unit cell (Rhodes 2000).

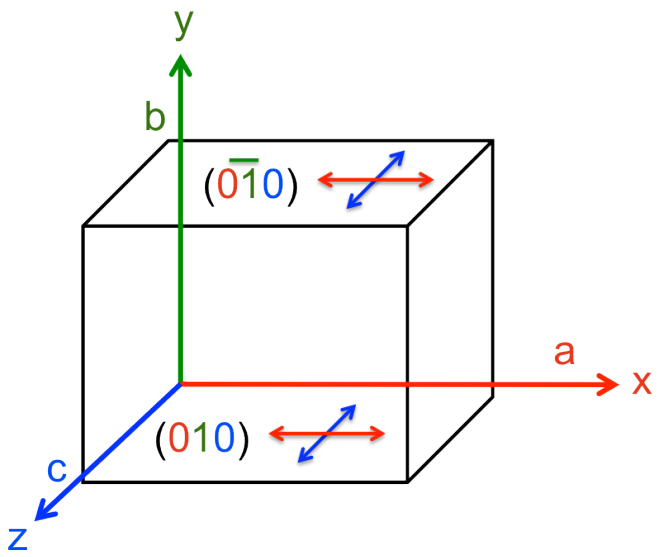


Figure 2.8: 3D representation of an example of a negative Miller index (0-10).
The orientation is the same, with respect to unit cell axes, as (010) but the plane is facing the opposite direction (Rhodes 2000).

2.1.3. X-ray Diffraction of Crystals

2.1.3.1. X-rays and X-ray Sources

X-rays where discovered by Wilhelm Conrad Röntgen in 1895, and are a type of high-energy, short wavelength electromagnetic radiation (Fig 2.8) (Frankel 1996). When electrons are accelerated towards a heavy metal anode, they dislodge electrons occupying lower orbitals in atoms of the anode. Electrons from higher orbitals travel to the lower orbitals to replace lost electrons in order to satisfy the energy state within the atom, and in the process they release energy in the form of X-rays. The energy that is emitted can be calculated from the difference between the energy of the higher orbital and the lower orbital, and will translate to a specific X-ray wavelength, described by Equation 2.1.

$$E = hv$$

Equation 2.1: The Energy Emitted by X-rays.
Where E = energy emitted, h = Planck's constant (6.6261×10^{-23}Js), v = frequency of the wavelength.

Depending on the orbital location of the dislodged electron and that of the 'replacement' electron, whether in the M, L or K level, and also the type of heavy metal used, the wavelength of the X-rays emitted will be different (Fig 2.8). For diffraction experiments with protein crystals, X-rays from electron transitions between level L to level K, referred to as K_{a1} and K_{a2} X-rays, are generally used as their wavelengths are approximately 0.7-1.5Å. As intra-molecular bonds in protein crystals are usually 0.8-3Å, X-rays of these wavelengths are able to penetrate and interact with these crystals during X-ray diffraction (Fig 2.8) (Bearden 1967, Krause & Oliver 1979).

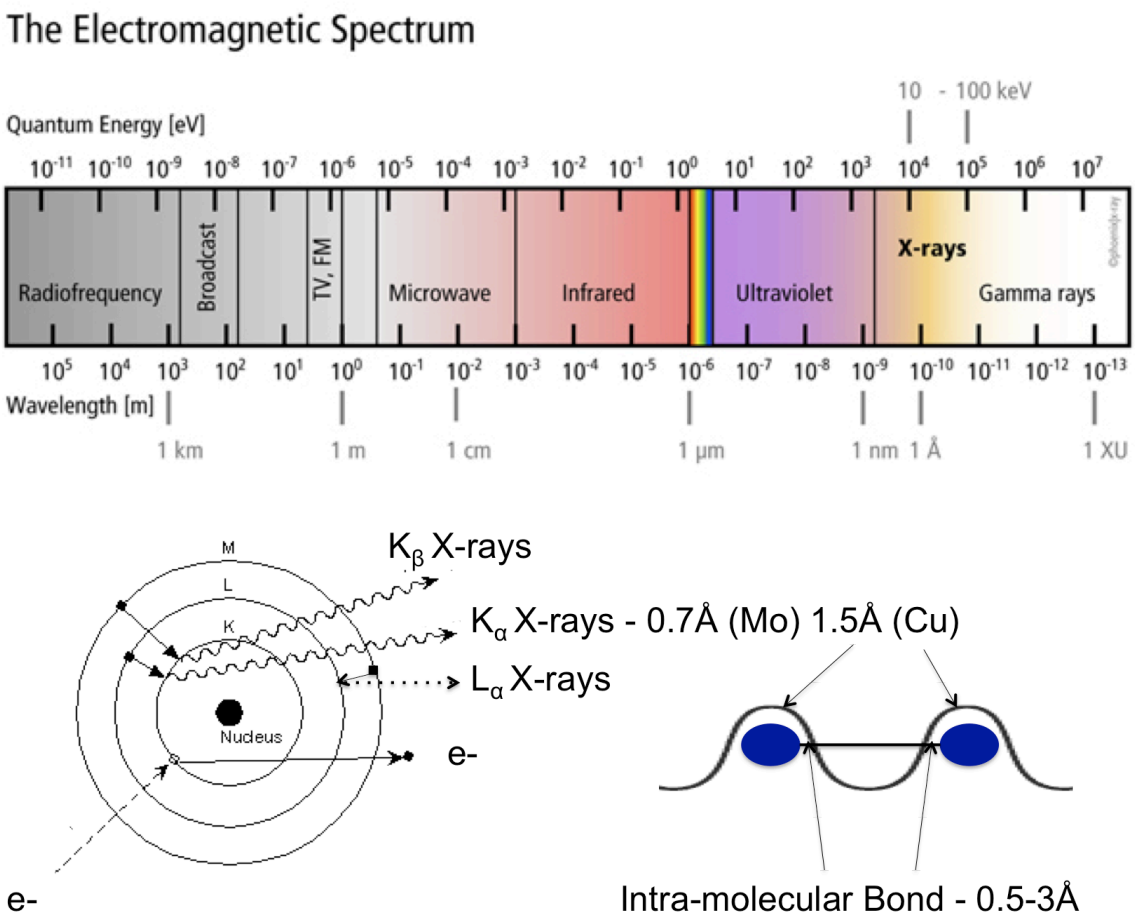


Figure 2.9: X-rays.
X-rays of the electromagnetic radiation spectrum are of high energy and short wavelength (0-100Å), generated by the bombardment of heavy metal atoms with accelerated, high energy electrons (e-). X-ray energy is released when electrons in high orbitals replace ones dislodged by the accelerated electrons. Electrons from L orbitals that replace those in K orbitals result in K_a X-rays, which are usually of a wavelength between 0.7-1.5Å (depending on the type of K_a X-ray, either type 1 or 2, or the heavy metal atom used), ideal for studying protein crystals that harbour inter-molecular bonds of usually 0.5-3Å in length (Bearden 1967).

In a laboratory, sealed X-ray and rotating anode tubes are used to produce X-rays. The latter is more powerful as heat dissipation from rotations of the anode is better than the water-cooling method in sealed tubes. However, much more powerful X-rays are produced by particle storage rings in Synchrotron facilities (e.g. the Diamond Light Source that was used for this project), in which electrons are accelerated at almost the speed of light. When forced to travel at a curve, these electrons emit X-ray energy, with Wiggler devices being used to bend beams in ways that can enhance X-ray beam intensity. Monochromators are also used in Synchrotrons, which can remove low-energy K_β X-ray wavelengths (Ooi 2009).

2.1.3.2. The X-ray Diffraction Experiment

The protein crystal is mounted on a rotating goniometer in between the source of X-rays and the X-ray detector, in a set-up in which the X-ray beam contacts the crystal as centrally as possible (see Figure 2.9). The detector is there to 'capture' the reflection spots, or Bragg reflections, that result from X-ray diffraction of the crystal. These reflections are collectively referred to as a diffraction pattern. Detectors are either image plates or charged-coupled devices (Rhodes 2000).

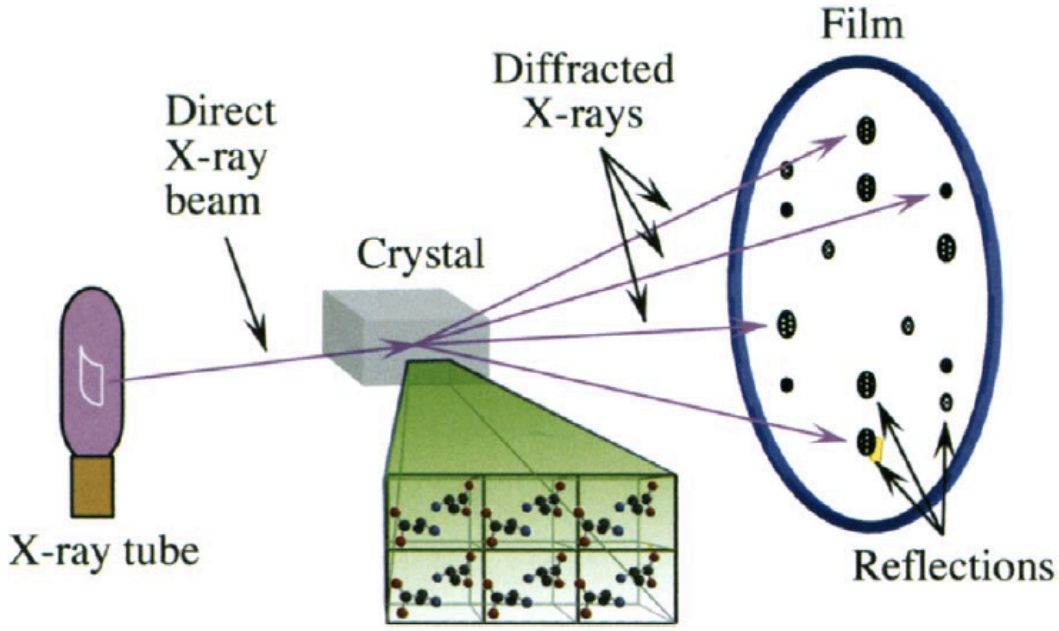


Figure 2.10: Experimental set-up for the X-ray diffraction of crystals. The X-ray beam hits the crystal, which is placed (usually inside a loop or capillary tube) on a rotating goniometer. The interaction of the X-ray beam with the crystal lattice results in X-ray diffraction and the capture of diffraction spots (or Bragg reflections) on an image plate or charged-coupled device (CCD) detector (image from Rhodes 2000).

2.1.3.3. Bragg's Law

The discovery that crystals could diffract X-rays is attributed to Max von Laue in 1912, a Nobel Prize winning physicist who experimented with copper sulphate crystals. He elucidated that diffraction can only occur when the X-rays that hit atoms in the planes of a crystal lattice have wavelengths of defined integral numbers. In 1915, the father-and-son team William Henry Bragg and Sir William Lawrence Bragg were awarded the Nobel Prize for their determination of NaCl and diamond crystal structures and their invaluable derivation of 'Bragg's Law' (Rhodes 2000). The Bragg equation (Equation 2.2) that relates to X-ray diffraction (Fig 2.10):

$$2d_{hkl}\sin\theta = n\lambda$$

Equation 2.2: The Bragg Equation.
Where *d* is the distance between two parallel planes, of index *hkl*, in a crystal lattice and nλ represents an integral number (*n*) of X-ray wavelengths (λ) that hit the crystal lattice planes.

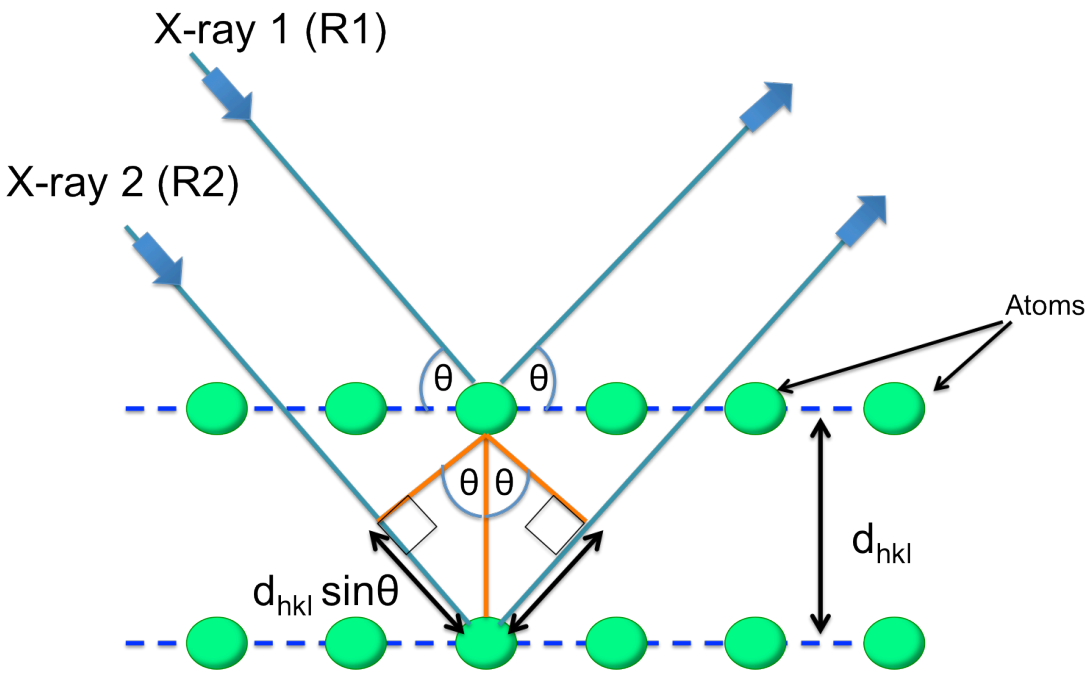


Figure 2.11: Diagram representing Bragg's Law.
The two incident X-ray beams R1 and R2, hit atoms of specifically spaced parallel planes by distance d_{hkl}. at an angle of θ. The R1 and R2 beams travel the same distance, but the R2 beam travels an additional distance of $2d_{hkl}$ sinθ with respect to the R1 beam (Blow 2005).

The θ angle of diffraction has an inverse relationship with the space between the parallel planes, d_{hkl}, meaning that large spaces created by large-sized unit cells with a greater number of atoms, will result in smaller diffraction angles and produce a greater number of reflections. Smaller unit cells result in larger diffraction angles, and then result in a smaller number of reflections (Elton & Jackson 1966).

Within a protein crystal, some lattice planes will not be able to result in diffraction of incident X-rays because they do not conform to Bragg's law, so to speak. For this reason, the crystal is rotated during the X-ray diffraction experiment in order to reveal planes of the crystal that are able to produce diffraction according to Bragg's law (Elton & Jackson 1966, Rhodes 2000).

2.1.3.3. Reciprocal Space, Bragg's Law and the Unit Cell

During the X-ray diffraction experiment with a crystal, it is the reflections of diffracted X-rays from the crystal lattice planes that are studied. A set of parallel planes of index *hkl* will diffract an X-ray, with the resulting reflection of the index *hkl*. The distances between *hkl* reflections on the diffraction pattern, have a reciprocal relationship with the unit cell axes a, b and c. In a sense, the X-ray diffraction pattern of a crystal is the crystal lattice but in reciprocal space (Rhodes 2000).

When an incident X-ray beam passes through a crystal, the crystal is the 'centre' of reflection within a 'sphere of reflection' also known as Ewald's sphere (Fig 2.11). The exiting X-ray beam passes through a point called the origin of reciprocal space lattice. The radius between the crystal and the origin of the reciprocal lattice represents the X-ray wavelengths in reciprocal space, i.e. 1/λ. When the crystal is rotated, other points in the reciprocal space lattice theoretically contact the surface of this sphere, which results in the diffraction of an X-ray from the centre of reflection, a lattice point in a real space lattice plane in the crystal, to the reciprocal lattice point at the surface of the sphere.

This reciprocal lattice point (P_{hkl}) is in fact, the *hkl* reflection and it results from X-rays diffracted from the set of parallel, real space planes of index (*hkl*), according to Bragg's Law (Rhodes 2000).

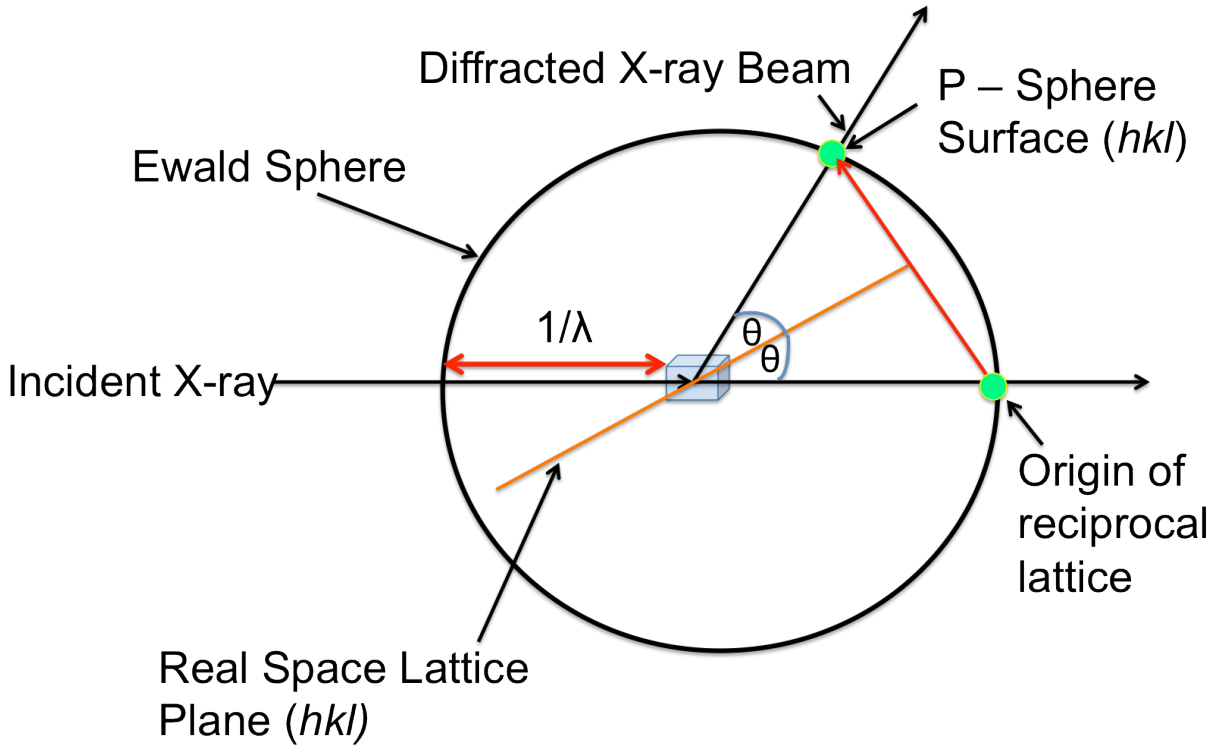


Figure 2.12: The Ewald Sphere.
The Ewald sphere shows that when an incident X-ray hits the real space lattice plane of a crystal (the 'centre' of reflection), then the X-ray beam will pass through a point (the origin of reciprocal lattice). The radius between the crystal and this origin represents the X-ray wavelengths in reciprocal space. As the crystal is rotated, other points (P) in the reciprocal space lattice will theoretically contact the Ewald Sphere surface, resulting in a hkl Bragg diffraction spot. This is according to Bragg's Law (Rhodes 2000).

2.1.3.4. The Diffraction Pattern

The Bragg reflection at the centre of the pattern is referred to as the origin of the reciprocal space lattice, and assigned the integer hkl indices of 000. The positions of each hkl reflection on the diffraction pattern are assigned coordinates, as it is these positions that result from the angle of the diffracted X-ray beam that diverged from the incident beam when it hit the real space crystal lattice plane. It is this angle of divergence that provides the information regarding the indices of each reflection, and from which the axes of the unit cell can be calculated.

The diffraction pattern represents reciprocal space, and that they represent the inverse of all dimensions (e.g. $Å^{-1}$) present in the real space crystal lattice (e.g. Å). The structural resolution that can be gained from a diffraction pattern is limited by the inverse of the distance from the reciprocal space origin to the quantifiable reflections of the outer edges of the pattern, for example 1/2.5Å may result in a molecular structure model of 2.5Å (Rhodes 2000).

2.1.3.5. From Reciprocal Space to Real Space

Every Bragg reflection in the diffraction pattern is actually a summation of many X-rays diffracted by one particular atom present in all unit cells of the crystal. X-rays are a form of radiation represented as waves, which are described by the components frequency, amplitude and phase. The X-ray wave frequency will be the same as the incident beam waves. However, the electron densities, or scattering factors, of atoms within the crystal lattice plane (hkl) will affect the amplitude of X-ray waves. Each diffracted wave can be assigned a structure factor F_{hkl} via the structure factor equation (Equation 2.3), which relates crystal lattice plane indexes with the amplitude and phase of X-ray waves:

$$F_{(hkl)} = \sum_{j=1}^{n} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)} e^{(-B_j \sin^2 \lambda^2)}$$

Equation 2.3: The Structure Factor Equation.
Where hkl refers to the crystal lattice plane indices and fj represents the electron density, or scattering factor, of the atom j within the unit cell, which has n number of atoms. In the first exponential part, the frequencies of the wave in x, y and z dimensions can be related to the hkl indices, with xj, yj and zj describing the atom's coordinates within the unit cell.

The Debye-Waller function is represented in the second exponential part of the structure factor equation, which refers to the 'B-factor'. The B-factor, or 'temperature' factor represents the extent of disorder and flexibility within the crystal lattice.

The real-space of the unit cell can be described using electron density representation that can indicate the positions of atoms in the unit cell. A Fourier transform equation (Equation 2.4) can be used to elucidate the electron density from amplitude and phase information from the Bragg reflections:

$$\rho_{(x,y,z)} = \frac{1}{v} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{-2\pi i(hx+hy+lz)}$$

Equation 2.4: The Fourier Transform Equation to Derive Electron Density from Amplitude Information from the Bragg Reflections and Calculated Phases.
Where p(x,y,z) represents the sum of the electron density and V is unit cell volume and F_{hkl} shows the sum of the total reflections in the diffraction pattern.

In the above equation, F_{hkl} is actually the sum of the total reflections in the diffraction pattern, and assumes that phase information is included for each structure factor (F_{hkl}) for each individual Bragg reflection. However, as information regarding the phases of the Bragg reflections is often missing (see Section 2.1.5), the equation can be re-arranged to highlight the relationship between phase information and the structure factor (Equation 2.5):

$$\rho_{(x,y,z)} = \frac{1}{v} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i(hx+hy+lz)+ia_{hkl}}$$

Equation 2.5: The Fourier Transform Equation Re-Arranged to Show the Relationship Between Phase Information and the Structure Factor.
Where $|F_{hkl}|$ represents the X-ray wave amplitude and $i\alpha_{hkl}$ represents phase information needed to calculate the structure factors for each wave (Rhodes 2000, Rupp 2009).

2.1.4. Data Processing

2.1.4.1. Indexing and Integration of the Diffraction Pattern Reflections

The Bragg reflections are assigned indices, and their distances from the origin of reciprocal space calculated in order to decipher unit cell dimensions. A small sample of images from an appropriate amount of rotations, depending on the crystal symmetry, are used to find the unit cell axes and also glance at theoretical space groups. Using information about the rotation axis angle (Φ) and the reflection of the direct beam as diffraction pattern 'origin', the coordinates of the other Bragg reflections can be measured, and then reciprocal lattice can be built.

An orientation matrix can relate the hkl indices of the reciprocal lattice to the space group and unit cell dimensions of the real space lattice. The matrix utilises all diffraction pattern images, unit cell axes, crystal mosaicity and orientation, plus information regarding the distance between crystal and detector, in order to assign reciprocal lattice plane indices to each Bragg reflection, and then integrate each of their intensities.

Problems with indexing can happen due to twinned crystals, which have two different crystal lattices of differing orientations, fused together. When used in X-ray diffraction, the Bragg reflections detected are then difficult to index due to the two different crystals having overlapping diffraction patterns. It can sometimes be possible to spot a twinned crystal prior to diffraction using a polarizer, as these often reflect light from one side whilst absorbing light on the other (Evans 2006, Rupp 2009).

2.1.4.2. Scaling and Truncation

Depending on the symmetry of a particular crystal, many Bragg reflections that are recorded will be related to others by symmetry and would be of equivalent intensities if it were not for variations in beam intensity, crystal integrity and detection. Scale factors are attributed to reflection intensities in each diffraction pattern image, which aim to elevate the effects of data collection errors. This is referred to as 'scaling' and the program SCALA of the CCP4 suite is often used simultaneously with MOSFLM in order to scale reflections alongside indexing and integration. If reflections are related to others in terms of symmetry, their intensities will be scaled so that their reflections will have identical intensities.

SCALA can also be used to check the quality of the diffraction data by examining the similarity of intensities of the reflections related to each other by symmetry. The resulting R_{sym} statistic for the entire dataset can indicate the relationship between resolution and significant structural data, and is defined by Equation 2.6:

$$R_{sym} = \frac{\sum_{hkl} \sum_i \left(I_i(hkl) - I_{mean}(hkl) \right)}{\sum_{hkl} \sum_i (I_i\ (hkl))}$$

Equation 2.6: The Equation to Derive R_{sym}.
Where Ii represents intensity of reflection (i) of indices hkl. Imean represents mean intensity of reflections of these particular indices.

The resolution limit for the data can also be indicated by whether the estimated mean intensity (I) decreases to twice the mean standard deviation of measured intensity, summarised by the I/σI ratio.

The program TRUNCATE in CCP4 is used to convert the scaled intensities to amplitudes. The B-factor of the molecular structure model is estimated using a Wilson plot. The Wilson plot yields structure factor amplitudes, plus <I> and σI values. Reflections are rejected if they have <-4σ values. Here, higher quality data can be gained from reflections of weak intensities, usually resulting from small, damaged or disordered protein crystals (Evans 2006, Rupp 2009).

2.1.4.3. Asymmetric Unit and Solvent Content Estimation with the Matthews Coefficient

Commonly, most of the protein crystal volume (approximately 40%) is composed of solvent molecules. The Matthews coefficient (V_m), the crystal volume per unit of protein molecular weight, can relate the solvent content with the number of asymmetric units within a protein crystal structure, using Equation 2.7.

$$x(\rho) = \frac{1.66 \cdot \bar{v}}{V_m} = \frac{1.23}{V_m}$$

Equation 2.7: The Matthews Coefficient Equation.
Where x(p) represents the fraction of the asymmetric unit that contains the protein molecule, and v refers to the partial specific volume of protein (Kantardjieff & Rupp 2003).

2.1.5. Solving the Phase Problem with Molecular Replacement

The positions of the Bragg reflections in the diffraction pattern can be indexed and then used to build a reciprocal-space lattice to calculate the axes of the unit cell, and the intensities of the reflections can determine the amplitudes of the X-ray waves, which can then be used to build the electron density map of atoms within the unit cell. However, the phase information of the X-ray waves is lost from the diffraction pattern. Without phase information, the structure factors (F_{kl}) of waves, and thus the electron density map, cannot be calculated. Fortunately there are methods used in order to overcome the phase problem; these include Molecular Replacement (MR), Isomorphous Replacement (IR) and Multi-Wavelength Anomalous Diffraction (MAD) (Rhodes 2000). MR is discussed in this section due to this being the strategy of choice for overcoming the phase problem during this project work.

2.1.5.1. Patterson Function Methods

In the past, Patterson function methods were used for molecular replacement within the program AMorRe (Navaza 1994), but now evolutionary algorithms and maximum likelihood functions are used in modern programs such as EPMR and Phaser, respectively (Kissinger et al 1999, McCoy 2006).

Patterson function methods involve three-dimensional rotation and translational searches using Patterson maps (see Figure 2.12). The spots on the Patterson map represents both intra-atomic distance vectors and (self-vectors) and inter-atomic vectors (cross-vectors), and are built from using the Patterson function equation (Equation 2.8), derived from the square of structure factor amplitudes measured from reflection intensities of the diffraction pattern (Patterson 1934). The Patterson map can also be described as the diffraction pattern convoluted with itself (Blow 2005).

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|^2\cos\left(2\pi(hu+kv+lw)\right)$$

Equation 2.8: The Patterson Function Equation.
Where (u,v,w) represent axes coordinates within the Patterson map, hkl are the experimentally determined indices from the diffraction pattern (Patterson 1934).

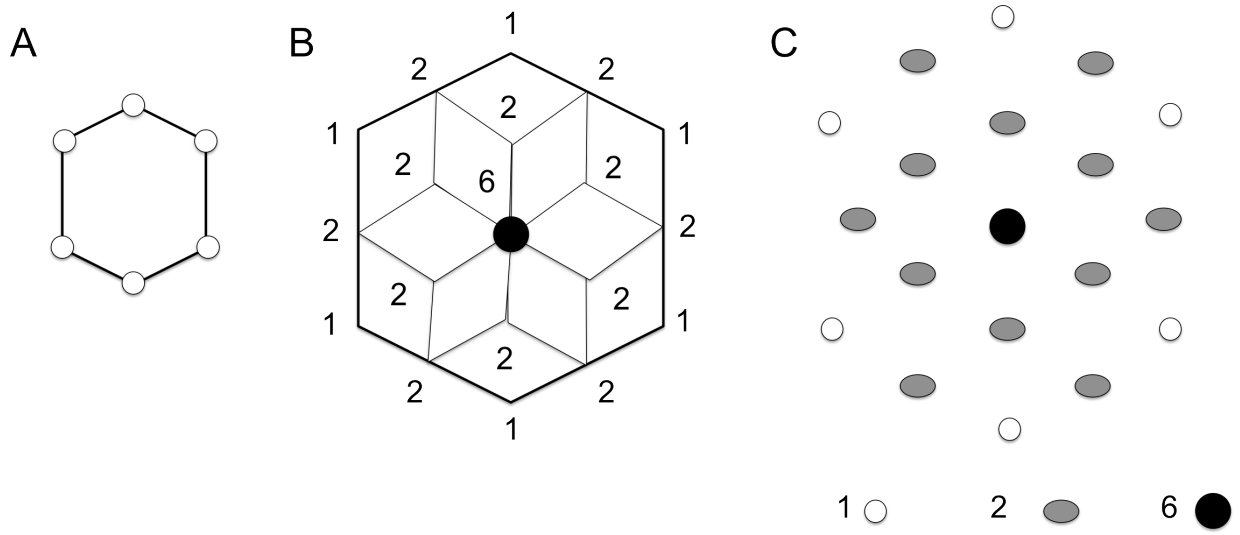


Figure 2.13: Example of a Patterson Map of a Molecule.
A) Benzene molecule represented as a hexagon shape, atoms shown as white circles. B) The benzene molecule convoluted with itself C) The Patterson vector map of the benzene molecule. Each atom has a zero length vector, so in total there are six zero-length vectors at the centre of both the convolution diagram (B) and the Patterson Map (C). Bonds between neighbouring atoms occur in six directions and additionally two parallel bonds, so a weight of 2 units is assigned to those inter-atomic vectors (B) and (C). The vectors between atoms on opposite sides of the benzene molecule have a weight of 1 unit, also represented on both the convolution diagram (B) and the Patterson map (C) (Modified figure from Blow 2005).

2.1.5.1. Molecular Replacement

MR has been used to solve approximately 80% of known protein structures, and involves the use of a search model from an already solved and phased crystal structure of a macromolecule with high sequence identity (at least higher than 40% sequence identity) to that of the macromolecule with the unknown structure. Computationally, the process involves superimposition of the search model onto the correct orientation and position of the unknown structure within the unit cell, via a series of three-dimensional rotational and translational searches originally using Patterson vector maps (Rossman 1990, Rossman & Blow 1962, Rupp 2009).

During the rotation search, the Patterson vector map of the search model is rotated in three-dimensions about the Patterson vector map constructed from experimental diffraction data of the unknown molecule in order to find a best fit between the self-vectors of the two different vector

maps. Cross-vectors, such as those from the symmetry operations within the crystal structure of the unknown molecule, and which do not match with the vector map of the search model, are an obstacle within this process, and the reason why the radius used for fitting is reduced to a size as small as possible. Rotations are specified by three Eulerian angles, α, β and γ (see Fig 2.13).

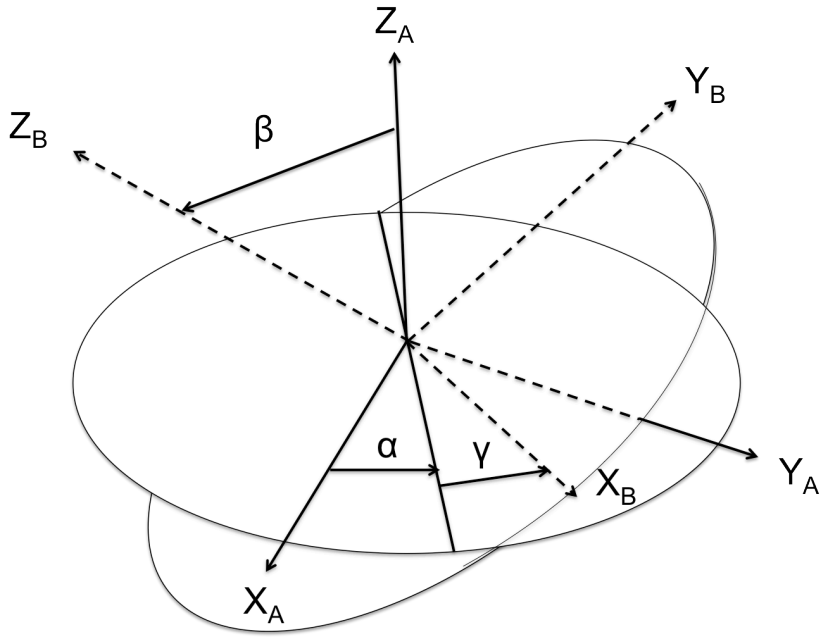


Figure 2.14: Eulerian Angles of Rotation used for the Rotation of the Search Model.
The Eulerien angles α, β and γ are used to rotate the perpendicular directions X_A, Y_A and Z_A to the different positions X_B, Y_B and Z_B, respectively (Modified figure from Blow 2005).

After each rotation, the fit between the two Patterson function maps is calculated via the rotation function equation (Equation 2.9).

$$R(\alpha, \beta, \gamma) = \iint_U P_M(u) \times P_X(u)_r du$$

Equation 2.9: The Rotation Function Equation.
Where the Patterson function of the search model is represented by PM(u) and the Patterson function of the unknown structure is shown as PX(u), with the fit defined by rotation angles (α, β and γ) of the rotation function (R) (Blow 2005).

A mean average rotation function value plus standard deviation is calculated from the many generated from the numerous searches. Rotation values are significant if they exceed four times the standard deviation (Blow 2005).

Once the rotation search has established a fit between the orientation of the search model and unknown structure within the unit cell, a translation search is carried out. The translation search uses the symmetry-related cross-vectors of the Patterson map. The space group of the crystal, specified by the diffraction pattern, is used to calculate such cross-vectors (Crowther & Blow 1967, Blow 2005. Translation functions form a comparison between the cross-vectors of the Patterson maps of search and unknown structures, related by symmetry operations. The resulting overlap can be described as a translation vector (t) between two symmetrically related molecules, with the translation function (T) having the highest value possible (see Equation 2.10).

$$T(t) = \int_v P_X(u,t) \cdot P_M(u) du$$

Equation 2.10: The Translation Function Equation.
Where the Patterson map of the search model is shown as PM(u), and the Patterson map of the unknown structure is represented as PX (u,t), translation vector as 't'. From the calculation of the fit between the cross-vectors of the maps, the maximum translation function 'T' value is found, indicating the most correct structure solution (Blow 2005).

Maximum likelihood methods are the latest improvement for molecular replacement phase solution, although still employ both rotation and translational searches similar to Patterson methods. Programs such as Phaser (McCoy 2006) rotate the search model with respect to the experimental structural data, and via Gaussian methods and probability calculations, the best fit of the model is attained with similar operations used for translation (McCoy 2004).

The rotation and translation searches result in many different model structure solutions, ranked by the correlation between estimated and experimental structure factor F_{hkl} wave amplitudes, determined by the correlation coefficient equation (see Equation 2.11).

$$CC = \frac{\sum_{hkl}\left(|F_{obs}|^2 - \overline{|F_{obs}|^2}\right) \times \left(|F_{calc}|^2 - \overline{|F_{calc}|^2}\right.}{\left[\sum_{hkl}(|F_{obs}|^2) - \overline{|F_{obs}|^2}) \times \sum_{hkl}\left(|F_{calc}|^2 - \overline{|F_{calc}|}\right)^2\right]^{1/2}}$$

Equation 2.11: The Correlation Coefficient Equation.
Where calculated and experimental structure factors for each Bragg reflection of hkl indices and averaged intensities (structure factor amplitudes with a bar above them), are expressed via the correlation coefficient (CC) (Blow 2005).

The various structure solutions will also have associated R-factors describing the quality of structure solution. The R-factor is calculated via Equation 2.12.

$$R_{work} = \frac{\sum_{hkl}\left||F_{obs}| - k|F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

Equation 2.12: The Correlation Coefficient Equation.
Where Fobs represents the unknown structure model and Fcalc represents the known (or 'search' model). A scaling factor, k, relates calculated structure factor amplitudes to the observed structure factor amplitudes (Rossman 1960, Rossman & Blow 1962, Rupp 2009).

2.1.6. Model Refinement

With phases determined, an electron density map is built from the electron density equation discussed in section 2.1.3.4, and a model molecular structure can be constructed. As the electron density map may contain phase errors, the refinement of both electron density map and molecular model must be carried out. Refinement is the process of reducing the differences between experimental structure factor amplitudes and those calculated from molecular replacement. Adjustment and improvement of the model can result in more accurate phase calculations, and build a more accurate electron density map, which can then be used to create a more reliable molecular structure model (Rupp 2009).

2.1.6.1. Model Bias and R_{free}

Despite favourable R-factor (R_{work}) statistics for a molecular structure model and its goodness-of-fit with the electron density map, these will not indicate whether the electron density map and model are incorrect as they may be highly similar. The R_{free} statistic is used to overcome such model bias (Brünger 1992, Kleywegt & Brünger 1996), in which a sample of reflections are separated from refinement calculations. These reflections (5-10% of the data) will be used to determine the R_{free}, summarized by Equation 2.13:

$$R_{free} = \frac{\sum_{h \in free} |F_{obs} - F_{calc}|}{\sum_{h \in free} F_{obs}}$$

Equation 2.13: The Equation to Derive R_{free}.
Where Fobs represents the unknown structure model and Fcalc represents the known (or 'search' model), whilst 'free' represents the random sample of reflections separated from the refinement process (Kleywegt & Brunger 1996).

2.1.6.2. Least Squares and Maximum Likelihood Refinement

The squared difference between the experimental and estimated structure factor amplitudes of the Bragg reflections are minimised during refinement by the 'least squares' method, based on the Gaussian distribution concept. It speculates that the phase is correct, but as we know, this may not be true. To attempt to counter this, the maximum likelihood method is used in which experimental amplitudes are expressed as probabilities, which will be similar to a reliable model (Pannu & Reed 1996). During refinement, adjustment of the structural molecular model can maximise these similar probabilities. The squared difference between the two different amplitudes can be represented by Equation 2.14:

$$Q = \sum_{hkl} w_{hkl} (|F_{obs}(hkl)| - |F_{calc}(hkl)|)^2$$

Equation 2.14: The Least Squares Equation.
Where Q represents the squared difference between experimental (F_{obx}) and estimated (F_{calc}) structure factor amplitudes (Pannu & Read 1996).

2.1.6.3. Rigid Body Refinement

After molecular replacement, the molecular model may not be in the correct position in the unit cell. Rigid body refinement speculates that the macromolecules are one rigid body in order to select the most optimal orientation between estimated and experimental structure factor amplitudes. This refinement strategy is employed for low-resolution data, e.g. of 6Å resolution (Afonine et al 2012).

2.1.6.4. Non-Crystallographic Symmetry (NCS) Refinement

This refinement strategy aligns all the molecule chains within the asymmetric unit, and if there is a high alignment score (>80%), this means local root-mean square (RMSD) can find domains and apply weights to atoms. Domains with similar RMSD values are restrained tightly (Afonine 2012).

2.1.6.5. TLS Group Refinement

TLS (translational, librational and screw-motion determination) groups can be used for B-factor refinement, as it can use three parameters to represent the translational, librational and screw motions of specific parts of a molecular model structure. The TLS algorithm can separate molecule chains into segments by comparing the least square residual of the structure to the TLS calculated model, and it is applied to each chain individually in order to reduce model bias (Winn et al 2001, Afonine et al 2012).

2.2. Nuclear Magnetic Resonance (NMR) Spectroscopy

If the protein of study is unable to crystallize, and it is of appropriate size, its atomic structure may be solved using NMR Spectroscopy. Advantages that NMR Spectroscopy has over X-ray crystallography is that it can also be used to track the solubility and folding of a sample, and also study the dynamic interactions of that protein with other ligands in 'real-time' due to the experimental sample being protein in solution. However, solid-state NMR has been developed in order to facilitate the study of larger, more difficult proteins such as those in complexes or cell membranes. In terms of solving protein structures, it is limited to those of 30-40kDa. However, modern TROSY-based methods can allow NMR structural biologists to study proteins of much larger molecular weights. This study used one-dimensional ^{1}H-NMR to check solubility and disorder of protein samples, and also TROSY-based backbone-assignment experiments (Pervushin 2000).

2.2.1. The NMR Spectrometer

Inside the NMR spectrometer, electricity passes through a large wire coil (or solenoid loop) that is kept cold (below 6K) with liquid helium (plus a 77K liquid nitrogen heat shield) in order to reduce its resistance. This superconducting magnet generates a stable magnetic field ranging between 6-24 Tesla. The wire itself is a matrix of metal alloys including niobium, tin and copper, and this set up is inside a vacuum. A vertical bore tube, where the sample is inserted, is in the centre of the magnet and is at room temperature (Fig 2.14). A small coil of wire, called the probe, is where the sample sits (Fig 2.14). During an NMR spectroscopy experiment, a computer instructs the probe to transmit a radiofrequency (RF) pulse towards the sample, and a detector monitors subsequent signals received from the sample. This probe coil is 'tuned' and 'matched' to enhance efficiency of radiofrequency transfer to the sample and improve the sensitivity of signals recorded. Prior to the experiment, shimming is done using shim coils that also surround the sample; in order to negate any inhomogeneity's from the external magnetic field (Keeler 2005).

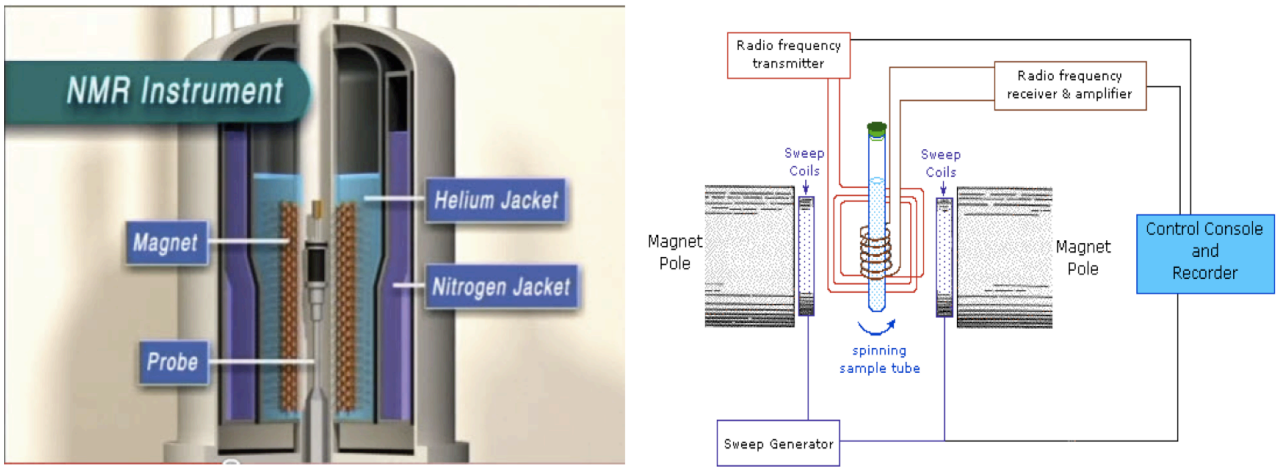


Figure 2.15: The Interior of an NMR Spectrometer.
NMR spectrometers include a solenoid loop of wire that is the superconducting magnet, kept cold using liquid nitrogen and helium. The sample, inserted into the central bore, sits in a small coil of wires called the probe, which emits and receives radiofrequencies according to, and detected by, the computer console (Images from Unkel et al 2012).

2.2.2. NMR Fundamentals

2.2.2.1. Nuclear Spin and Larmor Precession

Atomic particles within an atom; the electrons, protons and neutrons, all possess a fundamental property called 'spin'. This property can be thought of as the particle spinning on an axis, defined as spin angular momentum. For protons, which are positively charged particles, its spinning creates a magnetic moment (μ), or a small magnetic field, and becomes a magnetic dipole. The axis of the spin can orient with an external magnetic field (B_O) in one of two ways, parallel and anti-parallel depending on its nuclear spin quantum number (*m*), which can be either +½ (low energy, parallel) or –½ (high-energy, anti-parallel) (Fig 2.15). As a proton has a net nuclear spin of ½, it is a spin-half particle, and therefore, a nucleus comprised only of a proton, e.g. ^{1}H, is described as a spin-½ nucleus (Macomber 1998). When spin-½ nuclei are aligned either parallel or antiparallel to the external magnetic field ($\mathbf{B}_0$), their spin angular momentum causes them to rotate, or 'precess', like a spinning top around their axes (Fig 2.15).

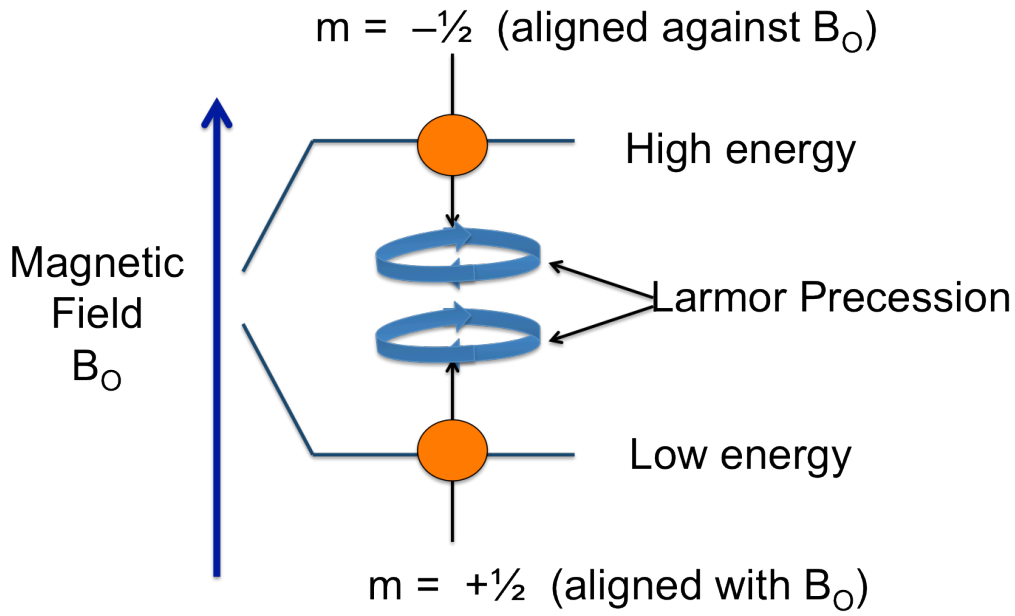


Figure 2.16: Spin ½ Particles, energy levels and Larmor Precession.
Spin-½ particles precess at a specific angular frequency called the Larmor frequency. These particles include protons and neutrons, and can be assigned one of two nuclear spin quantum numbers (m), -½ (higher quantum energy state) or +½ (lower quantum energy state). These two spin states have different orientations regarding their spin axes, when affected by an external magnetic field (B_O); m = - ½ particles have their spin axes aligned against B_O, m = +½ particles have their spin axes aligned with B_O (Macomber 1998).

Spin- ½ nuclei precess at a particular angular frequency, of radians per second, depending on the magnetogyric ratio of the particular nucleus and the strength of the magnetic field (independent of nuclear spin quantum states), shown by Equation 2.15:

$$\omega = \gamma B_0$$

Equation 2.15: The Nuclear Angular Frequency Equation.
Where ω represents the angular frequency, γ is the magnetogyric ratio and $\mathbf{B}_0$ is the external magnetic field (Macomber 1998).

The magnogyric ratio (γ) is the ratio between the magnetic moment of the magnet dipole nucleus, with its angular momentum. The precession frequency can be converted to hertz and referred to as the Larmor frequency, via this Equation 2.16:

$$v_0 = \frac{\gamma B_0}{2\omega} = \frac{\omega}{2\pi}$$

Equation 2.16: The Larmor Frequency Equation.
Where v_0 stands for the precession frequency, or Larmor frequency, in units of hertz (Macomber 1998).

Other nuclei apart from ^{1}H contain neutrons, which although are not charged particles, also have a net nuclear spin of ½. The number of neutrons and protons within a nucleus decides its net nuclear spin. Nuclei with an even number of both proton and neutrons will have a net nuclear spin of 0 due to both neutron and protons cancelling out each other's spin contributions – i.e. their spins are 'paired'. These nuclei are not useful for NMR spectroscopy, unlike isotopic nuclei with an uneven number of neutrons and/or protons, in which the net nuclear magnetic spin is ½, such as ^{13}C and ^{15}N. The axes of these spin-½ isotopic nuclei can also exist in the two quantum states, parallel or antiparallel to the external magnetic field. However, one isotope useful for NMR, but that is neither spin-0 nor spin-½, is deuterium (^{2}H), has one unpaired proton and one unpaired neutron, resulting in a net nuclear spin of 1 (Macomber 1998).

2.2.2.2. The Boltzmann Distribution

When radiofrequency radiation is applied to a spin-½ nuclei aligned either parallel (low energy) or anti-parallel (high energy) to the magnetic field, the radiation in the form of a photon promotes the low energy nuclei to high energy nuclei, effectively changing the quantum nuclear spin state of the low energy nuclei. However, the energy of the photon has to match the energy difference between the two quantum spin states (Fig 2.16), with the energy difference can be calculated from Equation 2.17.

$$\Delta E = \frac{\gamma h B_0}{2\pi} = h\nu_0$$

Equation 2.17: The Energy Difference Equation.
Where the energy difference (*ΔE)* is directly related to the magnetic field strength (B_0) and the magnetogyric ratio (γ). So the Larmor frequency, of the nucleus in question via Planck's constant (h = 6.6261 x 10^{-34} m^2 kg/s), also specifies the B_0.

The frequency of the radiation must also match the precessional frequency, and furthermore, there must be more nuclei of a lower energy spin state than nuclei of a higher energy spin state in the sample. If the sample is at thermal equilibrium, more than half of the nuclei will be in the lower energy state, aligned parallel with the external magnetic field, due to this alignment being more energetically favourable. So, the difference in energy is also related to the distribution of nuclei that are either in high energy or low energy state, i.e. the Boltzmann distribution, via Equation 2.18:

$$\frac{N\left(-\frac{1}{2}\right)}{N\left(+\frac{1}{2}\right)} = exp\left(-\frac{\Delta E}{kT}\right)$$

Equation 2.18: The Boltzmann Distribution Equation.
Where N refers to the number of nuclei in either high energy (-½) or low energy (+½) quantum spin states, *k* is the Boltzmann constant (1.381 x 10^{-23} J K^{-1}) (Keeler 2005).

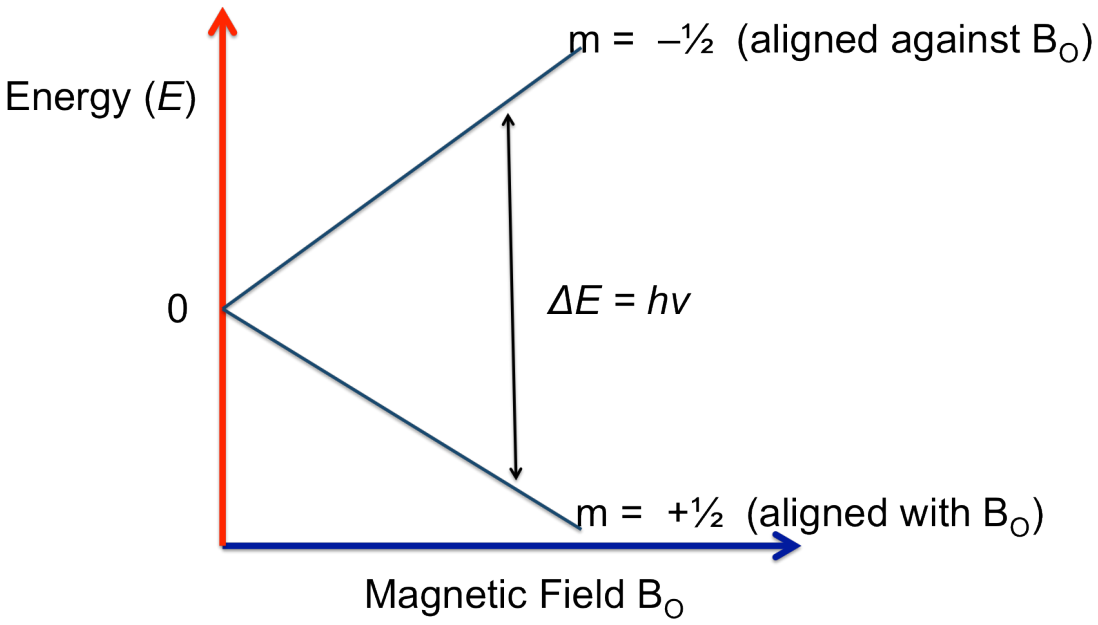


Figure 2.17: Energy difference between the two spin states.
A photon of radiofrequency excites low energy nuclei to high energy nuclei, but is has to match the energy difference between the two quantum spin states, with the energy difference related to magnetic field strength, and also the Larmor frequency of the nucleus. However, at thermal equilibrium, more than half of the nuclei will be in the lower energy state (Keeler 2005).

However, the ΔE between quantum nuclear spin states is relatively small. A larger ΔE is ideal as it means an increased population of lower energy spins over high energy spins. As the measurable NMR signal arises from this difference, more sensitive magnets with stronger magnetic fields are more desirable, as they can provide improved signal-to-noise ratios in NMR experiments. Furthermore, nuclei with a greater magnetogyric ratio and a low temperature can also result in a larger ΔE (Macomber 1998).

2.2.2.3. Pulse NMR to the NMR Spectrum

The process of recording an NMR spectrum will involve a fast pulse of radio-frequency radiation directed towards the sample. Prior to the pulse, nuclei in the sample will be at thermal equilibrium and each will precess at its angular frequency around the z-axis that has aligned either parallel or antiparallel to the external magnetic field. The orientation of this spin axis can be described as a 'vector', with its length equal to the spin's magnetogyric ratio. During thermal equilibrium, when the nuclei precess, the vectors of either low energy or high energy rotate with no net phase coherent – i.e. they are at random phases (Fig 2.17).

When the 90° RF pulse is applied, it will result in nuclei vectors that precess in the same direction at the same time on the y axis. Then, each nuclei will precess at its Larmor frequency with phase coherence, around the z- axis, with the individual vectors resulting a complete magnetic vector referred t o as the net magnetization. This net magnetization is results in an electrical voltage signal, or 'free induction decay' in the sample coils, and it is a collection of all the Larmor frequencies of all the nuclei in the sample (Macomber 1998).

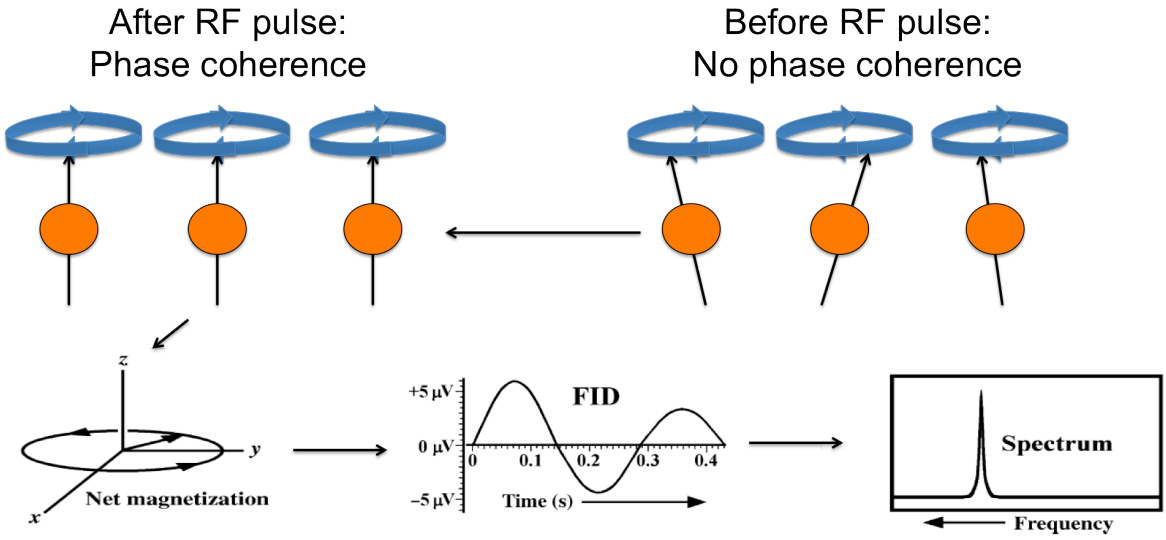


Figure 2.18: Phase coherence, net magnetization and the FID.
Nuclei in the sample at thermal equilibrium will precess at an orientation, or vector, around a z-axis that is aligned either with or against the external magnetic field, but these vectors will not be phase coherent until an RF pulse is applied. When the RF pulse irradiates the sample, nuclei vectors precess in the same direction at the same time on the y-axis, before then precessing with phase coherence around the z-axis, resulting in net magnetization. The net magnetization results in a signal, or FID (free induction decay), which will be a sum of all the nuclear Larmor frequencies in the sample. Via the Fourier transform, this FID is converted to an NMR spectrum: a graph of intensity vs. frequency (Modified image from Keeler 2005).

The FID is put through a Fast Fourier Transform, which converts it to a graph of intensity vs. frequency. This graph is the NMR spectrum, which shows each resonant frequency in the sample as peaks. The recording of the FID is often repeated a number of times, or 'scans', in order to increase signal intensities – this can improve the signal-to-noise ratio of the spectrum. The Larmor precession frequency of individual nuclei is related to their chemical environments within the molecule (Jacobsen 2007).

2.2.2.4. The Chemical Shift

An NMR spectrum is a graphical representation of the resonant frequencies achievable for a particular nucleus (see section 2.2.3.2). The number of peaks in a spectrum does not correlate with the number of that particular nucleus in a molecule as different nuclei may have the same chemical shift. Nuclei can be equivalent due to having the same chemical environments in symmetrical molecules.

The Larmor frequency of each nucleus in the sample will have subtle variations depending on its local molecular environment. Nearby bonding electrons and other neighbouring atoms will cause minute changes to the extent of the external magnetic field experienced by a particular nucleus.

Hence, the same nucleus can have different resonant frequencies; these differences can be designated chemical shifts (δ, ppm). They are independent of the external magnetic field strength, but describe the local chemical environment surrounding a particular nucleus (Macomber 1998, Jacobsen 2007).

The chemical environment of a particular nucleus, and thus the extent of the external magnetic field experienced by the nucleus, will be affected by nearby electronegative atoms (e.g. O and N), shielding from delocalised electrons (magnetic anisotropy), the sample solvent, temperature and pH. For example, delocalised electrons in an aromatic ring can move in a circle when the external magnetic field is applied, and thereby generate a tiny magnetic field itself that is opposed to the external field. The protons bound to aromatic rings will experience a stronger magnetic field than that of the external field, and resonate at approximately 7ppm. However, protons on other chemical groups nearby to aromatic rings in a folded and stable hydrophobic core structure, for example aliphatic groups, will be shielded from the external magnetic field and thus resonate at a frequency lower than their usual 1ppm. In fact, protons of aliphatic methyl groups that resonate at 0ppm indicate that a protein is stable and folded (Jacobsen 2007).

2.2.2.5. The Nuclear Overhauser Effect - NOE

The Nuclear Overhauser effect is seen when two nuclei interact directly via through-space (not through bonds) polarization transfer. However, this effect is usually seen only in nuclei of a distance of 5Å or less from each other. Even if they are not related to each other through bonding, if they are close in space then the effect will be seen (Anderson & Freeman 1962). The magnetization transfer between spins during NOE interactions occurs via the magnetic dipoles of the protons. For proteins of molecular weight >5kDa, the NOE intensity is approximately proportional to the distance between two interacting spins, to the inverse sixth power (r^{-6}) (Wider 2000).

2.2.3. Protein NMR Experiments

2.2.3.1. The NMR Sample

For isotopic labelling, a protein is expressed in bacteria grown in a medium with $^{15}NH_4Cl$ as the only nitrogen source, and with ^{12}C-glucose or ^{13}C-acetate as the only carbon source. A protein sample volume in a standard 5mm NMR tube should be approximately 500µl. For protein samples, minimal concentrations from approximately 100µM are normally required to gain a high-resolution ^{1}H NMR spectrum, however higher concentrations are needed for recording ^{13}C or ^{15}N spectra, with Shigemi tubes allowing for smaller volumes in order to reach higher concentrations.

The reason for replacing naturally abundant ^{12}C carbon with the ^{13}C isotope is because the nucleus of ^{12}C has an even number of both neutrons and protons in which the net nuclear spin quantum number is 0 due to paired spin contributions of these particles. Such a nucleus is not measurable using NMR spectroscopic methods. However, due to the ^{13}C isotope having an uneven number of neutrons (7) with an even number of protons (6), it has a spin-½ nucleus, measurable by NMR (Macomber 1997, Wider 2000).

The naturally abundant ^{14}N nucleus is measurable using NMR spectroscopy due to it having a net nuclear spin quantum number of 1. However, due to its large quadrupole moment, which interferes with the effects of the external magnetic field, ^{14}N nuclei result in weak and very broad chemical shifts, resulting in a less sensitive and low-resolution NMR experiment (Jeong et al 2002). This explains the replacement of the ^{14}N atoms of the protein with the ^{15}N isotope, also a spin-½ nucleus (Macomber 1997, Wider 2000).

The sample to be studied is usually dissolved in a liquid solvent. Biological NMR samples are usually dissolved in a solvent of 90% H_2O with 10% D_2O. Deuterium (^{2}H) is added to the sample so the magnetic field of the spectrometer is locked to a reference frequency. A continuous wave NMR experiment is used to monitor any changes in this deuterium signal, and thus any changes in the magnetic field (Keeler 2005).

2.2.3.2. 1D ^{1}H NMR Experiments

In a 1D ^{1}H-NMR spectrum of a protein it is possible to identify certain chemical shift positions as ^{1}H nuclei in particular chemical environments within a protein molecule (Fig 2.18). The signal-to-noise ratio is a good indication of whether the sample is of an appropriate protein concentration suitable for further experiments. Using ^{1}H NMR, it is possible to see whether a protein sample is unfolded or aggregated if there is minimal dispersion of resonances for different chemical groups, i.e. they are localised in the 'random coil' chemical shift range. In these spectra, a large peak at approximately 1ppm, representing methyls/methylene protons, indicates the presence of unstructured regions in a protein or aggregated material. However, the chemical shifts of 1D ^{1}H spectra of proteins that overlap significantly, these cannot be used for structural interpretation.

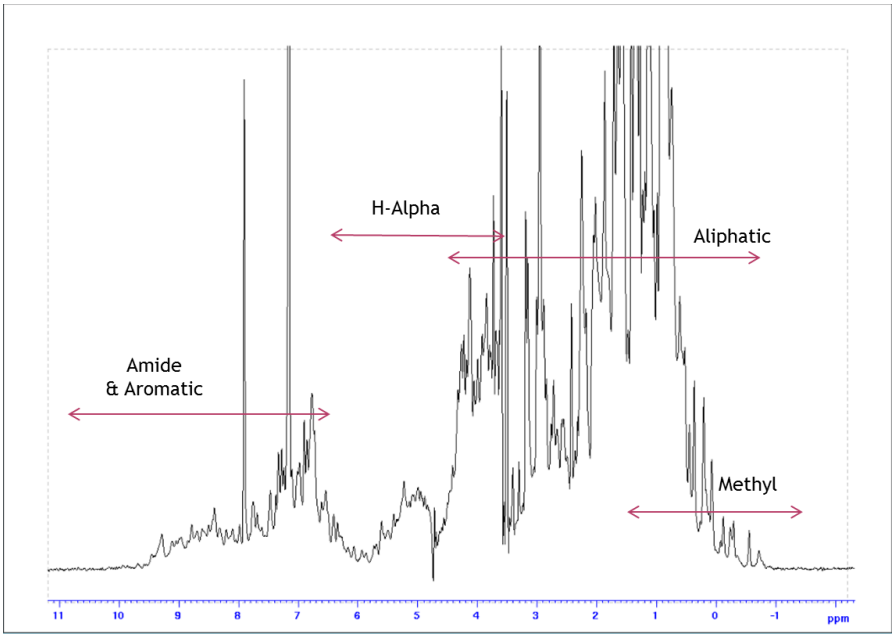


Figure 2.19: A 1D ^{1}H protein spectrum.
With regards to proteins, the ^{1}H nuclei that precess are connected to backbone nitrogen (HN), the α-carbon and atoms on the side chains (Hβ, Hγ etc). These particular positions will be represented by peaks at specific chemical shifts due to the effect of neighbouring atoms. Backbone amide ^{1}H nuclei will have shifts around 7-11ppm, side-chain amide and aromatic ^{1}H nuclei around 6.5-9ppm, α-carbon ^{1}H nuclei around 3.5-5.5ppm, side-chain ^{1}H nuclei between -0.5-3.3ppm and methyl groups between -0.3-1.3.

2.2.3.3. 2D NMR Experiments and TROSY

A 2D NMR experiment can be thought of as a magnetization 'jump' rom one nucleus to another either via J-coupling interaction (through-bond) or based on an NOE interaction (through-space), recorded via detecting nuclei in direct and indirect dimensions. 2D experiments involve two 90° RF pulses of frequencies which result in the transfer of magnetization from one nucleus to a second one. After the first pulse, nuclear spins precess for a time of t1, and this magnetization will be attributed to the chemical shifts of the first nucleus in question (E.g. ^{1}H). The second pulse gives rise to the magnetization transfer. Detection at the end of the experiment is also referred to as direct evolution time, t_2. The latter magnetization will be attributed to the chemical shifts of the second nucleus in question - this will result in a spectrum with two frequency axes (Macomber 1998).

Transverse relaxed optimised spectroscopy (TROSY) is used to improve both 2D and 3D NMR experiments by enhancing acquisition of sample FID via experimental tuning to a coherence at which there is a cancellation between dipole-dipole coupling and chemical shift anisotropy. In this way, the resonance frequency of the most slowly relaxing component is used for FID acquisition, which results in a signal of higher resolution (Pervushin et al 2000). TROSY experiments are able to negate the loss of signal that often occurs due to the fast tumbling and fast relaxation times seen in proteins larger than 30kDa. However, TROSY is only an option if a high magnetic field strength (>800MHz) is available (Gardner & Kay 1998).

2.2.2.4. Protein Assignment of Unlabelled Proteins with 2D NMR Spectroscopy

2D NMR experiments involve excitation of nuclear spins prior to detection of chemical shifts during an evolution period, t_1. Subsequently, during a mixing period, nuclear spins are correlated. This results in the mixing of chemical shift information from one nucleus with another nucleus, with the chemical shift of the latter being recorded in detection period t_2. These periods involve delays followed by radiofrequency pulses, named a pulse sequence. The NMR spectral assignment of protein residues can be carried out via 2D experiments. For small proteins (no more than approximately 80 residues long), assignment of residues can be achieved with 2D ^{1}H NMR techniques that either do not require the protein sample to be isotopically labelled, or may only require the use of a single isotopic label (usually ^{15}N or ^{13}C). These techniques include 2D Correlation Spectroscopy (2D-COSY), 2D Nuclear Overhauser Effect Spectroscopy (2D-NOESY) and 2D- Total Correlation Spectroscopy (2D-TOCSY).

The through-bond, J-coupling between ^{1}H resonances of peptide amide links with the resonances of protons linked to the Cα of the same residue can be correlated, and show a cross-peak in a 2D-COSY spectrum (Figure 2.19). These proton nuclei are thus interacting during the mixing time period of the experiment. The through-space NOE coupling between a Cα proton of one residue to an amide ^{1}H in the second residue can be seen with cross-peaks on a 2D-NOESY spectrum (Figure 2.20), as these protons would be 5Å close in space (Macomber 1998, Jacobsen 2007).

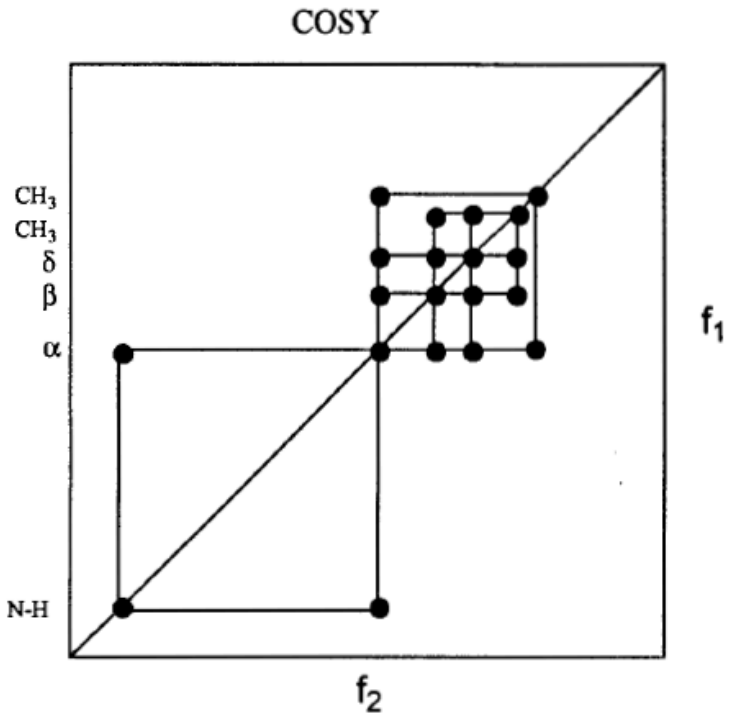


Figure 2.20: The overall layout of a 2D-COSY spectrum.
This correlates ^{1}H-nuclei with other ^{1}H-nuclei via through-bond, spin-spin (J-coupling) transfer of magnetization. Peaks relative to the diagonal represent NOEs (image from Macomber 1998).

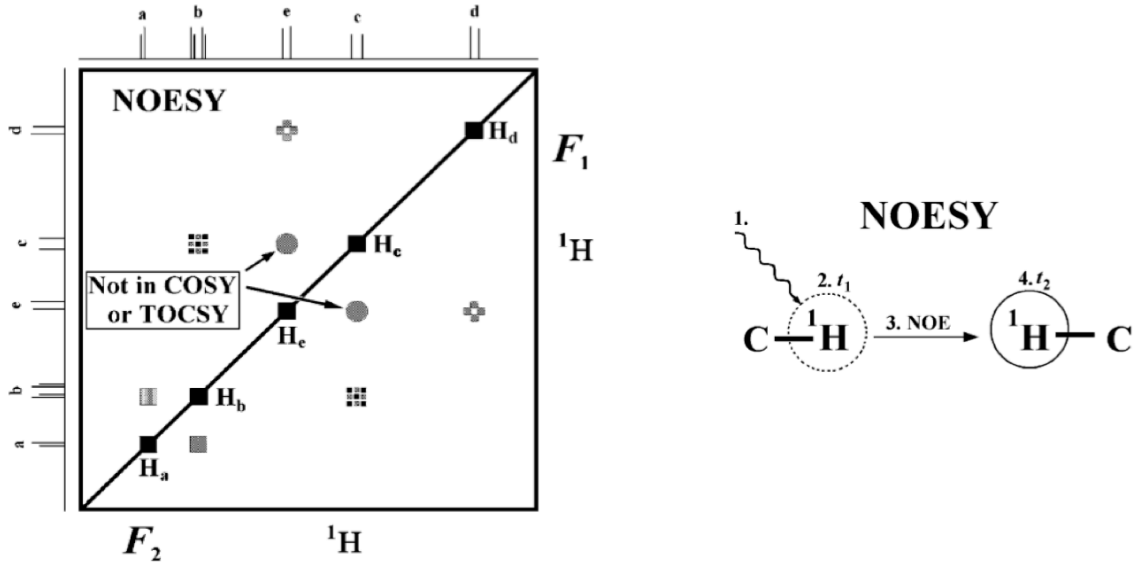


Figure 2.21: The overall layout of a 2D-NOESY spectrum.
This correlates ^{1}H-nuclei with other ^{1}H-nuclei via through-space (NOE) transfer of magnetization. Peaks relative to the diagonal represent NOEs (images from Jacobsen 2007).

When combining 2D-COSY and 2D-NOESY correlations, it is possible to assign residues down the polypeptide chain (Figure 2.21). The N-terminus should be assigned from the unique chemical shifts of the amine group protons. The residue side chains are then assigned using 2D-TOCSY spectra, in which cross-peaks represent both through-bond and through-space correlations (Figure 2.22).

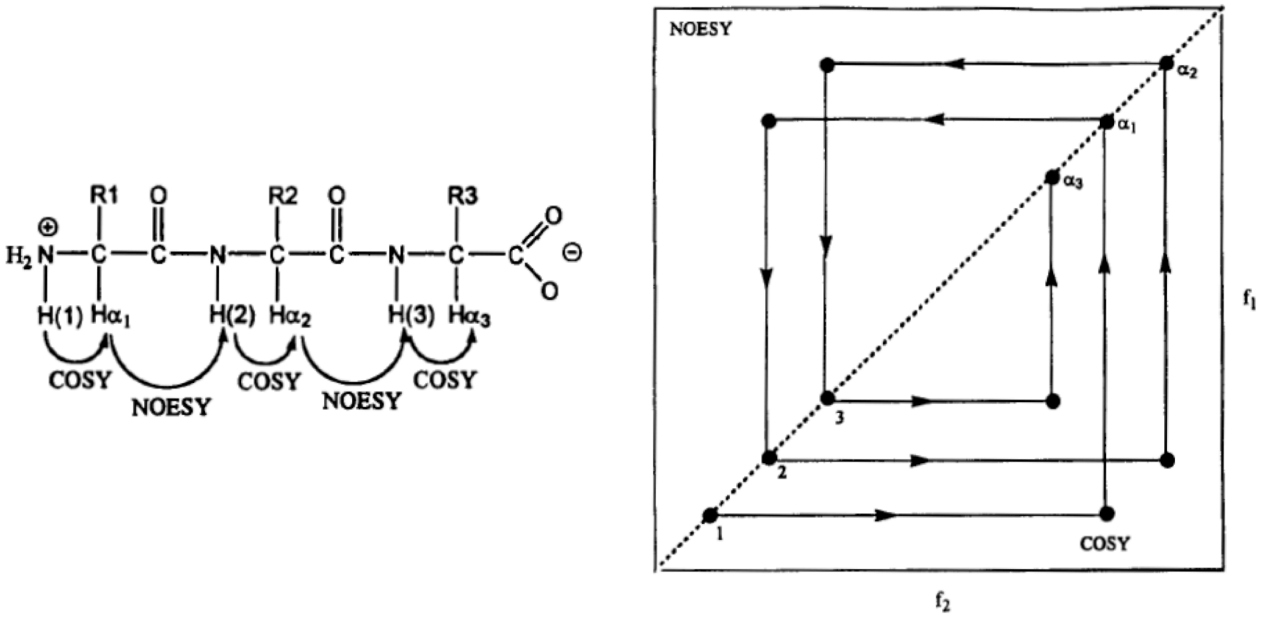


Figure 2.22: The Sequential Assignment of Residues using COSY and NOESY Combinations.
By combining 2D-COSY and 2D-NOESY spectra, the assignment of residues can be carried out from the correlation of through-bond J-coupling magnetization transfer between amide ^{1}H protons with Cα protons on the same residue (COSY) plus magnetization transfer through space (NOEs) between Cα protons and amide ^{1}H protons of different residues (image from Macomber 1998).

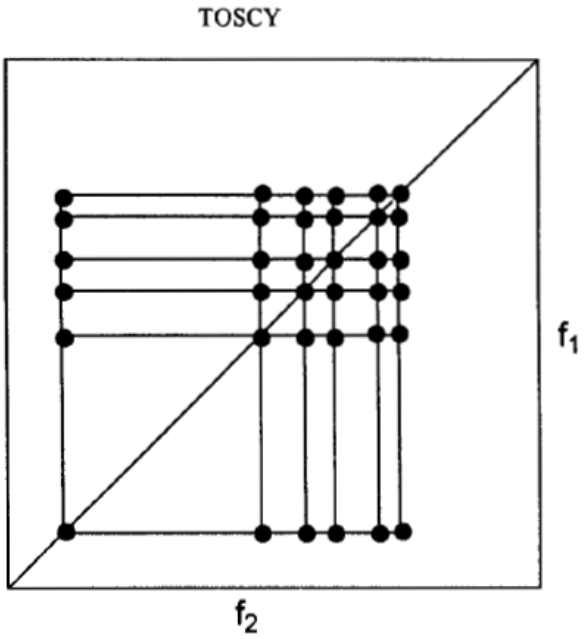


Figure 2.23: Example of a 2D-TOCSY spectrum.
Both J-coupling spin-spin correlations, and NOE cross-peaks of ^{1}H protons are shown on a 2D-TOCSY, in which residue side-chain information can be deduced from remaining unassigned cross-peaks (Image from Macomber 1998).

2.2.2.5. Protein Assignment of Isotopically Labelled Proteins with 2D and 3D NMR Spectroscopic Techniques

For proteins larger than 80 residues, 2D-COSY, 2D-NOESY and 2D-TOCSY spectra can become too crowded for correct assignment to be carried out and so such 2D techniques are at a disadvantage. Furthermore, mis-assignments can occur due to the fact that NOESY spectra can be influenced by secondary structure. For these reasons, heteronuclear triple resonance experiments that require only through-bond magnetization transfer correlations are used for large proteins, and for these techniques, proteins must be isotopically labelled with ^{13}C and ^{15}N. Plus, the magnetization transfer from J-couplings occurs more efficiently in comparison to that due to NOEs. As both relaxation and magnetization loss happens at a faster rate in larger proteins, the effect of efficient magnetization transfer rates puts such J-coupling reliant methods at an advantage (Sattler et al 1999).

Assignments from heteronuclear triple resonance experiments can be displayed on 2D-^{15}N-^{1}H Heteronuclear Single Quantum Correlation Spectroscopy (HSQC) Spectrum, which shows through-bond correlations between ^{15}N and ^{1}H nuclei in amide and amine N-H groups, from the chemical shift interactions of these nuclei during the mixing period of the 2D experiment. Cross-peaks in a ^{1}H-^{15}N-HSQC spectrum represent the backbone N-H group of each residue in the protein except proline (due to it having no amide proton), so is thought of as a unique 'fingerprint' for the particular protein (Figure 2.23). However, the ^{1}H-^{15}N-HSQC spectrum will also show cross-peaks corresponding to amine groups from particular side chain including glutamine and asparagine. ^{1}H-^{13}C-HSQC spectra can also be generated. No diagonal line of peaks is seen on a HSQC spectrum as the different frequencies corresponding to the different heteronuclei are recorded during evolution (t_1) and detection (t_2) periods (Wider 2000, Kwan et al 2011).

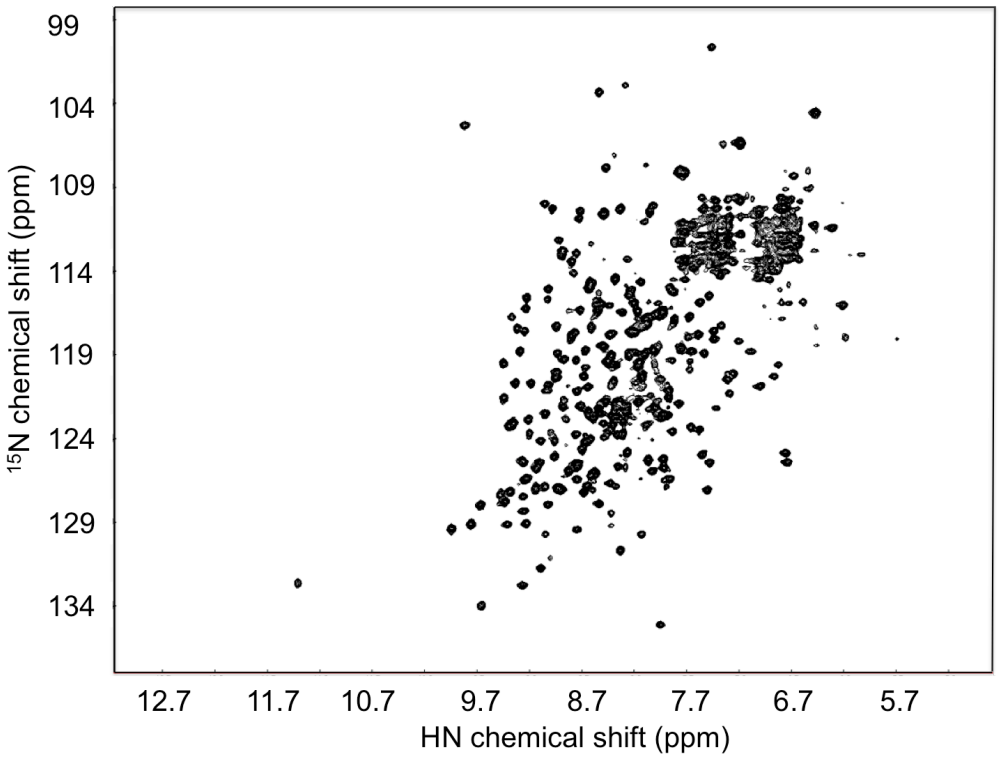


Figure 2.24: Example of a ^{1}H-^{15}N HSQC spectrum.
Each peak represents a correlation of the chemical shifts of ^{1}H and ^{15}N nuclei, for example in backbone amide groups. Each amino acid has its own amide group, but some residues (e.g. tryptophan and glutamine) will contribute two peaks due to their side-chain amine N-H groups.

For proteins larger than 80 amino acids, heteronuclear triple resonance experiments are used to assign H-N, ^{15}N, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei in proteins labelled with ^{1}H, ^{13}C and ^{15}N isotopes. These experiments involve a sequence of pulses interspersed by delays, with the effect of transferring magnetization via J-coupling between particular nuclei (Yamazaki et al 1994). These experiments can, for instance, correlate amide ^{1}H and ^{15}N chemical shifts with the chemical shift of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei in prior residues using the carbonyl carbon in between (see Figure 2.24) (Wider 2000).

Deuteration of the protein sample is often needed, especially in larger proteins in which slow tumbling rates and fast transverse relaxation can lead to peak broadening, greater overlap of peaks, signal loss and consequently an increase in FID acquisition time, especially for ^{13}C nuclei. Replacing ^{1}H with deuterium decreases the transverse relaxation rates of ^{13}C, and improves the sensitivity of the FID signal as well as influencing $^{13}C^{\alpha}$ chemical shifts in a way that is dependent on their position in the secondary structure (Gardner & Kay 1998, Wider 2000).

The triple resonance experiments include the HNCA, HNCO, HN(CA)CO, HNCACB and HN(CO)CACB experiments. The pulse sequences involve the excitation of the ^{1H} proton, correlated with the ^{15}N nucleus of an amide group via through-bond magnetization transfer, in which the chemical shift of the latter is acquired during evolution period t_1. The ^{15}N nucleus is correlated via carbonyl carbon with the C^{α} nucleus of the preceding residue, and in which the frequency of the C^{α} nucleus is acquired during a second evolution period t_2. Magnetization transfer from the ^{15}N and ^{13}C nuclei travels back to the amide proton originator of the sequence, and subsequently during t_3, the frequency of the proton is detected. Via Fourier transforms of this data, the result is a 3D spectrum with ^{1}H, ^{15}N and ^{13}C dimensions with cross-peaks representing the correlation between the chemical shifts of the ^{1}H-^{15}N group of one residue and the chemical shifts of the C^{α} nucleus from the preceding residue.

For example, the HNCA experiment involves the transfer of magnetization from the ^{1}H proton to the ^{15}N nucleus of an amide group, to the C^{α} nucleus on the 'i' (Same) and 'i-1' (preceding) residues, with magnetization returning back to the amide proton, with their resonant frequencies correlated (Fig 2.24). Whilst pulse-NMR is used for recording ^{1}H chemical shifts, chemical shifts of ^{15}N and ^{13}C nuclei are measured using constant-time NMR (Yamazaki et al 1994, Gardner & Kay 1998).

The HNCO experiment involves magnetization transfer from the ^{1}H (i) nucleus via the ^{15}N nucleus bonded to a ^{13}C C=O (i-1) carbonyl group and the transfer returns back to the ^{1}H nucleus (i). The HN(CA)CO experiment involves magnetization transfer from the ^{1}H (i) proton of the HN group (i), via the ^{15}N nucleus and the C^{α} (i) to the C=O ^{13}C (i) and C=O ^{13}C (i-1) atom, and back. The C^{α} atom is shown in brackets as its frequency is not detected – it is used as a transfer nucleus only, with the frequencies of HN, N and CO nuclei recorded (Yamazaki et al 1994).

The magnetization transfer in the HN(CA)CB proceeds in a similar way, except correlating frequencies between both the 13Cα and 13Cβ of both the 'i' and 'i-1' residues with the ^{1}H and ^{15}N nuclei, and the HN(CO)CACB correlates frequencies of the same but only for 'i-1' residues (Yamazaki et al 1994) (Fig 2.24). These experiments, along with the primary sequence of a protein, can elucidate the identity of the residues represented by the H-N correlation peaks of a ^{1}H-^{15}N-HSQC spectrum.

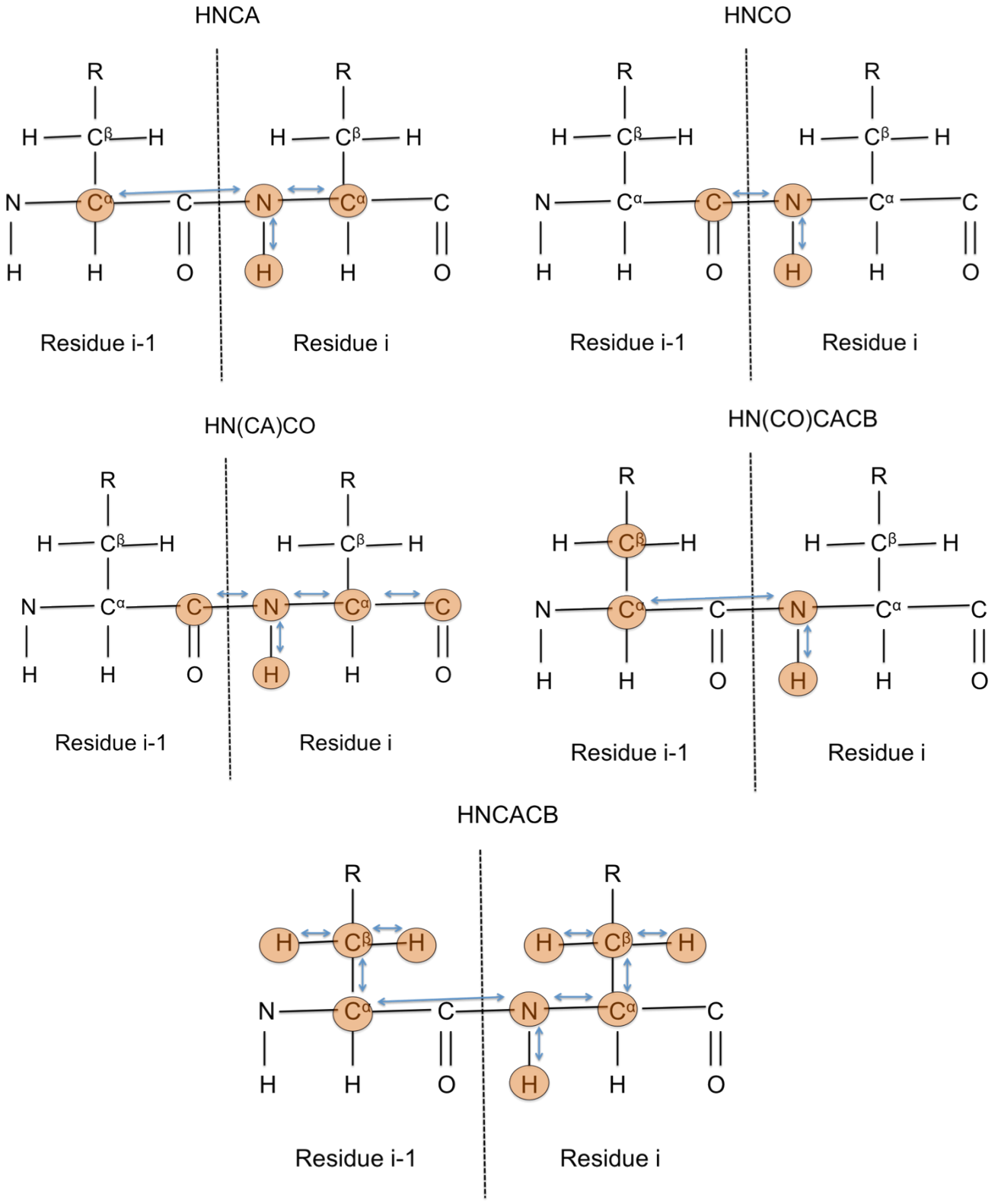


Figure 2.25: Magnetization transfer between nuclei in backbone assignment experiments.
The magnetization transfer (blue arrows) between nuclei that characterises each of the backbone resonance assignment experiments. Residue i corresponds to the same residue, residue i-1 corresponds to the previous one (Yamazaki et al 1994).

3. Materials and Methods

3.1. Cloning

The DNA sequences used for cloning of ALS genes were originally obtained courtesy of Lois Hoyer. In early stages of the work, sNT-Als1 and sNT-Als3 were produced as thioredoxin-fusion constructs and cleaved with Factor Xa. However, due to incomplete cleavage, some loss of protein via precipitation during the cleavage step, and the time consuming nature of this extra step, these constructs were re-cloned into pRSETa plasmids, sNT-Als3 also cloned into a pET46 plasmid.

3.1.1. Cloning of sNT-Als1 and sNT-Als3 into pET32 Plasmids

The cloning of sNT-Als1 (amino acids 1-300) was carried out by Dr Robert Yan, and the cloning of sNT-Als3 (amino acids 1-302) was carried out by Dr Jing Lin, using ligation independent cloning (Novagen). Short constructs of NT-Als1 and NT-Als3 had been designed in order to remove the flexible C-terminus and also the AFR, in order to make these more amenable for NMR spectroscopic ligand-binding studies.

3.1.2. Cloning of sNT-Als1 and sNT-Als3 into pRSETa Plasmids

3.1.2.1. Primer Design

Primers listed in Table 3.1 were designed so that the resulting DNA insert would have a stretches of 10 bases homologous with the terminal ends of the linearized pRSETa vector, contain recognition sequences for restriction enzymes that leave 5' overhangs after excision, and also have a suitable length of DNA insert sequence, according to instructions from the In-Fusion® HD EcoDry™ Cloning Kit User Manual. Melting temperatures and secondary-structures were checked using OligoCalc. They were synthesised by Invitrogen UK.

Table 3.1: Forward and Reverse Primers.
The primers were designed for the PCR amplification of DNA inserts coding for sNT-Als1 and sNT-Als3, to be cloned into pRSET plasmids.

Direction	Sequence
sNT-Als1 Forward	5'-GGTTCCGCGTGGATCCAAGACAATCACTGGTGTTTTTGATAGTTTTAAT-3'
sNT-Als1 Reverse	5'-CAGCCGGATCAAGCTTATTCTTGTATCCAGTCCATCTTAAAGT-3'
sNT-Als3 Forward	5'-GGTTCCGCGTGGATCCAAGACTATCACTGGTGTTTTCAACAG
sNT-Als3 Reverse	5'-CAGCCGGATCAAGCTTCTATATCCAGTCCATCTTAATGTGAAAGG
Key: Red – Sequences homologous with pRSETa Vector. Violet – Restriction sites. Underlined – Stop Codon. Blue – Insert Sequence	

3.1.2.2. Polymerase Chain Reaction (PCR)

The PCR reaction was used to amplify sNT-Als1 and sNT-Als3 sequences from template pET32 DNA plasmids, with reactions made up with reagents listed in the Table 3.2. Reagents were combined into PCR tubes on ice, with Phusion DNA polymerase added last. A final DMSO concentration of 10% was added as this can facilitate amplification of GC-rich sequences and prevent the formation of DNA secondary structures.

Table 3.2: Volumes and concentrations of PCR reagents used to amplify sNT-Als1 and sNT-Als3 DNA sequences.

Reagents and Concentrations	Final Concentration	Volume (µl)
5x Phusion HF buffer	1x	10
10mM dNTPs	200µM	1
100% DMSO	5%	5
10µM Forward primer	0.5µM	2.5
10µM Reverse primer	0.5µM	2.5
400ng/µl Template	20ng/µl	2.5
Nuclease Free Water	-	26
Phusion DNA polymerase	1.0 units	0.5
Total Volume		**50µl**

The PCR reaction was carried out in an Eppendorf Mastercycler Gradient thermocycler using the temperature steps listed in Table 3.3.

Table 3.3: Times and temperatures of PCR reaction steps performed on the thermocycler.

Step	Temperature (°C)	Time	
1. Initial Denaturation	98	30 seconds	
2. Denaturation	98	10 seconds	Repeat for 35 cycles
3. Annealing	72	30 seconds	
4. Extension	72	30 seconds	
5. Final extension	72	10 minutes	
6. Hold	4	Infinite	

3.1.2.3. DNA Agarose Gels

PCR products were analysed using 1% (w/v) agarose gels made with TAE buffer (40mM Tris-HCl pH8, 0.114% glacial acetic acid, 1mM EDTA) and 10,000x SYBR Safe for staining (Invitrogen). DNA samples were mixed with 6x loading dye (3.3mM Tris-HCl pH8, 2.5% ficoll 400, 11mM EDTA, 0.017% SDS and 0.015% bromophenol blue) and run at 100V for approx. 30 minutes alongside 1Kb and 100bp DNA ladders (New England Biolabs). A blue-light transilluminator was used to visualise the DNA bands.

3.1.2.4. DNA Extraction and Purification

DNA bands of the correct size were cut from the agarose gel using a scalpel. DNA was extracted and purified using protocol from the Gel Extraction Kit (Thermo Scientific) to remove impurities.

3.1.2.5. Digestion of pRSETa Plasmid

The pRSETa plasmid was digested with two restriction enzymes BamHI and HindIII (New England Biosciences) in order to insert the purified PCR products that coded for sNT-Als1 and sNT-Als3. This was carried out by adding the reagents listed in Table 3.4, into a 1.5ml Eppendorf tube which was placed in a 37°C incubator overnight. Digested products were visualised on a 1% agarose gel and purified using Gel Extraction Kit (Thermo Scientific) using the protocol detailed in sections 3.1.2.3 and 3.1.2.5.

Table 3.4: **Reagents and concentrations for the double digestion of pRSETa plasmid with Restriction Enzymes BamHI and HindIII.**

Reagent	Final Concentration	Volume (µl)
pRSETa Plasmid DNA	1µg	4
BamHI	10 units	1
HindIII	10 units	1
10x NEB buffer 4	1x	5
100x BSA	0.1x	0.5
Nuclease Free Water	-	38.5
Total Volume	**50µl**	
1x NEB Buffer 4 Components: 50mM Potassium acetate, 20mM Tris-acetate, 10mM Magnesium acetate, 1mM DTT, pH 7.9		

3.1.2.6. DNA Ligation

Purified DNA products were ligated to a digested and purified pRSETa plasmid using protocol and instructions from an In-Fusion EcoDry Cloning kit. 100ng of pRSETa plasmid and PCR product was added to each tube containing a reaction pellet, and mixed before being incubated for 15 minutes at 37°C and then for 15 minutes at 50°C using a thermocycler. The reaction mix was stored on ice prior to transformation.

3.1.2.7. The pRSETa Plasmid

The pRSETa plasmid was chosen for its T7 promoter for a high level of expression, N-terminal 6x histidine tag for Nickel purification later on and also for its relatively small size of 2.9kb (Fig 3.1). Smaller plasmids have been documented to have better transformant retention in comparison to plasmids of larger sizes, which would be an advantage for expression. Furthermore, plasmids of larger sizes are also less likely to be uptaken by cells. The pRSETa plasmid used in this study had been modified by members of the lab to not contain the Xpress Epitope as protein identification method used later on was Mass Spectrometry, not Western Blotting with antibodies that recognised this epitope.

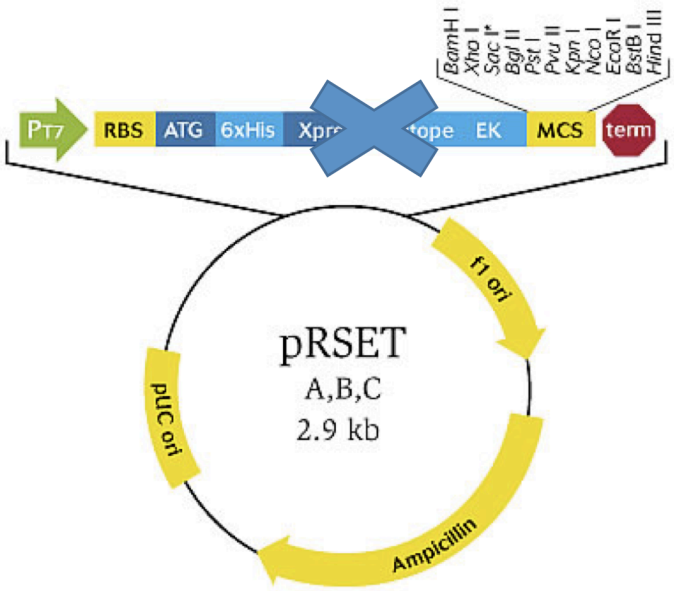


Figure 3.1: Diagram of the pRSETa plasmid.
This plasmid contains a T7 promoter, 6xHistine tag and ampicillin/carbenicillin resistance gene. The plasmid had been modified to not contain the Xpress epitope.

3.1.2.8. Transformation of Plasmid

To 50µl of DH5α supercompetent cells (New England Biolabs) in 1.5ml Eppendorf tubes, 1-5µl of In-Fusion cloning reaction was added, tubes were flicked 3 times, and the mixture left on ice for 30 minutes. The cells were heat-shocked at 42°C for 30 seconds and left on ice for 5 minutes, prior to adding 700µl of SOC media (New England Biolabs) and subjecting to shaking incubation at 250rpm and 37°C for 60 minutes. 120µl of mixture was spread onto LB agar plates with 100µg/ml carbenicillin antibiotic for selection, and incubated at 37°C overnight. As well as 120µl of the initial transformation mixture, the rest of the culture was spun at 6000rpm for 5 minutes, suspended in 120µl of fresh SOC media and applied to LB agar plates.

3.1.2.9. Plasmid Amplification and Purification

Colonies from transformation plates were grown in 5ml cultures of LB media (100µg/ml carbenicillin) at 37°C overnight and pelleted by spinning at 4500rpm for 10 minutes. Supernatant was removed and plasmids extracted and purified using materials and protocol from a Miniprep Kit (Thermo Scientific).

3.1.2.10. DNA Sequence Analysis

A sample of DNA was sent to Source Bioscience for commercial sequencing. They were analysed using both forward and reverse T7 primers. They were aligned using Clustal Omega with original gene sequence to confirm that there were no mutations.

3.1.2. Cloning of sNT-Als3 into pET46 Ek/LIC Plasmid

3.1.2.1. Primer Design

The sNT-Als3 was also cloned using Ligation-Independent Cloning into a pET46 Ek/LIC plasmid (Novagen) in order to test its expression as a his-tagged construct from a plasmid other than pRSETa. Primers were designed (listed in Table 3.5) in order for the sNT-Als3 DNA insert to anneal to the pET46 plasmid, as instructed in the Novagen Ek/LIC cloning manual, and synthesised by Invitrogen after checking annealing temperatures and secondary structure formation with OligoCalc. Amplification of the sNT-Als3 insert was carried out using the PCR method detailed in section 3.1.2.2, plus visualised and purified using the methods listed in section 3.1.2.3 and 3.1.2.4.

Table 3.5: Primers designed for the amplification of DNA insert coding for sNT-Als3 for cloning into pET46 Ek/LIC plasmid.

Direction	Sequence
Forward	5'-GACGACGACAAG AAGACTATCACTGGTGTTTTCAACAGT-3'
Reverse	5'-GAGGAGAAGCCCGGTTATATCCAGTCCATCTTAATGTG-3'
Key: Red – Plasmid annealing sequence. Underlined – Stop Codon. Blue – Insert Sequence	

3.1.2.2. DNA Ligation

The purified PCR products were treated with T4 DNA polymerase (New England Biosciences), as its 3'-5' exonuclease activity results in insert DNA having compatible overhangs for annealing with the pET46 Ek/LIC plasmid. The reagents in Table 3.6 were added to a 1.5ml Eppendorf tube.

Table 3.6: The reagents for T4 DNA polymerase treatment of the PCR product DNA insert of sNT-Als3.

Reagent w/ Concentrations	**Volume (µl)**
0.1pmol purified PCR product	x
10x T4 DNA Polymerase buffer	2
25mM dATP	2
100mM DTT	1
2.5U/µl T4 DNA Polymerase	0.4
Nuclease-free water	y
Total Volume	**20**
Final concentration of insert – 0.01pmol/µl	

T4 DNA polymerase was added last as this started the reaction, which was then mixed with a pipette-tip and incubated for 30 minutes at 22°C, with the enzyme reaction inactivated by a second incubation step of 20 minutes at 75°C. The T4 DNA polymerase treated insert at a volume of 2µl was mixed with 1µl of pET46 Xa/LIC vector (Novagen) and incubated in a sterile 1.5ml Eppendorf tube at 22°C for 5 minutes in order to anneal the insert to the vector. 1µl of 25mM EDTA (Novagen) was pipetted into the tube, which was incubated at 22°C for a further 5 minutes in order to stop the annealing reaction.

3.1.2.3. The pET46 Plasmid and Amplification

The pET46 EK/LIC vector contains a T7 promoter for high bacterial expression, an N-terminal 6x his-tag, selectivity for ampicillin/carbenicillin resistance and also an enterokinase (EK) site in order to cleave the 6x his-tag if needed (Fig 3.2). The annealed plasmid was transformed into competent cells, amplified, purified and sent for DNA sequence analysis using methods detailed in sections 3.1.2.8, 3.1.2.9 and 3.1.2.10.

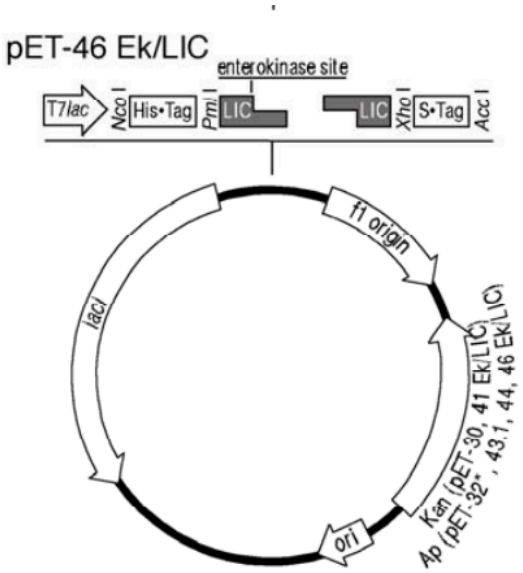


Figure 3.2: Diagram of the pET46 Ek/LIC plasmid.
This includes a T7 promoter, his-tag and EK site. Contains the gene for ampicillin and carbenicillin resistance.

3.2. Expression of sNT-Als1 and sNT-Als3 Proteins

3.2.1. Shuffle Cells

Both sNT-Als1 and sNT-Als3 constructs were expressed in *E. coli* Shuffle cells (New England Biolabs) as these engineered K12 cells constitutively express DsbC disulphide-isomerase. This enzyme aids the correction of mis-oxidised, disulphide-bond containing proteins, to correctly oxidized forms in the usually reducing environment of the bacterial cytoplasm. As both sNT-Als1 and sNT-Als3 proteins contain four disulphide bonds, this strain would be ideal for expressing such proteins.

3.2.2. Protein Expression in LB Media

Plasmids were transformed using the same method documented in section 3.1.2.8, except Shuffle cells were used. From the transformation plates, colonies were streaked and inoculated into 10ml cultures of LB containing 100µg/ml carbenicillin (an antibiotic known to be less susceptible to degradation in comparison to ampicillin) which were incubated overnight at 37°C. This start-up culture was added to 850ml LB media in 2.5L flasks, with the same carbenicillin concentration, which was then incubated at 37°C in an orbital shaker until OD_{600} of the cells reached 0.6-0.8. 1mM of IPTG (iso-propyl-b-1-thio-galactopyranoside) was used to induce protein expression from the plasmid, and the culture was incubated at 18°C overnight prior to harvesting. Lowering of the temperature was done in order to prevent overloading of protein folding components in the cells and facilitate production of properly folded cells.

3.2.3. Protein Expression in ^{15}N-Minimal Media

In order to produce protein labelled by ^{15}N, M9 minimal media was used instead of LB media and ^{15}N-ammonium chloride (Sigma) was used as the sole nitrogen source.

3.2.4. Protein Expression in ^{15}N-^{13}C-D_2O Minimal Media

To produce ^{15}N and ^{13}C labelled protein that would be more amenable to three-dimensional NMR spectroscopy studies, M9 minimal media was used but H_2O was replaced by D_2O (Sigma). ^{15}N-ammonium chloride and ^{13}C-D-glucose were used as the nitrogen and carbon sources (Sigma). As well as the 10ml start LB media starter culture (100µg/ml carbenicillin), a secondary starter culture of 100ml of 50/50 H_2O M9 minimal media/D_2O was used to inoculate the final culture.

3.3. Protein Purification

3.3.1. Cell Harvesting and Lysis

Cell cultures were harvested for 15 minutes at 4500rpm and 4°C. Pellet was resuspended in 25ml binding buffer (50mM $NaPO_4$ pH 8, 500mM NaCl, 10mM imidazole) protease inhibitors 1mM PMSF, 100µM Aprotinin, 1mM Leupeptin and 1mM Pepstatin were added to prevent protein degradation of proteases during cell lysis step. Cells on ice were lysed using sonication using a medium-sized probe at 70% power for 5 minutes. Clear cell lysates were obtained by centrifugation at 15000rpm for 50 minutes.

3.3.2. Nickel-IMAC Purification

Protein was purified from the supernatant using an Ni-NTA agarose (Qiagen) column equilibrated with binding buffer. Supernatant was added to the column, followed by washing with Wash Buffer (50mM $NaPO_4$ pH 8, 500mM NaCl, 10-20mM imidazole), Elution Buffer (50mM $NaPO_4$ pH8, 500mM NaCl, 300mM imidazole) and Resin Leach Buffer (50mM $NaPO_4$ pH8, 500mM NaCl, 100mM EDTA), with eluate and samples from each step collected. Resin Leach Buffer was often used because often not all protein was eluted from the column with Elution Buffer. To remove imidazole, EDTA, protease inhibitors and any traces of Nickel, samples collected after washing the column with Elution buffer and Resin Leach Buffer were dialysed separately in 4L of Dialysis Buffer (50mM $NaPO_4$, 500mM NaCl).

3.3.3. Factor Xa Protease Digestion

For Als proteins that had been expressed as thioredoxin fusions (using the pET32 plasmid), the eluted protein sample was buffer exchanged using dialysis into Factor Xa Cleavage Buffer (50mM Tris-HCl pH8, 50mM NaCl). Protein concentration of eluted sample was determined using a Nanodrop at 260 and 280nm and an amount of Factor Xa enzyme (approx. 1U/mg) used in relation to protein

concentration was added to the protein sample in a Spectra/Por® dialysis membrane (MWCO 5kDa). The mixture was incubated at room temperature for 16-24 hours with gentle agitation. The sample was added to equilibrated Ni-NTA resin in order to remove the his-tagged thioredoxin, and to collect his-free Als proteins.

3.3.4. Size Exclusion Chromatography

Als proteins were further purified using an ÄKTA HPLC (Amersham Biosciences) and Superdex^{TM}75 prep grade column (GE Healthcare), with Gel Filtration Buffer (50mM $NaPO_4$ pH8, 500mM NaCl, and using a flow-rate of 0.5ml/min. Protein samples were concentrated using Vivaspin 20 centrifugal concentrators (MWCO 3KDa) to 1ml prior to column application. Fractions were collected, analysed by SDS-PAGE, pooled and concentrated depending on whether they contained protein of expected Als protein size.

3.3.5. SDS-PAGE

Protein samples were combined in a 3:1 ratio with SDS sample buffer, heated at 100°C for 3 minutes and loaded onto commercially made 12% polyacrylamide glycine gels alongside a protein marker (Invitrogen). Samples ran for 45 minutes at 160V. Gels were stained by InstantBlue.

3.3.6. Protein Concentration Determination via UV Spectroscopy

UV Spectroscopy was used to determine the concentration of protein in each sample. A Nanodrop (Thermo Scientific) detected the absorbance of protein at a 280nm wavelength and the concentration was calculated using Beer's Law, where A represents absorbance at 280nm, ε is the extinction coefficient, c is the molar concentration of protein and 1 is the length of the light path. Protparam (ExPASy) was used to estimate the extinction coefficients of all proteins studied.

Beer's Law: $A_{(280)} = \varepsilon \times c \times 1$

3.4. NMR Spectroscopy

3.4.1. NMR Sample Preparation

Protein samples were concentrated to 200-600µM in a sample volume of 500µl using Vivaspin concentrators (MWCO 5kDa) at 4500rpm. They were buffer exchanged using the concentrators into NMR Buffer (50mM $NaPO_4$ pH 7, 50mM NaCl). 10% D_2O was added to the sample to ensure frequency locking of the NMR spectrometer. 0.1% sodium azide was added to prevent bacterial growth and a suitable amount of protease inhibitors added to prevent degradation of the sample.

The triple labelled sNT-Als1 had to be prepared for NMR spectroscopic backbone assignment studies in a different way from unlabelled or 15N-labelled sNT-Als1. After Nickel-IMAC purification and the imidazole-removal dialysis step, this sample was dialyzed in 4L of Unfolding Buffer (50mM $NaPO_4$ pH 8, 300mM NaCl, 4M urea) for 12 hours in order to unfold the protein and perform hydrogen-exchange of amine groups, and then dialyzed again with Re-Folding Buffer (50mM $NaPO_4$ pH 8, 300mM NaCl) for a further 12 hours. Afterwards, the sample was concentrated into 500µl and buffer exchanged with NMR Buffer, with the addition of 10% D_2O, 0.1% sodium azide and protease inhibitors.

3.4.2. Data Acquisition

NMR data was acquired from Bruker DRX magnets of 600MHz and 800MHz provided by the Cross Faculty NMR Centre (Imperial College), with either a TXI cryoprobe or room temperature probe, at 310K. 1D ^{1}H NMR experiments were recorded using Hahn-Echo with water pre-saturation or Excitation Sculpting with gradient water suppression. Experiments TROSY-HSQC, HNCACB, HN(CO)CACB, HNCO and HNCA(CO) using a ^{1}H spectral width of 16.97ppm, a ^{15}N spectral width of 24.0ppm and ^{13}C spectral width of 66ppm.

Topspin software (Bruker Biospin Ltd) was used to process 1D NMR data. NMRPipe and a modified version of NMRView was used to process and analyse 2D and 3D NMR data.

3.5. X-ray Crystallography

3.5.1. Pre-Crystallisation Screens

A pre-crystallisation test using a PCT kit (Hampton Research) was used along with protein concentrations of 5mg/ml, 10mg/ml and 20mg/ml in a Crystallisation Buffer (100mM Tris-HCl pH8, 50mM NaCl) in order to select the optimal concentration for further crystallisation trials.

3.5.2. Crystallisation Screens

Purified protein was used in a crystallisation screen in order to generate crystals and to select conditions for further optimisation. The sitting-drop vapour diffusion method was used to grow crystals and purified Als1 at 10mg/ml concentration in Crystallisation Buffer was mixed 1:1 with reservoir solution in 100nl drops, using a Mosquito® Crystallisation Robot (TTP Labtech) with commercial crystallisation screen plates and reagents (listed in Table 3.7). Plates were incubated at either 4°C or 20°C. In-house X-ray diffraction was used to differentiate protein from salt crystals. Altering of pH, concentration of precipitants, droplet size and addition of Hepta-T peptide, D_2O (Sigma) and additives (Hampton Research) was carried out to optimise crystallisation conditions (details in Chapter 5). As crystals that were produced were grown in wells containing the cryoprotectant ethylene glycol, this was then used as the cryoprotectant of choice for X-ray diffraction.

Table 3.7: The plates used for the large scale sNT-Als1 crystallisation screen.

Plate ID	Screen Name	Company
ICL1	Crystal Screens 1 & 2	Hampton Research
ICL2	Wizard Classic 1 & 2	Rigaku Reagents
ICL3	PEG/Ion & Natrix	Hampton Research
ICL4	Index	Hampton Research
ICL5	SaltRx 1 & 2	Hampton Research
ICL6	MemStart & MemSys	Molecular Dimensions
ICL7	PACT premier	Molecular Dimensions
ICL8	JCSG+	Molecular Dimensions
ICL9	MemGold	Molecular Dimensions
ICL10	Wizard 3 & PEG/ion 2	Rigaku Reagents & Hampton Research
ICL11	JBS Cryo 1-4	Jena Biosciences
ICL12	Proplex	Molecular Dimensions
ICL13	Morpheus	Molecular Dimensions
ICL14	PGA Screen	Molecular Dimensions

3.6. Differential Scanning Fluorimetry

In wells of blue-white RT-PCR plates, 2µM of purified protein in DSF buffer (50mM $NaPO_4$ pH 8, 500mM NaCl) was combined with 2mM of peptide or drug-molecule ligand (apart from negative controls in which no ligand was added) and Sypro Orange dye in a 1:1000 dilution (Invitrogen) to a total volume of 20µl. Plates were sealed with optical foil, centrifuged for 30 seconds at 2000rpm in order to remove any bubbles, and then subjected to 1°C/min temperature increments in an qPCR thermocycler (Agilent Technologies, Stratagene Mx3005p) from 25-95°C. Data was analysed using DSF analysis and transform Excel worksheets (Niesen et al 2007).

4. Production of sNT-Als Constructs for Crystallographic and NMR Spectroscopic Studies.

4.1. Production of sNT-Als1 for Crystallography

4.1.1. Expression Systems and Constructs for NT-Als1

Previous structural studies were carried out with the full length NT-Als1 expressed in plasmid pET32-Xa-LIC and the *E. coli* Origami(DE3) strain. This construct was produced as a fusion with a thioredoxin and a histidine tag, a proteolytic site for Factor Xa and the NT-Als1 adhesin encompassing amino acid residues 1 to 314 (Yan et al 2010, Salgado et al 2011). NT-Als1 contains four disulphide bonds and the *E. coli* Origami(DE3) strain contains mutations in the genes encoding thioredoxin reductase (trxB) and glutathione reductase (gor) to facilitate cytoplasmic disulphide formation. This NT-Als1 construct contained a flexible C-terminus that is recognised as a ligand by the PBC (Salgado et al 2011). A construct with a shortened C-terminus, sNT-Als1 (amino acids 1-300) had been cloned in this laboratory to obtain a version of the protein suitable for ligand binding experiments, which lacked the flexible C-terminus.

I used this construct to successfully express and purify sNT-Als1 protein at the beginning of the project (Fig 4.1), however some difficulties with this construct were experienced, including reduced protein yield due to incomplete cleavage from thioredoxin, protein precipitation following cleavage, the financial expense of the Factor Xa enzyme and the time-consuming nature of the cleavage step itself. During this time, the SHuffle K12 *E. coli* strain (NEB Biolabs) was made commercially available. This strain expresses a form of thiol-disulphide interchange protein DsbC which lacks the signal sequence, so the SHuffle *E. coli* cells have improved capabilities for cytoplasmic expression if disulphide-containing proteins over the *E. coli* Origami strain. Tests in our laboratory indicated that fully folded NT-Als adhesins with correctly formed disulphide bonds were obtained from this strain without the need of a thioredoxin tag. This led to my successful re-cloning of sNT-Als1 into a pRSETa plasmid (Invitrogen) encoding sNT-Als1 with an N-terminal polyhistidine (6xHis) tag. To improve expression and purification of sNT-Als1, a modified version of pRSETa was used, in which the Xpress epitope and the enterokinase site were replaced by a thrombin proteolytic site.

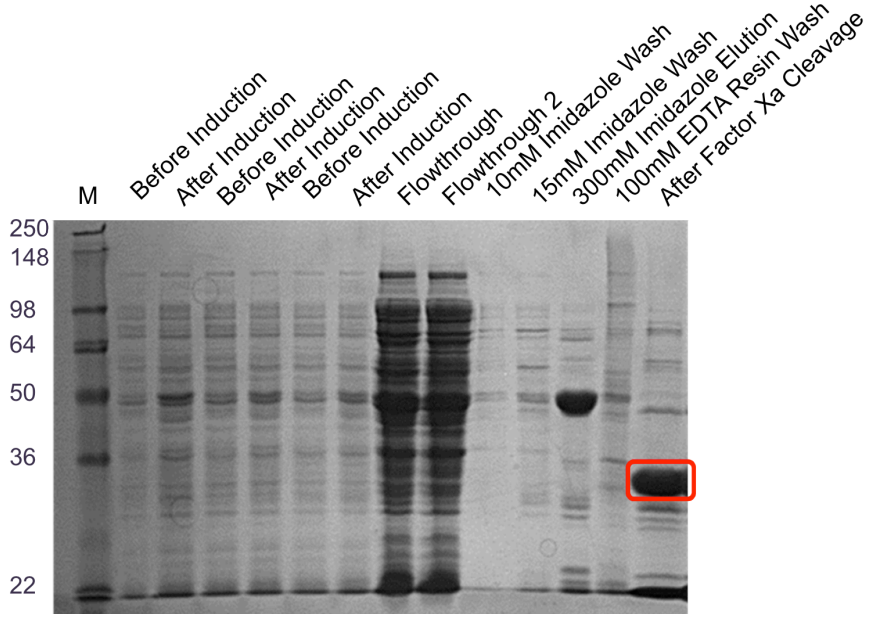


Figure 4.1: SDS-PAGE gel of samples from the induced expression and Nickel IMAC purification of the sNT-Als1-thioredoxin fusion protein, plus the Factor Xa cleavage of sNT-Als1 from thioredoxin.
The band encircled corresponds to the approximately 33kDa sNT-Als1 construct resulting from the cleavage step. The sNT-Als1-thrioredoxin fusion corresponds to the approximately 50kDa band in the labelled '300mM imidazole elution' lane.

As an additional feature of the Origami(DE3) strain, mutations in the trxB and gor genes are selectable on kanamycin and tetracycline, respectively. SHuffle K12 is also based on a trxB gor suppressor strain, but does not require these antibiotics to maintain that genotype. Possibly due to the fact that a single antibiotic is used to select the pRSETa plasmid, we observed that this led to a more stable expression system, with consistent protein yields per litre of culture. As a result, I adopted this expression system to produce sNT-Als1 in this study.

Approximately 2.5mg of sNT-Als1, was successfully purified from individual 850ml cultures, and approximately 3mg per litre from a 15L fermenter preparation. On reducing SDS-PAGE gels, bands at the expected size of 33kDa from induced cell samples indicated over-expression of a protein likely to be sNT-Als1. Bands were removed from a gel and sent for mass spectrometry analysis to confirm the identity of sNT-Als1 and that the molecular weight of the construct was 32.56kDa. This construct was purified successfully using both Nickel IMAC and size exclusion chromatography, and the gel filtration chromatogram showed a homogenous sample of high absorbance at 280nm, indicative of a high protein concentration (Fig 4.2). The pI of sNT-Als1 is 5.41.

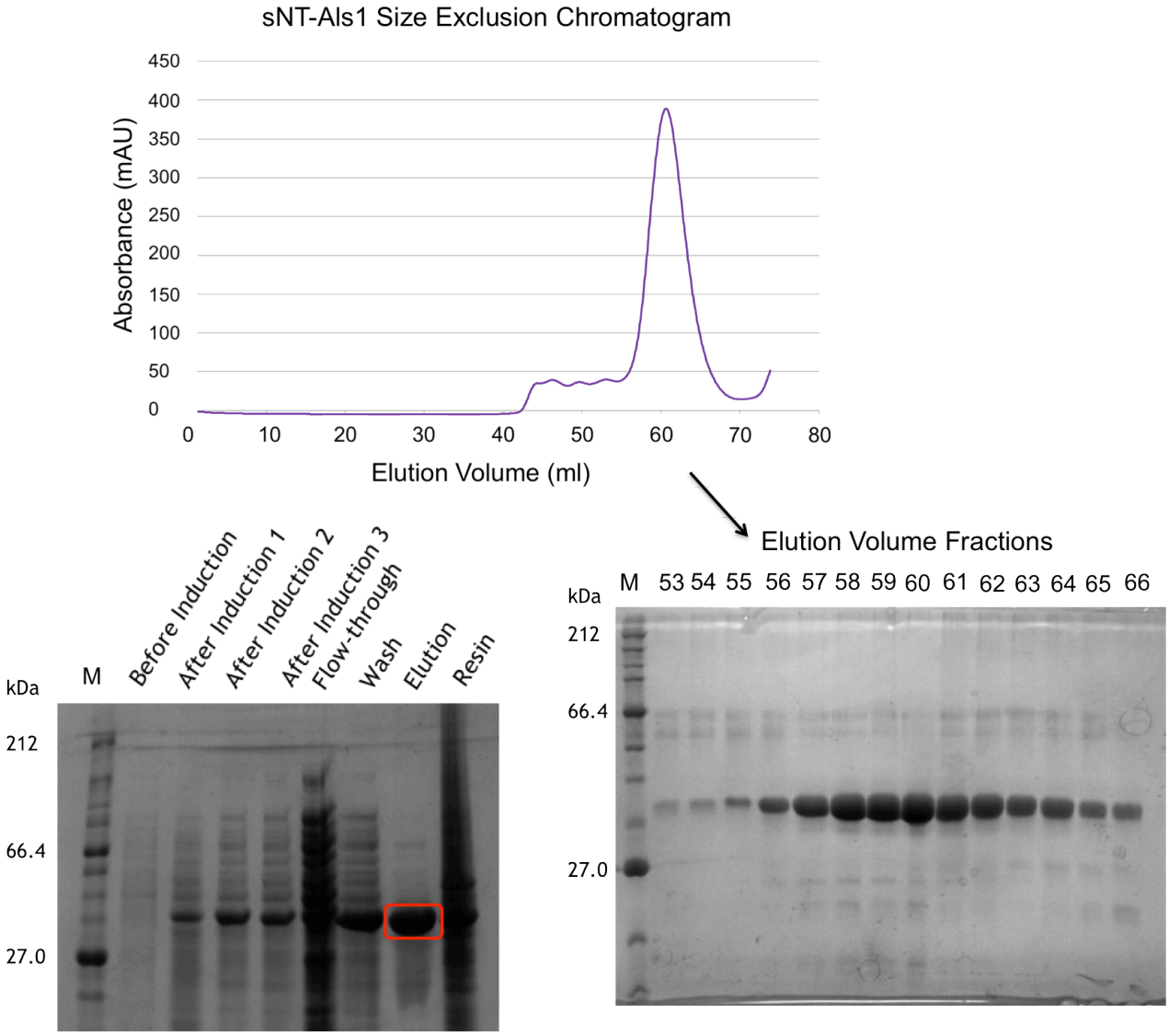


Figure 4.2: SDS-PAGE gels and a size exclusion chromatogram.
On the size exclusion chromatogram, a peak was seen with an absorbance at 280nm of more than 350mAu. This indicated a homogenous sample of high concentration and shows the successful expression and purification of the identified protein sNT-Als1. Gels below show samples from the initial Nickel affinity chromatography purification, elution band encircled around sNT-Als1 expected MW of 32kDa. The gel of elution volume fractions shows the result of size exclusion chromatography, with strong bands of expected size, with minimal impurities.

4.1.2. sNT-Als1 Sample Check with One-Dimensional ^{1}H NMR Spectroscopy

Recording ^{1}H and ^{1}H-^{15}N-HSQC NMR spectra is an efficient way to test whether a protein is folded properly, aggregated or disordered. An unlabelled sample of sNT-Als1 was checked with 1D ^{1}H-NMR spectra (Fig 4.3). The spectrum showed good peak dispersion, meaning that the ^{1}H nuclei of the protein construct were subjected to a variety of different chemical environments. In particular, the peaks in the region below 0.5ppm indicated that the protein contains ring current shifted methyl/methylene protons (i.e. below or above the plane of an aromatic side chain), which is typical of rigid, well-structured environments, such as hydrophobic cores, similar to previous construct NT-Als1 (Fig 4.4).

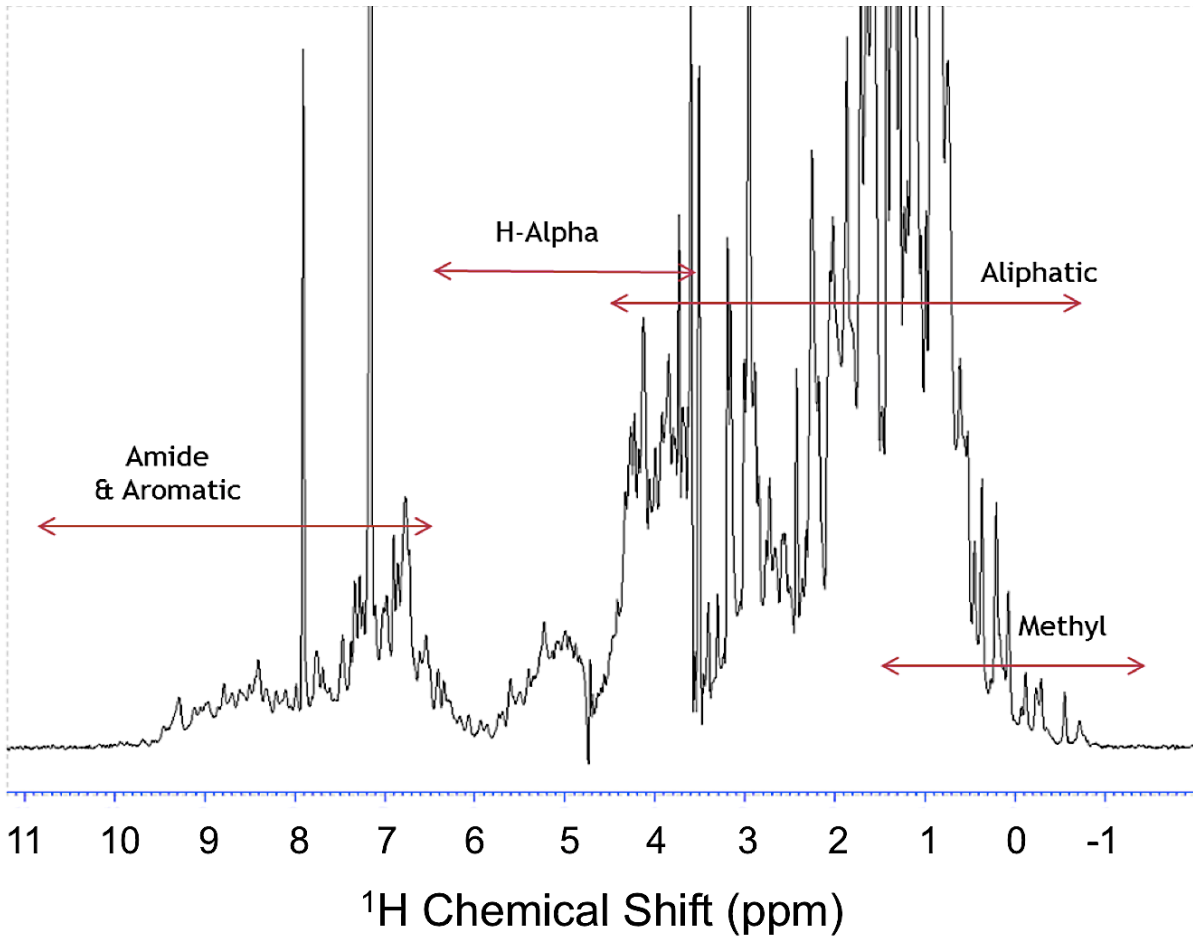


Figure 4.3: (1D) ^{1}H-NMR Spectrum of sNT-Als1.
The protein was in a buffer of 100mM Tris-HCL pH8, 50mM NaCl, with labels indicating the different environments of ^{1}H nuclei in the sample, including that of ^{1}H nuclei of methyl environments, below 0.5ppm. Peaks at 7.9 and 7.1ppm are characteristic of residual imidazole from earlier purification steps.

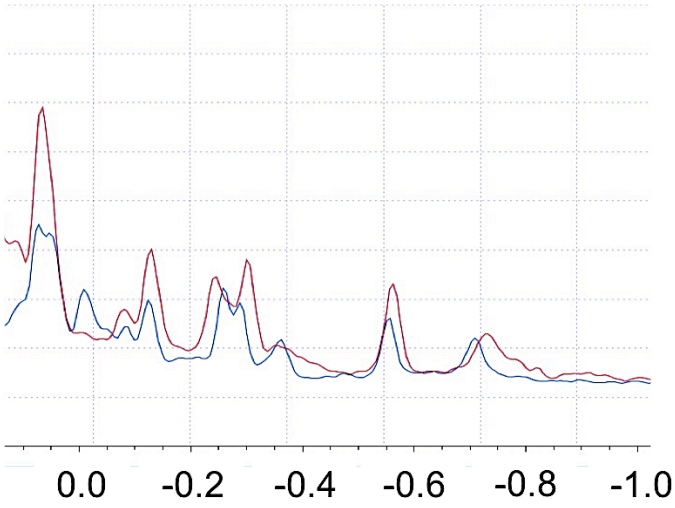


Figure 4.4: 1D ^{1}H spectral comparisons of ring-current shifted methyl peaks between NT-Als1 and sNT-Als1.
NT-Als1 spectrum is shown as coloured blue (Yan et al 2010) and the sNT-Als1 spectrum is coloured red. The peaks in the 0.5-0.0ppm regions correspond to ^{1}H chemical shifts found in methyl group environments whilst also affected by local magnetic field changes from to aromatic ring groups within the protein. These peaks in NT-Als1 and sNT-Als1 are similar, indicating that the hydrophobic cores of the two constructs are both stable and alike.

The same sNT-Als1 construct was labelled with ^{15}N, in order to get a ^{1}H-^{15}N-HSQC-TROSY spectrum (Fig 4.5). This confirmed good dispersion of N-H groups throughout the protein, with minimal overlap of signals, indicating that sNT-Als1 was stable and folded. This prompted me to set a sparse-matrix crystallisation screen.

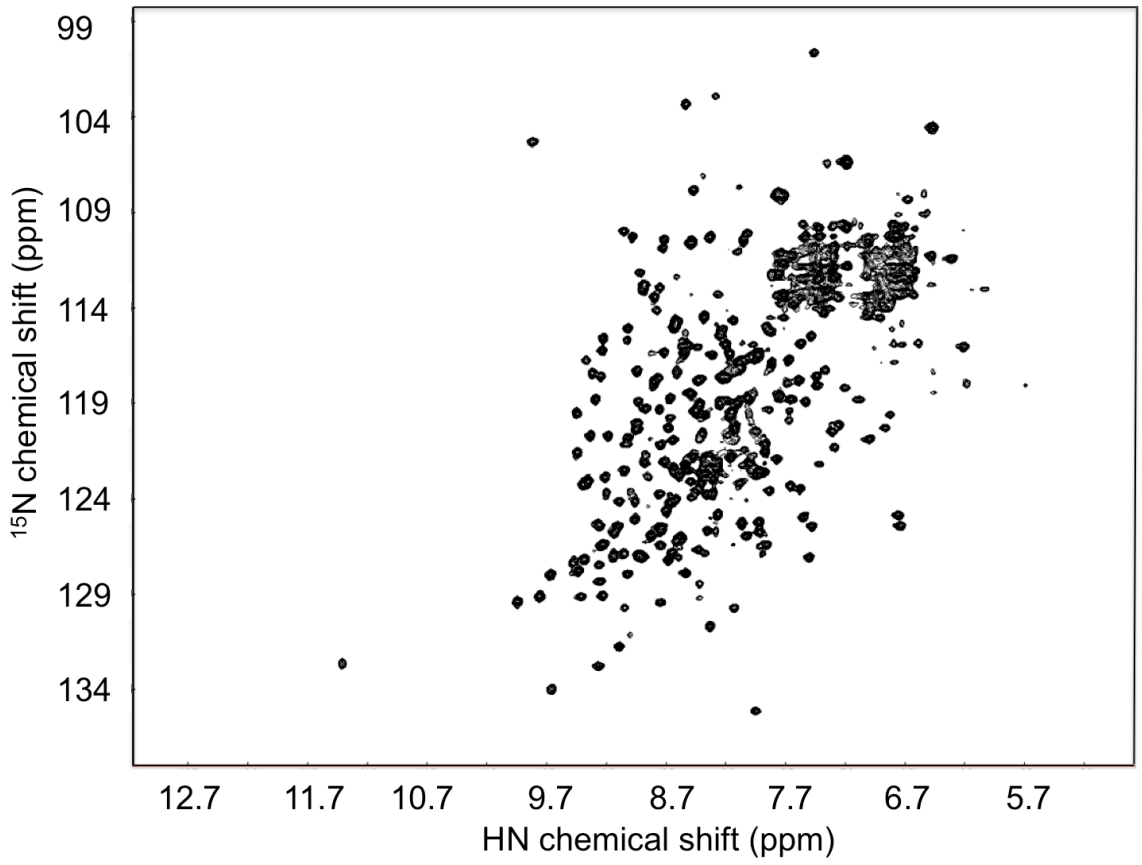


Figure 4.5: ^{1}H-^{15}N-HSQC-TROSY spectrum of sNT-Als1.
This shows well-resolved peaks (corresponding to N-H groups within the protein), with minimal overlap. This is indicative of dispersed N-H groups with minimal aggregation.

4.2. sNT-Als1 Crystallography

4.2.1. Initial Crystallisation Screens of sNT-Als1

Pre-crystallisation trials with a PCT Kit (Hampton Research) were used to decipher the ideal sNT-Als1 concentration and incubation temperature for a sparse-matrix crystallisation screen. Light precipitation was obtained with sNT-Als1 protein solution at a concentration of 10mg/ml at 20°C incubation temperature in the buffer 100mM Tris-HCl, 50mM NaCl. A large-scale trial was set up in these conditions, using 12 different pre-made plates utilising the sitting-drop method (Hampton Research). However, this screen produced only precipitation in the wells.

4.2.2. Co-Crystallisation of sNT-Als1 with Hepta-Threonine Ligand

NT-Als constructs were shown to interact with a Hepta-Threonine peptide N-TTTTTTT-C (Hepta-T), as shown by ^{1}H-^{15}N HSQC TROSY titration experiments for NT-Als1 (Yan et al 2010, Salgado et al 2011) and co-crystallisation and ITC studies with sNT-Als3 (Lin et al 2014). Using differential scanning fluorimetry (DSF), the potential for the Hepta-T peptide to interact with and stabilise sNT-Als1, and promote crystallization, was indicated. This experiment showed that the average ΔT_m shift of a sNT-Als1 melting curve after the addition of Hepta-T peptide was approximately 7°C ± SD 0.326, indicating a high probability of Hepta-T binding and sNT-Als1 construct stabilization (Fig 4.6).

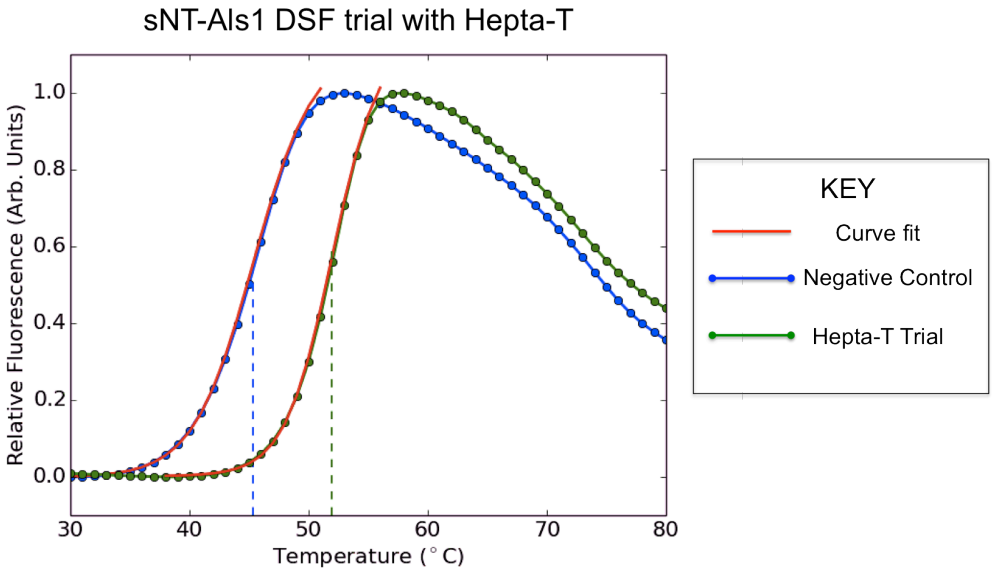


Figure 4.6: A comparison of the sNT-Als1 negative control melting curve with the melting curve of the Hepta-T peptide trial.
Shown here is the plotted relative fluorescence data for one of three sNT-Als1 negative control replicates (in blue) alongside the plotted relative fluorescence data for one sNT-Als1 plus Hepta-T peptide trial (in green). In order to derive the T_m values for each trial, a curve has been fitted to the data using the equation y = Imin + (Imax – Imin)/1+ exp (T_m – x/a). The dotted lines show how midpoint has been calculated, and been used to derive the T_m values. For the negative control, a T_m of 45.519°C ± 0.039 has been calculated, and for the Hepta-T peptide trial, a T_m of 52.891°C ± 0.065 has been derived.

Based on this result, a crystallisation screen was carried out with 10mg/ml sNT-Als1 in a 20:1 ratio with 10mM Hepta-T peptide. Again, precipitation was the main result, but spheroid crystals were obtained. Although these crystals were unsuccessful in producing X-ray diffraction patterns, the result suggested that Hepta-T aided crystallisation, and that optimisation of crystallisation conditions was necessary.

4.2.3. The Addition of Deuterium Oxide to the Co-Crystallisation of sNT-Als1 with Hepta-Threonine Peptide

It had been demonstrated by Dr Bing Liu in this laboratory that using a buffer containing deuterium oxide (D_2O) can facilitate crystallisation of proteins (Liu et al 2012). Dr Jing Lin also used this method for solving the structure of sNT-Als3 (Lin et al 2014). Another crystallisation screen was carried out with sNT-Als1, again with a 20:1 ratio with 10mM Hepta-T peptide, but these had been dissolved in a buffer containing 100% D_2O. From this screen, small, thin spiky crystals were seen in the F11 well of plate ICL11 (JBS Cryo Screens 1-4, Jena Biosciences) containing 25% w/v PEG 8000, 30% v/v ethylene glycol, 150mM ADA pH6.0 and 80mM ammonium sulphate (Fig 4.7).

These crystals were able to diffract, and gave a diffraction pattern that was characteristic of a protein. However, data collected could not be processed due to crystal twinning. Their small size and non-homogenous formation could be a disadvantage to data collection, so optimisation of the conditions was carried out in order to try to get larger, more suitably shaped crystals.

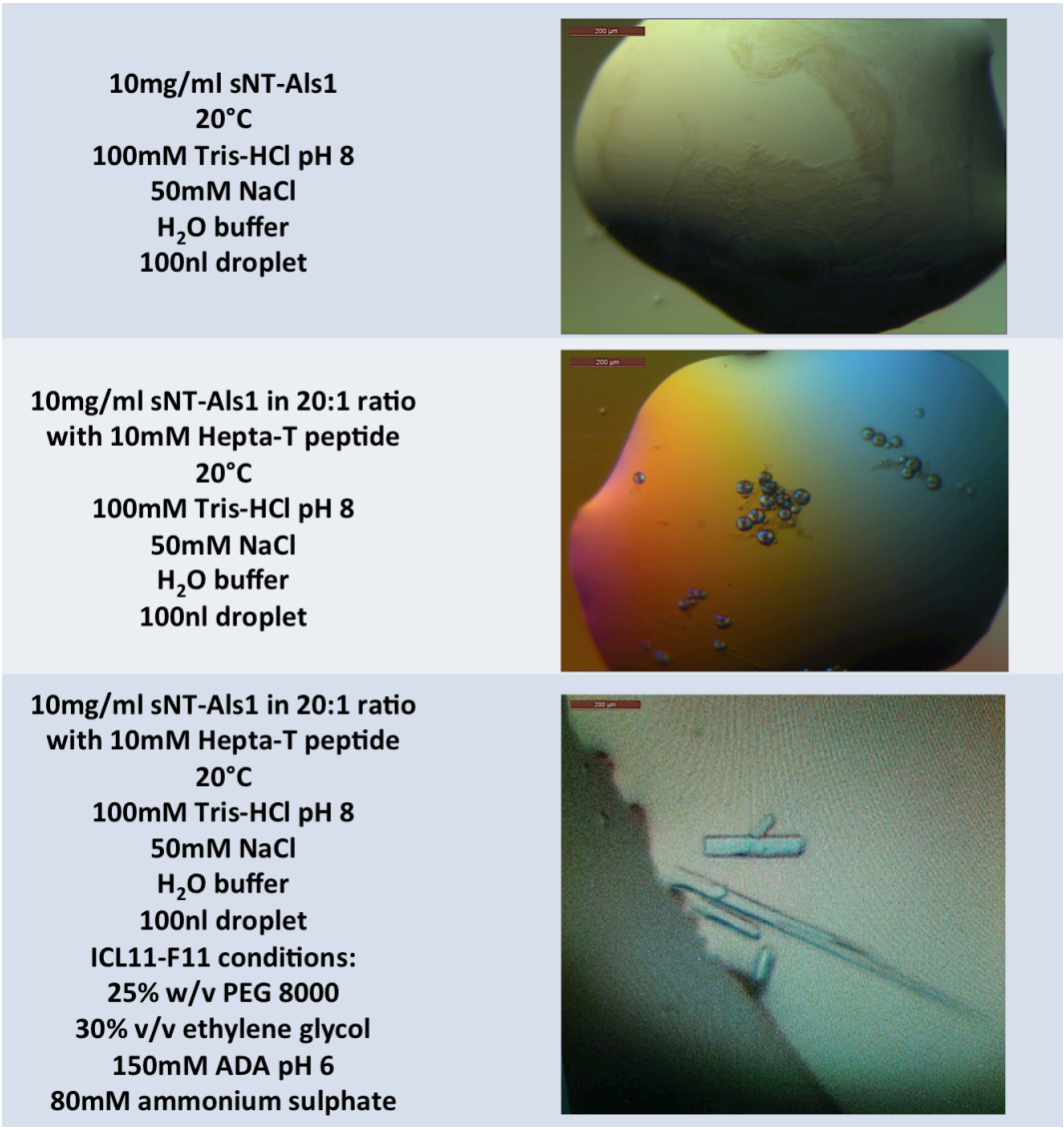


Figure 4.7: **Crystallisation plates of sNT-Als1 protein or in complex with Hepta-T peptide.**
Crystallisation plates including concentrated sNT-Als1 without ligand produced only precipitation. After the addition of Hepta-T ligand, this produced precipitation or crystals unsuitable for diffraction (e.g. the spheroid crystals shown in the figure). However, in the presence of D_2O buffer, diffractable protein crystals were produced, albeit with more optimization required.

4.2.4. sNT-Als1 Crystal Optimisation

The ICL11 plate was used again for crystallisation trials, but was used to test whether droplet size (100nl or 200nl), temperature (4°C or 20°C) and protein concentration (5, 10 or 20mg/ml) had an effect on the quality of crystals seen. A higher concentration of sNT-Als1 at 20mg/ml, regardless of temperature or well size, resulted in crystals that were smaller than 200µm and not suitable for diffraction. A lower concentration at 5mg/ml yielded minimal and disordered crystals. Optimal conditions were obtained at 10mg/ml protein at 4°C, with a well size of 200nl. This condition resulted in larger crystals (approximately 450µm) that were ordered and could be easily separated during the crystal fishing process (Fig 4.8). A lower temperature may have resulted in slower, more ordered crystal growth from nucleation.

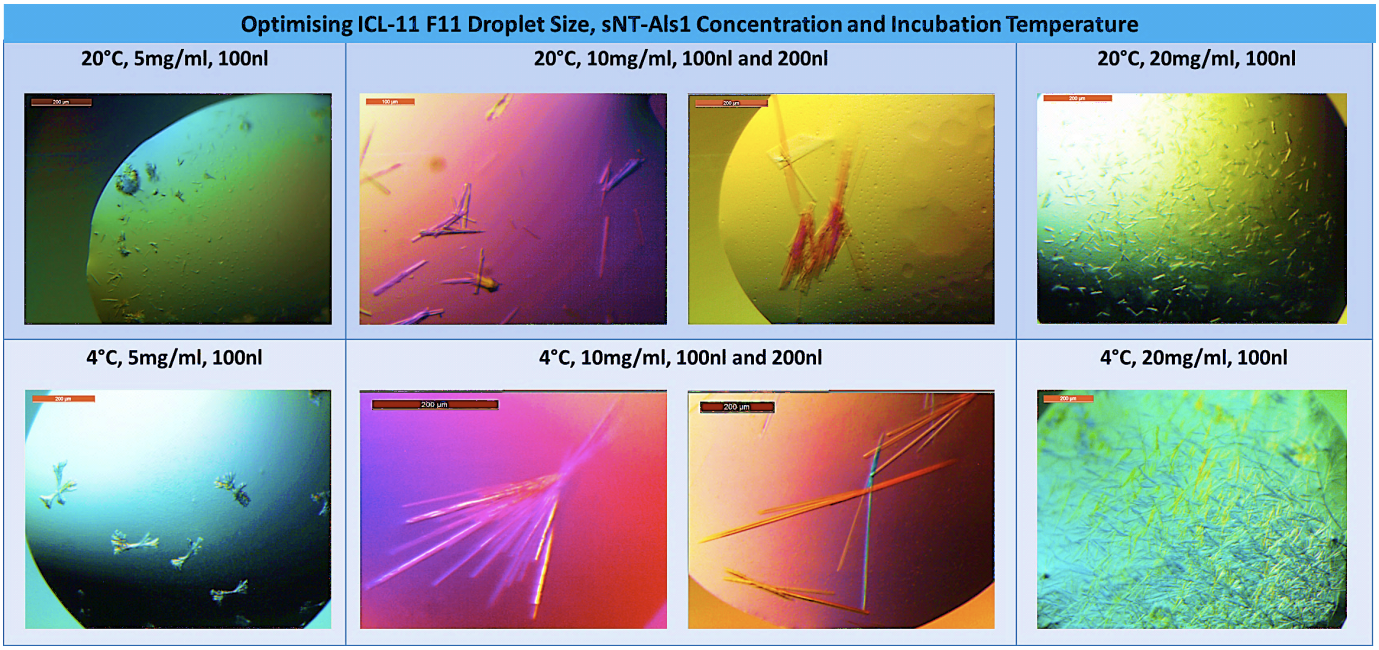


Figure 4.8: **Comparison of crystals from optimising crystallisation conditions in ICL11-F11**.
The droplet size (100nl vs. 200nl), concentration of sNT-Als1 protein (5mg/ml vs. 10mg/ml vs. 20mg/ml) and incubation temperature (4°C vs. 20°C) were optimized. Finally, an incubation temperature of 4°C, protein concentration of 10mg/ml and well size of 200nl gave larger, more ordered spike-shaped crystals.

Other optimisation plates were set up with an aim to improve the crystals produced by modifying the concentration of the PEG 8000 precipitant, pH of the ADA buffer, the concentration of ammonium sulphate precipitant salt and also increasing the droplet size to 2µl. A protein concentration of 10mg/ml and incubation temperature 4°C were used based on the results of the first optimisation test. After this optimisation step, crystals were obtained in 22% or 24% w/v PEG 8000, 30% ethylene glycol, 150mM ADA at pH 6.4 with an 80mM ammonium sulphate concentration. However, these crystals still grew as thin spikes (Fig 4.9).

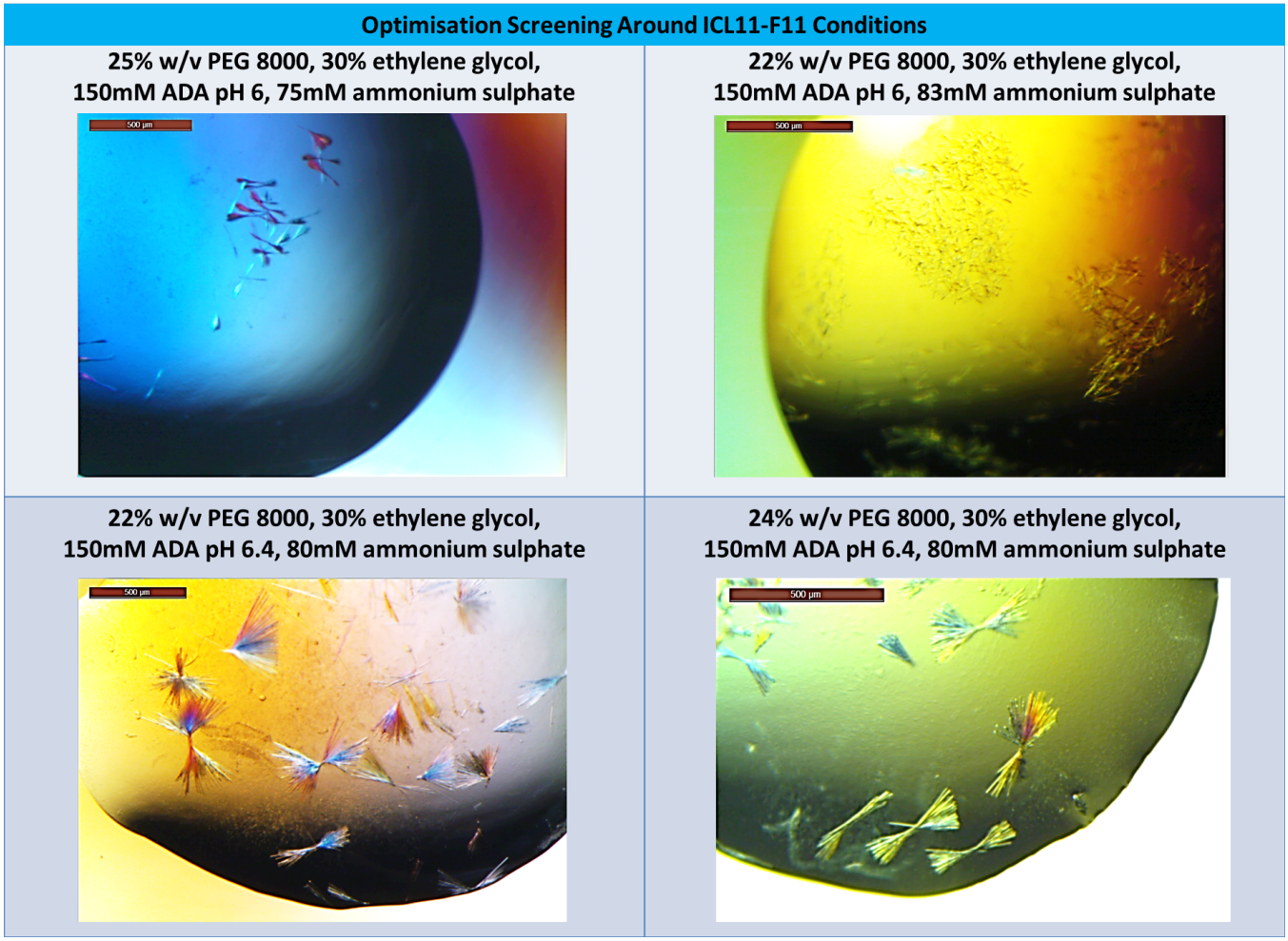


Figure 4.9: Comparison of the crystals produced from optimisation plate screens around ICL11-F11 conditions.
Optimisation of the PEG 8000 concentration, the pH of the ADA buffer and concentration of the ammonium sulphate salt were carried out. A larger droplet size of 2µl was also used, plus 4°C incubation temperature.

An additive screen (Hampton Research) was subsequently carried out, plus experiments testing the hanging drop crystallisation method instead of sitting drop – however these were unsuccessful at producing crystals. Crystals were taken from the best optimisation conditions, which included crystals from well F11 (4°C, 10mg/ml sNT-Als1 protein, 200nl droplet size) and from optimisation plate wells containing 30% ethylene glycol, 150mM ADA pH 6.4 and 80mM ammonium sulphate with either 22% or 24% PEG 8000.

4.3. sNT-Als1 Expression and Purification for Backbone Assignment and HSQC-Titrations using NMR Spectroscopy

4.3.1. ^{15}N-Labelled and Triple-Labelled sNT-Als1 Expression and Purification

Previously, Dr Robert Yan had carried out a backbone assignment of NT-Als1 1-314, to 95% completeness (Yan et al 2010). This construct, however, was unsuitable for ligand binding experiments because of self-complementation, as described above in section 1.7. In order to probe ligand-binding specificity, I aimed to assign the backbone resonances of sNT-Als1 and employ that information as a reference in chemical shift perturbation analyses using ^{1}H-^{15}N-TROSY-HSQC spectra. For this, sNT-Als1 (amino acids 1-300), was expressed and purified successfully from both ^{15}N-labelled minimal media and ^{15}N-^{13}C-^{2}H-labelled media. Small scale trials in 50ml cultures were used to check capability of expression and purification of sNT-Als1 from ^{15}N-^{13}C-^{2}H-labelled media before proceeding to use expensive litre-quantities of the same media, from which samples were analysed on SDS-PAGE gels (Fig 4.10).

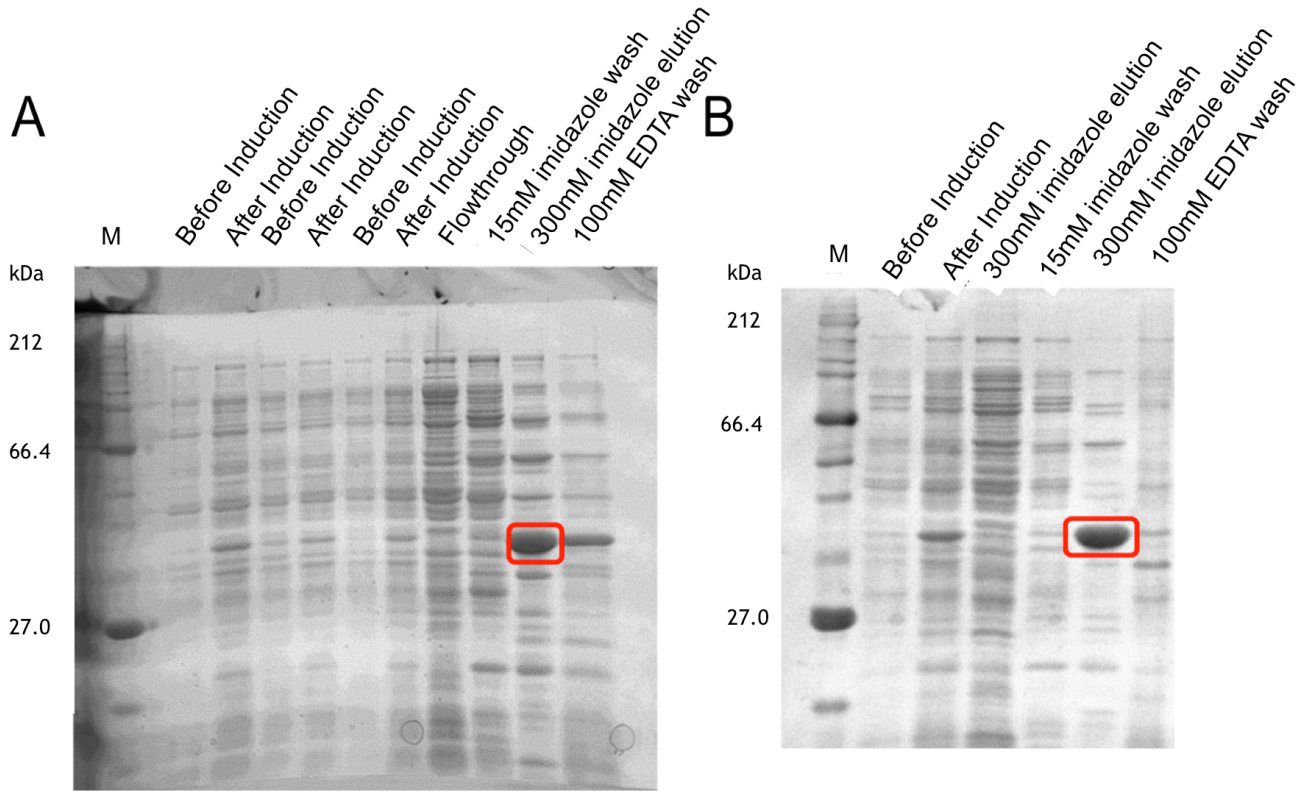


Figure 4.10: SDS-PAGE gels of labelled sNT-Als1 expression and purification.
A) SDS-PAGE gels of the expression and purification of sNT-Als1 from ^{15}N-minimal media. The band encircled red in the '300mM imidazole elution' lane corresponds to the expected sNT-Als1 MW of approximately 32kDa. B) Initial test of sNT-Als1 expression and purification from a small-scale D_2O-minimal media culture expression and purification trial. The band encircled red in the '300mM imidazole elution' lane again corresponds to the expected sNT-Als1 MW of approximately 32kDa.

Due to the expense of a triple labelled sNT-Als1 sample, plus the reduced yields compared to those produced from either LB or ^{15}N-minimal media (2.5mg per 850ml from LB media, 1.25mg per 850ml from ^{15}N-minimal media and 0.5-1mg per 850ml from ^{15}N-^{13}C-^{2}H-minimal media), and the fact that protein yield decreases with every purification step, and that a sample of high purity could be purified using Nickel-IMAC, size exclusion chromatography was not used in the production of triple labelled sNT-Als1. However, optimisation of the Nickel-IMAC purification process was carried out to maximise sample purity.

The final Nickel-IMAC purification included multiple high-volume buffer washes before the final elution; including a 20 column volumes of 15mM imidazole wash buffer followed by 20 column volumes of 20mM imidazole wash buffer, plus 5 column volumes of 50mM imidazole was buffer before the final 3 column volumes of 300mM imidazole elution buffer (plus a 2 column volumes amount of 100mM EDTA resin wash to recover any un-eluted sample). All the elution samples were pooled, dialyzed and concentrated to 2ml in 200mM Tris-HCl pH8.0, 300mM NaCl. To increase the rate of deuteron-to-proton exchange of amide groups, this sample was bought to 4M of urea and then dialyzed against the same buffer without urea. An extra overnight dialysis step to exchange the buffer to a low salt condition (50mM NaCl) was also performed. A sample was taken after these dialysis steps to check the integrity of the protein by SDS-PAGE gels before proceeding to NMR spectroscopic experiments (Fig 4.11). Some precipitation was seen after transfer to a low salt buffer, however the final concentration of sNT-Als1 was approximately 0.6-0.7mM.

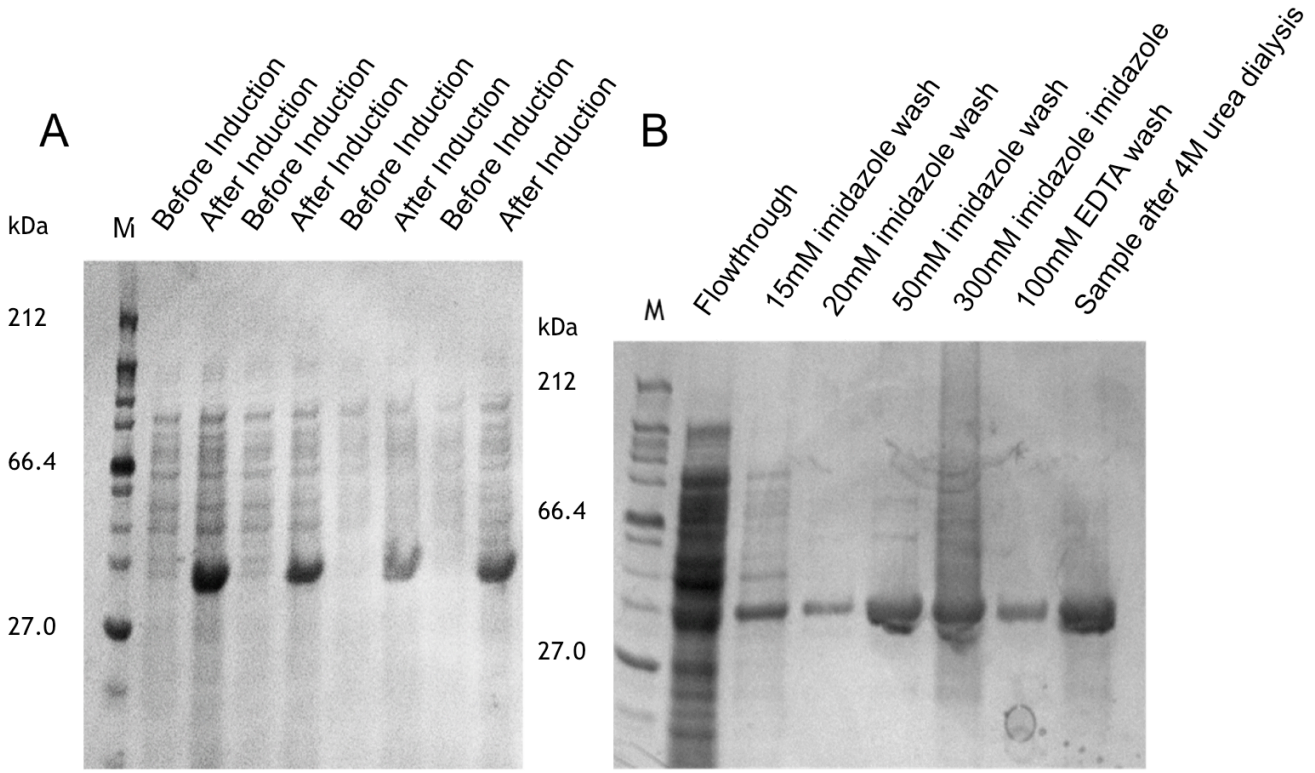


Figure 4.11: **SDS-PAGE Gels of triple labelled sNT-Als1 expression and purification**.
A) ^{15}N-^{13}C-^{2}H-labelled sNT-Als1 induced expression and B) purification from ^{15}N-^{13}C-^{2}H labelled media. Samples were taken from each wash condition of the optimised Nickel-IMAC purification. This was in order to conserve as much protein as possible, via avoiding a size-exclusion chromatography purification step. Washes using 20 column volumes of 15mM imidazole wash buffer, 20 column volumes of 20mM imidazole wash buffer and 5 column volumes of 50mM imidazole wash buffer were employed to produce a sample as pure as possible.

4.3.2. Optimisation of Sample Conditions

The same buffer conditions for the crystallography of sNT-Als1 were not optimal for NMR studies of sNT-Als1. A buffer kept at pH8 using 100mM Tris-HCl was used for the purification of sNT-Als1, as this construct had a pI of approximately 5.41 (ProtParam), with this buffer being optimal for purifying the protein for the purposes of crystallisation. However, the buffer molecule Tris has a very sharp peak at approximately 4ppm which results in ADC overflow problems with regards to the NMR spectrometer. The exchange of the buffer from 100mM Tris-HCl pH 8 to 50mM Sodium phosphate pH8 was monitored using 1D ^{1}H-NMR spectra in order to check if any changes to stability or solubility had occurred, of which any were negligible. Plus, the NaCl salt concentration decided to be reduced

to 50mM from 100mM as lower salt concentrations resulted in increased sensitivity and stronger peaks recorded from the NMR spectrometer (Fig 4.12).

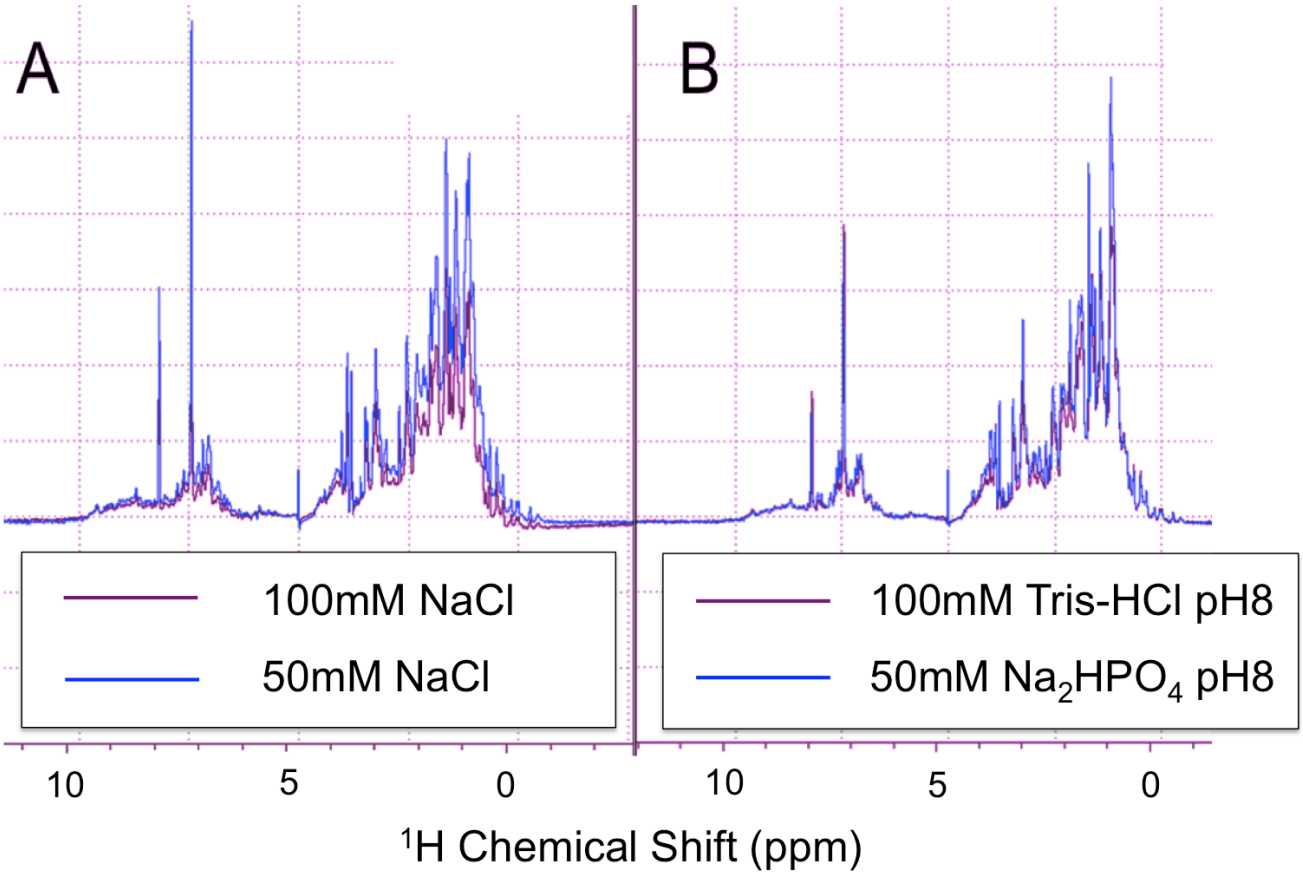


Figure 4.12. Comparisons of 1D ^{1}H-NMR spectra for sample buffer optimization.
A) Comparison of 1D ^{1}H-NMR spectra of sNT-Als1 with a Tris-HCl buffer containing 100mM NaCl (violet) vs. 50mM NaCl (blue). Albeit with the same protein concentration, the peaks recorded were stronger with a buffer containing 50mM NaCl. B) Comparison of 1D ^{1}H-NMR spectra of sNT-Als1 with a buffer of 100mM Tris-HCl pH8.0 (violet) vs. 50mM Sodium Phosphate pH 8.0 (blue), both with 50mM NaCl concentration. As the spectra are almost identical, although slightly sharper peaks can be seen in the sample in 50mM sodium phosphate buffer, this was decided as the superior sample buffer.

Samples of unlabelled sNT-Als1 were buffered to different pH's and incubated at 35°C for four days to identify conditions that are optimal to record 3D NMR spectra. 1D spectra of the sample were recorded before and after incubation (Fig 4.13). At pH8.0, a reduction in signal to noise was observed, especially in the amide region, due to increased rate of exchange of protons with solvent water, whereas the difference in the intensity of signals between pH 6.0 and 7.0 was negligible overall. However, a higher amount of precipitation was seen for the protein sample at pH 6.0 at the later stages of incubation, as the conditions approached the predicted pI of 5.41 for this construct. Therefore, a buffer containing 50mM sodium phosphate pH7 with 50mM NaCl was chosen to carry out the backbone assignments. This buffer is different from the buffer of 50mM sodium phosphate pH6, 50mM NaCl, that was used by Dr Robert Yan (Yan et al 2010), but the difference in pH is useful to avoid precipitation after titration of sNT-Als1 with a number of small acidic ligands (i.e. containing carboxylates such as e.g. dinitrosalicylic acid), especially in cases where TROSY-HSQC data had to be recorded over long periods of time at 35°C.

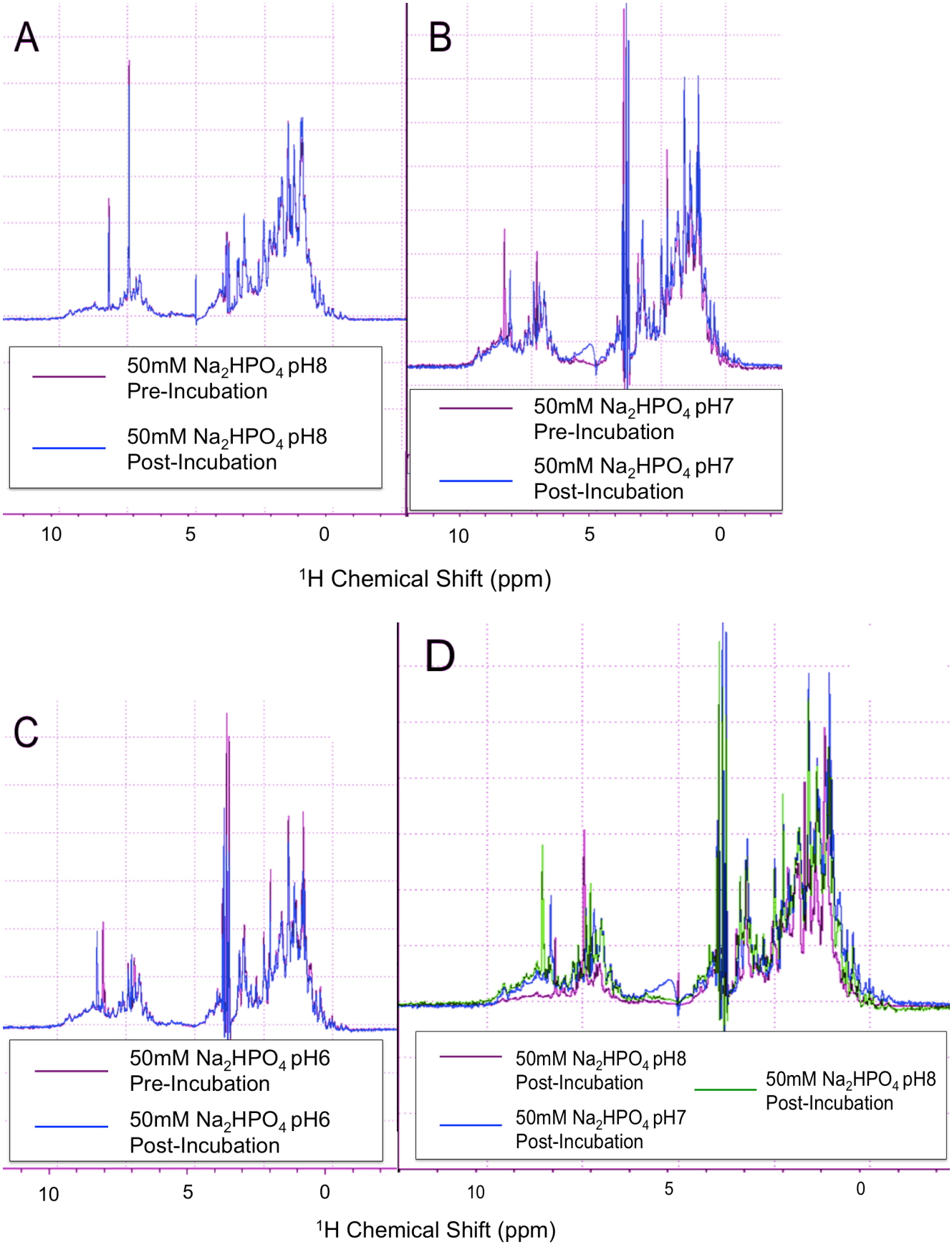


Figure 4.13: **Comparison of 1D 1H-NMR spectra of sNT-Als1 for pH optimization**.
A) Comparison of 1D ^{1}H-NMR spectra of a 250µM sample of sNT-Als1 buffered to pH8.0 using a 50mM sodium phosphate buffer, B) and buffered to pH7.0 in the same way C) plus pH6.0 also buffered using 50mM sodium phosphate buffer, both before (purple) and after (blue) 35°C incubation in a water bath for four days. D) Shows a final comparison of post-incubation 1D ^{1}H-NMR spectra of sNT-Als1 for pH8.0 (purple), pH7.0 (blue) and pH6.0 (green).

A sample of sNT-Als1 in a buffer of 50mM sodium phosphate pH7.0 and 50mM NaCl was incubated with 4M urea for 24 hours at room temperature, with 1D ^{1}H-NMR spectra being taken before and after incubation. The 1D NMR spectra comparisons showed that the sNT-Als1 construct was intact after this treatment (Fig 4.14), showing that the same treatment can be used in triple labelled samples suitable for 3D NMR experiments to obtain the backbone resonance assignments.

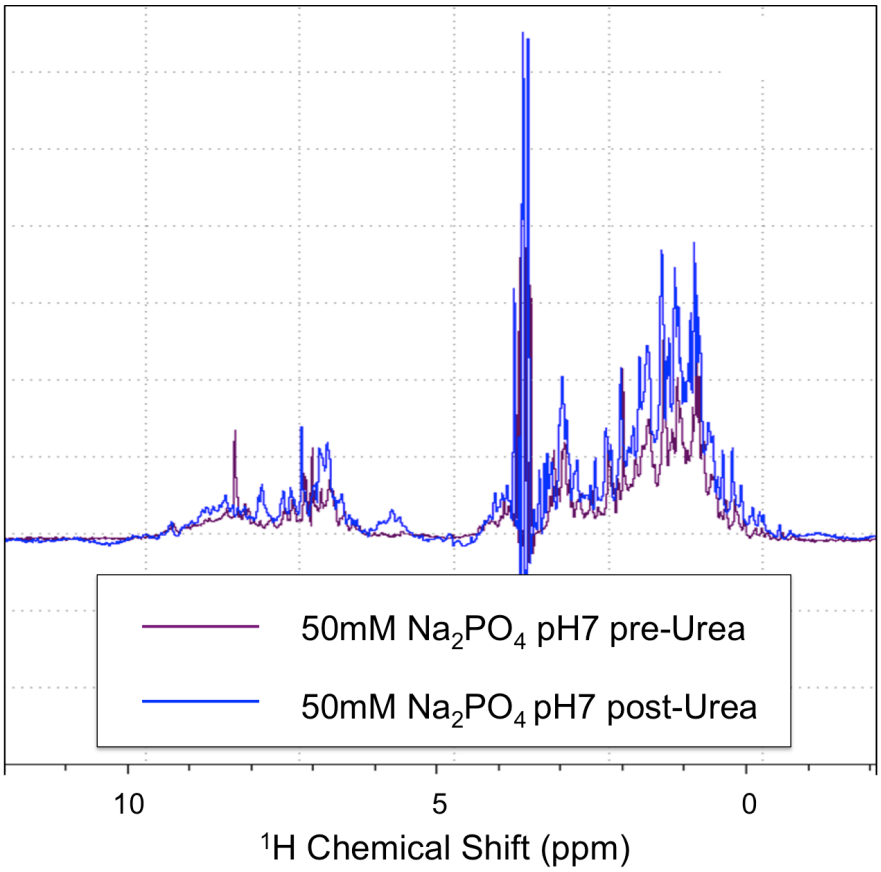


Figure 4.14: **Comparison of 1D ^{1}H-NMR spectra of sNT-Als1 before and after incubation with 4M urea**. The sNT-Als1 spectrum prior to incubation is shown as purple, and after incubation is shown as blue. Good sample dispersion and peak intensity before and after urea treatment showed that the construct remained intact, and that this step could be used in the final triple labelled sample for backbone resonance assignment.

Previous experiments showed that 4M urea is sufficient to induce full exchange of backbone amides (Yan et al 2010). Despite this, incomplete exchange of protons was observed for sNT-Als1, as several peaks were missing on recorded TROSY-HSQC spectra. Originally the sNT-Als1 was incubated in both 3M and 4M urea for shorter hours (4-5 hours) at room temperature, but an overnight incubation with 4M urea (still at room temperature) with fast dialysis improved the intensity of the peaks that could be observed, although did not result in the complete number of peaks expected for sNT-Als1.

4.4. sNT-Als3 Expression and Purification for Backbone Assignment Studies with NMR Spectroscopy

4.4.1. ^{15}N-Labelled sNT-Als3 Expression and Purification

In parallel to this work, structure determination of sNT-Als3 by X-ray crystallography was carried out in this laboratory (Lin et al 2014). To complement these studies, I also attempted expression and purification of triple labelled sNT-Als3 (amino acids 1-302) in order to record 3D NMR spectra for backbone resonance assignments. Again, to avoid the time-consuming and expensive process of cleaving sNT-Als3 from its thioredoxin tag from expression in the pET36-Xa-LIC plasmid, sNT-Als3 was cloned into a pET46 for expression with an N-terminal his-tag. However, purity of the sample after Nickel-IMAC purification using wash buffers (as detailed in the Materials and Methods section 3.3.2) was not homogeneous, and significant precipitation/aggregation was observed after concentrating the sample for size exclusion chromatography. Even whilst attempting to optimise purification using wash buffers containing increasing concentrations of imidazole (5mM, 10mM, 15mM, 20mM, 30mM, 50mM imidazole, 500mM NaCl, 50mM sodium phosphate pH 8.0, 10 column volumes each), this aggregation and precipitation was the same result (Fig 4.15).

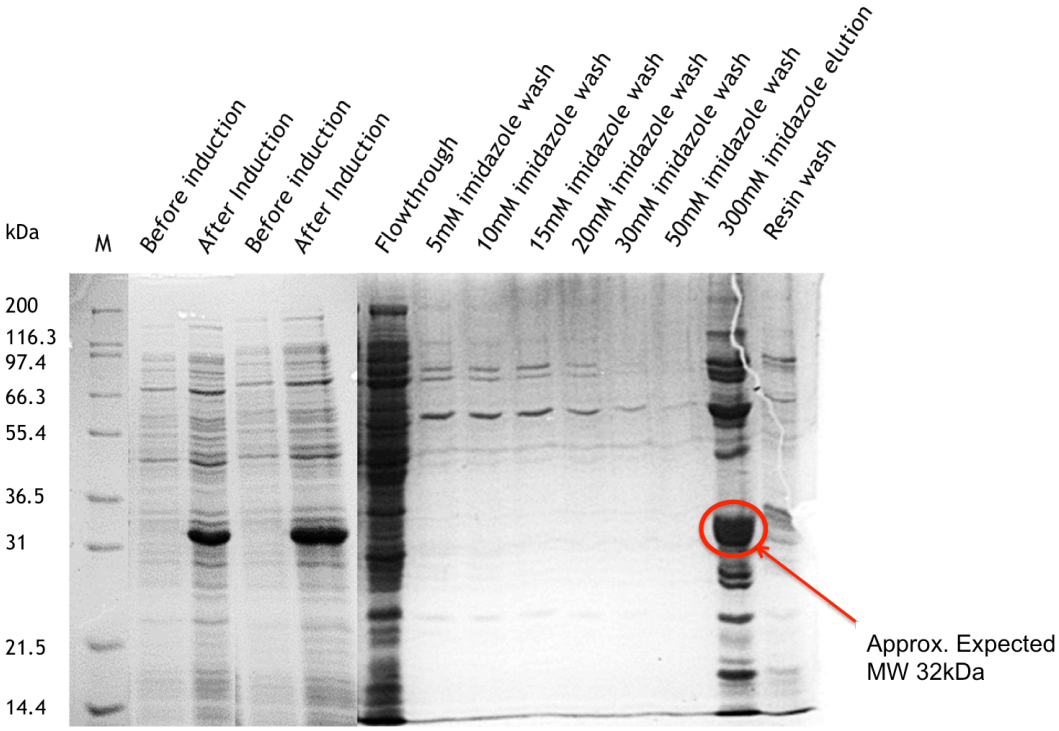


Figure 4.15: SDS-PAGE Gel of sNT-Als3 expression and purification using a pET46 plasmid.
The construct was expressed using Shuffle cells and Nickel-IMAC purification with washes containing increasing concentrations of imidazole from 10mM-300mM. The band encircled in the 300mM imidazole elution lane corresponds to that of the expected MW of sNT-Als3 of approximately 32kDa. However, in the same lane, contaminants and/or aggregates can be seen.

Due to these issues, the sNT-Als3 construct expressed in pET32-Xa-LIC was used for subsequent studies. This could be expressed and purified from SHuffle cells in LB and ^{15}N-labelled minimal media with much less aggregation but with some contamination (Fig 4.16). Gel filtration after cleavage with Factor Xa was used to remove the cleaved thioredoxin and most of the contamination – albeit some contamination could be seen afterwards. An ^{15}N-labelled sample of sNT-Als3 was used to collect a 1D ^{1}H-spectrum and a ^{1}H-^{15}N TROSY-HSQC-spectra (Fig 4.17 and 4.18). From these results, the sNT-Als3 construct was thought to be suitable for expression in triple labelled media.

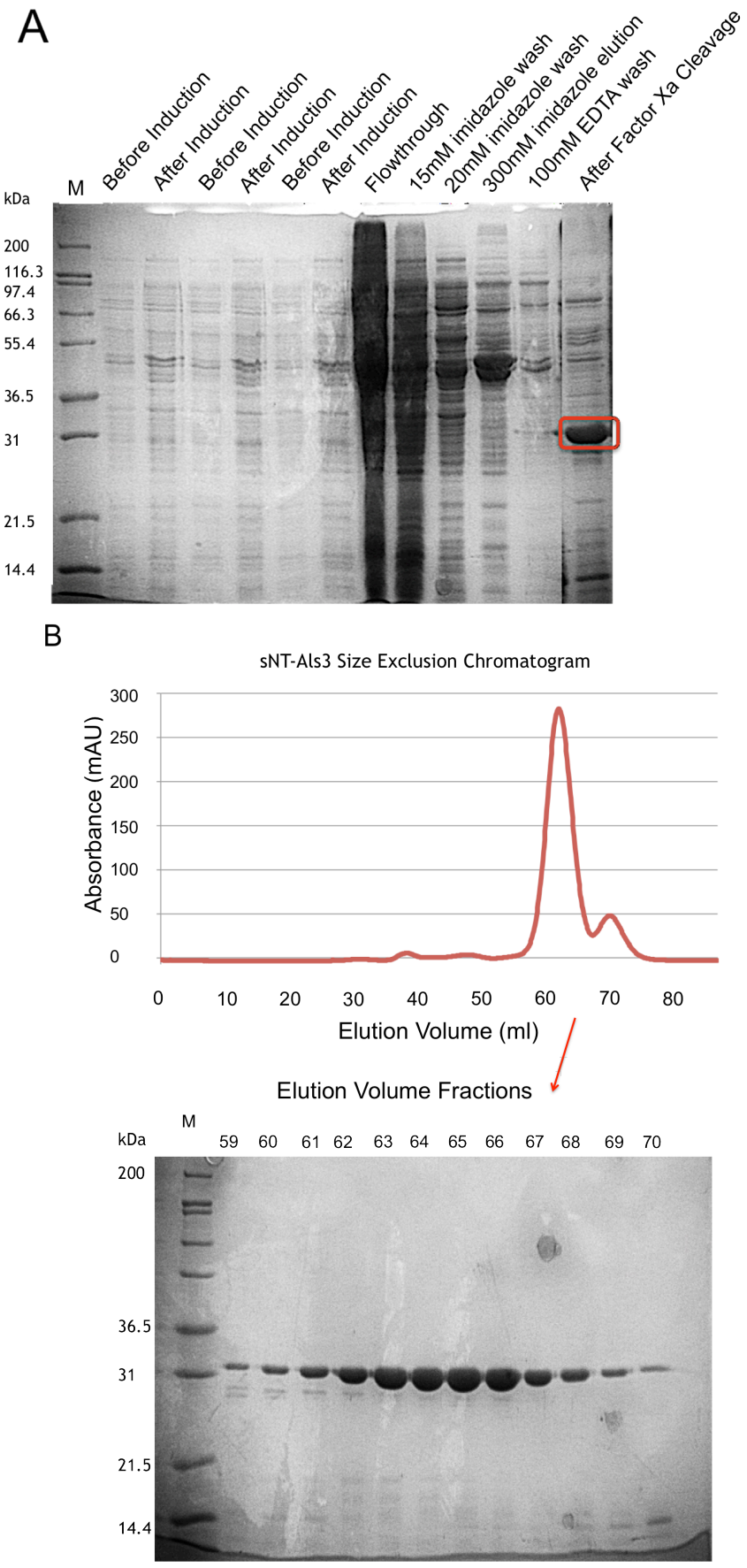


Figure 4.16: **SDS-PAGE gels for expression and purification of sNT-Als3-TRX using pET32 plasmid**.
A) SDS-PAGE Gels of sNT-Als3-TRX expression, purification and cleavage with Factor Xa (sNT-Als1 construct, of the expected MW of approximately 32kDa is encircled red) and B) Gel filtration by Size Exclusion Chromatography.

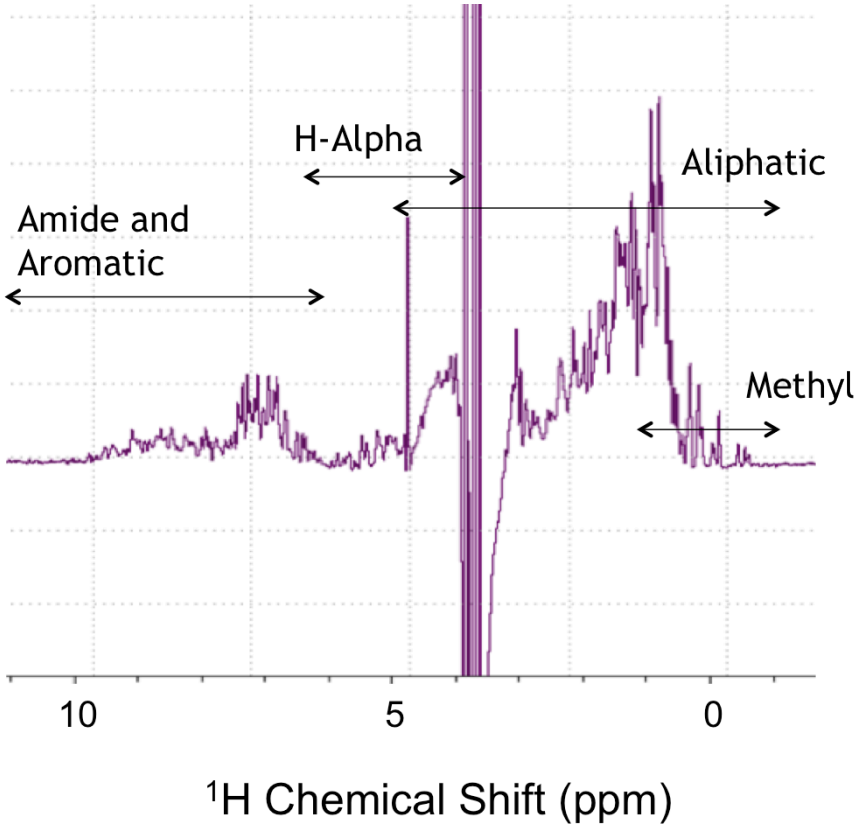


Figure 4.17: **Checking the sNT-Als3 construct with ^{1}H-NMR spectroscopy.**
The 1D ^{1}H-NMR spectrum of sNT-Als3 construct in a buffer of 100mM Tris-HCl pH8, 50mM NaCl. The concentration of the sample was approximately 100µM. The spectrum showed good peak dispersion, indicative of ^{1}H nuclei within the protein experiencing a variety of different chemical environments.

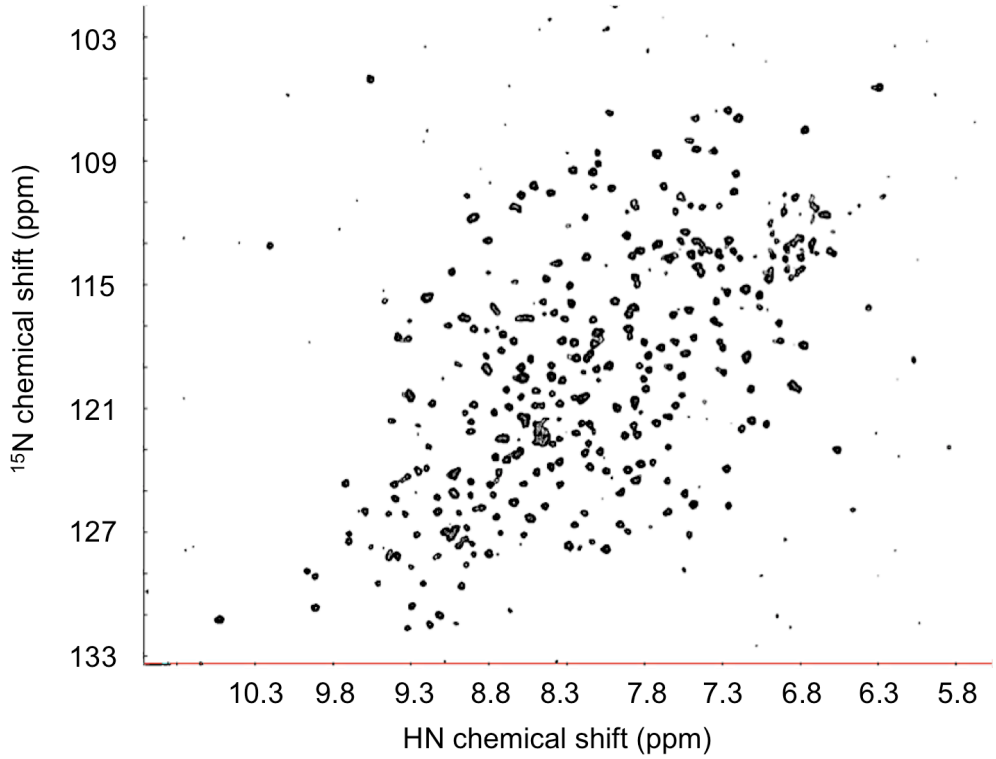


Figure 4.18: **Checking the sNT-Als3 construct with ^{1}H-^{15}N-TROSY-HSQC-NMR spectroscopy.**
^{1}H-^{15}N-TROSY-HSQC spectrum of sNT-Als3 construct showed well-resolved peaks with minimal overlap, indicative of dispersed N-H groups with minimal aggregation.

4.4.2. Triple-Labelled sNT-Als3 Expression and Purification

The sNT-Als3-TRX construct was first expressed in ^{15}N-^{13}C-^{2}H-minimal media as a small-scale 50ml trial, followed by a large two-litre preparation. The construct was successfully cleaved using Factor Xa with the same conditions used for sNT-Als1 (Fig 4.19).

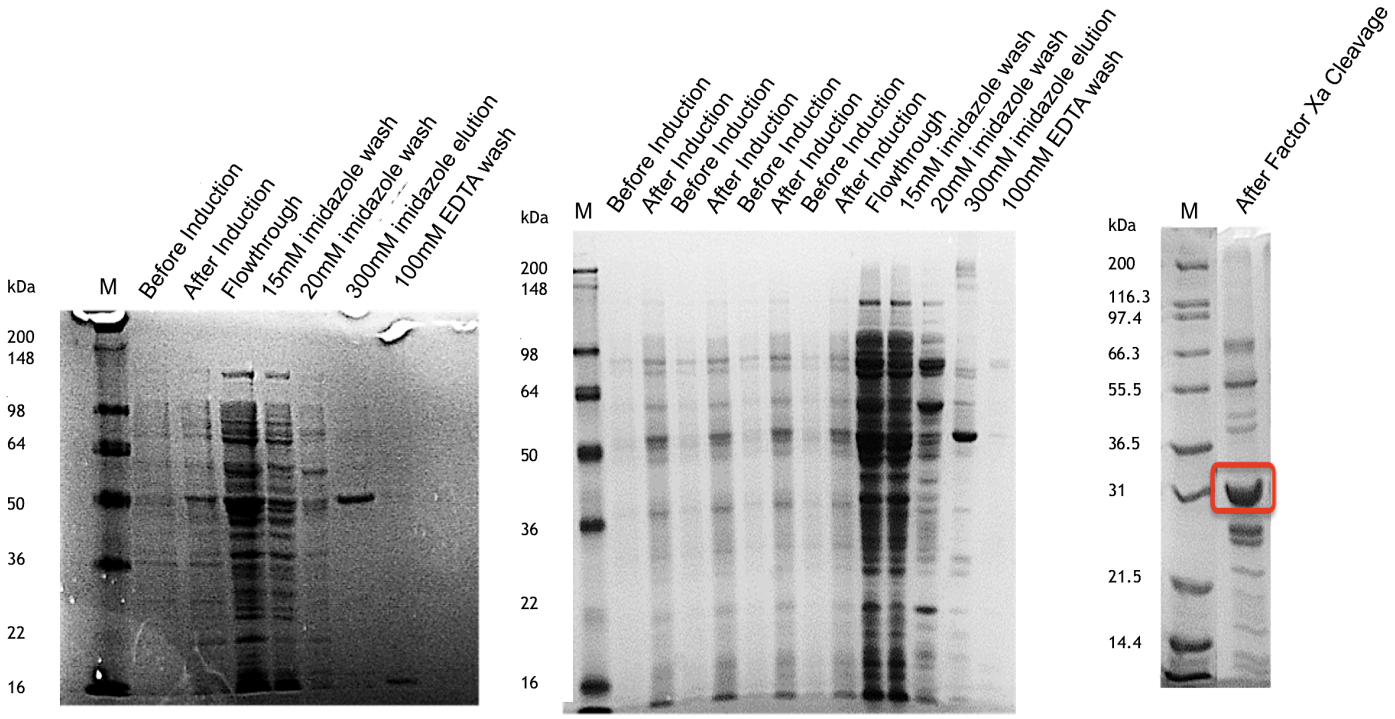


Figure 4.19: **SDS-PAGE gels of sNT-Als3-TRX expression and purification tests from D_2O-minimal media**.
A) SDS-PAGE gels of a small scale trial of sNT-Als3-TRX expression in D_2O-minimal media, B) large scale expression and purification of ^{15}N-^{13}C-^{2}H-labelled sNT-Als3-TRX from ^{15}N-^{13}C-^{2}H-minimal media and C) after Factor Xa cleavage of ^{15}N-^{13}C-^{2}H-labelled sNT-Als3 from TRX. The sNT-Als3 construct of expected size approx. 32kDa is encircled red.

However, although expression of triple labelled sNT-Als3 from ^{15}N-^{13}C-^{2}H-labelled media was successful, significant contamination was seen in size exclusion fractions. Analysis by SDS-PAGE showed that a distinct band, corresponding to the cleaved portion of the thioredoxin tag in pET32-Xa-LIC, co-eluted with the sNT-Als3 fraction, despite their differences in molecular weight, even after a second Nickel-IMAC step. (Fig 4.20) Interestingly, this effect was not observed in protein expressed in non-deuterated, minimal media.

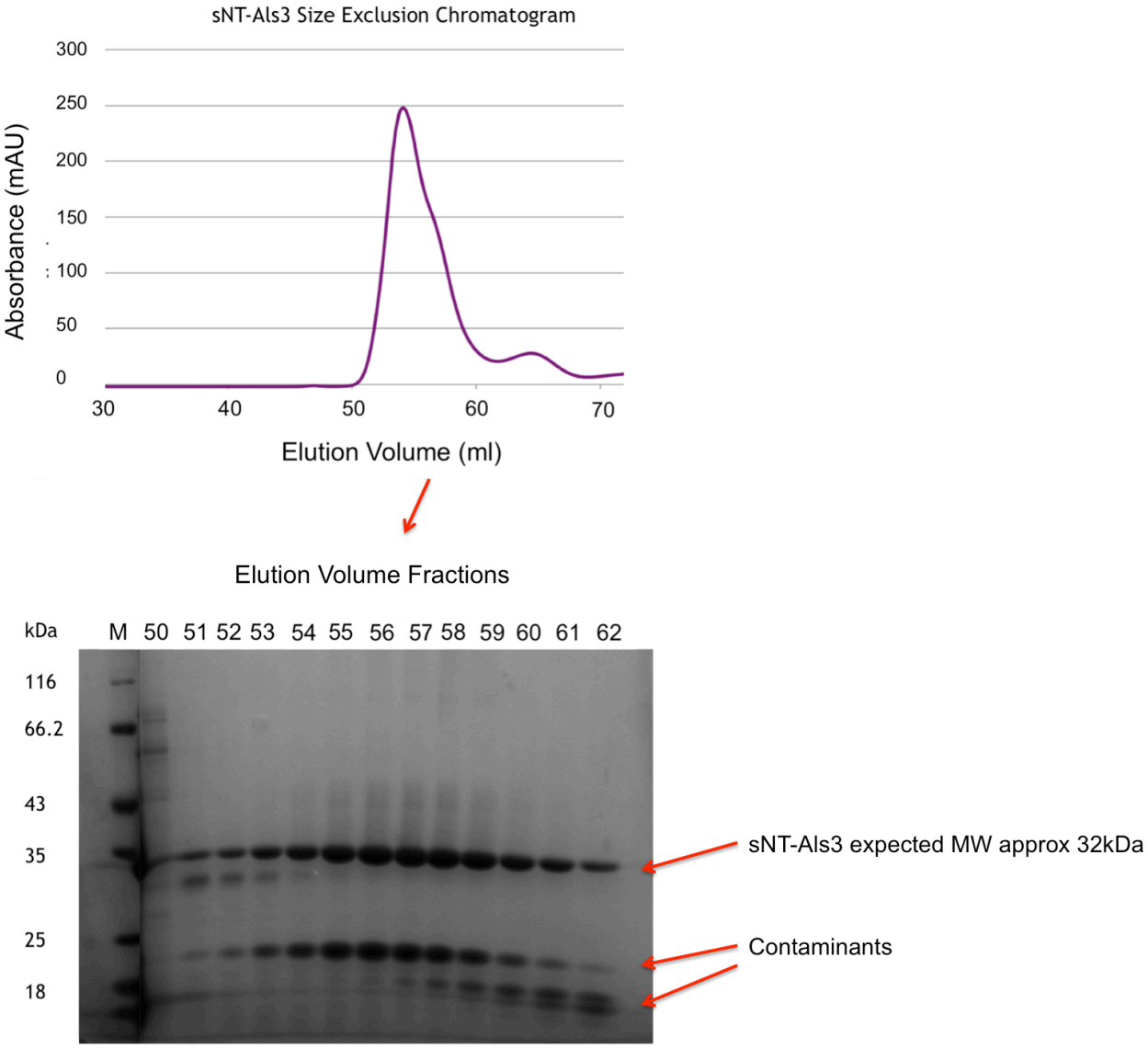


Figure 4.20: sNT-Als3 and Sample Contamination Problems.
SDS-PAGE gel and chromatogram of contaminated sNT-Als3 size exclusion chromatography fractions. Whilst a band corresponding to the expected sNT-Als3 size (approx. 32kDa) is seen, a contaminant band appears at approximately 25kDa.

To remove this contaminant, purification experiments involving different concentrations of sodium chloride, the addition of 4M urea or 1mM DTT were carried out by size exclusion chromatography using an analytical S75 column. Some improvement in purity was seen for samples containing 1mM DTT and 500mM NaCl, however this was difficult to reproduce when purifying sNT-Als3 with a larger S75 column (Fig 4.21).

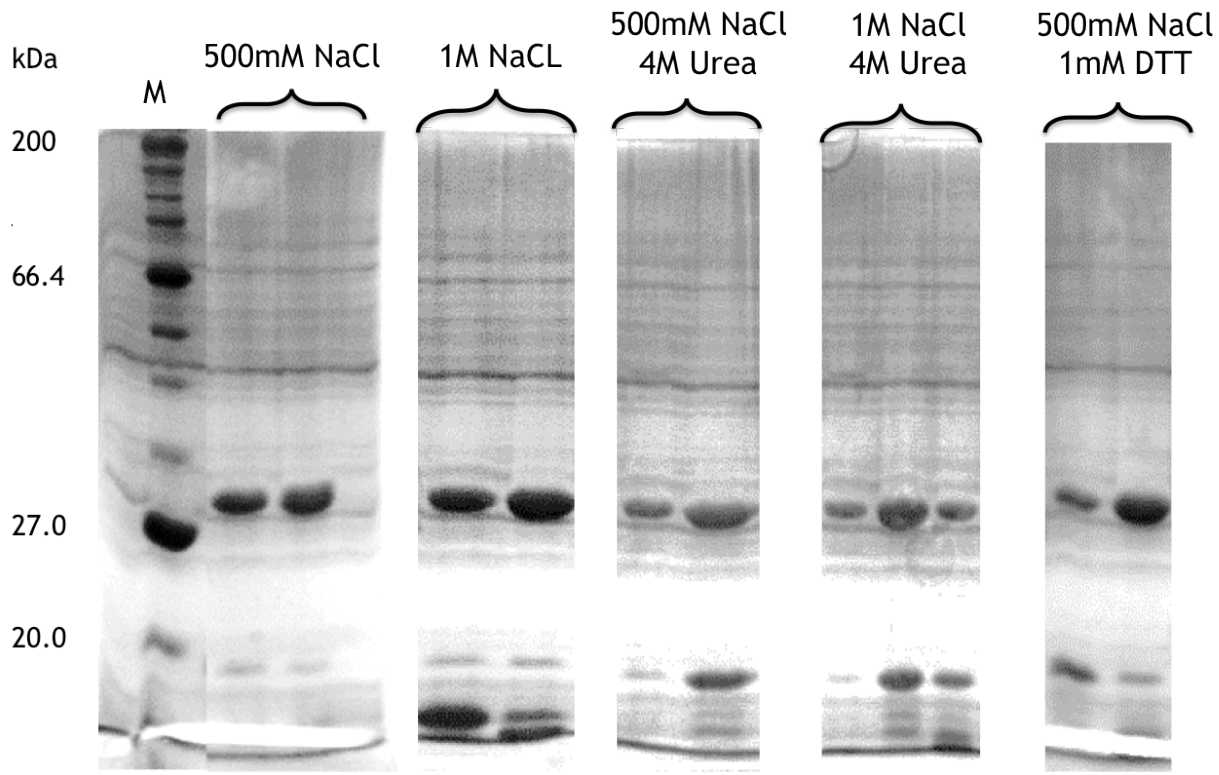


Figure 4.21: SDS-PAGE gels of size exclusion chromatography optimisation trials.
Optimisation included NaCl concentration modifications (500mM and 1M NaCl), addition of 4M urea, and addition of 1mM DTT. Some improvement in impurity is seen for the samples with 1mM DTT and 500mM NaCl, however contaminants are still present.

4.4.3. Further sNT-Als3 Expression and Purification Experiments

After this, expression of sNT-Als3 in the pET46 plasmid was revisited in order to see if modified expression methods could stop the protein from aggregating. It was thought that 'leaky' expression of sNT-Als3 (produced at 37°C prior to IPTG-induction) could promote aggregation of newly synthesised protein after induction. To solve this problem, the pLemo titratable-expression system was employed for production of sNT-Als3. The pLemo plasmid was extracted from *E.coli* Lemo21 cells and transformed into SHuffle cells along with the pET46 plasmid expressing sNT-Als3. Small scale culture expression tests were carried out which included various levels of L-rhamnose included with the addition of the inoculated colonies. The rationale behind this was that the E. coli T7 RNA polymerase could be inhibited at pre-induction temperatures by the T7 lysozyme expressed from the pLemo plasmid under the L-rhamnose-induced rhaBAD promoter. Unfortunately, this test showed negligible expression and purification yields of sNT-Als3 on SDS-PAGE gels.

A final expression test, whereby SHuffle cells were both grown and induced to express sNT-Als3 (again from pET46 plasmid) at a constant temperature of 18°C. This was then purified using Nickel-IMAC in a 4°C cold room, with 5 column volumes of a wash buffer containing 3M urea plus a 10 column volumes of 20mM imidazole wash buffer. Many contaminants were seen in the final elution, both before and after removing imidazole via dialysis (Fig 4.22). The elution after dialysis was re-purified on Nickel IMAC and subjected to multiple washes of varying imidazole concentrations, including 20 column volumes of 20mM imidazole wash buffer and 10 column volumes of 50mM imidazole wash buffer prior to the final 5 column volumes of 300mM imidazole wash buffer. This sample was then dialysed again to remove imidazole, concentrated to 5ml and another SDS-PAGE sample was taken. However, on further concentration of sample to 1ml, the precipitation and aggregation observed in previous attempts was seen again.

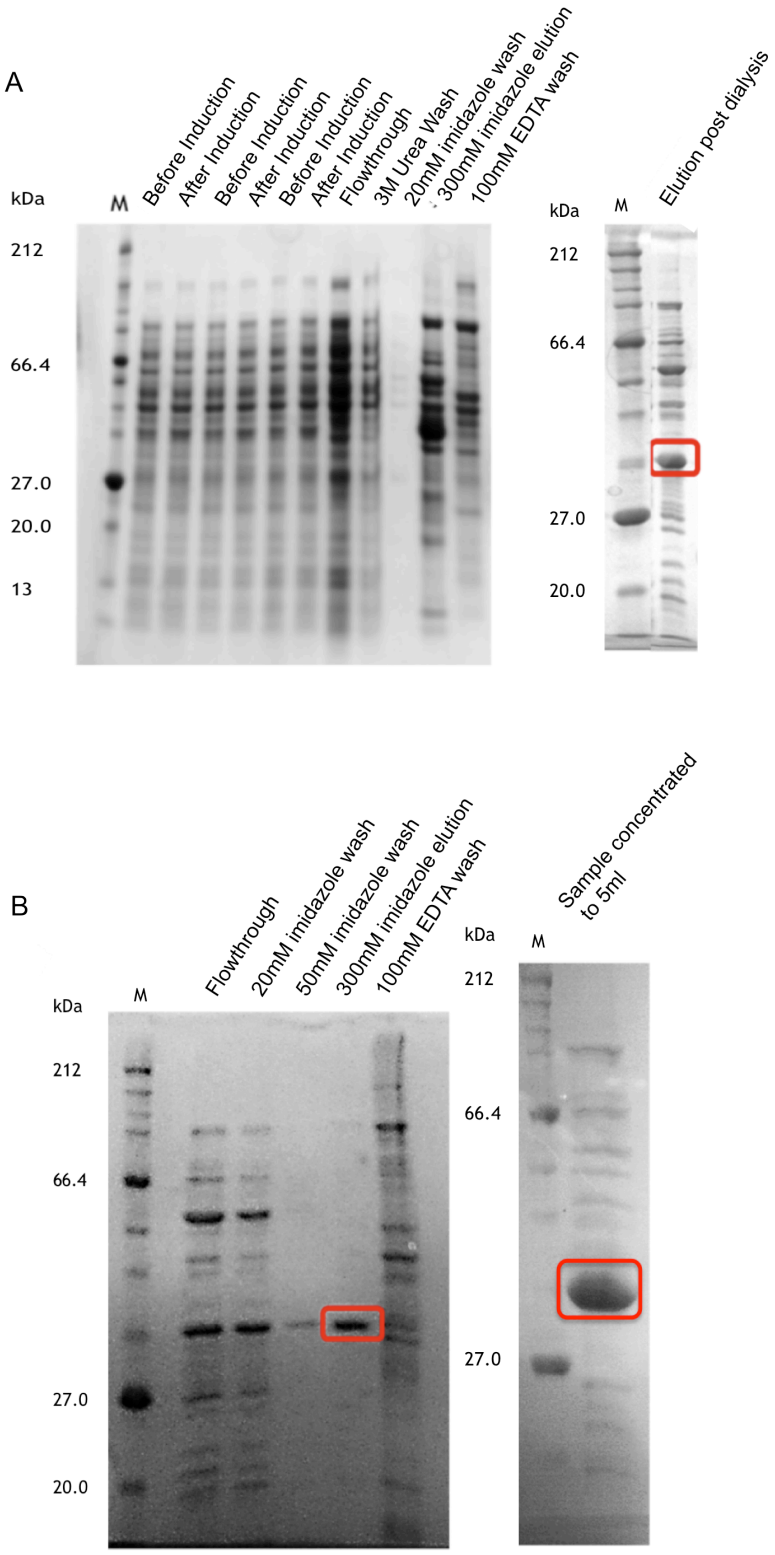


Figure 4.22: **SDS-PAGE gels from sNT-Als3 low temperature expression and purification.**
A) SDS-PAGE gels of sNT-Als3 expression and purification after Shuffle cells were grown and induced at constant temperature of 18°C, plus a 4°C purification including 5 column volumes of a 3M urea wash volume followed by 10 column volumes of 20mM imidazole wash buffer. B) SDS-PAGE gel of an optimised Nickel purification using 20 column volumes of 20mM imidazole wash buffer, 10 column volumes of 50mM imidazole wash buffer before the 300mM imidazole elution buffer. Another SDS-PAGE sample was taken after the sNT-Als3 elution sample was subjected to dialysis to remove imidazole. The sNT-Als3 sample of expected size approximately 32kDa is encircled in red and although some contaminants are seen, it was hypothesized that they could be removed by size exclusion chromatography.

The sNT-Als3 backbone assignment project was put on hold whilst efforts were focussed on solving the crystal structure and carrying out backbone assignment experiments for sNT-Als1. Later on, the production of mutant constructs of sNT-Als3 was carried out by other students of the Cota lab, which showed promise in solving the construct instability and aggregation problems, and which should be used to generate a NMR backbone assignment of sNT-Als3 in the near future.

5. Structural Determination of sNT-Als1 with a Hepta-Threonine Peptide

The short Als1 N-terminal construct (sNT-Als1) construct was produced using methods described in Chapter 4 and then applied for structural studies by X-ray crystallography. This construct was used because previous attempts to crystallise the full length NT-Als1 were unsuccessful due to aggregation. Furthermore, attempts to generate diffractable crystals from native sNT-Als1 had been unsuccessful. However, co-crystallisation of sNT-Als1 with Hepta-T peptide ligand with the additional use of deuterium oxide additive and further crystallization condition optimization lead to the production of diffractable crystals. Hepta-T was chosen as a ligand due to it being documented as able to interact with Als proteins in previous literature (Gaur et al 2002), was demonstrated with previous structural studies to interact with sNT-Als3 and NT-Als9.2 (Salgado et al 2011, Lin et al 2015), and shown to have a high probability of binding to sNT-Als1 from DSF experiments (see Chapter 4, section 4.2.2).

The crystal structure of sNT-Als1 with the Hepta-T peptide (sNT-Als1/Hepta-T), described more in detail in section 5.2.1 (Fig 5.1), will not only be useful for elucidating knowledge of the Als1 PBC shape, structure and interaction with the Hepta-T ligand, but also indicate PBC differences and thus ligand-binding specificity between sNT-Als1 and sNT-Als3. Furthermore, knowledge of how particular PBC residues interact with the residues and chemical groups of the Hepta-T peptide will form a basis for the design of high-affinity ligands from fragment libraries. The result of such information shall drive the development of therapeutic drug-like molecules.

5.1. From Data to Structure

5.1.1. X-Ray Diffraction Data Collection

Crystals from co-crystallisation screens of sNT-Als1 with Hepta-T peptide (sNT-Als1/Hepta-T crystals) were fished from well F11 of plate ICL11 (Jena Biosciences), and from optimisation plate wells. The wells included ethylene glycol as both a precipitant and cryoprotectant, meaning that the crystals could be immediately flash frozen using liquid nitrogen without prior bathing in a separate cryoprotectant solution. X-ray data collection was performed using synchrotron radiation with a 0.96861Å wavelength from beamline I-24 (Diamond Light Source, UK). The crystal that resulted in the most optimal diffraction was from the 200µl ICL11-F11 well. This crystal resulted in a diffraction pattern of Bragg spots beyond 2Å, but 2.2Å resolution corresponded to data with good statistics (Fig 5.15). Data was collected for 360°, at 100K, with an oscillation angle of 0.5° per frame and with a Pilatus 6M (6-Megapixel) detector at a distance of 347.037mm.

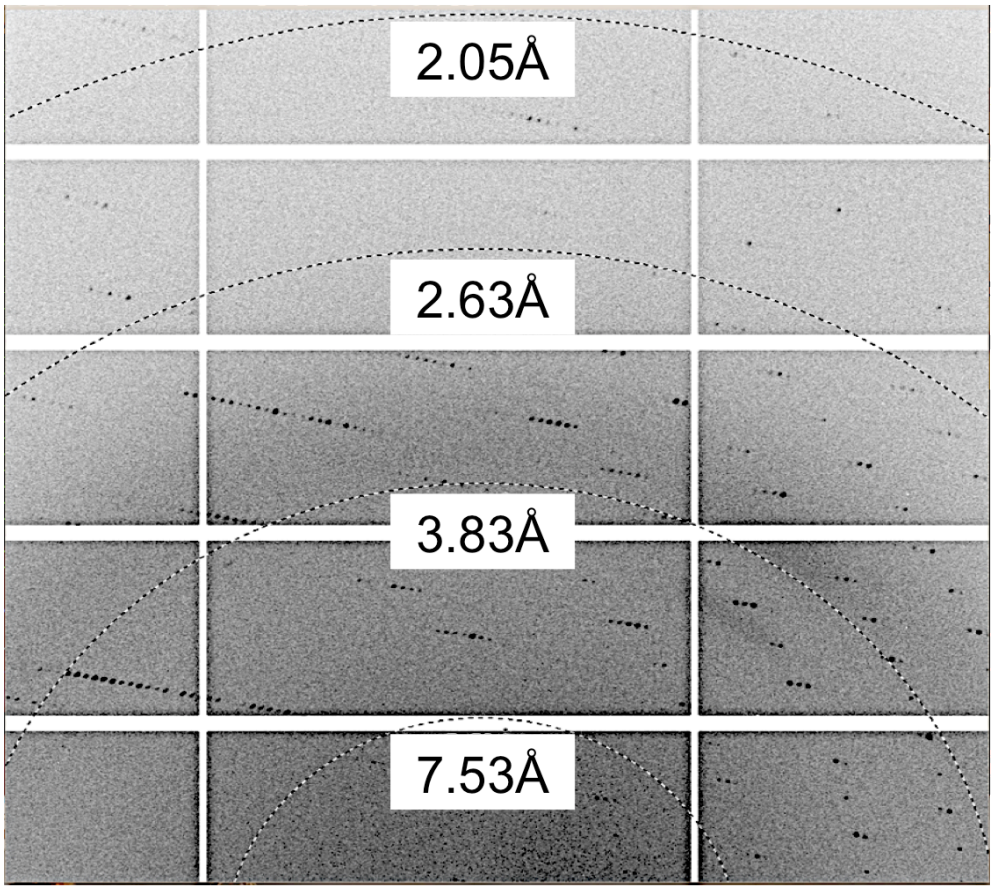


Figure 5.1. Diffraction image from an sNT-AIs1/Hepta-T crystal.
Diffraction pattern obtained from synchrotron radiation with a 0.96861Å wavelength from beamline I-24, and using a Pilatus 6M (6-Megapixel) detector at a distance of 347.037mm (Diamond Light Source, UK). Bragg Diffraction spots can be seen at the outermost edges of the pattern, corresponding to 2.2Å resolution.

5.1.2. Data Processing

As the quality of the diffraction images significantly decayed after 180° of data collection due to radiation damage, images 1-180 were indexed and processed using the data reduction program Xia2 (Winter 2010) and then scaled using the program Scala (Evans 2006), which estimated a point group of P222 for this particular crystal. The point group refers the Bravais lattice type (in this instance P, for primitive lattice type), and the most prominent symmetry operations along the three crystallographic axes (in this instance 222, and Laue group 'mmm' corresponding to rotational symmetry). The analysis of systematic absences between Bragg diffraction spots within the images was carried out using the program Pointless (Evans 2006, Evans 2011), which confirmed that the crystal had a space group of $P2_12_12_1$. The space group notation '$P2_12_12_1$', of space group number 19, refers to a collection of mathematical generators describing the symmetry relationships between the atoms of the crystal. For this particular acentric crystal, the symmetry operations are described as 'screw axes', in that the positions of identical atoms can be extrapolated from a rotation plus a translation operation.

The program Truncate (French & Wilson 1978) from the CCP4 (Collaborative Computational Project No. 4) suite (Winn et al 2011) was used to calculate structure factor amplitudes from intensities. Overall statistics for data collection and processing are listed in Table 5.1. Matthews coefficient calculations using the Matthews Coefficient program (Matthews 1968, Kantardjieff & Rupp 2003), also from the CCP4 program suite (Winn et al 2011), estimated the presence of two molecules per asymmetric unit with a Matthews coefficient (V_M) value of 2.85Å^3/Dalton and P(total) of 0.70, with solvent content of 56.88% (see Table 5.2).

Table 5.1: Summary of Overall Data Collection, Processing and Scaling Statistics for sNT-Als1/Hepta-T crystal.

Statistic		**sNT-Als1/Hepta-T**
Space Group		$P2_12_12_1$
Unit Cell Dimensions (Å)	*a*	51.83
	b	58.14
	c	227.07
Unit Cell Angles (°)	α	90
	β	90
	γ	90
Wavelength (Å)		0.96861
Beamline		I-24, Diamond
Number of reflections (total)		132428
Number of reflections (unique)		40632
I/sigI (mean)		7.8
Multiplicity		3.3
Anomalous multiplicity		1.8
R_{sym}(%)		0.030
R_{merge} (%)		0.146
R_{merge} in top intensity bin (%)		0.030
R_{meas} (within I+/I-)		0.174
R_{meas} (all I+ and I-)		0.174
R_{pim} (within I+/I-)		0.094
R_{pim} (all I+ and I-)		0.094
Half-data set correlation coefficient		0.993
Completeness (%)		99
Anomalous completeness (%)		88.9
Low resolution limit (Å)		51.83
High resolution limit (Å)		2.10
Δ_{anom} Correlation between half		0.068
Fractional partial bias		0.000
Average mosaicity		0.000
Average Wilson B Value ($Å^2$)		22.51
Number of molecules in Asymmetric Unit		2
Matthews Coefficient/V_M ($Å^3$/Dalton)		2.85
Solvent Content (%)		56.88

Table 5.2: Summary of Matthews Coefficient values for the sNT-Als1/Hepta-T Crystal.

Number of Molecules in Asymmetric Unit	Matthews Coefficient (V_M) (A^3/Dalton)	Solvent Content (%)	P (2.10)	P (tot)
1	5.70	78.44	0.00	0.01
2	2.85	56.88	0.63	0.70
3	1.90	35.33	0.37	0.29
4	1.43	13.77	0.00	0.00

5.1.3. Structural Determination

The structure of sNT-Als1/Hepta-T was solved using the Molecular Replacement method and using the program Phaser 2.5.0 (McCoy et al 2007) with phase information calculated from the structure factors of the known crystal structure of sNT-Als3 (amino acids 1-302), used as the search model due to an approximately 82% amino acid sequence identity between the two proteins (Lin et al 2014, PDB code: 4LEB). A resolution range of 2.10-51.75Å was used for the molecular replacement. The result was a definite solution with two molecules per asymmetric unit and space group $P2_12_12_1$ (screw axes 2_1), with final Z-scores of 71.8 and log likelihood gain of 9864 (see Table 5.3). The packing check ensured no clashes between asymmetric units of the unit cell.

Table 5.3: Molecular Replacement Statistics for sNT-Als1/Hepta-T Crystal (Z-scores)

R-Factor Z-Score	Translational Z-Score	Packing Check (PAK)	Log Likelihood Gain
20.2	71.8	0	9864

5.1.4. Structural Refinement and Model Building

The refinement of atomic coordinates and isotropic B-factors was carried out using both refinement programs Refmac5 (Murshudov & Dodson 1997, Vagin et al 2004) and Phenix (Afonine et al 2012) from the CCP4 program suite (Winn et al 2011). The goal of refinement is to gain the most optimal compliance between observed reflection amplitude structure factors (F_{obs}) and the structure factor amplitudes calculated from the model (F_{calc}). The similarity of both amplitude values is expressed using the R_{work} number, which is a measure of the errors between model and the electron density calculated from experimental reflection data (Weiss & Hilgenfeld 1997). Another measure is also used to judge the model – the R_{free}, which is a value calculated from a random selection of approximately one thousand reflections not used for the model refinement process, and which checks for the over-fitting of the model (Brünger 1992). Ideally, the R_{work} should be as low as possible, and the R_{free} should not exceed the R_{work} value by more than 7% (Wlodawer et al 2007).

The initial R_{work} and R_{free} values for the structure were 0.3798 and 0.3897 respectively. The final R_{work} and R_{free} values of 0.2115 and 0.2581 were achieved by repeated cycles of automated refinement using maximum-likelihood statistical algorithms plus the manual adjustment of the model to improve its correlation with the electron density. Manual corrections of the model involved using the Real-Space Refinement, Regularize, Rotate/Translate, Rigid Body, Simplex, Rotamer and Edit Chi Angles Refinement tools of the program Coot (Emsley et al 2010) in order to improve the geometrical positions of atoms within the model, the ψ and ϕ backbone torsion angles and optimize the chi angles and rotamer type of each residue. Refmac5 (Vagin et al 2004) was used initially for once cycle of refinement, but the program Phenix was used for the rest of the refinement cycles due to its ability to coordinate with COOT for ease of atomic adjustment, plus its intuitive graphical user interface, which improved the efficiency of model refinement.

Within the automated refinement cycles using Phenix, refinement strategies included the calculation of TLS parameters to refine anisotropic displacements for large collections of atoms such as chains, the refinement of atomic occupancies, XYZ coordinates to correlate the refinement of the geometry of the model structure with the original X-ray data and the refinement of restrained, individual atomic B-factors (Afonine et al 2012).

Because the asymmetric unit of the sNT-Als1/Hepta-T crystal structure contains two identical sNT-Als1/Hepta-T molecules, non-crystallographic symmetry (NCS) restraints could be used for the refinement strategy so that the root-mean-square distance (RMSD) of both asymmetric unit molecules were compared. Corresponding atoms from both molecules with similar RMSD values were assigned tight NCS restraints. Both global and torsion-angle NCS restraint calculations were carried out during refinement steps, but torsion-angle restraints were preferred as it allowed for

molecular flexibility whilst also conserving specific conformations. Ordered water molecules were specified at first, but many water molecule coordinates were modified or added manually.

X-ray/stereochemistry weights and X-ray/ADP (atomic displacement parameters) weights were optimized. Simulated annealing using both Cartesian and torsion angles as parameters, and which adds molecular dynamic predictions in order to bypass phase bias, was used to improve model building. The program Reduce (from command 'Automatically correct N/Q/H errors') was employed to correct and flip back side-chains (Afonine et al 2012). The final refinement resulted in an R_{work} of 0.212 and an R_{free} of 0.258, with the complete refinement statistics listed in Table 5.4.

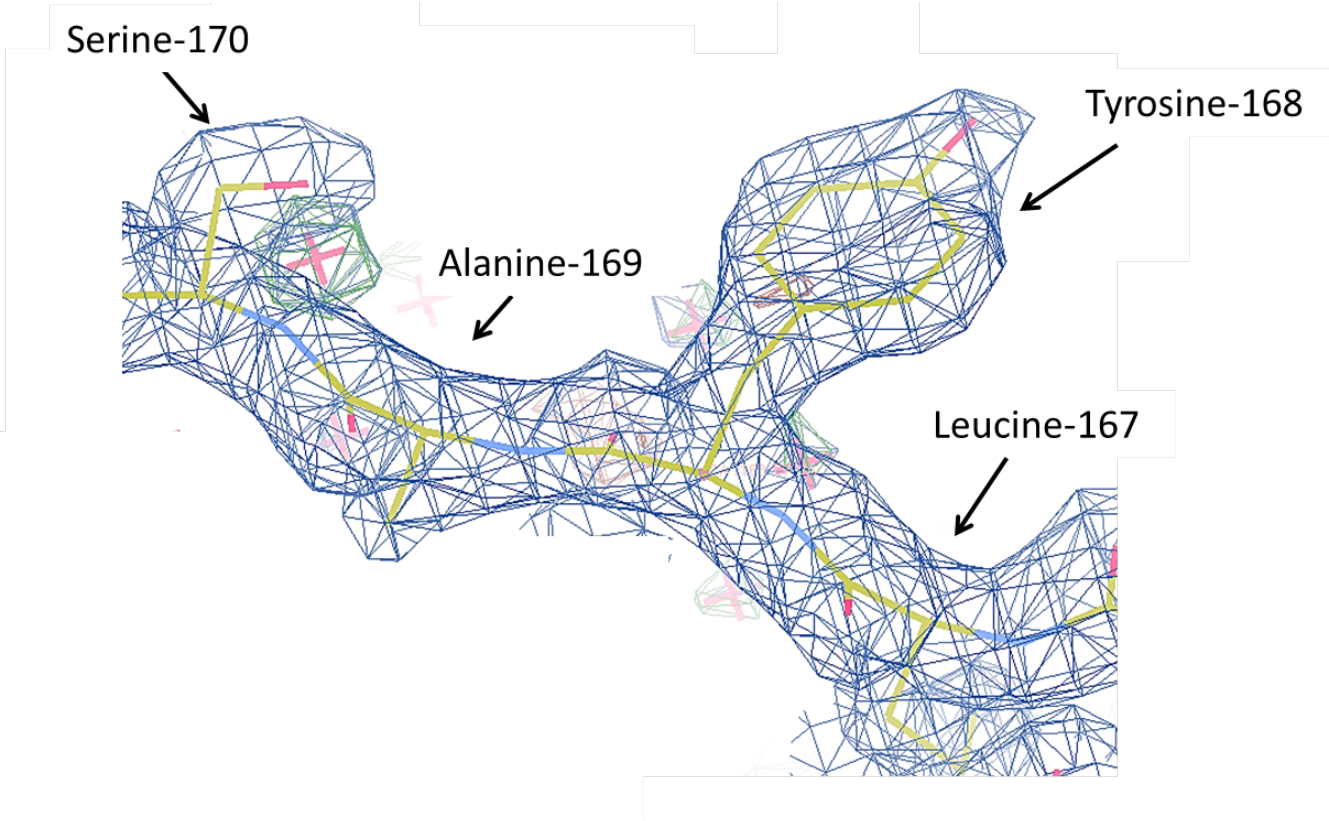


Figure 5.2: An example of an $2F_{obs}$-F_{calc} (Electron Density Map) of the sNT-Als1/Hepta-T crystal structure and Model Building.
The 2Fo-Fc map (shown in blue) is contoured at 1 Sigma and displayed over a stick model (yellow, cyan and pink) of the sNT-Als1/Hepta-T model to demonstrate the quality of data. The stretch of residues displayed includes L167, Y168, A169 and S170 (Modelled using Coot).

Table 5.4: Refinement Statistics for Final Refinement of sNT-Als1/Hepta-T Model

Refinement Statistic		**Value**
Resolution (Å)		51.8 – 2.1
Reflections used in Refinement		40593
Reflections used in R_{free} Calculation		40593
Total Protein Residues in Asymmetric Unit model		613
Number of Protein Molecules in Asymmetric Unit		2
Number of Ions		4
Number of Water Molecules		357
Number of Non-Hydrogen Atoms		4793
Number of Protein Atoms		4421
RMS Deviations from Ideal	Bond Length (Å)	0.011
	Bond Angles (°)	0.836
	Dihedral Angles (°)	11.612
R_{work}		0.2115
R_{free}		0.2581

5.1.5. Model Validation

Following model building and refinement, validation was carried out using COOT (Emsley et al 2010), MOLPROBITY (Chen et al 2010), PROCHECK (Lasowski et al 1993) and the PDB server (Berman et al 2003). Coot validation tools were used to facilitate the adjustment of side chain orientation via rotamer analysis, and check for unexplained sections of electron density. No unexplained sections of electron density were pinpointed once the model building was complete, and this suggested that no surplus ligands, water molecules or ions were present that had not been modelled.

PROCHECK validation tools to check main chain (Figure 5.3 and Table 5.5) and side chain parameters (Figure 5.4 and Table 5.6) were carried out. These validation tools suggested that the sNT-Als1/Hepta-T crystal structure compared favourably or inside expected values, with other well-refined structures of similar resolutions when comparing factors such as Ramachandran plot statistics, peptide bond planarity, bad non-bonded interactions, alpha carbon tetrahedral distortion, hydrogen bond energies, overall G-factors and side-chains..

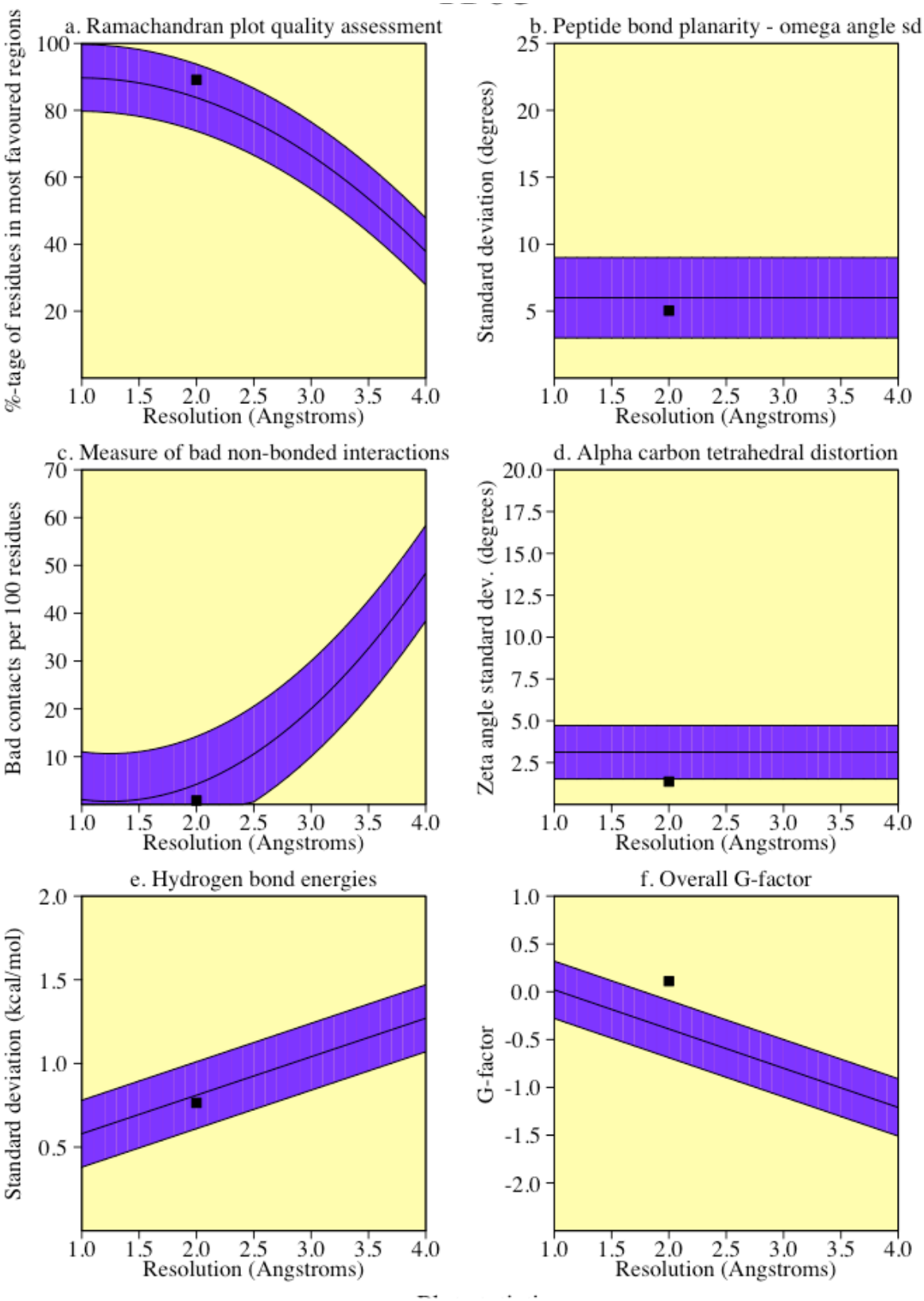


Figure 5.3: Main Chain Parameters (PROCHECK).
Six graphical representations of the main-chain parameters display how the structure of sNT-Als1/Hepta-T compares with other PDB structures that have been well-refined with a similar resolution. The parameters correspond to a) Ramachandran plot quality, b) peptide bond planarity, c) bad non-bonded interactions, d) Calpha tetrahedral distortion, e) main-chain hydrogen bond energy, f)G-factor. The violet band in each graph corresponds to results from well-refined structures, the central black line represents a least-squares fit to the mean as a function of resolution, the width of the band on either side corresponds to a one standard deviation variation from the mean. The G-factor value is a measure of the normality of the structure, taken from the averaging of the G-factors of individual residues within the structure (Kabsch & Sander 1983).

Table 5.5: Main Chain Parameters (PROCHECK) Statistics for sNT-Als1/Hepta-T Structural Model

Stereochemical paramter	Number of data points	Parameter Value	Comparison values		No. of band widths from mean	Rank
			Typical value	Band width		
a) Percentage of residues in favourable Ramachandran positions A, B, L	533	89.1	83.8	10.0	0.5	Inside
b) Omega angle standard deviation	607	5.0	6.0	3.0	-0.3	Inside
c) Bad contacts/100 residues	5	0.8	4.2	10.0	-0.3	Inside
d) Zeta angle standard deviation	565	1.4	3.1	1.6	-1.1	Better
e) H-bond energy standard deviation	373	0.8	0.8	0.2	-0.2	Inside
f) Overall G-factor	617	0.1	-0.4	0.3	1.7	Better

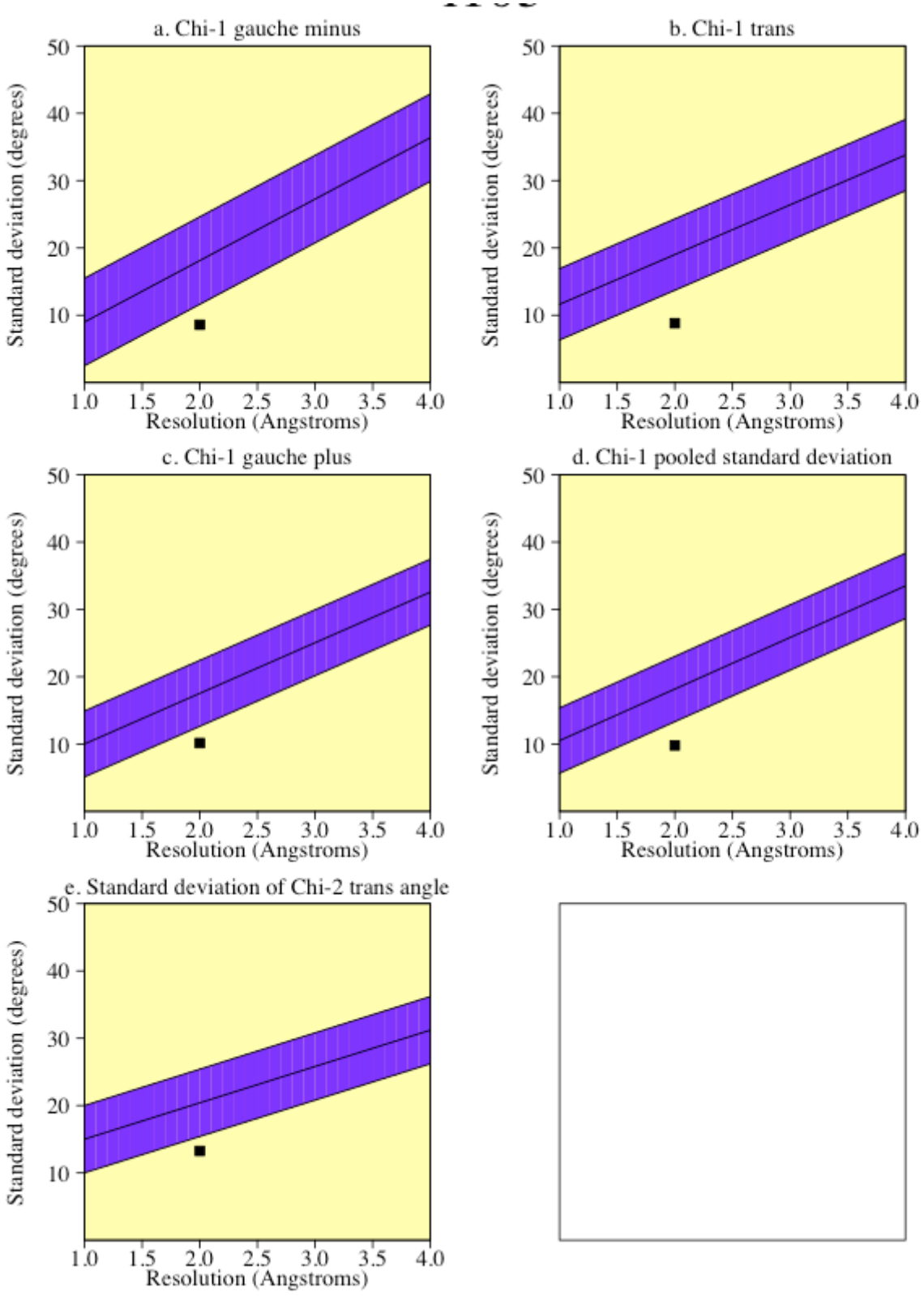


Figure 5.4: Side Chain Parameters (PROCHECK).
Five graphical representations of the side-chain parameters of the sNT-Als1/Hepta-T structure and how these compare with other well-refined structures of a similar resolution. Parameters displayed include a) Standard deviation of the chi-1 guache minus torsion angles, b) Standard deviation of the chi-1 trans torsion angles, c) Standard deviation of the chi-1 guache plus torsion angles, d) Pooled standard deviation of all chi-1 torsion angles. e) Standard deviation of the chi-2 trans torsion angles. The violent band represents the results from well-refined structures, central black line shows a least-squares statistical fit to the mean as a function of resolution, and the width of the band either side of the black line represents a variation of one standard deviation from the mean.

Table 5.6: Side Chain Parameters (PROCHECK) Statistics for sNT-AIs1/Hepta-T Structural Model

Stereochemical paramter	Number of data points	Parameter Value	Comparison values		No. of band widths from mean	Rank
			Typical value	Band width		
a) Chi-1 guache minus Standard Deviation	152	8.6	18.1	6.5	-1.5	Better
b) Chi-1 trans Standard Deviation	113	8.8	19.0	5.3	-1.9	Better
c) Chi-1 guache plus Standard Deviation	231	10.1	17.5	4.9	-1.5	Better
d) Chi-1 pooled Standard Deviation	496	9.8	18.2	4.8	-1.7	Better
e) Chi-1 pooled Standard Deviation	94	13.3	20.4	5.0	-1.4	Better

The Ramachandran plot, developed by Sasiekharan and Ramachandran in the early 1960s, is a graphical representation of amino acid ϕ and ψ torsion angles and is a very useful validation tool to show the available torsion angles for particular amino acids in a protein and pin-point residues that have been modelled in a sterically improbable conformation (Ramachandran et al 1963). Ramachandran plots from MOLPROBITY (Chen et al 2010) (Fig 5.5 and Table 5.7) revealed that 97.19% (588/605) of the residues were found in favoured regions and 2.81% (17/605) of residues were in allowed regions. No outliers were found in disallowed regions.

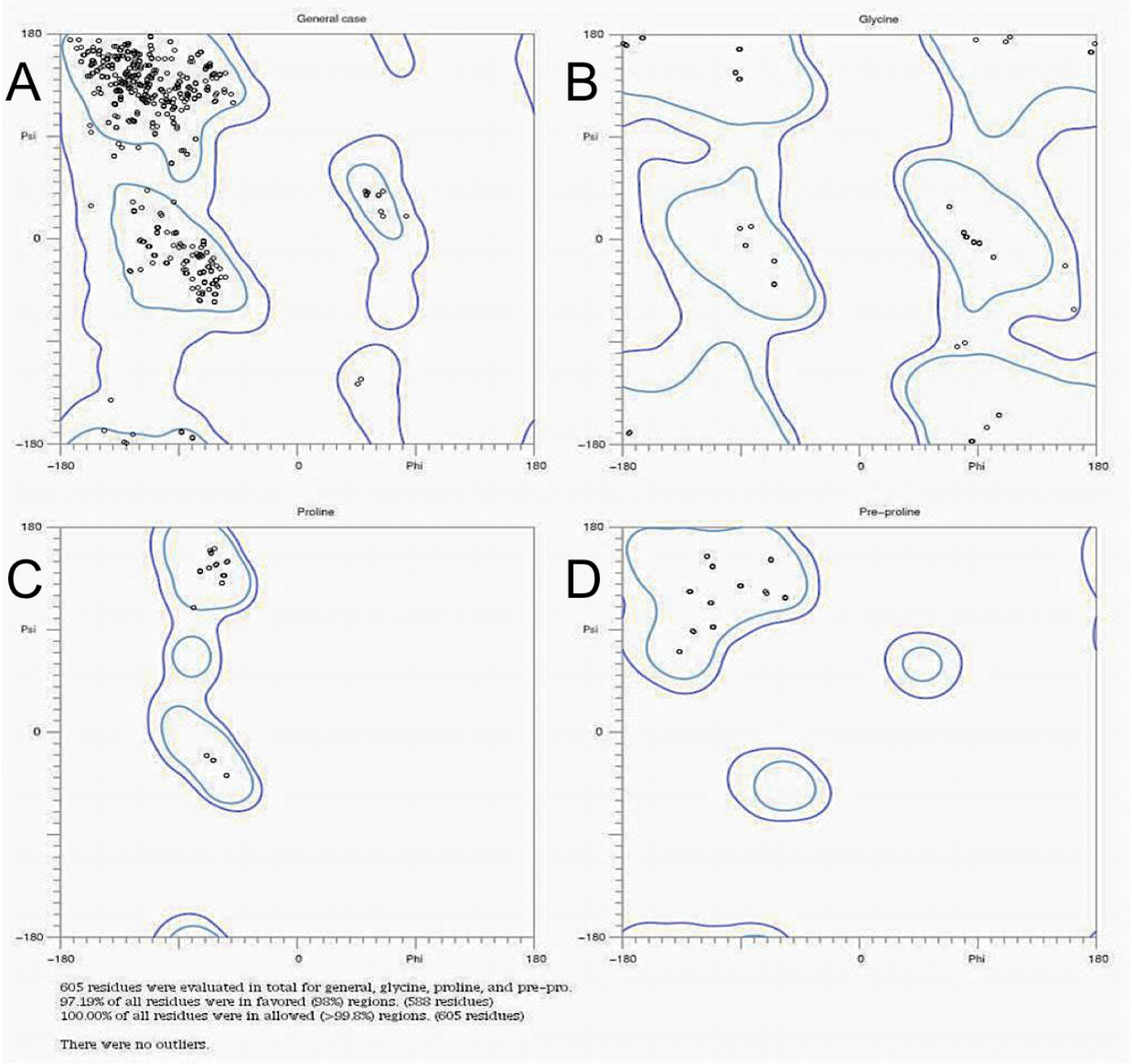


Figure 5.5: Ramachandran Plots from MOLPROBITY Validation Tool.
These are for validation of A) the overall structure, B) glycine residues, C) proline residues and D) Pre-proline residues for the sNT-Als1/Hepta-T Model Structure (Molprobity).

Table 5.7: MOLPROBITY Ramachandran Plot Statistics for sNT-Als1/Hepta-T Crystal

Residues in Favoured Regions	588 residues	97.19%
Residues in Allowed Regions	605 residues	100%
Residues in Disallowed Regions	0 residues	0%

PROCHECK analysis (Lasowski et al 1993) for further validation of the model also included main chain Ramachandran plot combined with the checking of the torsion angles and chiral centres of individual amino acids. Figure 5.6 shows the Ramachandran plot produced by PROCHECK, and Table 5.8 shows the accompanying statistics.

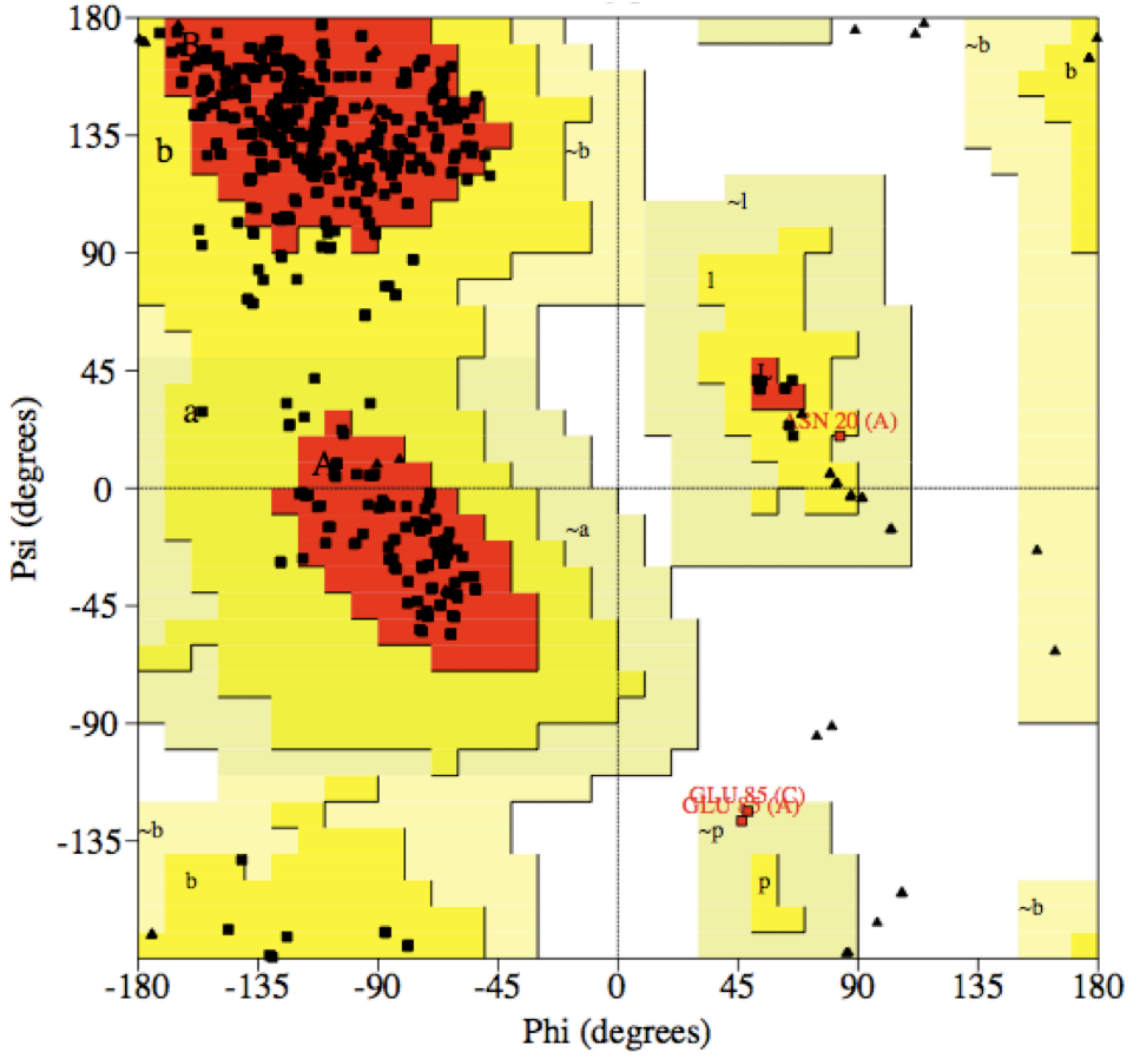


Figure 5.6: Main chain Ramachandran Plots from PROCHECK (Lasowski et al 1993) analysis of the sNT-Als1/Hepta-T Crystal structure.
A graphical representation of the phi-psi torsion angles for all residues in the structure (except those at N-termini or C-termini of the chains). Glycine residues shown as triangles, these residues are not restricted to sections of the plot favourable for other sidechains. Plot colouring represents different regions according to Morris et al (1992). Red shows the 'core' regions of the most favourable phi-psi combinations.

Table 5.8: MOLPROBITY Ramachandran Plot Statistics for sNT-Als1/Hepta-T Crystal

Residues in Most Favoured Regions (A, B, L)	475 residues	89.1%
Residues in Additional Allowed Regions (a, b, l, p)	55 residues	10.3%
Residues in Generously Allowed Regions	3 residues	0.6%
Residues in disallowed regions	0 residues	0.0%
Number of non-glycine and non-proline residues	533 residues	
Number of end-residues (excluding Gly and Pro)	12	
Number of glycine residues (triangles)	48	
Number of proline residues	24	

Ideally, 90% of residues within the structure should be in the most-favourable 'core' regions coloured red, or in the additional allowed regions of the Ramachandran plot (Figure 5.6), according to an analysis of 118 structures (of 2.0A resolution). PROCHECK marked residues N20 from chain A and E85 from both Chain A and Chain C as residues in generously allowed regions.

Both PROCHECK and the PDB validation server flagged up a collection of residues as being rotamer outliers. These residues include A19, N20 and E85 in both molecules. However, the modelling of these residues fit the electron density (Fig 5.7). N250 was flagged to have an unexpected chiral centre – however this residue also fits the electron density, albeit with some flexibility in both molecules.

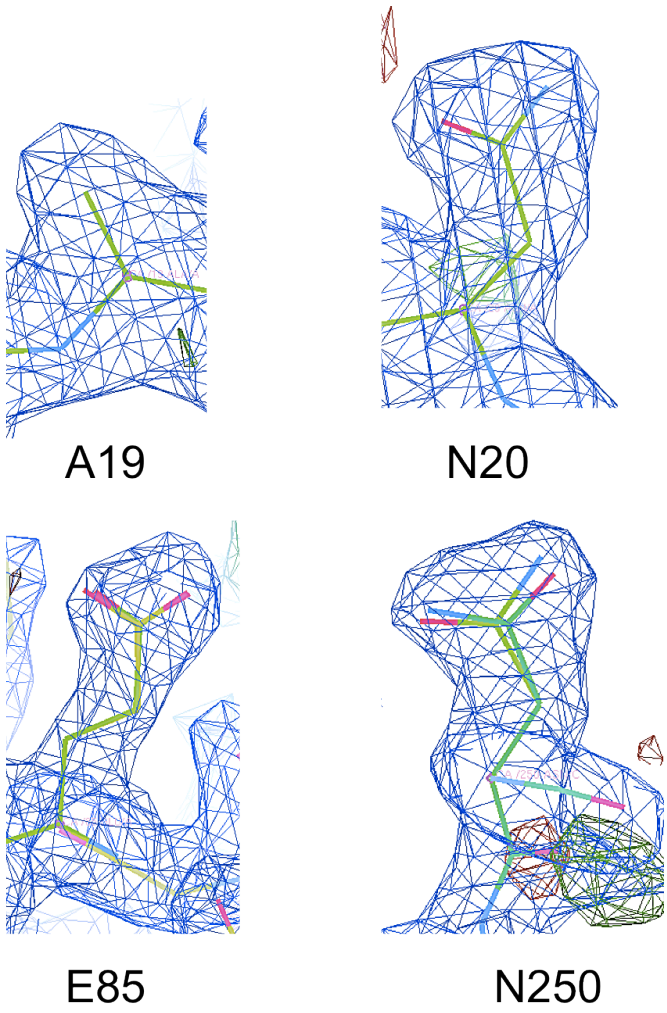


Figure 5.7: 2Fo-Fc Map Comparison with the PDB validation server check.
Shown are sections from the 2Fo-Fc map (coloured blue), contoured at 1 Sigma, displayed over two superimposed stick models; the stick model coloured yellow represents the model from Chain A, the stick model coloured green represents the model from Chain C. The superimposition was achieved using the 'NCS Ghost Coordinates' function of Coot. The PDB validation server pin-pointed residues within the models from both sNT-Als1 chains, namely A19, N20 and E85 as rotamer outliers, and N250 to have an unexpected chiral centre. However, as can be seen from the models vs. the 2Fo-Fc map, the rotamers chosen in refinement fit the electron density.

The final crystal structure at 2.1Å of sNT-Als1/Hepta-T is comprised of two protein molecules (chains A and C), each with their own seven-residue Hepta-T ligand (chains B and D), resulting in four chains, with a total of 613 residues. Chain A has 299 residues and Chain C has 300 residues. There are 357 water molecules and 4 sulphate ions. The first amino acid of the protein, Lysine-1, was clear in both molecules. Electron density preceding this N-terminal lysine corresponds to remnants of a histidine tag that was not cleaved prior to crystallisation. However, the expected C-terminal residues Lysine-299 and Arginine-300 were not seen in the electron density in Chain A – similarly Arginine-300 was not seen in Chain C. Using PyMOL, Chain A and Chain C can be aligned with an RMS of 0.175(269

to 269 Cα atoms), showing a high degree of model similarity between the two sNT-Als1 molecules of the asymmetric unit (Fig 5.8). The Hepta-T molecules, Chain B and Chain D, can also be aligned to an RMS of 0.182 (7 to 7 Cα atoms), also showing a high degree of model similarity between these chains.

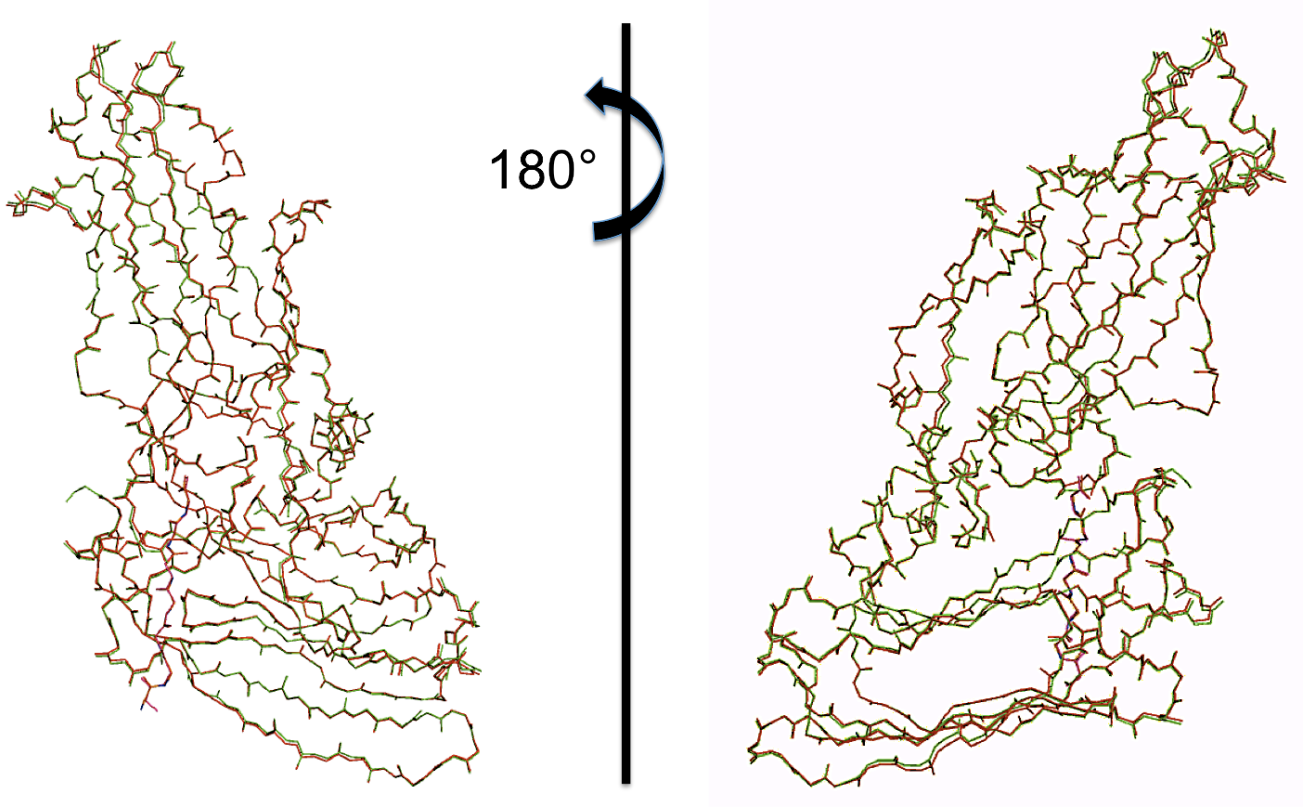


Figure 5.8: **The alignment of sNT-/Als1 Chains A and C, and Hepta-T Chains B and D, shown as a Cα Trace.**
The alignment of sNT-Als1 molecule Chains A (green) and C (red), of 269 Cα of Chain A with 269 Cα, with an RMS of 0.175, shows structural similarity. Hepta-T Chains B (pink) and D (orange), of which 7 Cα were aligned with 7 Cα, also show high structural and positional similarity, with an RMS of 0.182. Alignment carried out using PyMOL.

The distribution of atomic B-factors (displacement parameters) can be used to indicate whether or not a structure as a whole is ordered or highly mobile, with high B-factors (more than 40-50Å^2) corresponding to overall high flexibility of atoms within the structure (Karplus & Schulz 1985). The overall Wilson B-factor for the structural model is 22.51Å^2, which is similar to average B-factors (listed in Table 5.8) calculated using the Baverage program from the CCP4 suite (Winn et al 2011). The average B-factor for Chain A is 25.07 Å^2 and for Chain C, 25.00 Å^2. These B-factors indicate that overall the structure has a high degree of order, despite some regions having high B-factors. A plot of the average B-factors for each residue of sNT-Als1 Chains A and C are displayed in Figure 5.9. The residues N145, G146 and D147 in both sNT-Als1 molecule Chains A and C have high B-factors; N145 has an average B-factor of 60.7Å^2 in Chain A and 58.0 Å^2 in Chain C, G146 has an average B-factor of 72.9 Å^2 in Chain C and 86.6 Å^2 in Chain C and D147 has an average B-factor of 75.6Å^2 in Chain A and 61.0 Å^2 in Chain C. However, this stretch of residues with high B-factors occurs in an unstructured loop region of the protein, which is expected to possess a high degree of flexibility (see Figure 5.10). Additionally, a histidine residue prior to K1 of sNT-Als1 Chain A, also present on an unstructured loop region, has a high B-factor of 99.8 Å^2.

Table 5.9: The number of atoms and average B-factors for each chain, as calculated using the program Baverage (CCP4).

Chain	Atoms	Average B-factor Value (Å^2)
A (sNT-Als1)	2273	25.07
B (Hepta-T)	50	35.98
C (sNT-Als1)	2281	25.00
D (Hepta-T)	50	37.79

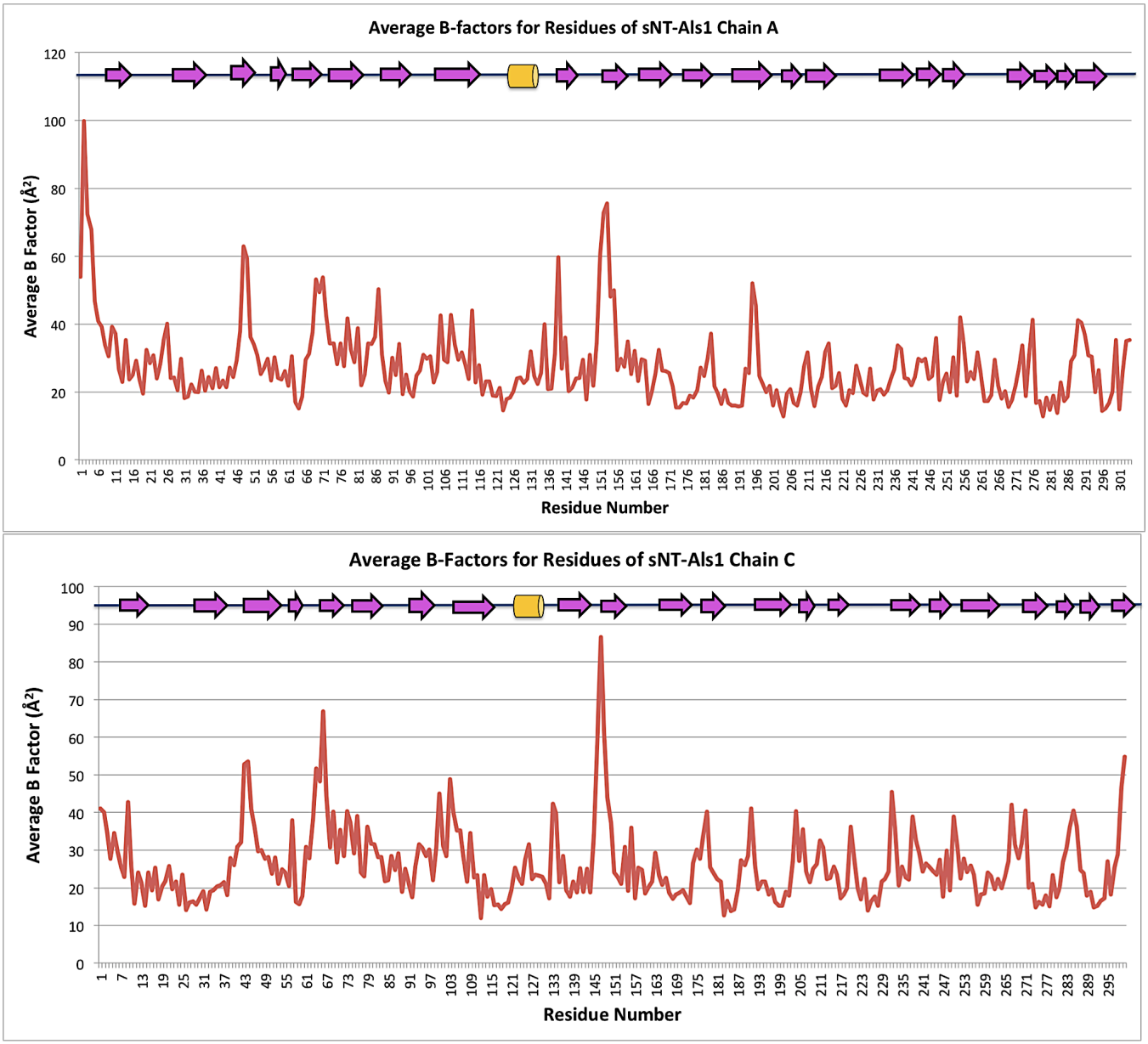


Figure 5.9: Charts for Average B-factor values calculated for each residue of sNT-Als1 molecules Chain A and Chain C.
The region responsible for the high B-factor values in both Chains A and C corresponds to residues N145, G146 and D147, which have high B-factor values. N145 has an average B-factor of 60.7Å2 in Chain A and 58.0 Å2 in Chain C, G146 has an average B-factor of 72.9 Å2 in Chain C and 86.6 Å2 in Chain C and D147 has an average B-factor of 75.6Å2 in Chain A and 61.0 Å2 in Chain C. Additionally, a histidine residue prior to K1 of sNT-Als1 Chain A, also present on an unstructured loop region, has a high B-factor of 99.8 Å2. Charts constructed in Excel, with average B-factor values calculated using the Baverage tool of CCP4 (Winn et al 2011).

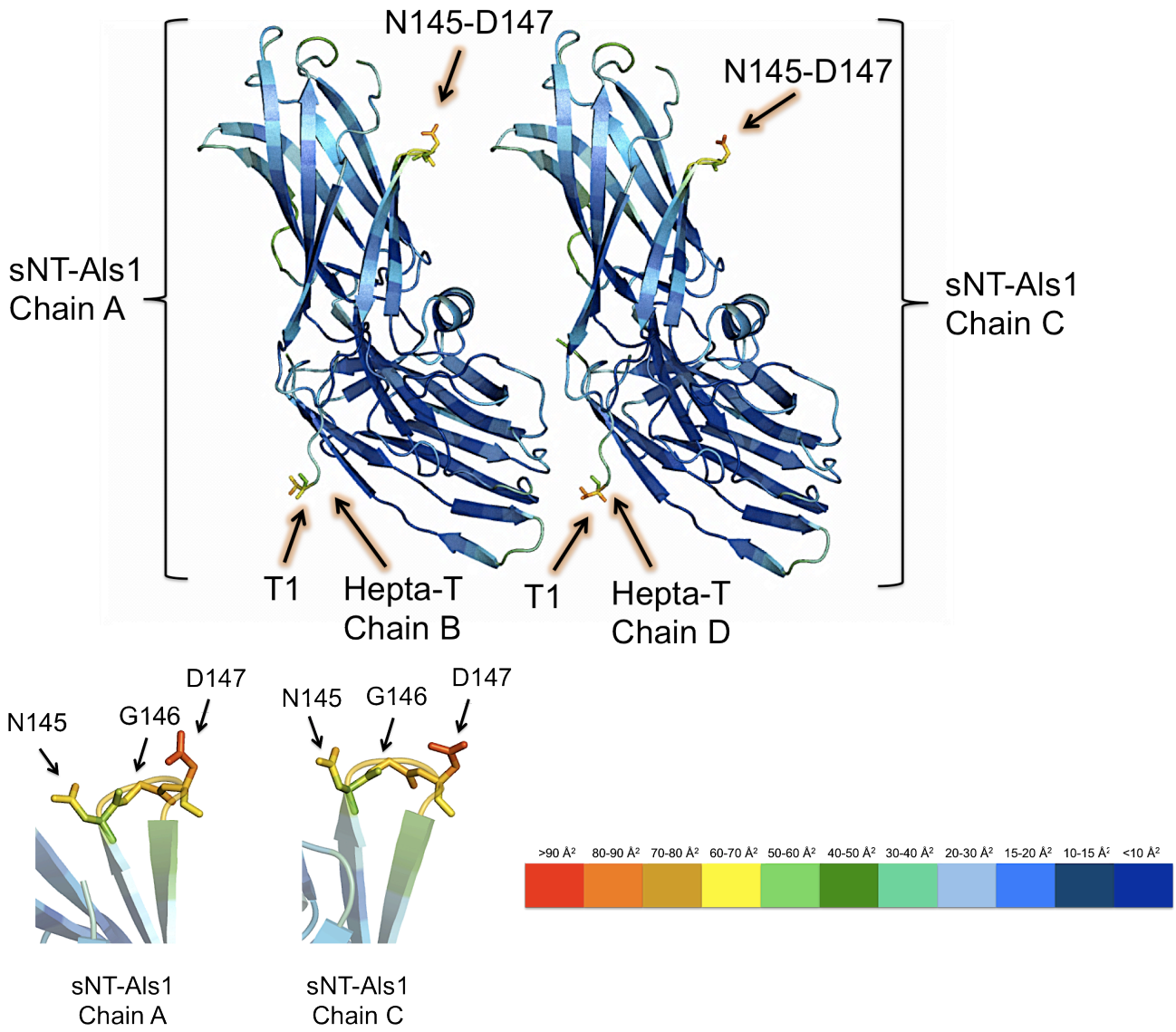


Figure 5.10: Locations of residues with high average B-factors on cartoon representations of sNT-Als1 molecules Chain A and Chain C.
The representations of sNT-Als1 molecule chains A and C have been coloured according to an average B-factor scale. Yellow-red correspond to high average B factors, dark green-dark blue corresponds to low average B-factors. The residues N145, G146 and D147 in both sNT-Als1 Chains A and C occur in an unstructured loop region of the protein, which may exhibit a higher degree of flexibility in comparison to other regions and could account for the high B-factors. Also shown in this picture is Hepta-T chains B and D, in which the T1 residue from both Hepta-T chains exhibits an overall moderately high average B-factor of 55.7Å^2 in Chain B and 51.3 Å^2 in Chain D, but has side chain atoms with B-factors ranging from 73-82 Å^2 in Chain B and from 76-86 Å^2 in Chain D.

Average overall B-factors for the residues of Hepta-T peptide chains B and D were 35.98 Å^2 and 37.79 Å^2, respectively (see Table 5.8). Average B-factors for individual residues of Hepta-T peptide chains B and D are shown in Figure 5.11. The average B-factors for chains B and D are higher than the average B-factors for sNT-Als1 molecule Chains A and C, reflecting the higher level of flexibility available for atoms within the Hepta-T peptide ligands. Threonines at position 1 of both Hepta-T ligand Chains B and D have the highest average B-factors from these chains (see Figure 5.11), with an average B-factor value of 55.7 Å^2 for Hepta-T Chain B and an average B-factor value of 51.3Å^2. Additionally, the B-factor values of T1 side-chain atoms range from 73-82 Å^2 in Chain B and 76-86 Å^2 in Chain D. The position of T1 is 'outside' of the PBC and so would possibly be less affected by the conformational restraint effects of peptide-ligand binding in comparison to Hepta-T T2-T7 residues.

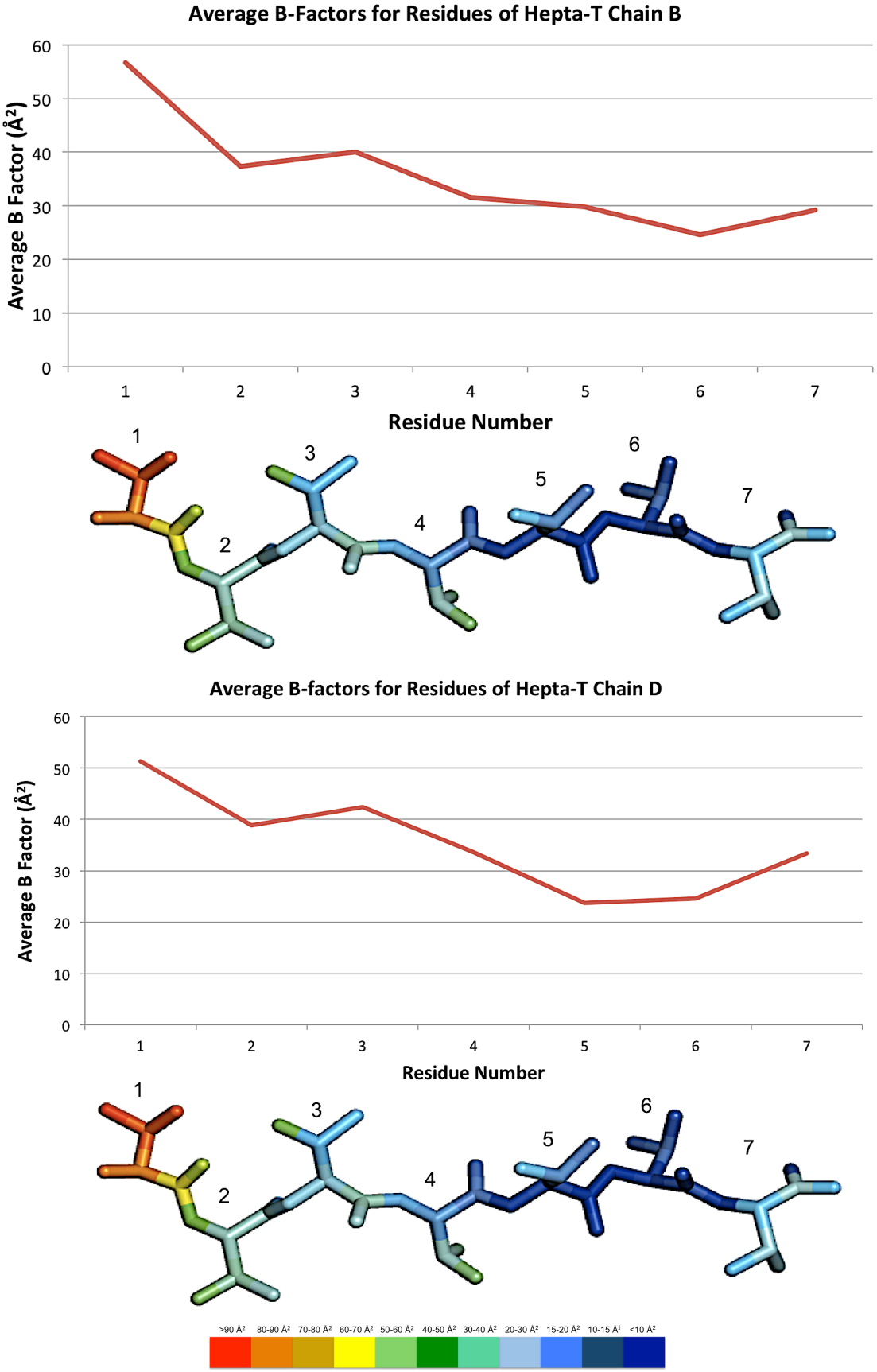


Figure 5.11: The Average B-factors for each residue of Hepta-T Peptide Chains B and D, alongside Stick Representations of the Hepta-T Peptide Chains, coloured according to a B-factor scale.
T1 in both Chains have the highest average B-factors than T2-T7; 55.7Å^2 in Chain B and 51.3 Å^2 in Chain D. Furthermore, the side-chain atoms of T1 have B-factors ranging from 73-82 Å^2 in Chain B and from 76-86 Å^2 in Chain D.

5.2. The Crystal Structure of SNT-Als1 with Hepta-T Peptide

5.2.1. Overall Structure and Topology of sNT-Als1

Similarly to the other N-terminal Als adhesins that have been solved (NT-Als9.2 and sNT-Als3), the overall sNT-Als1 protein structure has two DEv-type immunoglobulin-like domains (N1 and N2) and four conserved disulphide bonds (Salgado et al 2011, Lin et al 2014) (Fig 5.12). A similar fold is found in ClfA from *S. aureus* (Deivanayagam et al 2002) and SdrG from *S. epidermidis* (Ponnuraj et al 2003). Although these latter proteins lack the conserved disulphide bonds and specific ligand-binding cavity components of sNT-Als1, they contain an exposed cleft for peptide binding using a 'dock, lock, and latch' mechanism (Bowden et al 2008, Salgado et al 2011). The N1 domain contains 10 β-strands with one α-helix, whilst the N2 domain is comprised of just 12 β-strands with no α-helices (Fig 5.13 and Fig 5.14). The function of the alpha-helix α1 is currently unknown – whether it has a role in stabilizing the adhesion should perhaps be explored further.

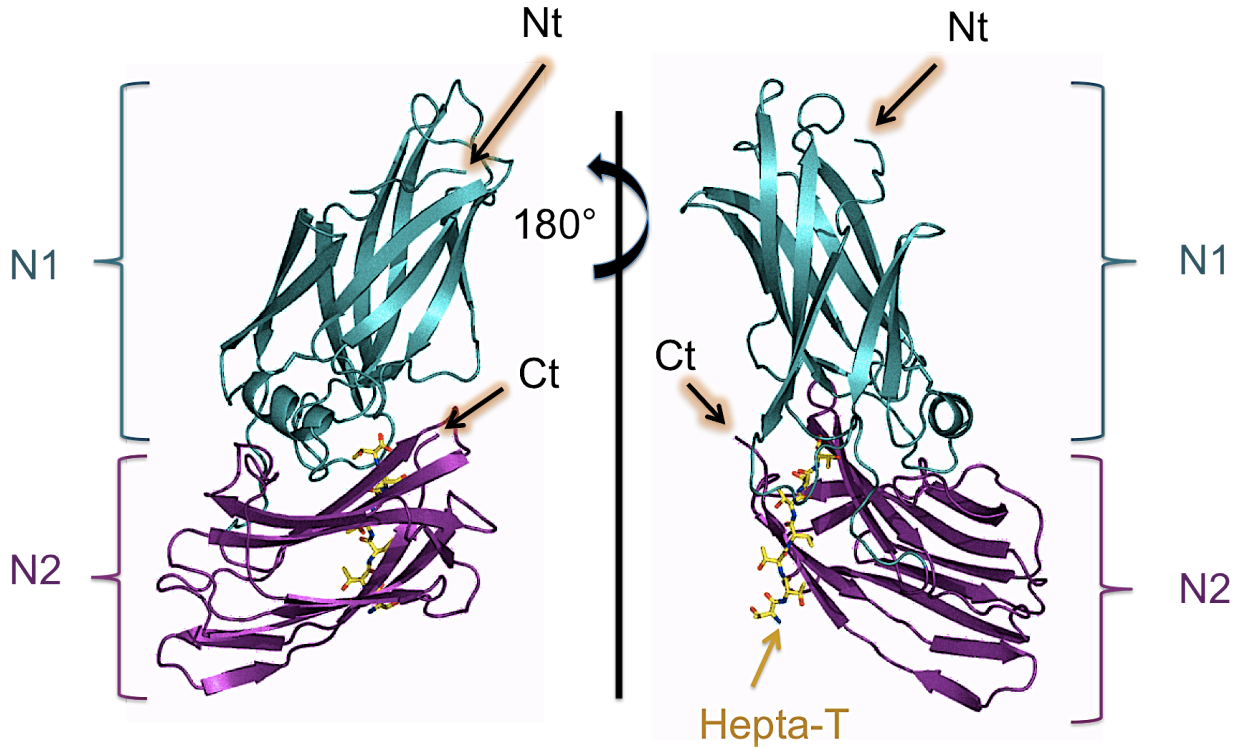


Figure 5.12: Simplified representation of the crystal structure of an sNT-Als1 molecule (Chain A) bound to the Hepta-T peptide ligand (Chain B), from two stereo views.
The N1 and N2 immunoglobulin domains of sNT-Als1 Chain A have been coloured in cyan and violet, respectively. Hepta-T peptide has been coloured yellow. N-terminus (Nt) and C-terminus (Ct) are indicated with black arrows. Figure constructed in PyMOL.

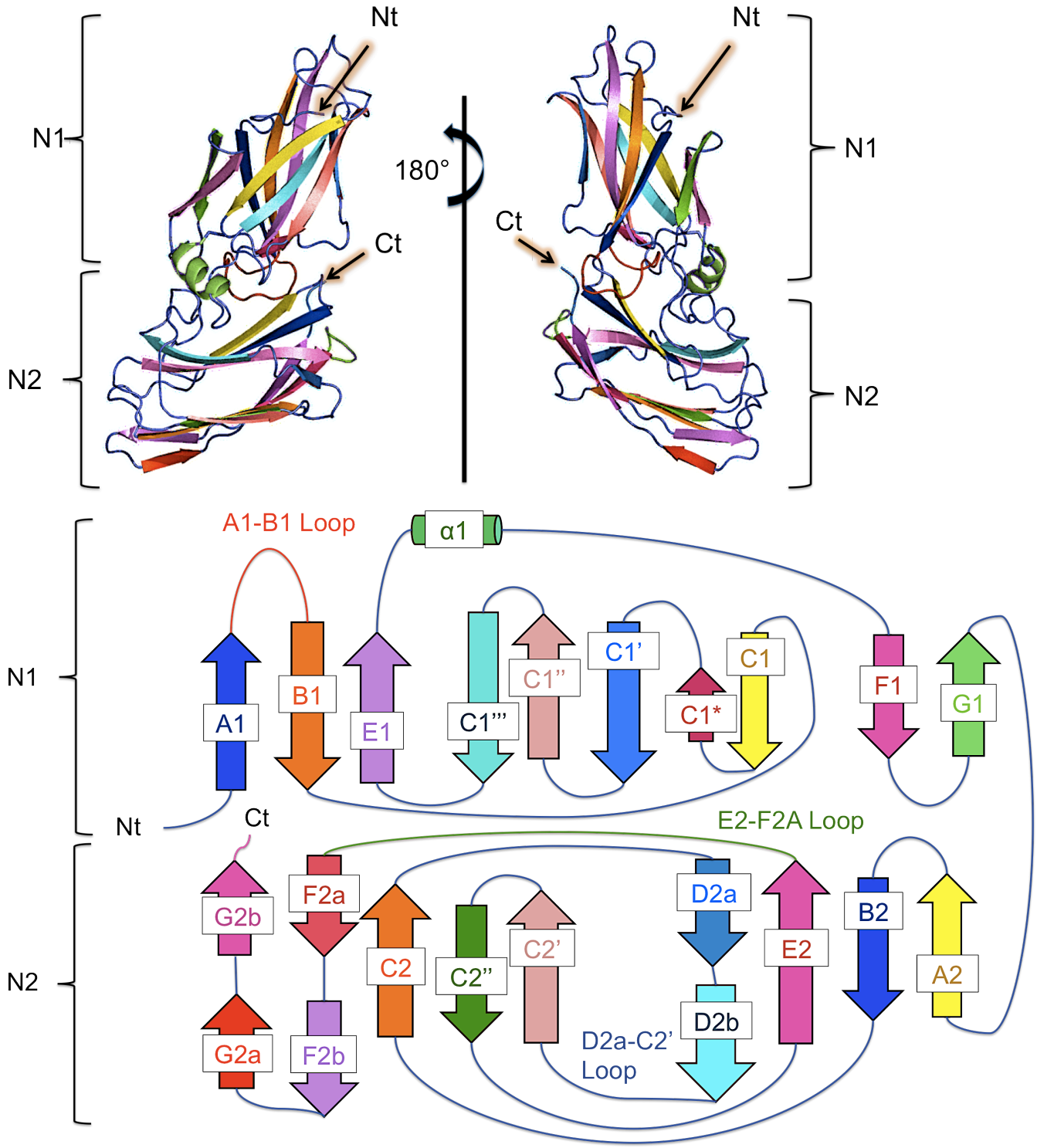


Figure 5.13: **Individually coloured secondary structure elements in an sNT-Als1 cartoon diagram from two stereo views and the labelling of secondary structural elements within a topology diagram**.
The individual secondary structure elements of the sNT-Als1 molecule (Chain A) cartoon diagram (Top) have been coloured in such a way that they match the topology diagram shown (Below). B-sheet secondary structure elements are represented as arrows, alpha helices are represented as cylinders, and the annotations of the secondary structure are according to those specified in the supplementary material of Salgado et al 2011. N1 and N2 domains have been labelled in both representations, and N-terminus (Nt) and C-terminus (Ct) have been labelled. In the topology diagram, the A1-B1 loop (red). E2-F2A loop (green) and D2a-C2' Loop (violet) have been labelled as these loops have residues important for the functioning of the sNT-Als1 PBC structure.

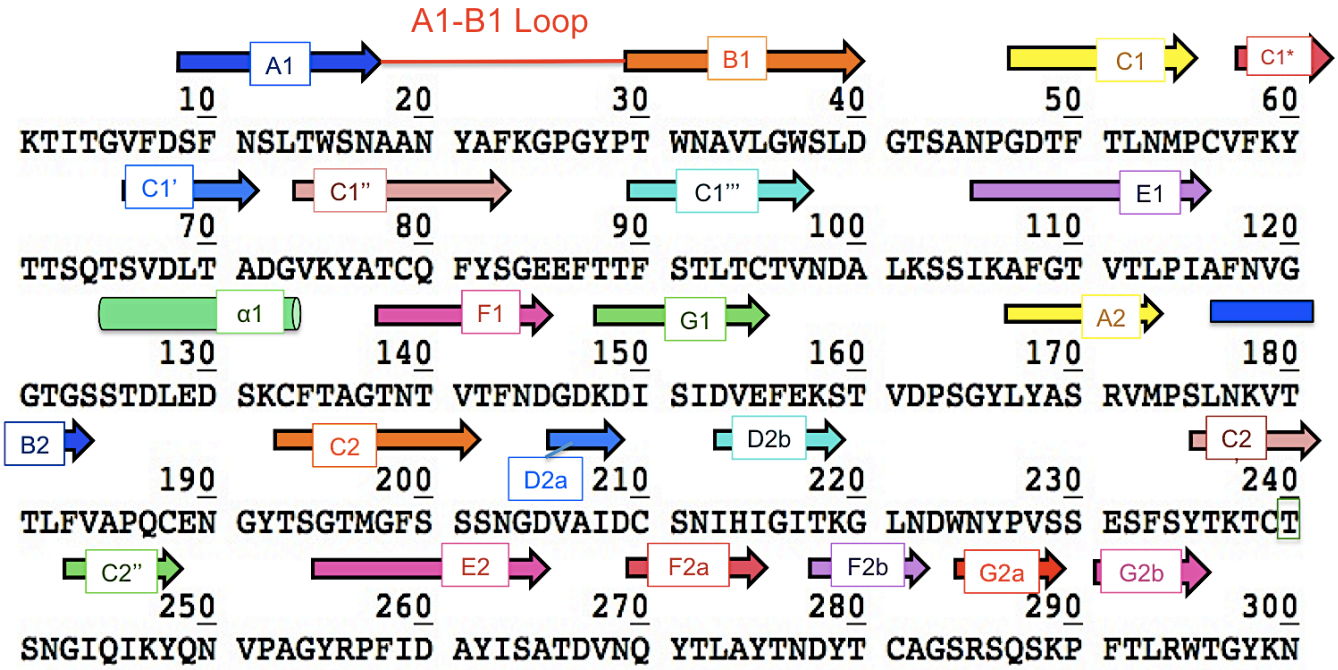


Figure 5.14: **The primary sequence of sNT-Als1 alongside the labelled secondary structure elements**. The amino acid sequence of sNT-Als1 is shown alongside labelled secondary structural elements, colour coded in a way that they match those shown in Figure 5.13. β-sheet elements are shown as arrows, the α-helix is shown as a cylinder. A1-B1 loop has been shown as a red line. The secondary structure elements have been labelled according to those from Salgado et al 2011.

Overall, positive and negative electrostatic charge is evenly distributed throughout the surface of the sNT-Als1 protein (see Figure 5.15). This is expected for a globular domain of an adhesion protein that resides in an aqueous extracellular environment.

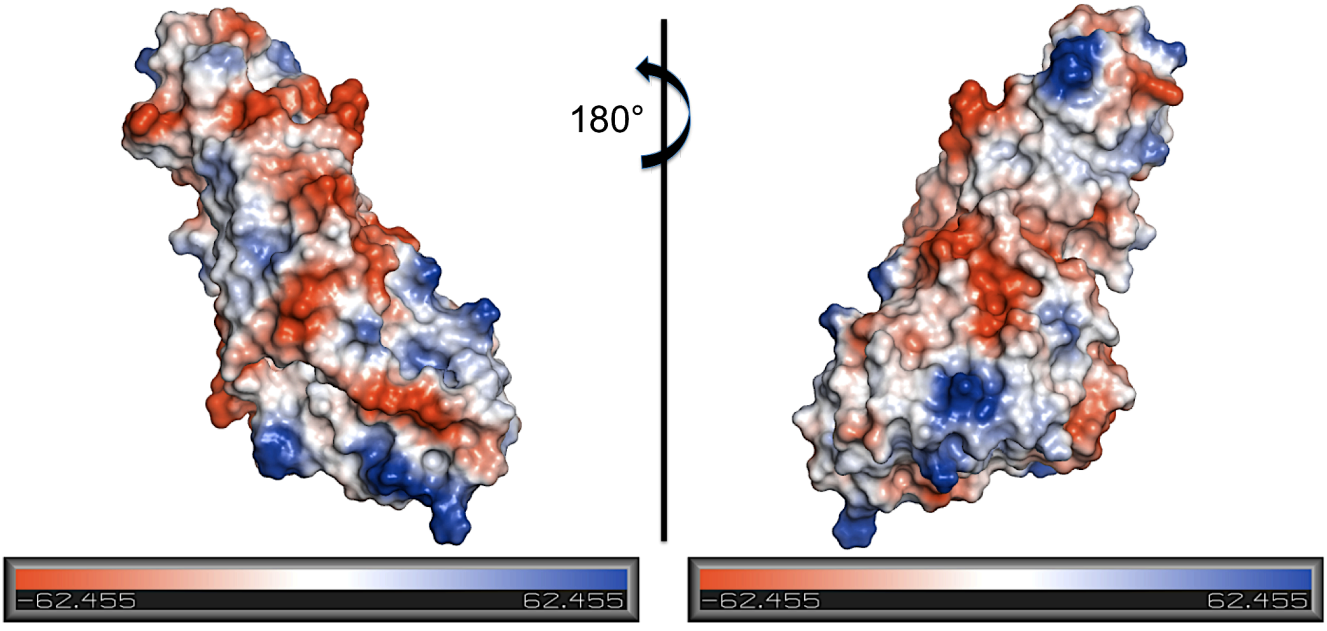


Figure 5.15: **The Overall Surface Electrostatic Profile of sNT-Als1 (Chain A) from Two Stereo Views**. Negative charge has been coloured red, neutral charge is colourless and positive charge coloured blue, according to the scales shown below. Figures constructed using the vacuum electrostatics tool in PyMOL.

5.2.2. Crystal Packing

The asymmetric unit is comprised of two sNT-Als1/Hepta-T molecules (see Figure 5.16), which pack together to form a crystal via a well-connected lattice. Although the sNT-Als1/Hepta-T crystal structure reveals a crystallographic dimer, it is important to note that sNT-Als1 is not dimeric in solution. It had been shown previously that NT-Als1 is monomeric in solution from NMR relaxation and SAXS experiments (Salgado et al 2011), therefore, the crystallographic dimer shown here is not biologically relevant.

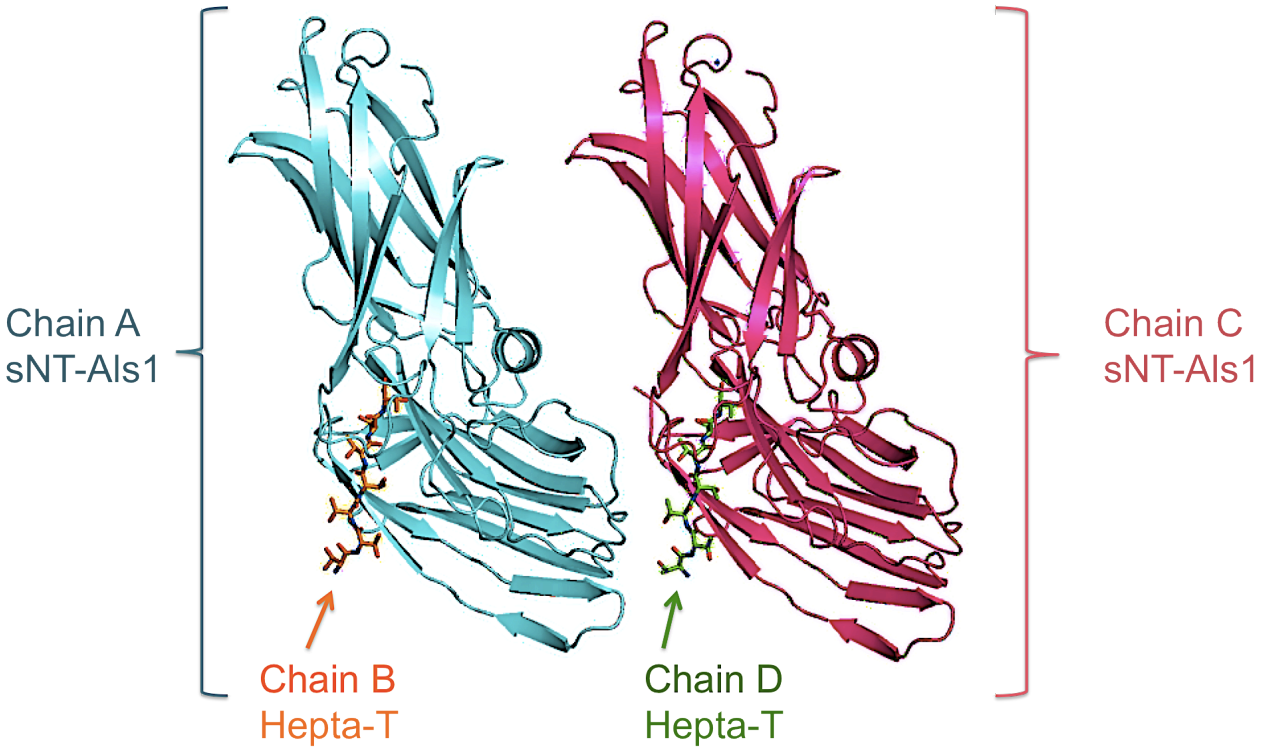


Figure 5.16: The Asymmetric Unit of the sNT-Als1/Hepta-T Crystal Structure.
The asymmetric unit of the sNT-Als1 is comprised of two sNT-Als1 molecules (Chain A, shown in cyan and Chain C, shown in magenta) bound to a Hepta-T peptide each (Chain B, shown in orange and Chain D, shown in green). Although a crystallographic dimer, it is not a biologically relevant dimer, as confirmed with previous SAXS and NMR relaxation experiments with NT-Als1 (Salgado et al 2011). Image made in PyMOL.

The orthorhombic unit cell contained four asymmetric units with eight sNT-Als1/Hepta-T molecules in total (see Figure 5.17), of the same colouring as from Figure 5.15. The representation in Figure 5.17 can also be thought of as a visual demonstration of how the asymmetric unit fits the crystallographic space-group $P2_12_12_1$ symmetry operations assigned to the sNT-Als1/Hepta-T crystal structure. The asymmetric units relate to each other via rotation and then translation operations for each unit cell axes, as is described by the $2_12_12_1$, or screw axes, notation. As the crystal lattice is described as a Bravais primitive lattice, the lattice points pertain to the four corners of the unit cell only.

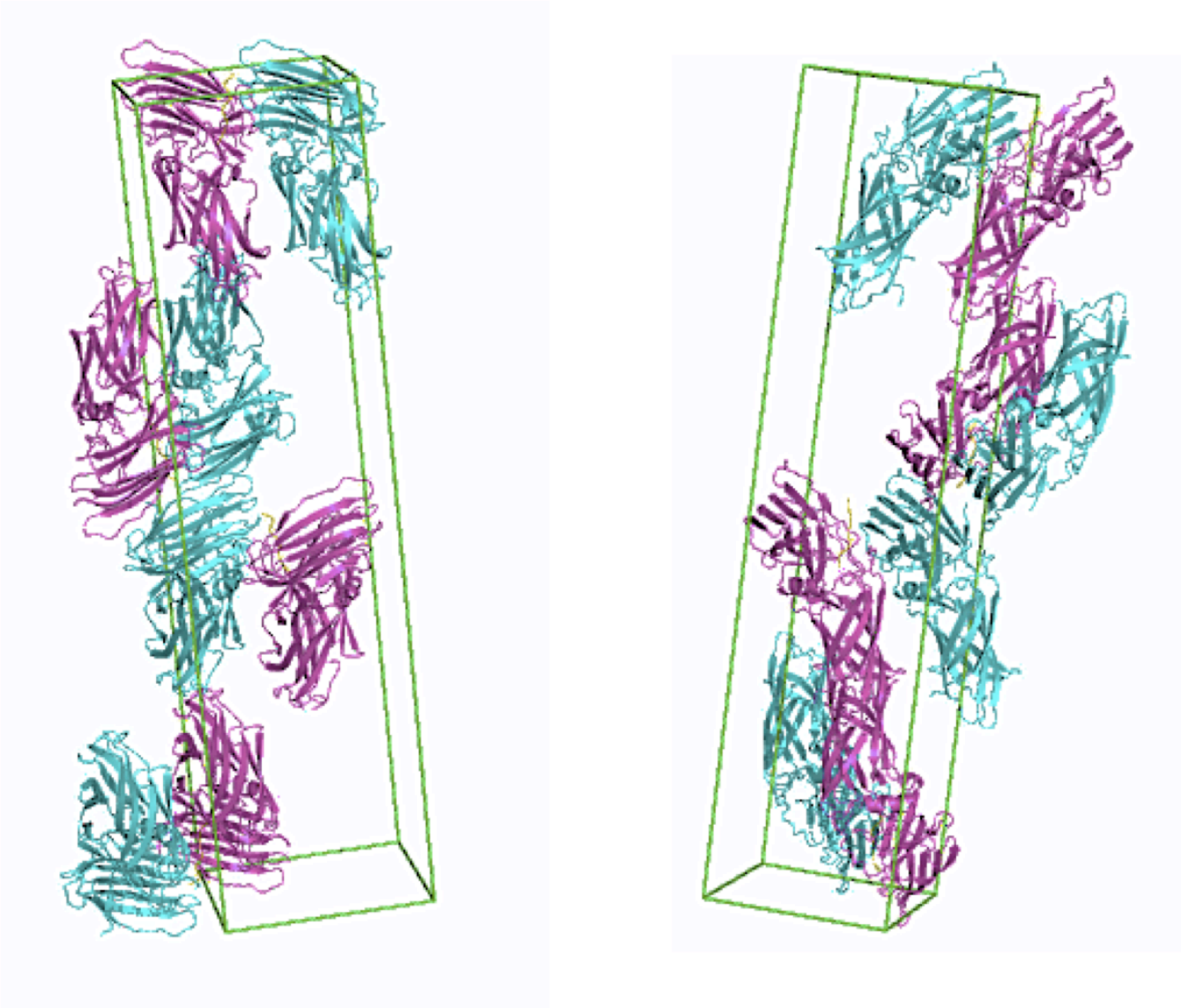


Figure 5.17: Crystal Packing Arrangement of sNT-Als1/Hepta-T asymmetric units within the Orthorhombic Unit Cell from Two Stereo Views.
The asymmetric unit, comprised of two sNT-Als1 molecules, Chain A (cyan) and Chain C (magenta), and Hepta-T peptides Chain B (orange) and Chain D (green) are shown with respect to the crystallographic orthorhombic unit cell (green oblong). Four asymmetric units are present, with eight sNT-Als1/Hepta-T molecules in total. The asymmetric units are related via $2_12_12_1$, or screw axes, symmetry operations, of a rotation followed by a translation. The unit cell shown is of primitive Bravais lattice type, as the four corners of the unit cell are the only lattice type. Figure constructed using Symmetry Mates function of PyMOL.

5.2.3. Crystal Contacts

It is important to highlight crystal contacts as these may help in solving the structures of sNT-Als1 in the apo form, and also other sNT-Als protein constructs in the future. The residues that contribute to crystal contacts between the two sNT-Als1/Hepta-T molecules, as shown using Ligplot+ (Lasowski & Swindells 2011) are displayed in Figure 5.18. Hydrogen-bond interactions are formed between K219 of Chain A with A19 of Chain C and between Q249 and N250 of Chain A with R294 of Chain C. Crystal contact interactions mediated via coordinated water molecules are between A253 of Chain A with N20 of Chain C and also Y235 of Chain A with E270 of Chain C.

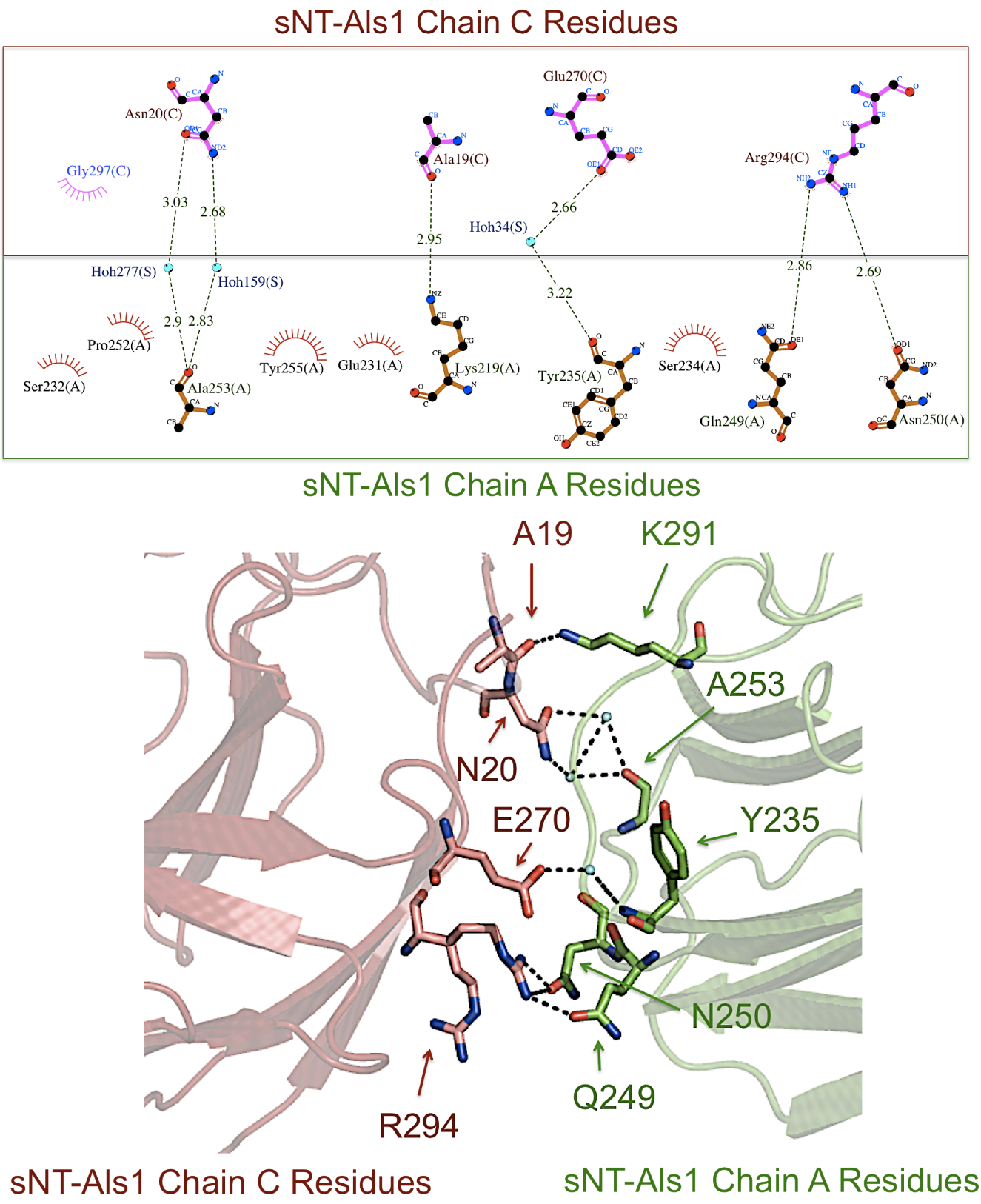


Figure 5.18: Ligplot+ and PyMOL Representations of Interactions between sNT-Als1 molecules Chain and Chain C that pinpoint potential crystallography contacts.
sNT-Als1 molecule Chain A residues are shown in the green box, sNT-Als1 Chain C residues are shown in the dark red box. These were predicted using the DIMPLOT function of Ligplot+ (Lasowski & Swindells 2011). Hydrogen-bonds with their distance measurements in Å are shown as dotted green lines in the Ligplot+ diagrams, and dashed black lines in the PyMOL figure. Water molecules are depicted as cyan circles in both figures. These figures show that residues K219 from Chain A form a direct hydrogen-bond with A19 from Chain C, and Q249 and N250 from Chain A form direct hydrogen-bonds with R294 from Chain C. A253 of Chain A interacts via hydrogen bonds with water molecules with N20 of Chain C, and the same for Y235 of Chain A with E270 of Chain C.

When mapping the interactions of sNT-Als1 residues with the Hepta-T ligand using Ligplot+, an interaction was seen between the side-chain amine group of N250 from sNT-Als1 Chain A with the side-chain hydroxyl group of T1 of the Hepta-T peptide (Chain D) that is bound to the PBC of sNT-Als1 Chain C (Fig 5.19). On speculation, this is a potential explanation for why the inclusion of the Hepta-T peptide allowed for the crystallisation of sNT-Als1/Hepta-T rather than sNT-Als1 on its own, by allowing for an extra crystal contact interaction.

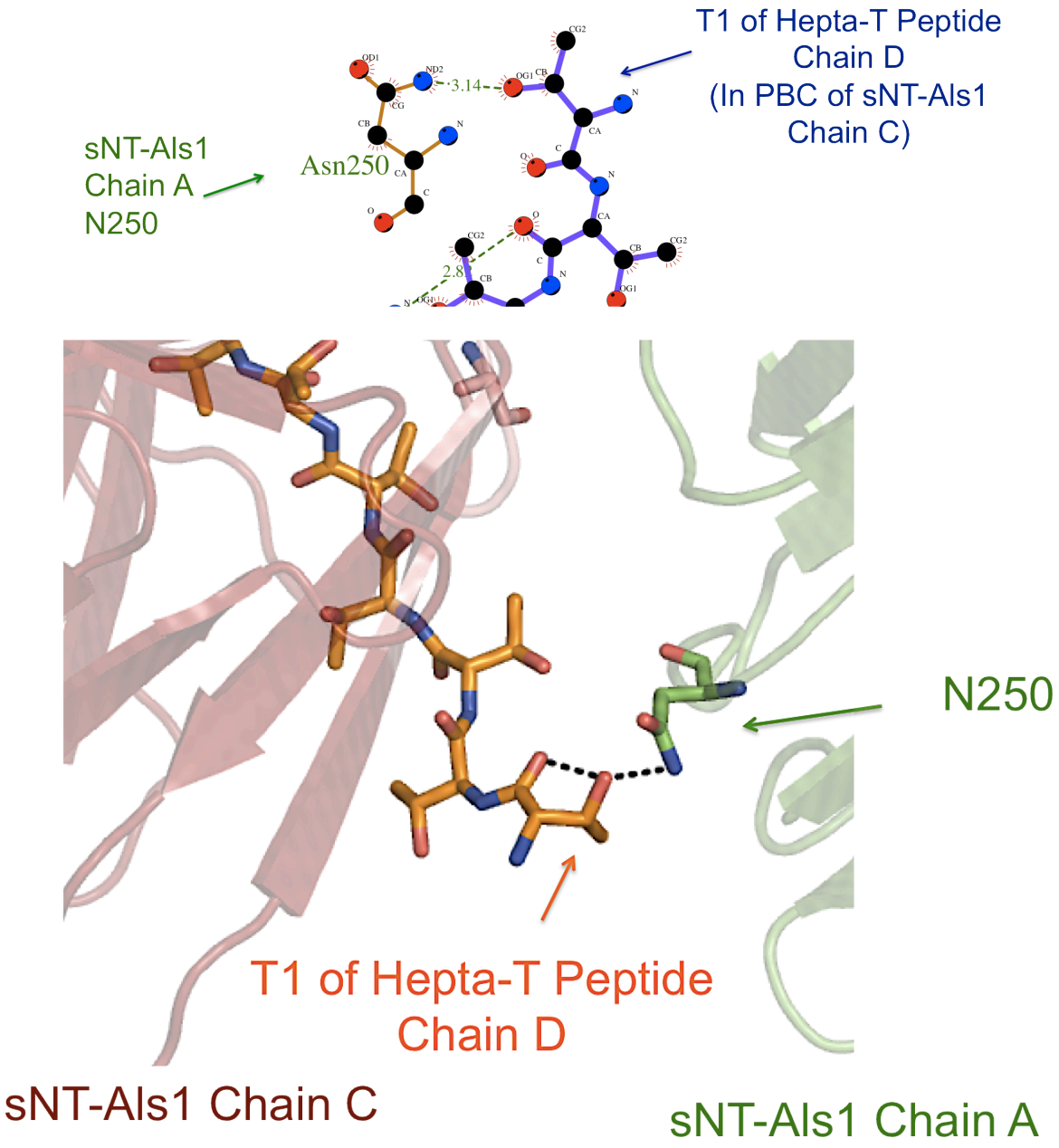


Figure 5.19: **Hydrogen-bond crystal contact between N250 of sNT-Als1 Chain A and T1 of the Hepta-T Peptide in the sNT-Als1 Chain C PBC, predicted with Ligplot+ and Depicted in PyMOL**.
Shown in the Ligplot+ diagram (Lasowski & Swindells 2011) is the 3.14Å length hydrogen-bond (green dashed line in Ligplot+ diagram, black dashed line in the PyMOL figure) between the side chain amine group of N250 of sNT-Als1 molecule Chain A and the hydroxyl group of T1 of the Hepta-T ligand within the PBC of sNT-Als1.

5.2.4. The Peptide Binding Cavity (PBC)

The sNT-Als1 PBC is a cavity formed by residues from both N1 and N2 immunoglobulin domain components. The A1-B1 loop, plus the C1* strand, form the 'cap' of the PBC (Fig 5.20). The C1* strand contains the conserved K59 residue that interacts with the Hepta-T peptide via a salt bridge, with the C-terminal carboxylic acids of peptide ligands, and is also seen in sNT-Als3 and NT-Als9 crystal structures (Salgado et al 2011, Lin et al 2014). The A2 and G2b strands form the walls of the PBC, whilst residues of the F2a strand, D2b-C2' loop and E2-F2a loop complete the PBC structure. Lengthwise, the binding pocket is able to accommodate a peptide sequence of up to six amino acid residues.

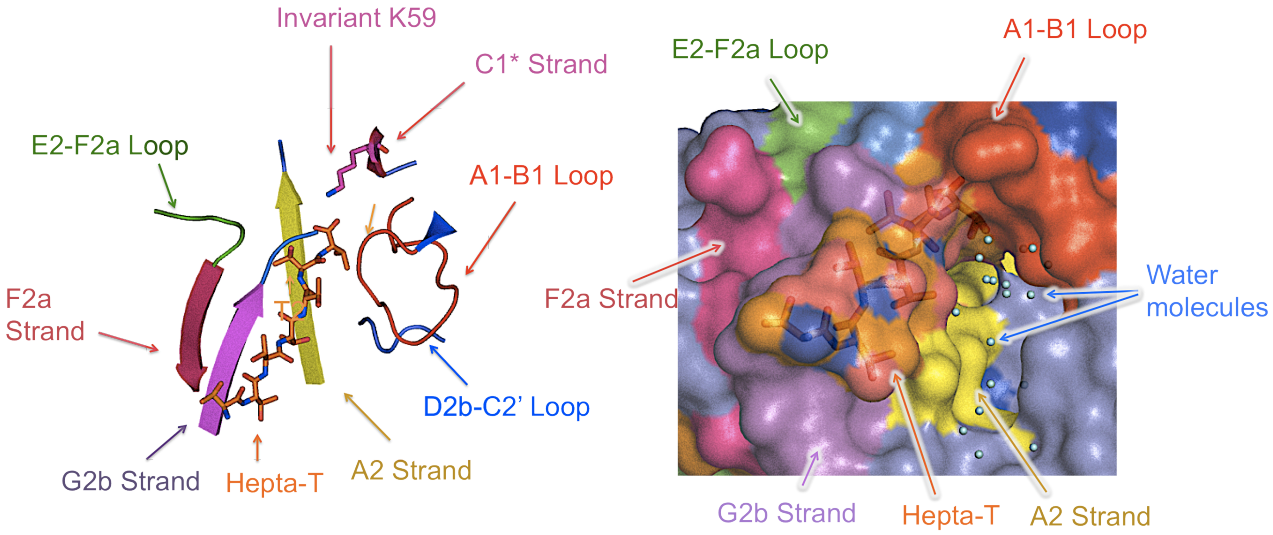


Figure 5.20: Cartoon and Surface Representations sNT-Als1 PBC with the Hepta-T peptide.
Cartoon secondary structural elements shown are the A1-B1 loop (coloured red), the C1* strand with invariant K59 (magenta), A2 strand (yellow), E2a-F2a loop (green), F2a strand (magenta) and G2b strand (pink). Hepta-T peptide is shown as an orange-coloured stick representation in the cartoon figure, and its slightly transparent surface representation is shown in the Surface figure. The colours of the cartoon representations match those from the surface representation beside it. Water molecules are shown as small cyan circles. Images constructed using PyMOL.

The approximate volume of the sNT-Als1 PBC (without Hepta-T peptide ligand) as calculated by the Schrodinger Sitemap program is 448.99Å^3 (see Figure 5.21). This program fills the PBC with a virtual 'grid' that is specified to occupy sites that are relatively shielded from solvent (Halgren 2007, Halgren 2009). The Sitemap program also calculated that for the sNT-Als1 PBC, its relative hydrophobic character had a value of 0.373, and its relative hydrophilic character had a value of 0.631. The electrostatic charge map of the PBC, plus its volume, hydrophobicity and hydrophilicity was calculated using Sitemap after preparation of the protein structure model via the addition of virtual hydrogen atoms.

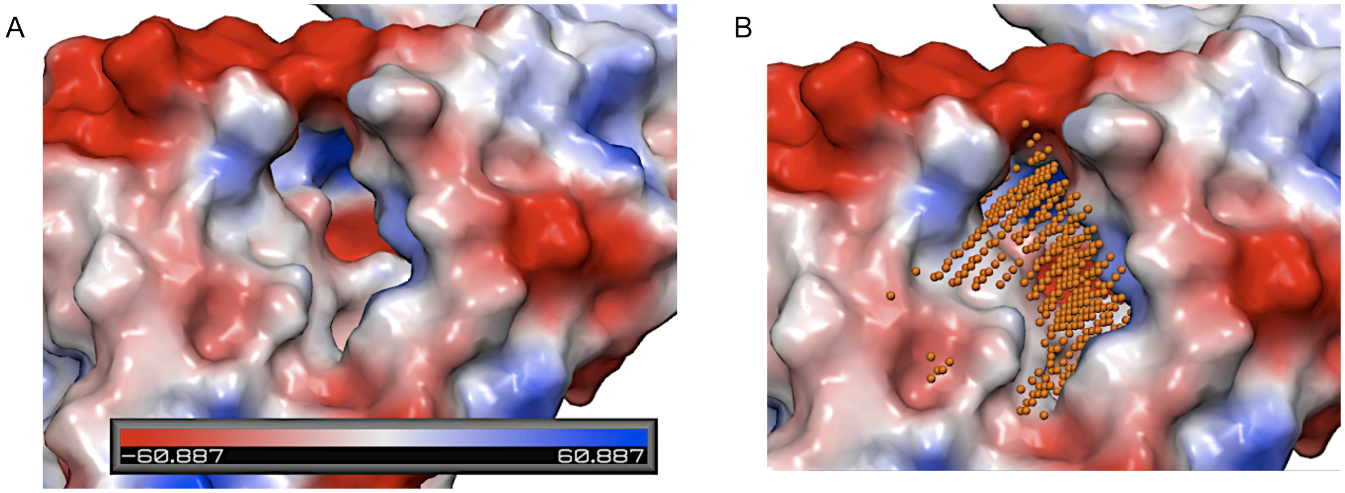


Figure 5.21: Electrostatic profile of the sNT-Als1 PBC without Hepta-T and Representations of the virtual 'grid' used to estimate PBC volume.
A) The electrostatic profile of sNT-Als1 PBC was calculated using PyMOL. B) Schrodinger Sitemap (Halgren 2007, Halgren 2009) was used to fill the sNT-Als1 volume with a virtual 'Grid' in order to estimate a relative volume of 448.99Å^3 for the PBC.

5.2.5. Interactions of sNT-Als1 PBC with Hepta-T

The PBC's of sNT-Als1 Chain A and Chain C were analysed using Ligplot+ in order to decipher the residues that were involved in the binding of sNT-Als1 to Hepta-T. Three types of interactions are seen between the sNT-Als1 PBC residues from both chains and the Hepta-T peptide; firstly there are direct hydrogen-bonds (H-bonds), secondly there are hydrophobic interactions and thirdly, hydrogen-bonds mediated by the network of water molecules. As the complete PBC-Hepta-T interaction profile is complex, each of the three types will be discussed separately, and PyMOL and Ligplot+ graphics here will use sNT-Als1 Chain A due to ease of viewing the Hepta-T/PBC interaction without any influence from crystal contacts. For ease of relating the Hepta-T peptide to the structures of Als protein PBCs, the numbering of the Hepta-T residues will be in reverse. From now on, T1 will refer to the carboxyl-terminal end of the peptide that contacts K59, and T2-27 follow this rule.

5.2.5.1. Residues that Interact Directly with the Hepta-T Peptide

Residues of the sNT-Als1 PBC 'cap' that are important for directly contacting the Hepta-T peptide via H-bonds are the side chains of Y21 on the A1-B1 loop and K59 on the C1* strand. The side chain amine group of K59 contacts the C-terminal carboxylate of the Hepta-T peptide (residue T1) via a salt bridge, whilst the side chain hydroxyl group of Y21 forms a direct H-bond with the carbonyl oxygen of T2 of the Hepta-T peptide (Fig 5.22 and Fig 5.23).

From the A2 strand of the sNT-Als1 PBC 'wall', the backbone carbonyl oxygen from residue S170 is able to form direct H-bond interactions with the backbone amide and also the side-chain hydroxyl group of T1 of the Hepta-T peptide. The backbone amide of residue V172, also from the A2 strand, forms H-bond interactions with the C-terminal carboxylate of Hepta-T T1.

On the G2b strand, the other 'wall' of the sNT-Als1 PBC, the backbone carbonyl oxygen of residue T296 forms a H-bond interaction with the backbone amide of the Hepta-T T2 residue, plus the backbone amide of T296 can form another H-bond interaction with the backbone carbonyl oxygen of T4 of the Hepta-T peptide. Also from the G2b strand, the backbone carbonyl oxygen of R294 forms a direct H-bond interaction with the backbone amide nitrogen of T4, plus the amide nitrogen of R294 can form a H-bond interaction with the carbonyl oxygen of T6 from the Hepta-T

peptide. These interactions between the G2b strand and the Hepta-T are characteristic of hydrogen-bond interactions between β-sheets.

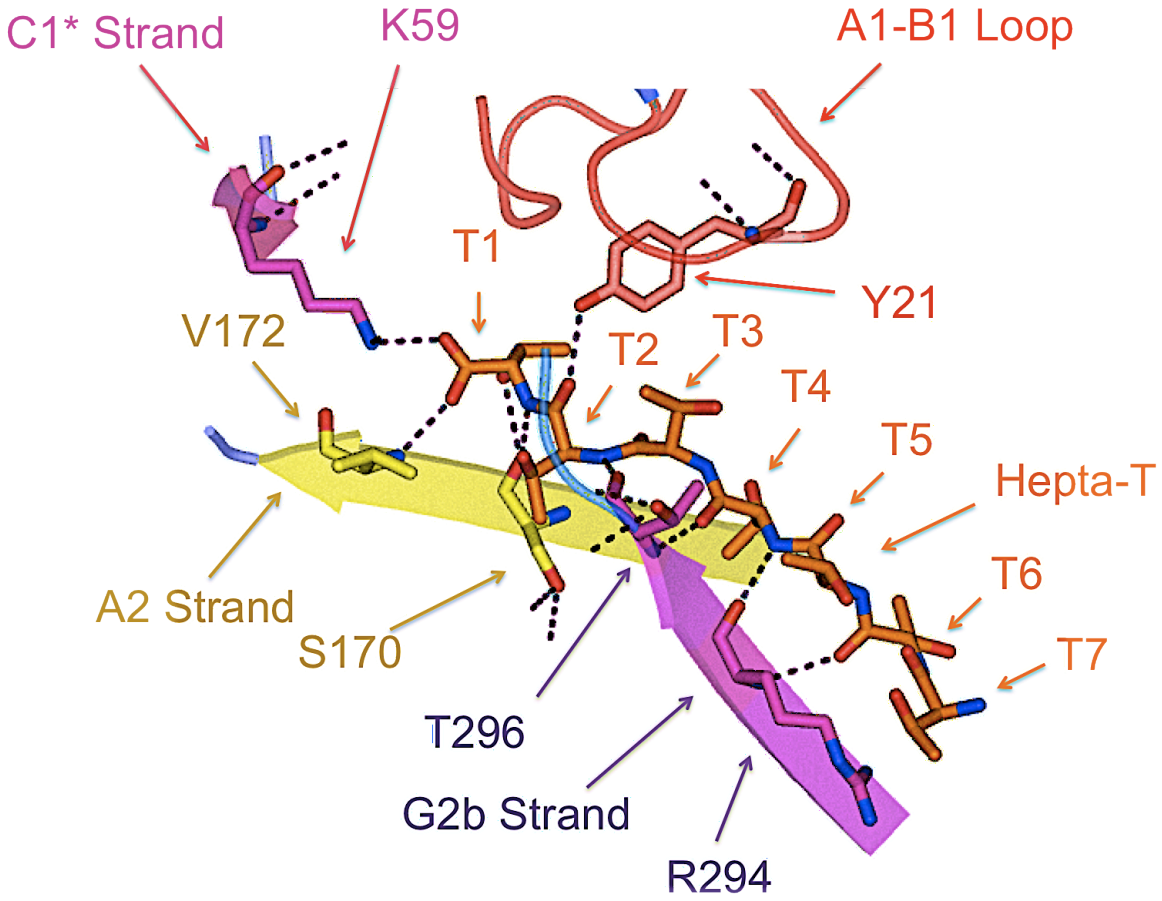


Figure 5.22: Cartoon and Stick Figure of the sNT-Als1 PBC residues that make direct Hydrogen-Bond Contacts with the Hepta-T Ligand.
Oxygen and nitrogen atoms are shown as blue and red, respectively, on the residues shown as sticks. Hydrogen bonds are shown as black dashed lines. The Hepta-T ligand backbone is coloured orange. The amine group of K59 on the C1* strand, coloured magenta, is shown forming a hydrogen-bond salt-bridge with the C-terminal carboxylate of the Hepta-T. The A1-B1 loop – the 'cap' of the PBC, is coloured red, as is the Y21 residue whose side chain hydroxyl group forms a direct H-bond with the carbonyl oxygen of Hepta-T T2. S170 and V172 of the A2 strand are coloured yellow; the backbone carbonyl oxygen of residue S170 forms direct H-bond interactions with the backbone amide and also side-chain hydroxyl group of the Hepta-T T1, and V172 forms H-bond interactions with the C-terminal carboxylate of Hepta-T T1. Of the G2b strand (coloured pink), the backbone carbonyl oxygen of residue T296 forms a H-bond interaction with the backbone amide of the Hepta-T T2 residue, plus the backbone amide of T296 forms a H-bond interaction with the backbone carbonyl oxygen of T4 of the Hepta-T peptide. R294 also forms a H-bond interaction with the backbone amide nitrogen of Hepta-T T4, plus the amide nitrogen of R294 forms a H-bond with the carbonyl oxygen of Hepta-T T6. Figure constructed in PyMOL.

5.2.5.2. Hydrophobic and Water Molecule Network Interactions of sNT-Als1 PBC Residues

Residues of the sNT-Als1 that provide hydrophobic interactions between the PBC and sNT-Als1 include residues A19, Y28 and P29 of the A1-B1 Loop, R171 and the Y168 aromatic ring of the A2 strand, Y271 of the F2a strand and L293, W295, G297 and Y298 of the G2b strand, according to the Ligplot+ representation seen in Figure 5.23. As incomplete electron density was seen for Y298 in sNT-Als1 Chain A, the side chain is not represented.

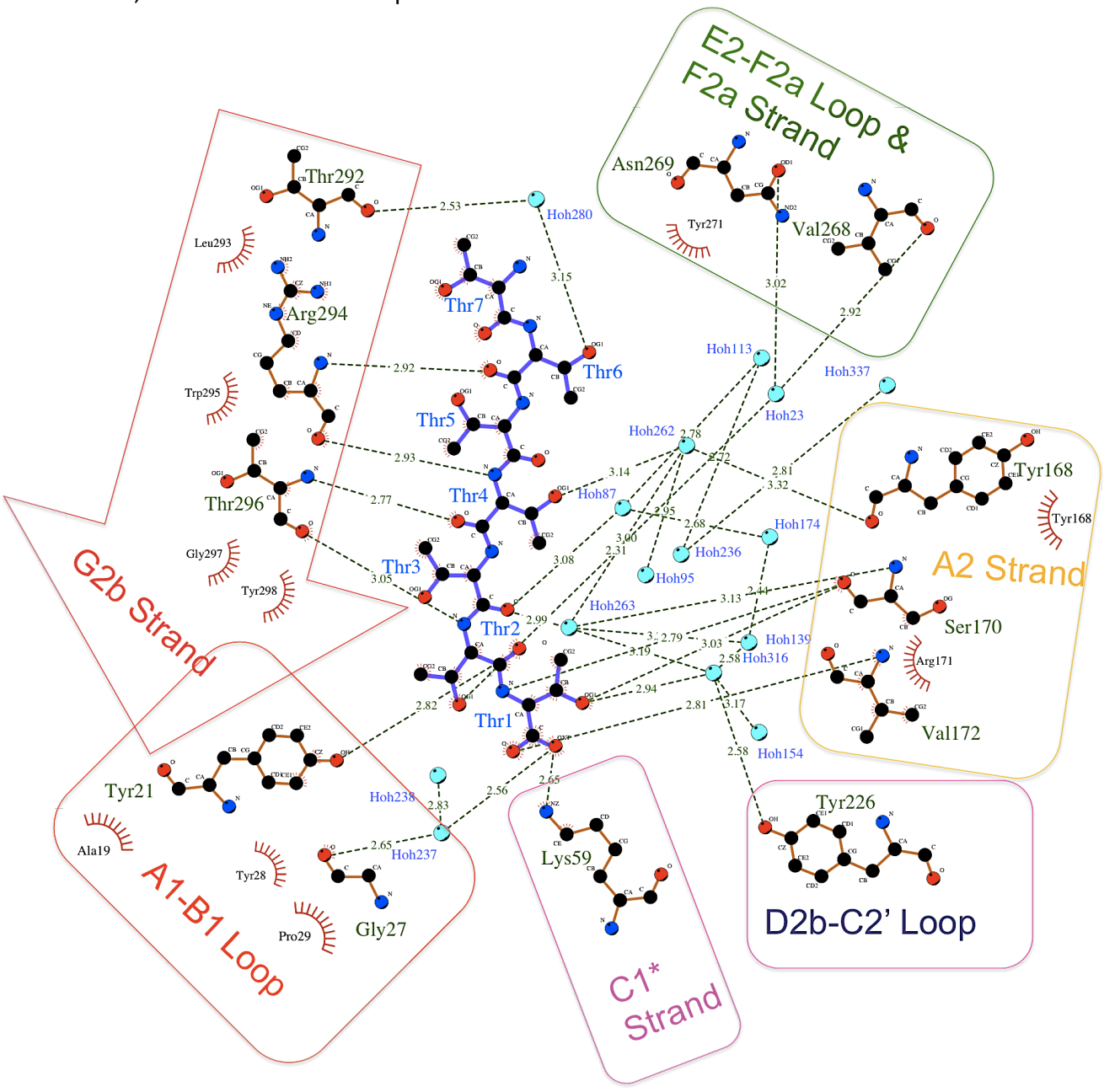


Figure 5.23: **Residues of the sNT-Als1 PBC that interact with the Hepta-T peptide.**
Hepta-T residue positions (indicated with blue labels) are shown as 1-7 counting from the C-terminus. The PBC residues of sNT-Als1 interact with Hepta-T directly or indirectly through the water molecules (cyan circles) via H-bonds (dotted green lines). 15 water molecules are present in the PBC. The secondary structure locations of the residues are also shown; for instance K59 on the C1+ strand, A19, Y21, Y28, G27 and P29 of the A1-B1 loop, Y168, S170, R171 and V127 of the A2 Strand, Y226 of the D2b-C2' Loop, V268, N269 and Y271 of the E2-F2a Loop and F2a strand and T292, L293, R294, W295, T296, G297 and Y298 of the G2b Strand. Residues surrounded by red, dashed semi-circles contribute hydrophobic interactions. Figure made using Ligplot+ (Lasowski & Swindells 2011).

There are 15 water molecules present in the PBCs of both sNT-Als1 Chain A and Chain C, which form a H-bonding network between PBC and Hepta-T residues. As can be viewed from the Ligplot+ representation in Figure 5.23, and the PyMOL picture in Figure 5.24, PBC chemical groups that interact with the Hepta-T peptide via this network, include the carbonyl oxygen from G27 of the A1-B1 loop which interacts via a water molecule to the Hepta-T carboxylate hydroxyl group. From the A2 strand, the backbone amide of S170 interacts via a water molecule with the backbone carbonyl oxygen of Hepta-T T2, and the backbone carbonyl oxygen of Y168, which interacts similarly with the hydroxyl group of the Hepta-T T4 side-chain. On the D2b-C2' Loop, Y226 interacts via its side-chain hydroxyl group, to the backbone carbonyl oxygen of Hepta-T T2 via a water molecule. On the E2-F2a Loop, the backbone carbonyl oxygen of V268 and side-chain hydroxyl of N259 interacts via the same water molecule, with the Hepta-T T2 side-chain hydroxyl group. Finally, the carbonyl oxygen of residue T292 of the G2b strand can interact via a water molecule with the side chain hydroxyl group of Hepta-T T6. Although 15 water molecules are present in the PBC, six individual water molecules mediate sNT-Als1 PBC-Hepta-T hydrogen-bond connections.

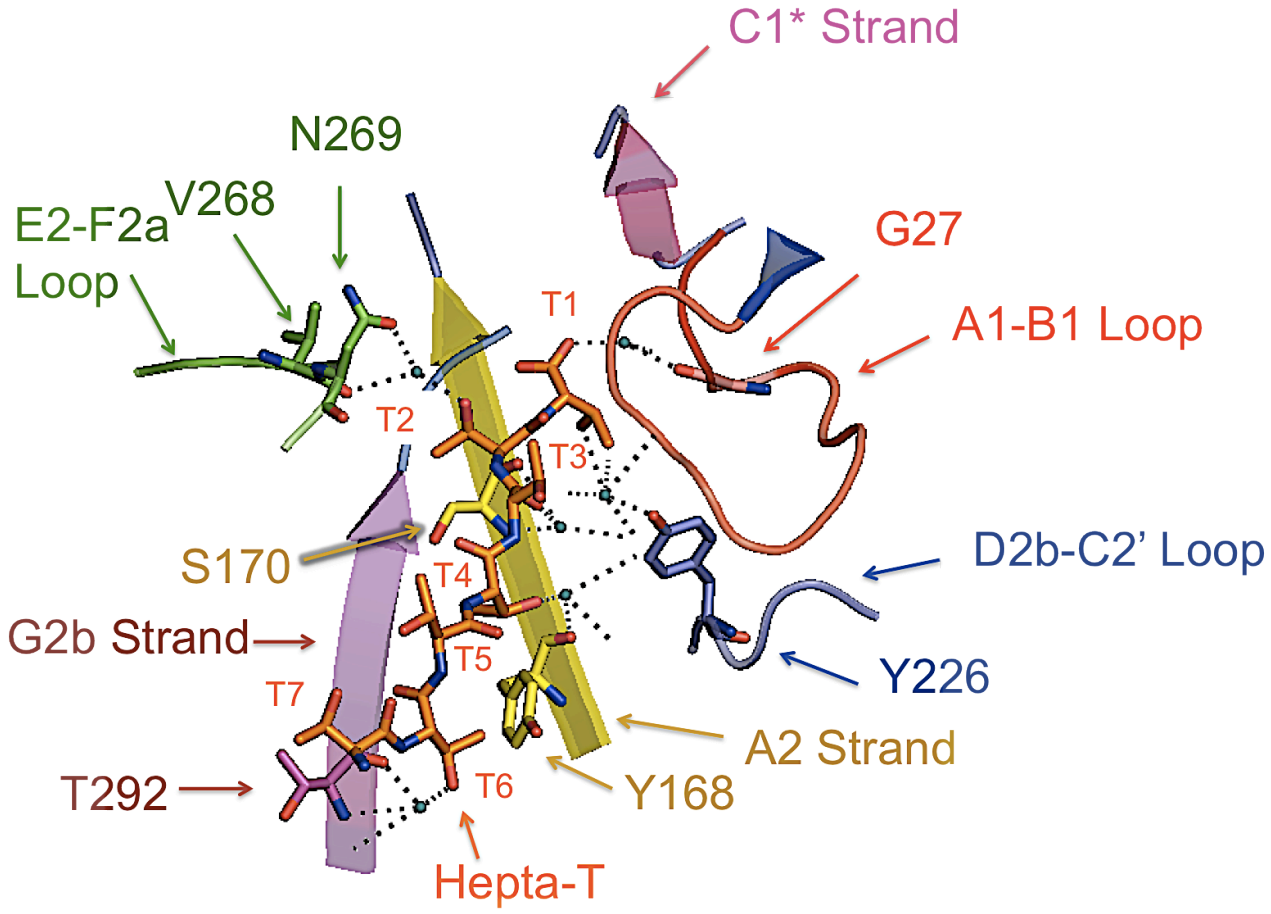


Figure 5.24: Cartoon and Stick Figure of sNT-Als1 PBC residues that interact with the Hepta-T peptide through water-molecule-mediated H-bonds. Oxygen and nitrogen atoms are shown as blue and red, respectively, on the residues shown as sticks. Hydrogen bonds are shown as black dashed lines. Water molecules are shown as blue circles. Six individual water molecules mediate hydrogen-bond interactions between the sNT-Als1 PBC residues and the Hepta-T peptide (orange-coloured backbone). These include the interaction between the carbonyl oxygen from G27 of the A1-B1 loop (coloured red) with the backbone carbonyl of Hepta-T T2. Also, the backbone amide of S170 on the A2 strand (coloured yellow) interacts similarly with the backbone carbonyl oxygen of Hepta-T T2, and the backbone carbonyl of Y168 interacts via a water molecule with the hydroxyl group of T4 of Hepta-T. The side-chain hydroxyl group of Y226 from the D2b-C2' loop (coloured violet) interacts via a water molecule with the backbone carbonyl oxygen of Hepta-T T2. The backbone carbonyl oxygen of V268 and the side-chain hydroxyl of N259 of the E2-F2a loop (coloured green) interact via the same water molecule, with the side-chain hydroxyl of T2 of Hepta-T. On the G2b strand (coloured pink), the carbonyl oxygen of T292 interacts via a water molecule with the side-chain hydroxyl of Hepta-T T6. Figure constructed with PyMOL.

To conclude, the co-crystal structure of the sNT-Als1/Hepta-T protein-peptide complex was solved to a resolution of 2.1Å. This structure was able to provide information regarding the PBC residues important for either directly contacting the Hepta-T peptide via H-bonds, via coordinated water-molecules or through hydrophobic interactions. These residues can be compared with Hepta-T-contacting residues of the sNT-Als3 PBC (Lin et al 2014), and those of the NT-Als9.2 PBC that contact the fibrinogen-γ C-terminal peptide (Salgado et al 2011). Plus, analysis of sNT-Als1 Chains A and C of the asymmetric unit is highly useful for extracting crystal-contact information, with which should be useful for solving structures of other sNT-Als constructs in the future. The structural sNT-Als1/Hepta-T model, and prior crystallisation optimisation conditions (noted in Chapter 4) should also be a valuable tool for the co-crystallisation of sNT-Als1 with small-compound ligands that have potential for development into therapeutic drug molecules against the adhesion of Als1 and Als3 *C. albicans* proteins, to medical devices and host cells.

5.3. The PBC Residue Differences of sNT-Als1 with sNT-Als3

An important use for the solved crystal structure of sNT-Als1 with the Hepta-T peptide is that it can be used to compare the interactions of the PBC with the peptide with those of the crystal structure of the sNT-Als3/Hepta-T complex solved by Lin et al in 2014. As sNT-Als1 has a residue sequence identity of approximately 82% with sNT-Als3, and also both could be co-crystallised with Hepta-T, this suggests that they possibly interact with Hepta-T, and potentially a peptide-like drug molecule, in a similar way. However, it is apparent that Als1 and Als3 have subtle differences in ligand binding specificity, for instance Als3 has been shown to interact with ferritin whilst Als1 has shown not to bind. Furthermore, their PBC's differ in terms of five amino acids (see Table 5.9), which brings the question of whether this small collection of differing amino acids are responsible for specificity differences between the two adhesion proteins.

The structural data indicate that there is a close overlap in the backbone conformations of sNT-Als1 and sNT-Als3. Alignment of sNT-Als1 molecule Chain A crystal structure with the crystal structure of sNT-Als3 shows an RMS value of 0.486Å (250 to 250 atoms aligned) (Fig 5.25). This high structural similarity suggests that it is the side chains of five differing amino acids in the PBC that may determine ligand binding specificity differences between the two proteins (See Table 5.9).

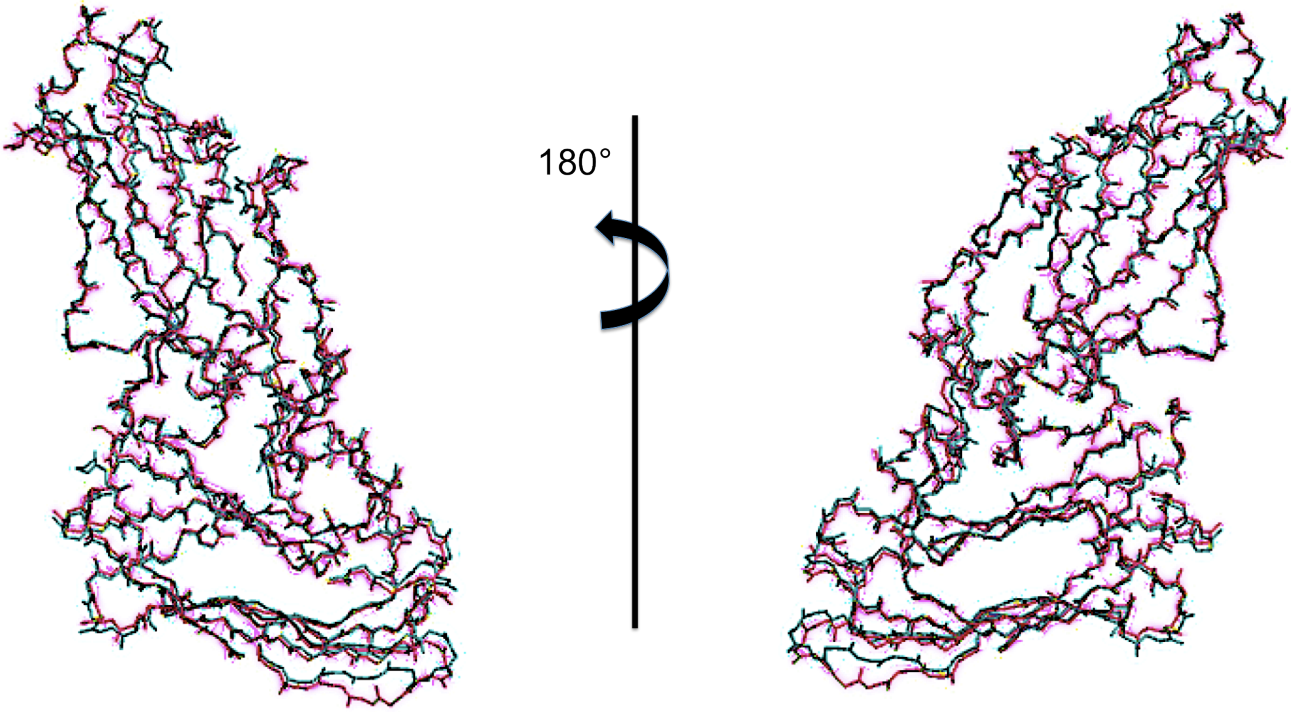


Figure 5.25: **Alignment of sNT-Als1 molecule Chain A with the sNT-Als3 structure from two stereo views.** sNT-Als1 is coloured magenta, sNT-Als3 has been coloured cyan. RMS of 0.486 was achieved after an alignment of 250-250 Cα atoms, indicating high structural similarity. Alignment was carried out in PyMOL using sNT-Als3 crystal structure (Lin et al 2014, PDB code 4LEB)

Table 5.10: PBC Residues of sNT-Als1 and sNT-Als3 that have been identified by Ligplot+ to interact with the Hepta-T ligand.
Ligplot+ was used to plot residues that interact with the Hepta-T ligand in both sNT-Als1 and sNT-Als3 (Lasowski & Swindells 2011). PDB file (code 4LEB) of sNT-Als3 was used to find those for Als3 (Lin et al 2014). The residues that differ between Als1 and Als3 are highlighted with yellow.

Location	Residue Number	Als1	Als3
A1-B1 Loop	19	A	A
	20	N	T
	21	Y	Y
	22	A	N
	27	G	G
	28	Y	T
	29	P	P
C1* Strand	59	K	K
	61	T	T
E1 Strand	116	A	A
A2 Strand	166	Y	Y
	168	Y	T
	169	A	D
	170	S	S
	172	V	V
D2b-C2' Loop	226	Y	Y
E2-F2a Loop	268	V	V
	269	N	N
F2a Strand	271	Y	Y
G2b Strand	292	T	T
	293	L	L
	294	R	R
	295	W	W
	296	T	T
	297	G	G
	298	Y	Y

When aligning the PBC of the sNT-Als1/Hepta-T structure with the sNT-Als3 structure in PyMOL, the binding pockets display a highly similar structure (Fig 5.26). The cap of both PBCs are comprised of similarly positioned A1-B1 Loops the C1* strands with the 'walls' of both cavities formed from A2 and G2b strands. Conserved residues in sNT-Als1 and sNT-Als3 that make direct contact with the Hepta-T peptide via hydrogen-bonding are Y21 of the A1-B1 loop, K59 of the C1* strand, S170 and S172 of the A2 strand plus R294 and T296 from the G2b strand. These similarities reflect the high conservation of peptide-binding profiles between Als1 and Als3 N-terminal adhesion domains.

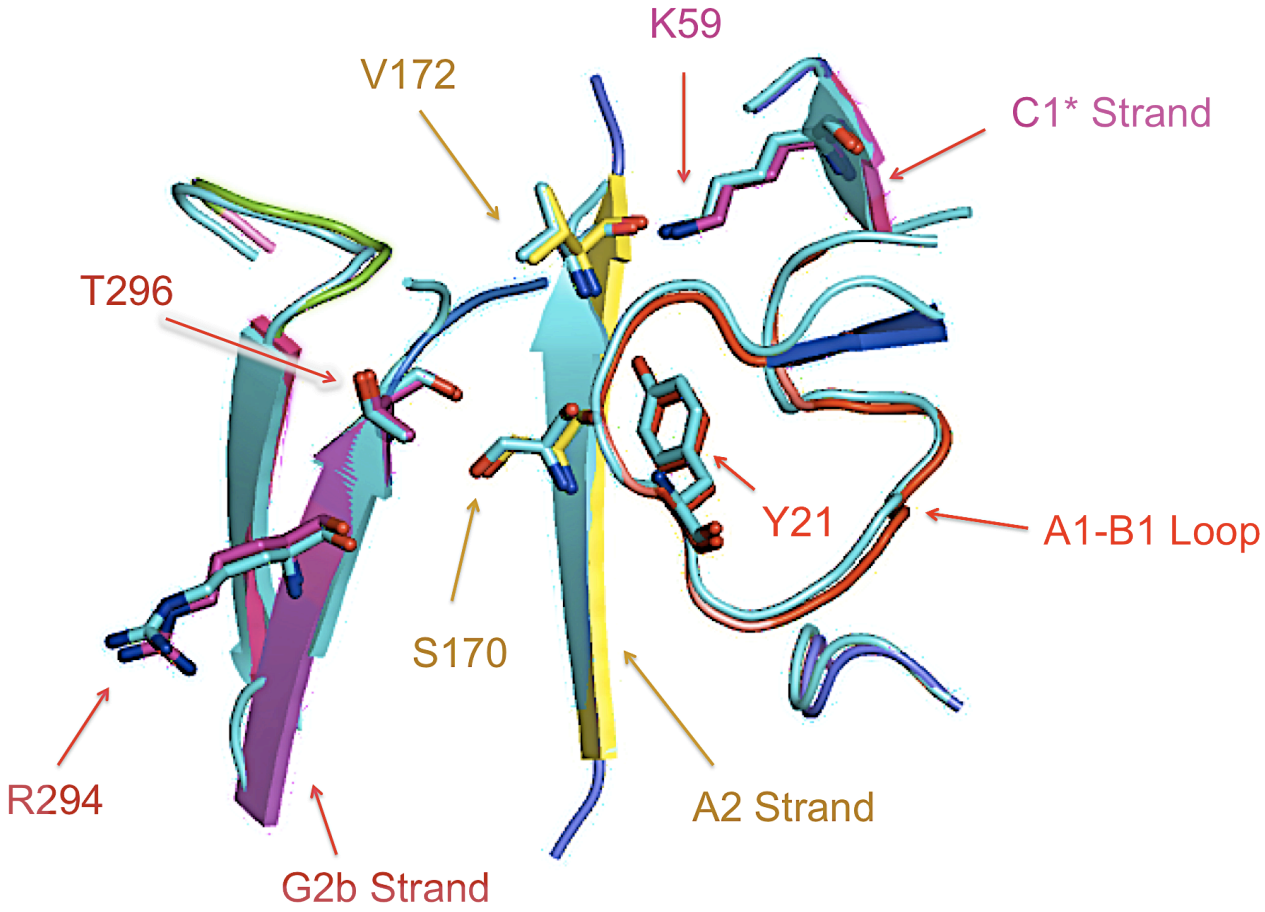


Figure 5.26: **The sNT-Als1 PBC secondary structure elements and residues that form direct H-bonds with the Hepta-T aligned with those of the sNT-Als3 PBC**.
The labelled secondary structure elements of sNT-Als1, which include C1* Strand (magenta), A1-B1 loop (red), A2 strand (yellow) and G2b strand (pink), were aligned with those of sNT-Als3 (all coloured cyan). The same residues (represented as sticks) that form direct H-bonds with the Hepta-T ligand in sNT-Als1 are the same as those for sNT-Als3, labelled K59, Y21, S170, V172, R294 and T296, and also align. Figure constructed with PyMOL using sNT-Als3 crystal structure (Lin et al 2014, PDB code 4LEB).

Using Ligplot+ to compare PBC ligand-binding interactions of sNT-Als1/Hepta-T with sNT-Als3/Hepta-T, the conserved residues that make direct contact with the Hepta-T ligand, apart from the invariant K59, do so via conserved chemical group interactions (Fig 5.27 and Fig 5.28). For example, the side chain hydroxyl group of the Y21 in the A1-B1 loop H-bonds with the backbone carbonyl oxygen of T2 in the Hepta-T in both sNT-Als1 and sNT-Als3, and the same interaction mechanism for V172 and S170 of the A2 strand and also T296 of the G3b strand is seen in both constructs as well. One difference is that the backbone carbonyl of R294 H-bonds to the backbone amide group of Hepta-T T4, and the backbone amide of R294 H-bonds to the carbonyl of Hepta-T T6 in sNT-Als1, whereas in sNT-Als3, the carbonyl of R294 H-bonds only to the side chain hydroxyl group of Hepta-T T5. This may suggest differing positions of T4, T5 and T6 of the Hepta-T peptide when in either the sNT-Als1 or sNT-Als3 PBC.

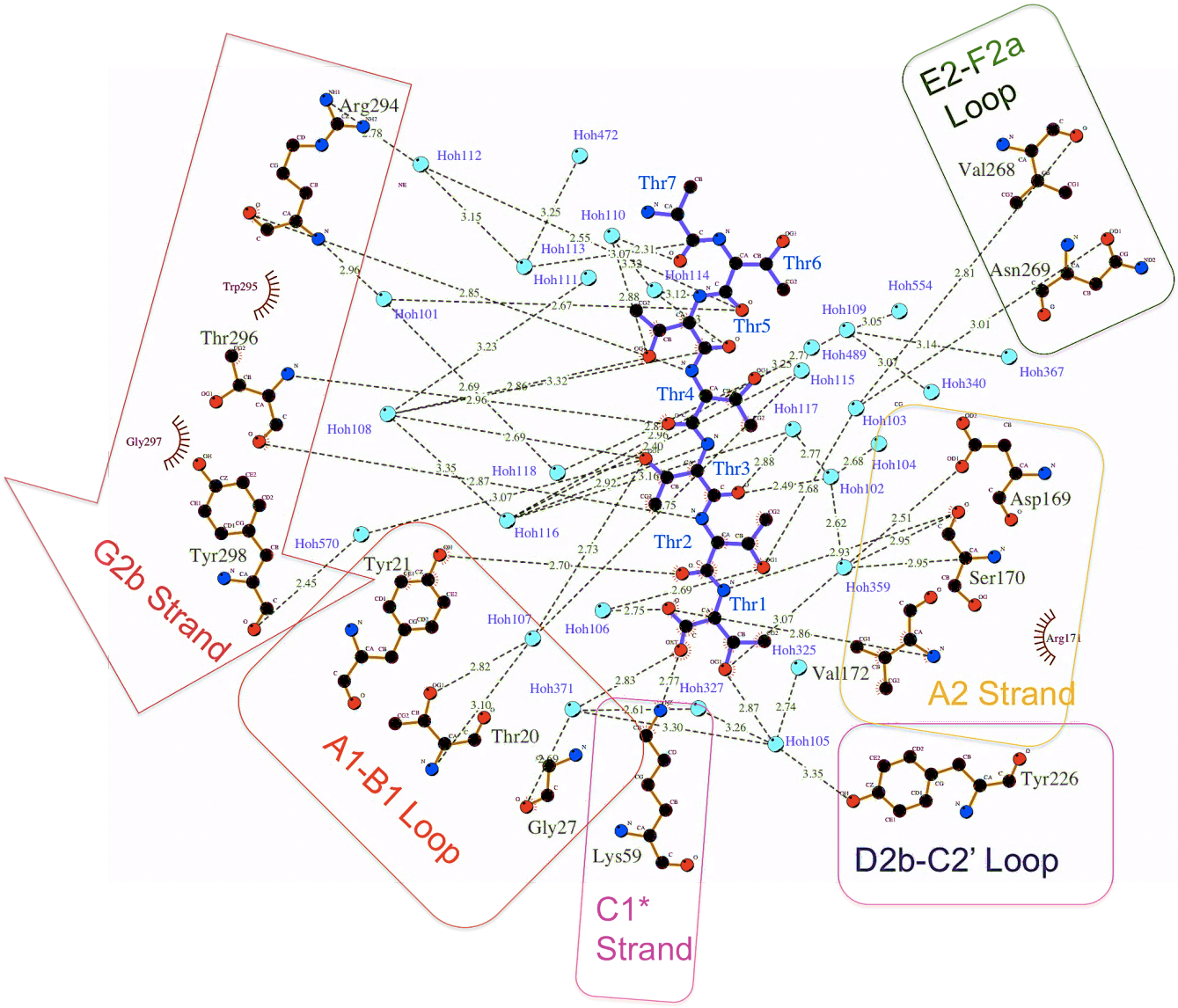


Figure 5.27: **A detailed view of sNT-Als3 PBC residues that interact with the Hepta-T peptide**. Residues (Green labels), their locations with respect to the PBC (matching the topology diagram in Chapter 5), and their hydrogen bond interactions (green dotted lines) with either the Hepta-T peptide ligand (Blue residue labels) and/or water molecules (cyan circles). Hydrophobic interactions from residues are represented by red curves with dashes. Image produced using Ligplot+ (Lasowski & Swindells 2011) using the sNT-Als1/Hepta-T model from the crystal structure solved by Lin et al 2014, PDB code 4LEB.

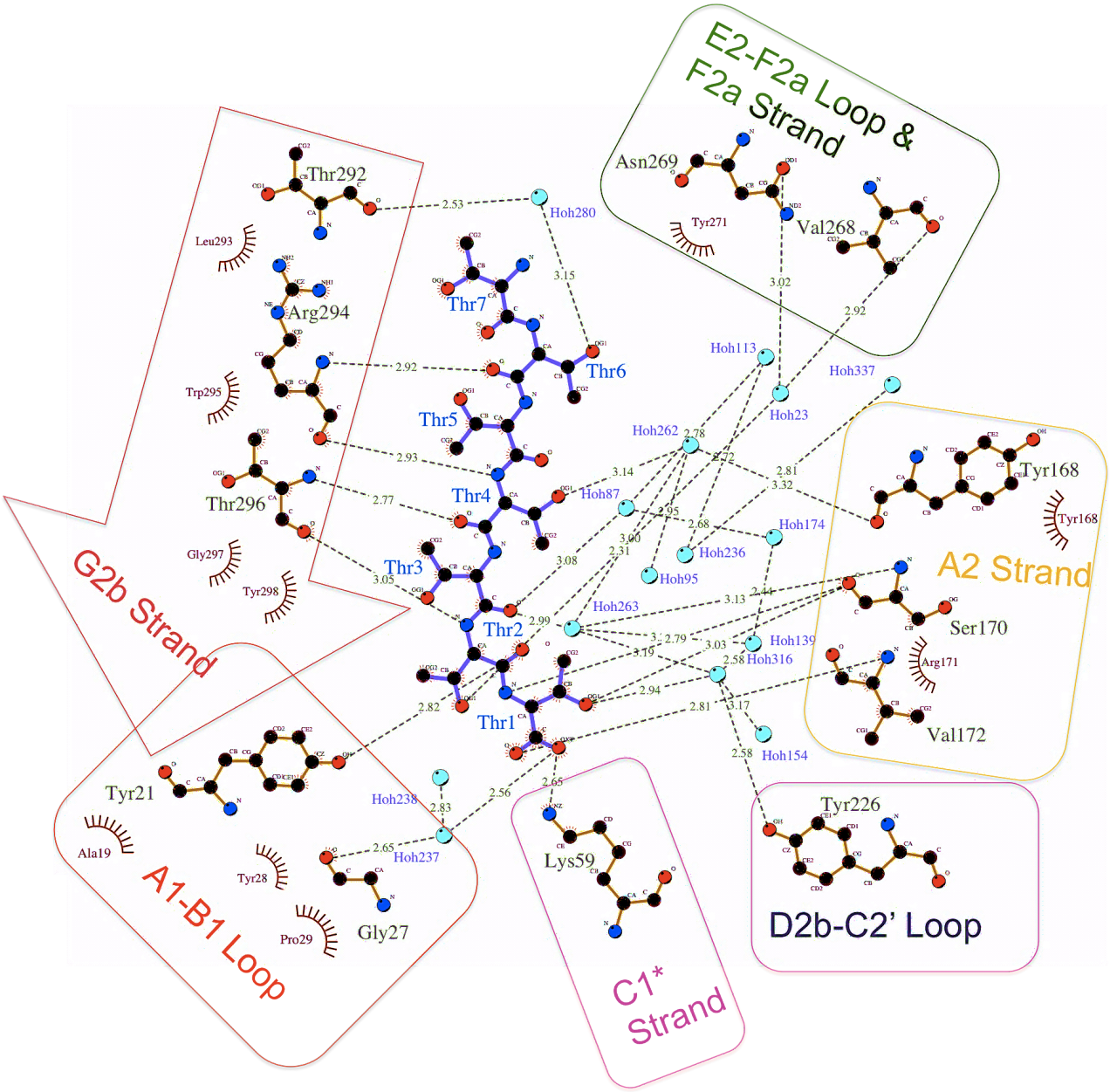


Figure 5.28: A detailed view of sNT-Als1 PBC residues that interact with the Hepta-T peptide.
Residues (Green labels), their locations with respect to the topology diagram in Chapter 5, and their hydrogen bond interactions (green dotted lines) with either the Hepta-T ligand (Blue residue labels) and/or water molecules (cyan circles). Hydrophobic interactions are represented by red curves with dashes. Image produced using Ligplot+ using the sNT-Als1/Hepta-T model from the crystal structure solved in Chapter 5 (Lasowski & Swindells 2011).

The same Ligplot+ diagrams (Fig 5.27 and Fig 5.28) show that differences can be seen in PBC residues that provide hydrophobic interactions with the Hepta-T in both sNT-Als1 and sNT-Als3 proteins. Whilst residues R171 of the A2 strand, and W295 and G297 of the G2b strand provide hydrophobic interactions in both constructs, extra hydrophobic interactions are noticed in sNT-Als1 from A19, Y28 and P29 of the A1-B1 loop (T28 in Als3), Y168 of the A2 strand, Y271 of the F2a strand and L293 and Y298 of the G2b strand. This may suggest differences in hydrophobicity within the sNT-Als1 and sNT-Als3 PBCs.

A network of 15 water molecules inside the PBC is observable in the crystal structure of sNT-Als1 (Fig 5.28), whilst that of sNT-Als3 includes 28 molecules (Fig 5.27). Residues that establish H-bonds with water molecules that interact with the Hepta-T in the same way in both sNT-Als1 and sNT-Als3 PBCs include the backbone carbonyl of G27 on the A1-B1 loop, the backbone amide of S170 on the A2 strand, the backbone carbonyl of V268 and the side-chain hydroxyl of N269. Differences include the interactions, via H-bonds with water molecules, of sNT-Als3 T20 with the Hepta-T T3 side

chain hydroxyl, that of the D169 side chain carboxylate with the Hepta-T T1 side chain hydroxyl, the Y298 backbone carbonyl with the Hepta-T T3 backbone amide, and both the R294 side chain amine and backbone amide groups with the Hepta-T T5 backbone carbonyl. Furthermore, S170 in sNT-Als3 interacts via a water molecule, with the Hepta-T T1 side chain hydroxyl group, via both its backbone carbonyl and amide groups, whereas in sNT-Als1, the backbone amide of S170 interacts via a water molecule, with the Hepta-T T2 backbone carbonyl. Plus, whilst the side chain hydroxyl of Y226 in sNT-Als1 interacts via a water molecule with the Hepta-T T2 backbone carbonyl group, the same Y226 side chain hydroxyl in sNT-Als3 interacts with the Hepta-T T1 side chain hydroxyl. Finally, an interaction between T292 and the side chain hydroxyl of Hepta-T T6 via a water molecule can be seen in the sNT-Als1 PBC, but this interaction was not pin-pointed by Ligplot+ in sNT-Als3.

These minute differences, including that connections between sNT-Als1 PBC residues and Hepta-T residues are mediated by 6 individual water molecules, whilst in sNT-Als3 these connections are mediated by 8 individual water molecules. Additionally, 28 water molecules are located in the sNT-Als3 PBC, whereas 15 water molecules occur in the sNT-Als1 PBC. This suggests a wider, more accommodating PBC in terms of number of water molecules, and possibly ligand size in sNT-Als3 in comparison to sNT-Als1, which may offer an explanation regarding Als3's greater contribution to *C. albicans* adhesion in comparison to Als1 (Hoyer et al 2008).

The tyrosine at position 28 makes the A1-B1 loop of the sNT-Als1 PBC sterically bulkier than in sNT-Als3. Plus, π-π stacking occurs between the aromatic rings of Y21 and Y28 in this A1-B1 loop, contributing to a potentially less flexible A1-B1 loop in sNT-Als1 in comparison to sNT-Als3 (Fig 5.29). However, the sNT-Als3 A1-B1 loop does contain another tyrosine in the form of Y23, but this is not as close to Y21 in space as Y28 is in sNT-Als1 to form π-π interactions.

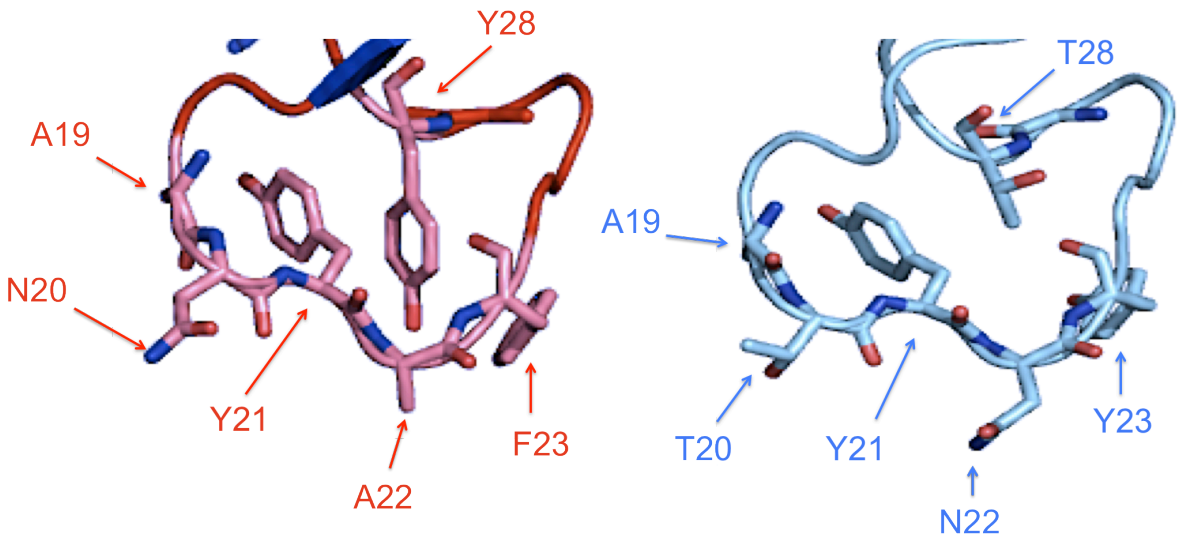


Figure 5.29: **A1-B1 loop differences between sNT-Als1 and sNT-Als3.**
The A1-B1 loop of sNT-Als1 (shown in pink) has N20 instead of T20 in the sNT-Als3 A1-B1 loop (shown in cyan), A22 instead of N22 and Y28 instead of T28. The side chain of Y28 of sNT-Als1 forms π-π interactions with Y21, which could lead to a less flexible and more rigid PBC A1-B1 loop in comparison to sNT-Als3. A1-B1 loop representation from the crystal structure of sNT-Als3 (Lin et al 2014, PDB code 4LEB).

For positions 20 and 22 in the A1-B1 loop, sNT-Als1 contains asparagine and alanine residues, whereas sNT-Als3 contains threonine and asparagine, respectively (Fig 5.30). It is possible that these residues change the shape of the PBC opening, with T20 in sNT-Als3 offering a slightly wider PBC opening than N20 in sNT-Als1.

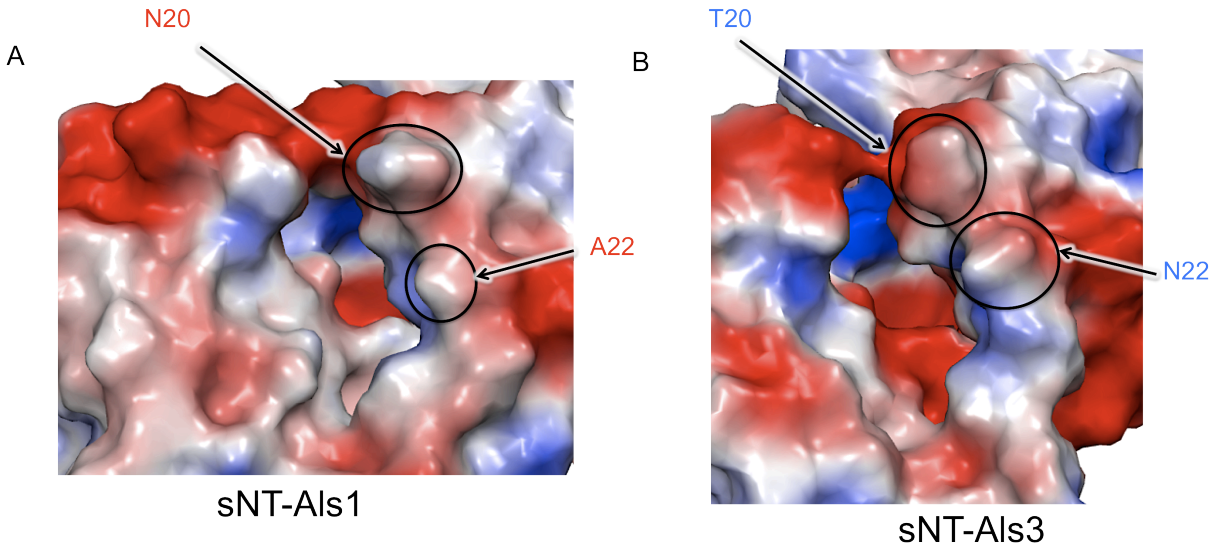


Figure 5.30: Comparison of the Surface representations of A1-B1 loop residues at positions 20 and 22 in sNT-Als1 and sNT-Als3.
A) Shows how the residues N20 and A22 of the sNT-Als1 A1-B1 loop contribute to the PBC surface. B) Shows how the residues T20 and N22 of the sNT-Als3 A1-B1 loop contribute to the PBC surface (sNT-Als3 crystal structure from Lin et al 2014, PDB code 4LEB).

Interestingly, Y168 in sNT-Als1 is occupied by a threonine in sNT-Als3. The side-chain of Y168 occludes part of the entry to the PBC, and may restrict peptides with bulky side chains from entering the rest of the cavity and interacting with the innermost K59. However, the side-chain oxygen from Y168 may be able to form hydrogen bond interactions with side chains of appropriate size that inhabit position 4 (P4) of the peptide ligand, if we consider the C-terminal residue of the peptide as position P1. Residue T168 in sNT-Als3 does not exhibit the same steric hindrance at the opening of the PBC, suggesting that sNT-Als1 may display narrower ligand specificity towards peptides with shorter side chains in P4 than sNT-Als3 (Fig 5.31).

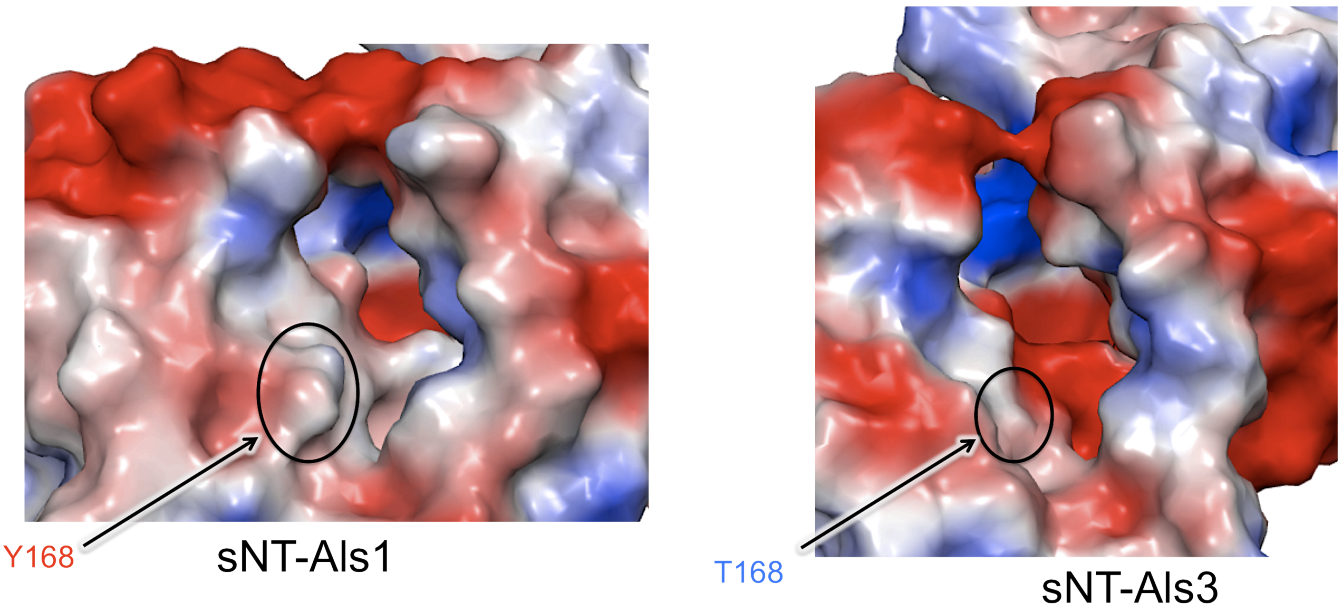


Figure 5.31: Comparison of the Surface Representations of the A2 Strand residue at amino acid position 168 in sNT-Als1 and sNT-Als3.
The bulky-side chain of residue Y168 of sNT-Als1 can be seen as more prominent in defining the outer surface of the PBC in comparison to the smaller side chain of T168 of sNT-Als3 (sNT-Als3 crystal structure from Lin et al 2014, PDB code 4LEB).

In summary, the accommodation of more water molecules in the sNT-Als1 PBC in comparison to sNT-Als1, the effect of a less flexible A1-B1 loop in sNT-Als1 due to Y28 and the occlusion of the PBC entry via Y168 in sNT-Als1 suggests at a wider PBC in sNT-Als1 which may be more versatile with regards to binding differently sized ligands, especially with large side chains at peptide position 4. All these factors may explain why Als3 has a greater contribution to *C. albicans* adhesion than Als1 (Zhao et al 2004, Oh et al 2005, Hoyer et al 2008).

5.4. PBC Residue Differences of sNT-Als1 and NT-Als9.2

Structure determination of NT-Als9.2 in complex with a C-terminal peptide from fibrinogen-γ was the first to highlight the importance of the invariant K59 and the PBC in the interaction of Als adhesins with peptide ligands (Salgado et al 2011). Using PyMOL, the crystal structure sNT-Als1 molecule Chain A was aligned with the NT-Als9.2 crystal structure (Fig 5.32), with a RMS of 0.550Å for 263 to 263 atoms aligned. The RMS deviation for this alignment is greater than that for the alignment of sNT-Als1 with sNT-Als3, correlating with the fact that sNT-Als1 has a higher residue sequence identity with sNT-Als3 (approximately 82%) than with NT-Als9.2 (approximately 69%).

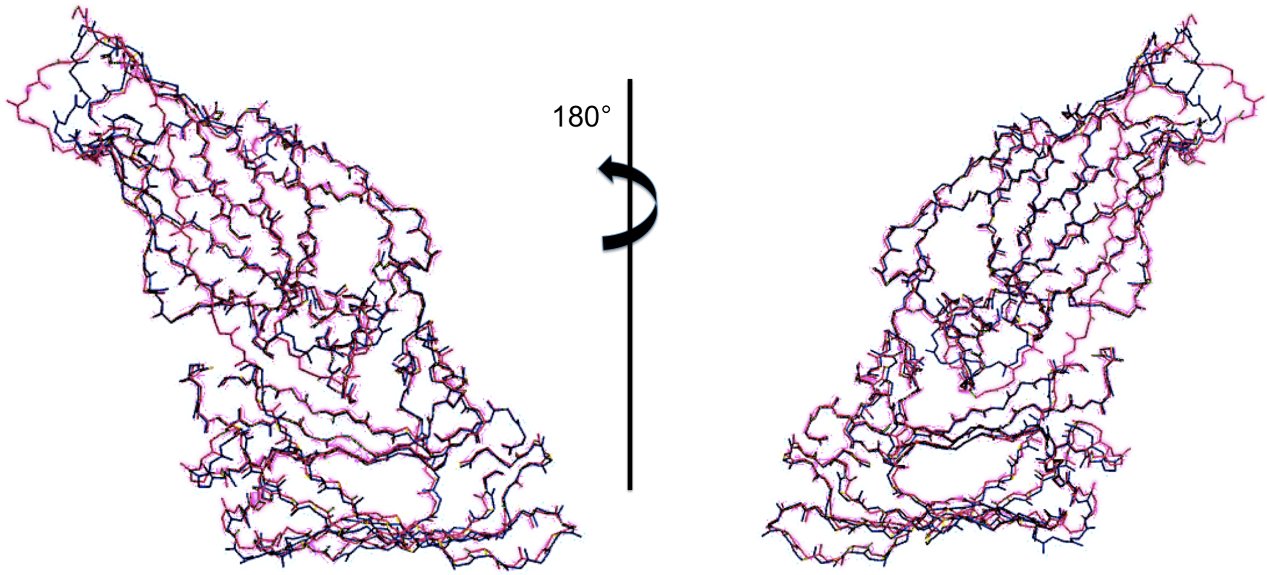


Figure 5.32: **Alignment of sNT-Als1 molecule Chain A with the NT-Als9.2 structure from two stereo views.**
sNT-Als1 is coloured magenta, sNT-Als3 has been coloured magenta, NT-Als9.2 has been coloured dark violet. An RMS of 0.550 was achieved after an alignment of 263-263 Cα atoms, indicating a relatively high structural similarity. Alignment was carried out in PyMOL, using the NT-Als9.2 crystal structure (Salgado et al 2011, PDB code 2Y7L).

The PBC residue differences between Als1 and Als9.2 are greater than the differences between Als1 and Als3. There are 11 PBC residue differences between Als1 and Als9.2 (See Table 5.10).

Table 5.11: PBC Residues of sNT-Als1 and NT-Als9.2 that have been identified to interact with either the Hepta-T ligand, or Fib-γ C-terminal peptide, respectively.
Ligplot+ was used to identify both sNT-Als1 and NT-Als9.2 residues that interact with their respective peptide ligands. (Lasowski & Swindells 2011). The NT-Als9.2 PDB model (PDB code 2Y7L) was used to find such residues (Salgado et al 2011). The PBC residues that differ between Als1 and Als9.2 are highlighted with yellow.

Location	Residue Number	Als1	Als9.2
A1-B1 Loop	19	A	V
	20	N	E
	21	Y	Y
	22	A	A
	27	G	E
	28	Y	T
	29	P	P
C1* Strand	59	K	K
	61	T	I
E1 Strand	116	A	T
A2 Strand	166	Y	Y
	168	Y	T
	169	A	S
	170	S	S
	172	V	I
D2b-C2' Loop	226	Y	Y
E2-F2a Loop	268	V	V
	269	N	N
F2a Strand	271	Y	Y
G2b Strand	294	T	T
	295	L	L
	296	R	T
	297	W	W
	298	T	W
	299	G	G
	300	Y	Y

When comparing the PBC's of sNT-Als1 and NT-Als9.2, the main differences are located in the A1-B1 loop. Alike to sNT-Als3, threonine is at position 28 in NT-Als9.2, in comparison to Y28 in sNT-Als1, meaning that π-π are also unlikely to occur in the A1-B1 loop of this construct (Fig 5.33).

With regards to direct interactions of the ligand-contacting PBC residues, due to different peptides, Hepta-T and fibrinogen-γ being involved in co-crystallisation of sNT-Als1 and NT-Als9.2, respectively, it is difficult to accurately compare and contrast interacting residues. However, it is apparent that contacts between K59, S170 and T296 are conserved between the two constructs (Fig 5.33 and 5.34). I172 and T294 in NT-Als9.2 replace V172 and R294, respectively, in sNT-Als1. Y21 is conserved in both constructs, but contributes to hydrophobic interactions, not hydrogen-bonding with the fibrinogen-γ peptide.

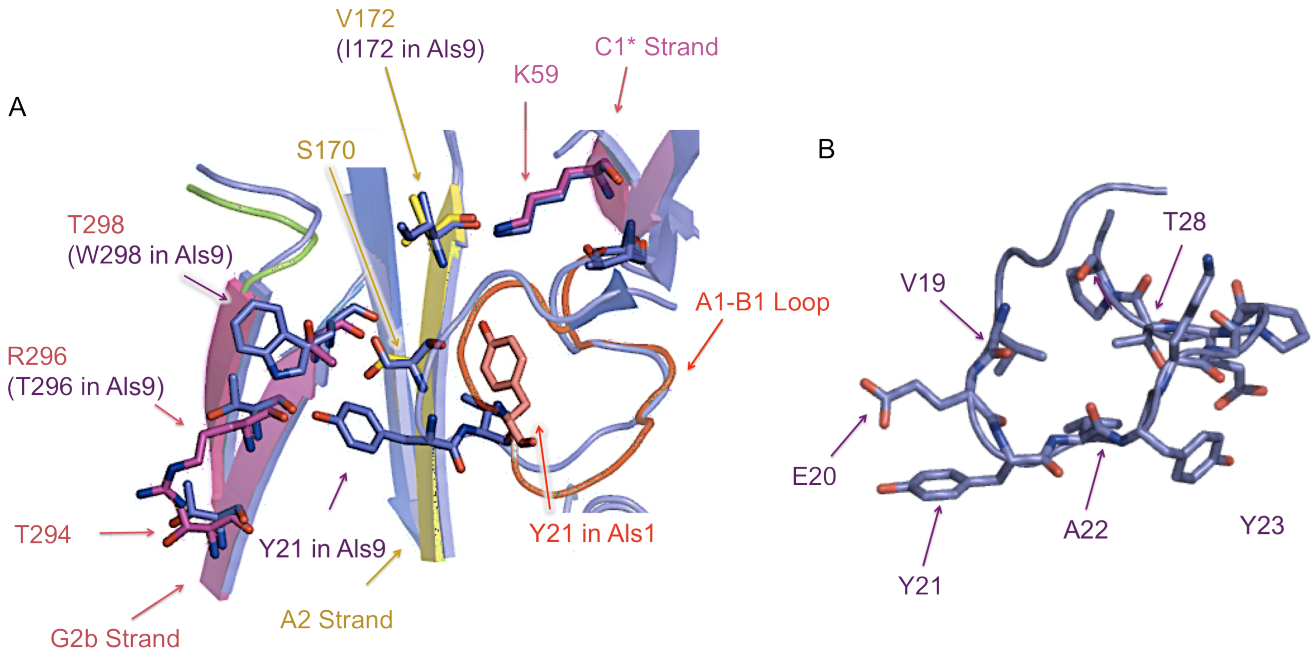


Figure 5.33: The sNT-Als1 PBC secondary structure elements and residues that form direct H-bonds with the Hepta-T ligand aligned with those of the NT-Als9.2 residues that interact with fibrinogen-gamma and a Cartoon and Stick representation of the A1-B1 Loop of NT-Als9.2.
A) The labelled secondary structure elements of sNT-Als1 which include the C1* strand (magenta), A1-B1 loop (red), A2 strand (yellow) and G2b strand (pink) were aligned with those of NT-Als9.2 (dark violet). The same residues K59, S170, T294 and Y21 in both constructs interact with peptide ligands, however Y21 does not interact with fibrinogen-gamma peptide with a hydrogen-bond, but interacts hydrophobically instead. The PBC configuration of Y21 is different from Y21 of sNT-Als1, due to structural differences in the A1-B1 loop. V172 in sNT-Als1 is instead I172 in NT-Als9.2, R296 in sNT-Als1 is instead T296 in Als9, and T298 in Als1 is instead W298 in Als9. The A1-B1 loop (B), includes T28 instead of the Y28 found in sNT-Als3. Figure constructed with PyMOL, NT-Als9.2 structure from Salgado et al 2011, PDB code 2Y7L.

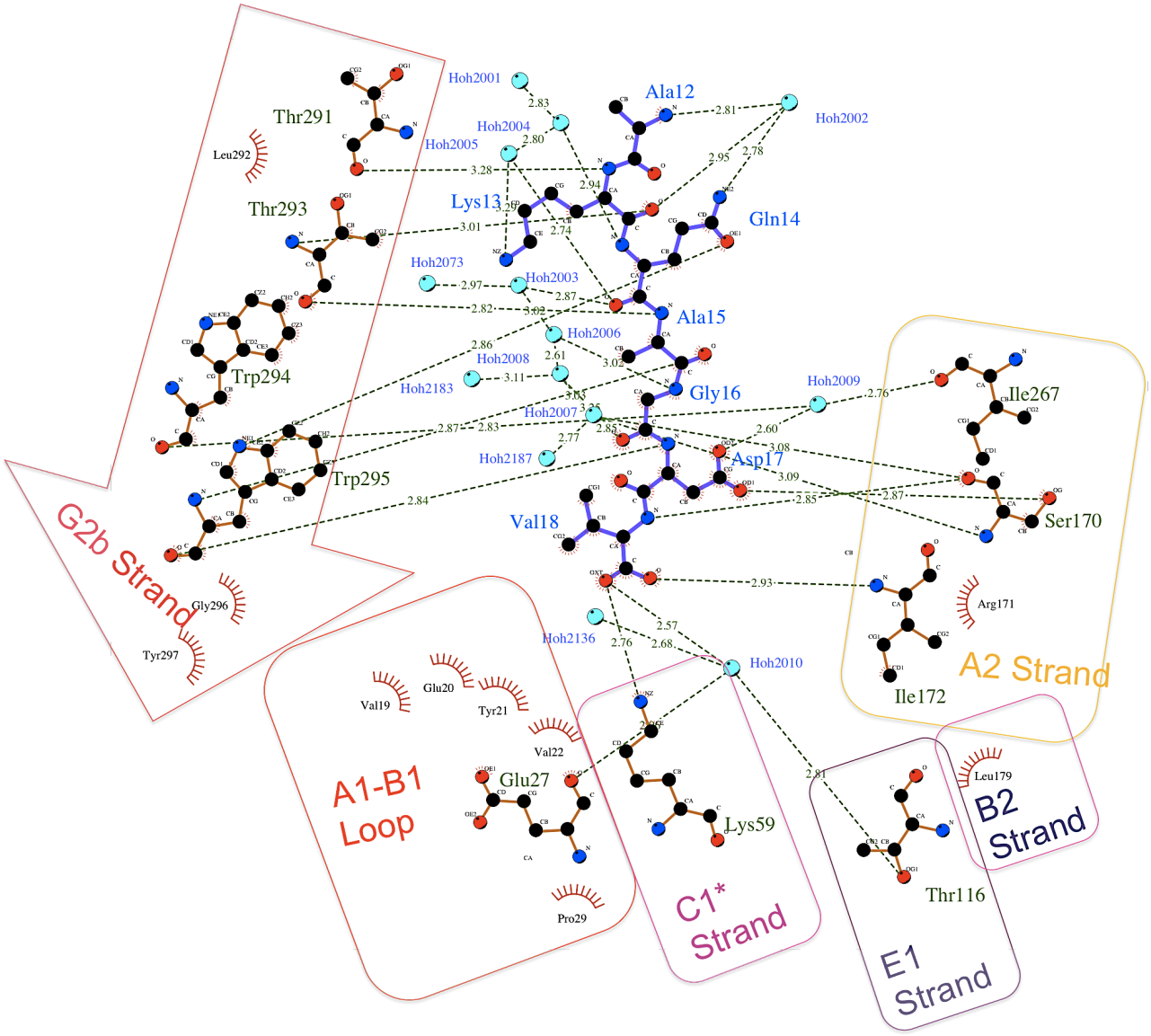


Figure 5.34: A detailed view of NT-Als9.2 PBC residues that contact the fibrinogen-γ peptide.
Residues (Green labels), their locations, and their hydrogen bond interactions (green dots) with either the fibrinogen-γ peptide ligand (Blue residue labels) and/or water molecules (cyan circles). Hydrophobic interactions represented by red curves with dashes. Image produced using Ligplot+ (Lasowski & Swindells 2011) using crystal structure solved by Salgado et al 2011, PDB code 2Y7L).

Both water molecule networks of sNT-Als1 and NT-Als9.2 contain 15 water molecules. The residues that interact with the water molecule network in the NT-Als9.2/fibrinogen-γ structure include E27, I116, T168 and I267. Whilst G27 and Y168 in sNT-Als1 interact with the water molecule network, residues at positions 116 and 267 in NT-Als9.2 were not shown to. Again, similarly to sNT-Als3, threonine is at position 168, in comparison to Y168 in sNT-Als1.

5.5. Overall Summary of PBC Comparisons between sNT-Als1, sNT-Als3 and NT-ALS9.2

As detailed in sections 5.3 and 5.4, there are some PBC residue differences between sNT-Als1, sNT-Als3 and NT-Als9.2. However, when looking at these three different PBCs from the wider perspective of their overall shape and volume, it is possible to see how minute residue differences may impact on their overall accommodation of peptide ligands.

For instance, on comparison of the water molecule networks between the three proteins, both PBCs of sNT-Als1 and NT-Als9.2 accommodate a network of 15 water molecules, as seen from structural data. The water molecule network of sNT-Als3 is much greater, in that its PBC contains 28 water molecules. However, when looking at the hydrogen bonds between PBC residues and peptide residues mediated by individual water molecules, 6 water molecules are utilized in the sNT-Als1 PBC/Hepta-T interaction and 8 water molecules for the sNT-Als1 PBC/Hepta-T interaction. This seems to suggest that sNT-Als3 may have a wider PBC in comparison to sNT-Als1. Two water molecules are seen to mediate hydrogen-bonds in the NT-Als9.2 PBC/fibrinogen-gamma interaction, suggesting that this PBC may be narrower than both sNT-Als1 and NT-Als9.2, or that water molecules offer variable interaction compensations depending on the shape and size of the peptide ligand's residue side chains.

In order to estimate approximate volumes of the three PBCs as well as relative hydrophobicity and hydrophilicity ratings (see Table 5.11), the program Schrodinger Sitemap was used (Halgren 2007, Halgren 2009). This program placed a virtual 'grid' within each PBC, from the structural data of all three proteins prepared with virtual hydrogen atoms added (see Figure 5.35). These grids incorporated the volumes of each of the peptide ligands, and were cut according to the increase in solvent exposure at the outer surface of the PBCs.

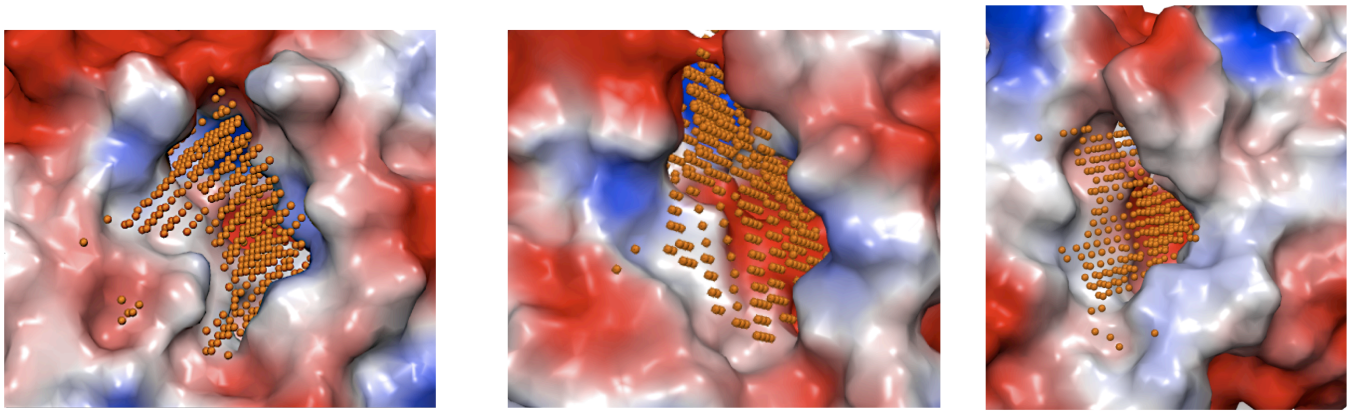


Figure 5.35: Volume Estimations of the PBCs of sNT-Als1, sNT-Als3 and NT-Als9.2.
A virtual 'grid', represented as orange non-bonded spheres, was used to estimate approximate volumes of the PBCs of sNT-Als1, sNT-Als3 and NT-Als9.2. This grid was calculated using the Schrodinger Sitemap program (Halgren 2007, Halgren 2009). (sNT-Als3 structure from Lin et al 2014, PDB code 4LEB, and NT-Als9.2 structure from Salgado et al 2011, PDB code 2Y7L).

Table 5.12: Estimated volumes of sNT-Als1, sNT-Als3 and NT-Als9.2 PBCs.
Volume estimates were carried out using Schrodinger Sitemap (Halgren 2007, Halgren 2009).

Protein PBC	Volume (Å^3)	Relative Hydrophobicity	Relative Hydrophilicity
sNT-Als1	448.99	0.373	0.631
sNT-Als3	514.84	0.136	0.846
NT-Als9.2	365.29	0.605	0.631

The approximate volume results from Sitemap suggests that the PBCs may be ranked in terms of volume, with sNT-Als3 possessing the largest estimated volume (approximately 514.84Å^3), sNT-Als1 with an estimated PBC volume of approximately 448.99Å^3 (less than sNT-Als3, but a greater than NT-Als9.2) and NT-Als9.2 PBC with the smallest estimated volume of 365.29Å^3.

Interestingly, when comparing the atomic B-factors of the residues for each PBC peptide ligand, it is possible to speculate how the shape and size of each PBC may have an effect on the relative flexibility and average B-factor of each peptide ligand residue. Very high B-factors (>80A^2) can be seen only for the side-chain atoms of the residue at Position 7 (P7) of the Hepta-T peptide bound to sNT-Als1, whereas similarly high B-factors are shown for both P7 and P6 of the Hepta-T bound to sNT-Als3 (see Figure 5.36). Medium high B-factors (>60A^2) are also seen for residues at P5 and P4 of the Hepta-T bound to sNT-Als3. A very high B-factor (>80A^2) is seen for the side chain of P6 residue, plus medium-high B-factors (>60A^2) for the P7 residue of the fib-γ peptide bound to NT-Als9.2. This suggests that the sNT-Als3 PBC shape may impose less conformational restraints on its bound Hepta-T at peptide positions P4, P5, P6 and P7 in comparison to sNT-Als1 and NT-Als9.2. Furthermore, the shape of the PBC of NT-Als9.2 may impose more conformational restraints on its bound peptide ligand in comparison to the PBC's of sNT-Als1 and sNT-Als3.

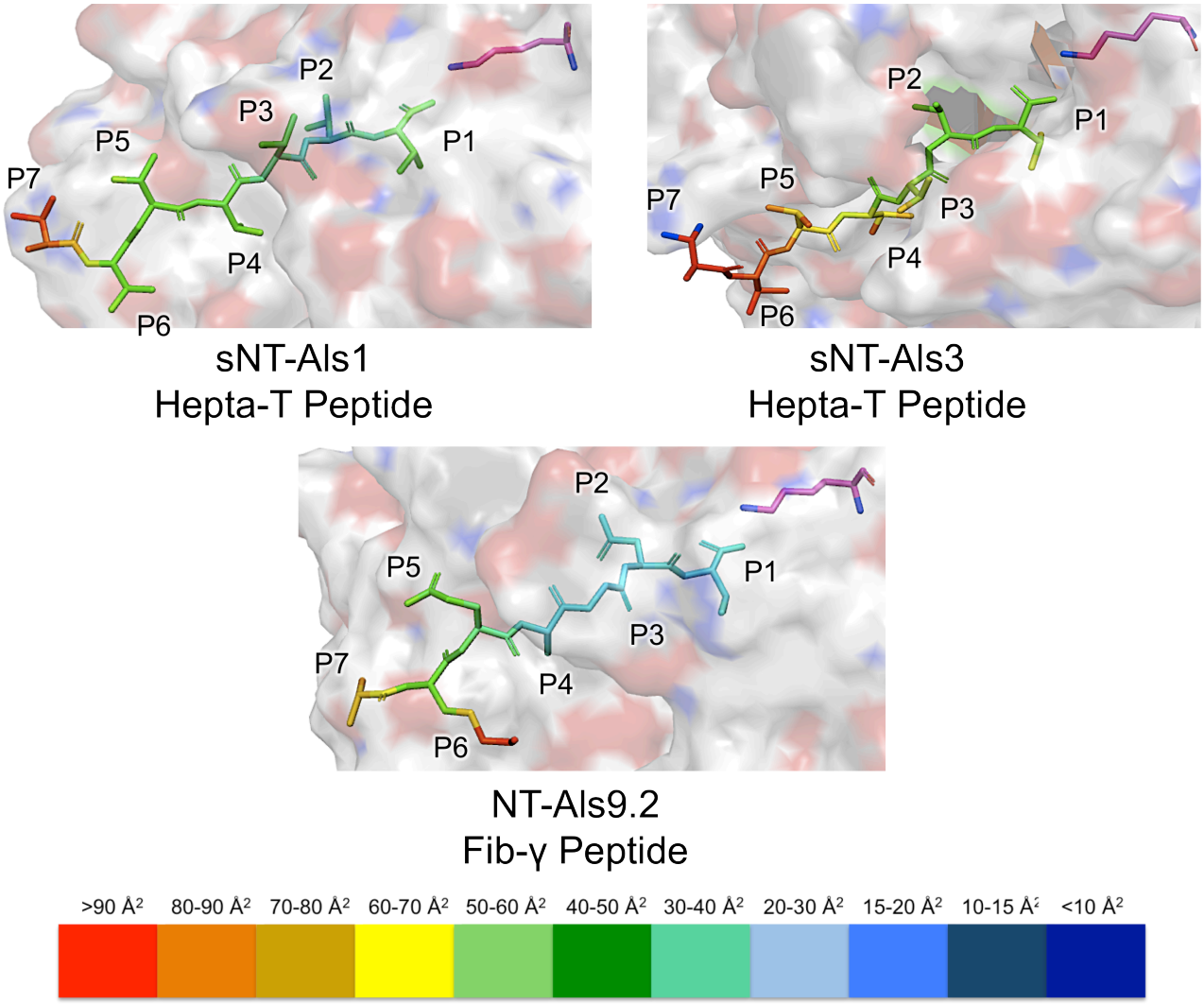


Figure 5.36: Average B-factors of Peptide Ligand Residue Positions Whilst Bound to the PBC's of sNT-Als1, sNT-Als3 and NT-Als9.2, Alongside a B-factor Scale.
The peptides bound to the PBCs of sNT-Als1, sNT-Als3 and NT-Als9.2 have their residues labelled according to the order of binding, and coloured with respect to the B-factor scale. The stick-figure side-chain of K59 (shown in magenta) is shown to complement the representation of peptide orientation. The residue at position 1 (P1) is closest to K59, and the residue at position 7 (P7) is the one furthest from the K59. (sNT-Als3/Hepta-T structure from Lin et al 2014, PDB code 4LEB and NT-Als9.2/Fibγ structure from Salgado et al 2011, PDB code 2Y7L).

Another question regarding PBC specificities of these proteins remains whether there are electrostatic charge differences between them. On comparison of the relative electrostatic profile of the outer surface of the sNT-Als1, sNT-Als3 and NT-Als9.2 PBCs, they do not differ extensively in terms of charge (see Figure 5.37). The opening of the cavity can at best be described as having neutral charge. A 'roof' of negative charge above the cavity occurs for sNT-Als1 and sNT-Als3, which differs from a more neutral region of NT-Als9.2. Within the cavity, the invariant K59 that is conserved between the three proteins can possibly account for the blue, basic region of the PBC.

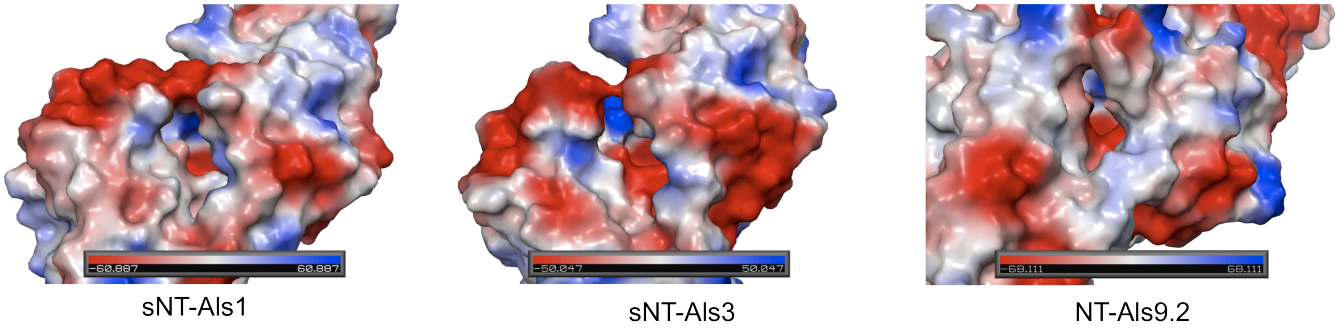


Figure 5.37: Comparison of the Surface Electrostatic Charges between the PBCs of snT-Als1, sNT-Als3 and NT-Als9.2.
The more negatively charged, acidic regions are coloured red, the more positively charged, basic regions are coloured blue. (sNT-Als3 structure from Lin et al 2014, PDB code 4LEB. NT-Als9.2 structure from Salgado et al 2011, PDB code 2Y7L).

From the information provided here, it can be concluded that the PBCs of sNT-Als1, sNT-Als3 and NT-Als9.2 may differ more in terms of cavity volume and the steric effects of residues that line the opening of the cavity, rather than electrostatic differences. The binding mechanism of the peptides with all three PBCs is highly similar, in that the salt-bridge interaction between the peptides' carboxyl termini with the invariant K59 amine group is common to all. However, peptides containing residues with bulky side-chains that produce steric clashes with the cavity opening, may have more trouble reaching and interacting with K59 in the PBCs of sNT-Als1 and NT-Als9.2 than in the PBC of sNT-Als3.

6. sNT-Als1 Backbone Assignment with NMR Spectroscopy

6.1 – Background

Structure determination of NT-Als9.2 was obtained by X-ray crystallography (Salgado et al 2011). However, this construct included a C-terminus with a 6-residue, amyloid-forming region (AFR). In purified NT-Als adhesins, this region is flexible, and can be recognised as a ligand by the peptide-binding cavity (PBC) of the protein in inter- or intramolecular interactions, resulting in self-binding (See Chapter 1, section 1.6.3) (Salgado et al 2011). To avoid this effect, smaller constructs lacking the last thirteen residues of full-length NT-Als adhesins (or sNT-Als, amino acids 1-299 or 1-300 of the mature sequence) were designed. Tests were carried out to ensure that the stability and the structural properties of sNT-Als3 are not affected by this deletion (Lin et al 2014). Importantly, these constructs retain the ability to bind peptides and should allow the measurement of accurate binding constants for different ligands.

Previously, assignment of backbone resonances of NT-Als1 (amino acids 1-313) was carried out in this laboratory. Using NMR chemical shift perturbation (CSP) analysis, the assignment confirmed that ligands interact specifically with the PBC in solution. It also provided valuable per-residue information on the relaxation properties of the protein (Yan et al 2010, Salgado et al 2011). Furthermore, ^{1}H-^{15}N-TROSY-HSQC spectra were used to monitor the titration of NT-Als1 with Hepta-T identified chemical shifts attributed to residues inside but also outside the PBC (Fig 6.1), an indication that ligand binding induces changes in the conformation of the flexible C-terminus containing the AFR, as observed between the free- and ligand-bound forms of NT-Als9.2 or other regions of the protein.

In this project, the sNT-Als1 construct (1-300) was used to determine the crystal structure in complex with Hepta-T (see Chapter 5). To complement the structural work and facilitate ligand-binding assays by NMR, I endeavoured to re-assign the backbone resonances of this construct. We hypothesised that lack of the flexible C-terminal region would minimise the chemical shift changes from residues outside the PBC, which would simplify the identification of ligands in titration experiments using NMR spectra. Although the binding mode for peptides in NT-Als adhesins is well established, it is less clear whether small chemical fragments identified as ligands will invariably occupy the PBC or other regions of the protein. If successful, the same method can be extended to ligand binding studies of the remaining NT-Als adhesins.

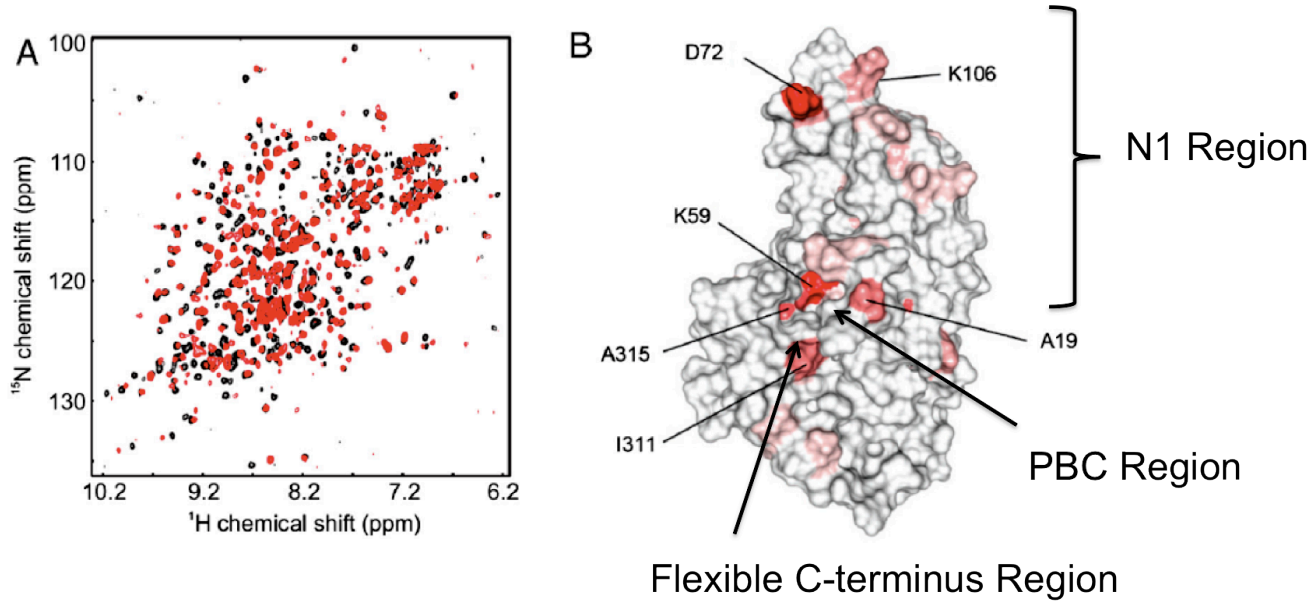


Figure 6.1: **^{1}H-^{15}N-TROSY-HSQC titration of NT-Als1 with Hepta-T peptide ligand and how the chemical shift changes of the assigned peaks correspond to residues on the NT-Als1 protein structure.**
The chemical shift changes seen in the ^{1}H-15N-HSQC titration (A) as NT-Als1 before ligand (black) and NT-Als1 after ligand titration (red) are attributed to residues K59 and A19 of the PBC, I311 and A315 of the flexible C-terminus and also amino acids D72 and K106 regions of the protein surface at the top of the N1 domain (B). This was explained as due to changes in the conformation of the flexible C-terminal region during its binding to the N1 domain of the protein (Modified figures from Salgado et al 2011).

6.2 – The ^{2}H^{15}N^{13}C Triple Labelled sNT-Als1 Sample.

A ^{2}H-^{13}C-^{15}N (triple) labelled sNT-Als1 sample was produced using methods discussed in Chapter 4, based on a protocol detailed in Yan et al, 2010. As the relative natural abundance of the ^{15}N and ^{13}C nuclei is low, bacteria expressing the sNT-Als1 protein were grown using ^{15}N-labelled NH_4Cl as the sole source of nitrogen, and ^{13}C glucose as the sole source of carbon.

Due to sNT-Als1 being larger than 30kDa, it was necessary to deuterate the sample (Gardner & Kay 1998). The atomic structure of proteins larger than 30kDa, are often difficult to study using 3D NMR spectroscopic experiments due to higher amounts of spectral overlap, plus low resolution and sensitivity. Slow tumbling rates and fast transverse relaxation lead to peak broadening and a decrease in FID acquisition time, especially for ^{13}C nuclei. Replacing the ^{1}H proton with deuterium can improve resolution of NMR spectra by decreasing the transverse relaxation rates of ^{13}C (and also ^{15}N) and improve peak line-widths, thus improving the FID signal that is acquired. Deuteration of the sample in this way can especially enhance the sensitivity of triple resonance backbone assignment and also NOESY experiments (Kay & Gardner 1997, Yamazaki et al 1994, Wider 2000). Interestingly, deuterium can influence the chemical shifts of ^{13}Cα's in a way that is dependent on their position in the secondary structure of the protein – this artefact of sample deuteration can also give valuable structural information (Gardner & Kay 1998).

Deuteration of sNT-Als1 was carried out by using D_2O instead of H_2O for the minimal growth media used for protein expression in bacterial cultures. However, as deuterated ^{13}C glucose was not used during the triple labelling of the sNT-Als1 sample, this resulted in incomplete deuteration of the sample. Incomplete deuteration would have been especially seen in aromatic groups due to the utilisation of glucose for their synthesis by bacteria (Gardner & Kay 1998).

However, for the deuterated sNT-Als1 sample to be fully useful for NMR spectroscopic experiments, the exchange of ^{2}H nuclei in backbone and side-chain NH groups with ^{1}H protons had to be carried out, as recording of three-dimensional spectra to obtain resonance assignments require irradiation of protons on NH groups in order to initiate the magnetization transfer from the amide proton, which is required for the correlation of 15N and 13C chemical shifts. A H_2O buffer containing

50mM sodium phosphate pH8.0, 300mM NaCl and 4M urea was used to accelerate the rate of exchange of protons, a method successfully used to achieve complete protonation of backbone amides in NT-Als1 (Yan et al 2010).

In order to view whether the production of deuterated sNT-Als1 was successful, a 1D ^{1}H NMR spectrum of the concentrated 0.5mM sample was recorded at 310K after exposure to conditions that induce the back exchange of amide protons (Fig 6.2).

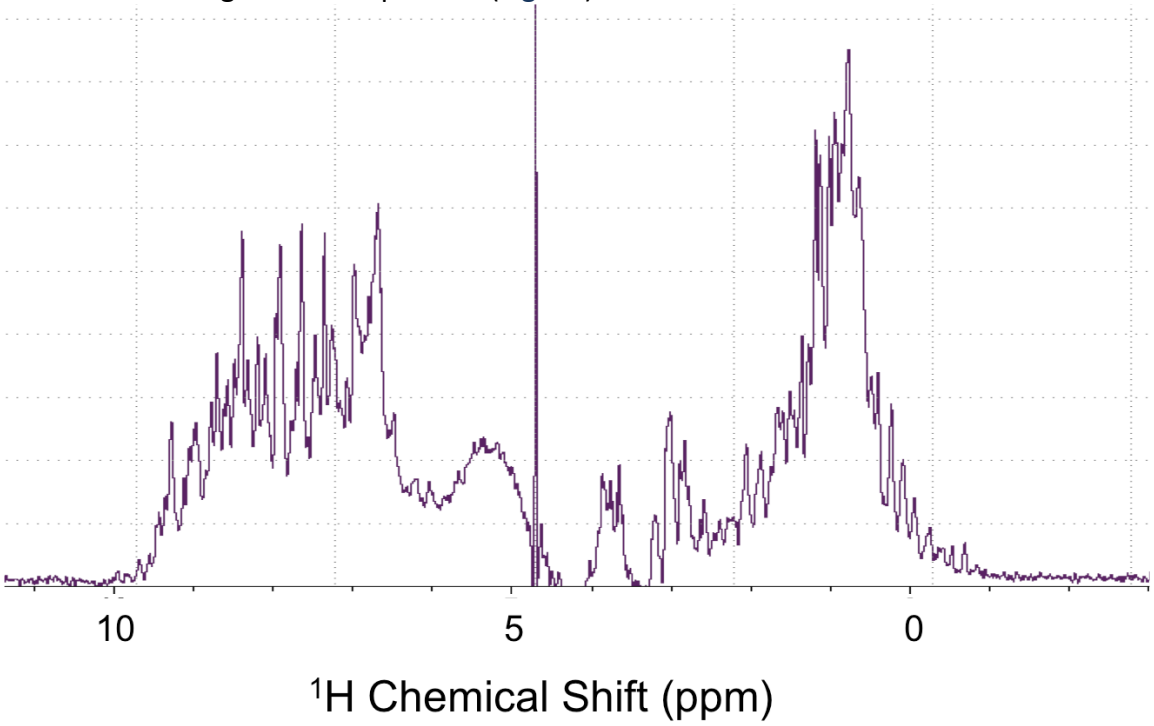


Figure 6.2: **^{1}H 1D spectrum of partially deuterated ^{13}C-^{15}N-sNT-Als1**.
Good peak dispersion showed that production of a soluble, non-disordered sNT-Als1 sample was successful. A reduced density of peaks is mainly observed in the region corresponding to aliphatic carbons (2-5ppm), showing that ^{2}H exchange with ^{1}H had occurred.

The 1D ^{1}H spectrum of the deuterated ^{15}N-^{13}C sNT-Als1 sample was compared with a 1D proton NMR spectrum of an unlabelled sample (Fig 6.3) in order to test the effect of deuteration on the spectral peaks of the protein construct.

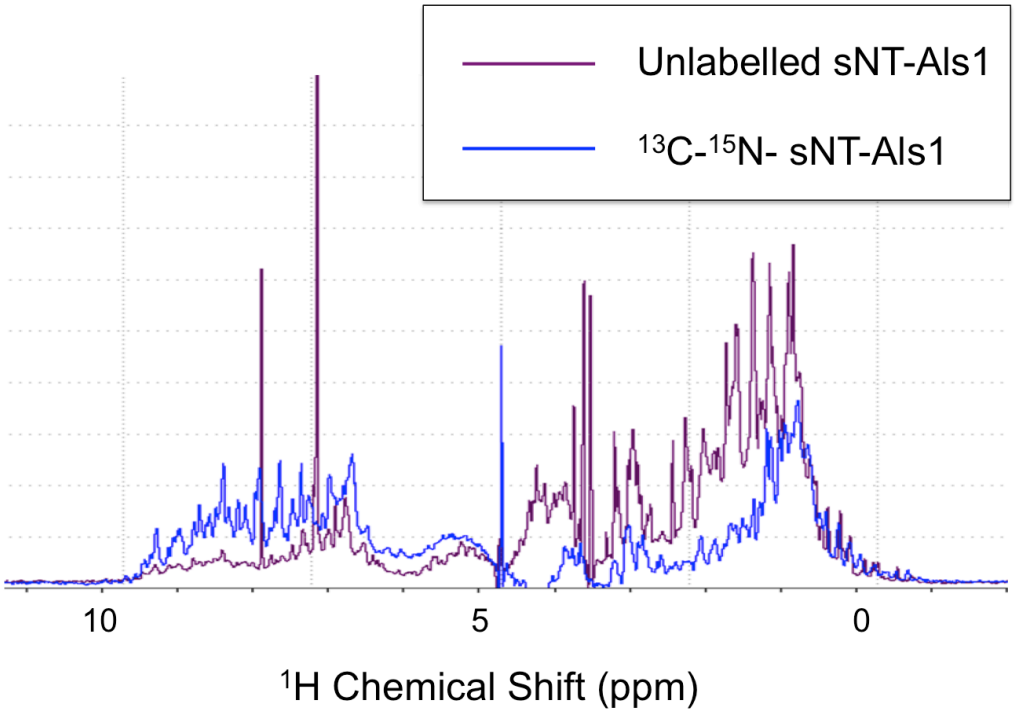


Figure 6.3: **A comparison of the 1D ^{1}H one-dimensional spectra of unlabelled sNT-Als1 sample with that of the triple-labelled sNT-Als1.**
The triple labelled sNT-Als1 sample (blue) has more prominent peaks than the unlabelled sample (violet) that correspond to ^{1}H nuclei on the side-chain and backbone N-H groups. However, a couple of peaks in this region are more prominent in the unlabelled sample, suggesting that there was incomplete back exchange of ^{1}H protons.

This, and also the overlay of deuterated sNT-Als1 with non-deuterated NT-Als1 (Fig 6.6), indicated incomplete back exchange of ^{1}H protons, although the same back exchange method used successfully for NT-Als1 (Yan et al 2010) was employed in the preparation of the triple labelled sNT-Als1 sample. It is possible that an increase in stability seen from shortening the Als1 N-terminal construct also results in greater difficulty in ^{1}H protons from reaching amide NH groups in order to exchange with ^{2}H.

To increase the rate of exchange of protons, sNT-Als1 samples were exposed to an extended dialysis time of 12 hours at room temperature, high salt concentrations of 0.5M NaCl and pH variations (pH 7.0-pH 8.0), but the maximal urea concentration was kept at 4M to avoid protein denaturation. Work is underway to obtain sNT-Als1 samples ideal for tridimensional NMR spectra, by exposing the sample to temperatures and higher denaturant concentrations that promote the back exchange and do not unfold the protein.

6.3 – Backbone Assignment of sNT-Als1 with NMR Spectroscopy

6.3.1. NMR Backbone Assignment Experiments

The NMR spectra that were recorded in order to assign the backbone residues of sNT-Als1 included TROSY-HSQC, TROSY-HNCACB, TROSY HN(CO)CACB, TROSY-HNCO, TROSY-HN(CA)CO and ^{15}N-NOESY-HSQC, at 310K on a Bruker 800 MHz Avance II spectrometer using a TXI Cryoprobe and also Room Temperature Probe.

The magnet imposing an external magnetic field strength of 800Mhz was used because the use of transverse relaxation optimised spectroscopy (TROSY) was needed in order to perform these experiments for the sNT-Als1 protein sample as it was greater than 30kDa. The TROSY method

combined with sequential assignment experiments and 2D correlation experiments enhance acquisition of sample FID via experimental tuning to a coherence at which there is cancellation between dipole-dipole coupling and chemical shift anisotropy. In this way, the resonance frequency of the most slowly relaxing component is used for FID acquisition, which results in a higher resolution signal (Pervushin et al 2000). It is able to negate the loss of signal that often occurs due to the fast tumbling and fast relaxation times seen in proteins larger than 30kDa. However, it is only able to be used if a high magnetic field strength (800 or more MHz) is available (Gardner & Kay 1998).

6.3.2 Backbone Assignment Data

NMR spectra were processed using NMRPipe (Delaglio et al 1995). Peak picking from a ^{15}N-TROSY-HSQC spectrum was performed using NMRView (Johnson & Blevins 1994). An NMRview-based script (Marchant et al 2008) was used to compile a list of pseudoresidue (PR) chemical shifts from backbone atoms: HN_i, N_i, C^{α}_i, C^{β}_i, $C^{\alpha}_{i-1,}$ C^{β}_{i-1}, C'_I and C'_{i-1}.

The program MARS was used to automatically assign part of the backbone (Jung & Zweckstetter 2004). It was set to run with a maximum length of 5 PR fragments, with 0.25ppm tolerance for C' and HA and 0.5ppm tolerance for C^{α} and C^{β} shifts with correction for perdeuterated proteins. Input files used included a fixed assignment file for glycine PRs, a primary sequence of sNT-Als1 (1-300) in FASTA format, a secondary structure prediction PSIPRED file and PR list (chemical shift table).

MARS assigned 72 residues of medium-to-medium-high confidence in total (80-100% confidence), 33 with medium confidence (50-80% confidence) and 29 with low confidence (less than 50% confidence). These assignments were manually checked using NMRview, with a typical strip view of TROSY-HNCACB (i-1 & i) and TROSY-HN(CO)CACB (i-1) data as shown in Figure 6.4.

In short, MARS had assigned approximately 24% of the backbone residues, with 228 residues left unassigned. However, many peaks from the backbone assignment experiments were weak, with MARS having difficulty with pin-pointing these, also meaning that the data was quite fragmented. As MARS cannot assign segments fewer than 3-5 residues in length, this would have also posed some difficulty with the program.

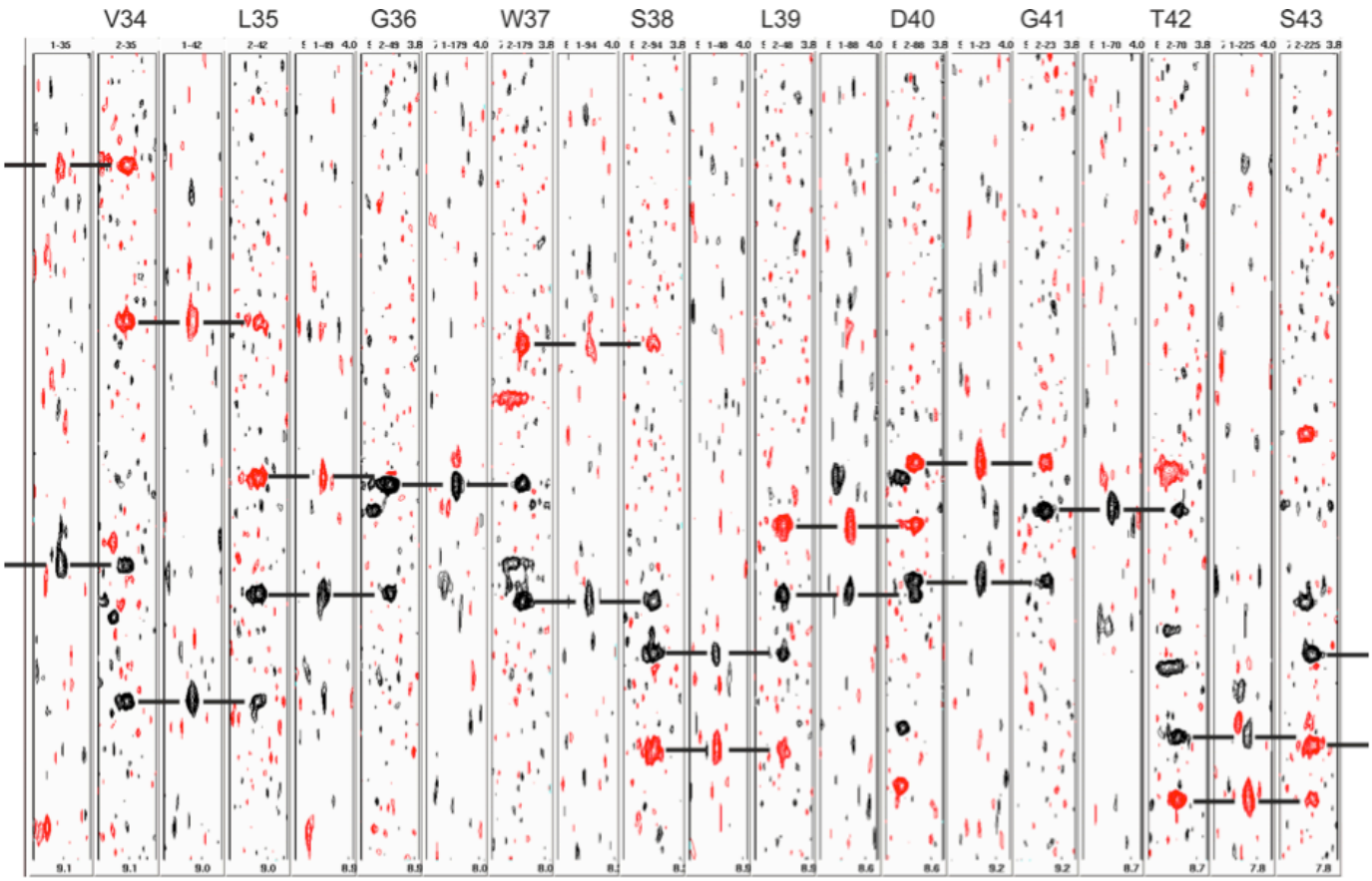


Figure 6.4: **A typical strip plot of TROSY-HNCACB (i-1 & i) and TROSY-HN(CO)CACB (i-1) data.** C^{α} peaks are shown in black, and C^{B} peaks are shown in red. The lines connecting the peaks represent sequential connectivity for the stretch of amino acids V34 to S43 of the sNT-Als1 sequence.

6.4.2. Manual Assignments

After the MARS run, manual assignment was carried out using supplementary peak information from the $^{1}H^{15}N$-NOESY-HSQC experiment, which was useful to verify positions of either weak, missing or overlapping peaks from backbone assignment experiments (Fig 6.5). This NOESY experiment detects NH-NH nuclear Overhauser effects (NOEs), in which in which ^{1}H nuclei are coupled through-space if they are within a distance of <5Å (Anderson & Freeman 1962). The cross-peaks on the diagonal part of the spectrum represent correlation between NH protons that are in that close proximity, regardless of whether they are close due to their position in the protein's primary sequence.

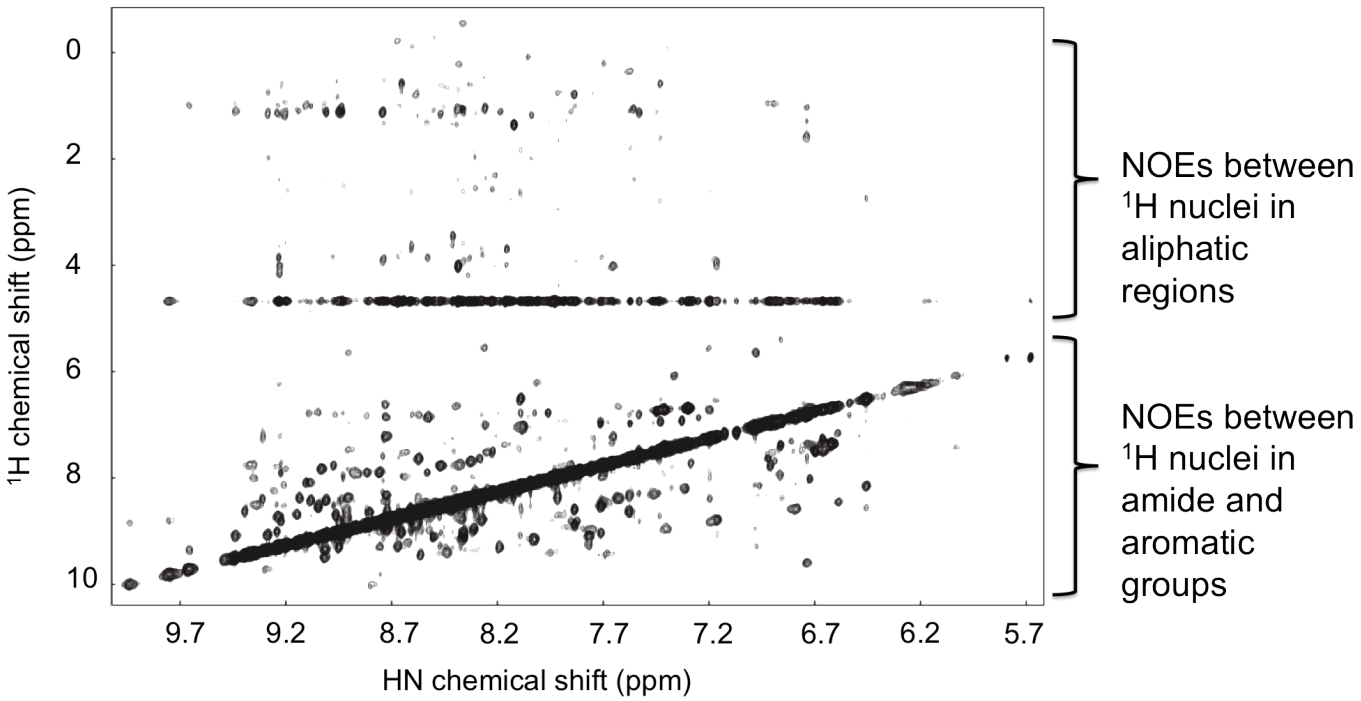


Figure 6.5: **The $^{1}H^{15}N$-NOESY spectrum of sNT-Als1.**
The cross-peaks on the diagonal represent through-space correlations between ^{1}H nuclei of backbone HN groups with other ^{1}H nuclei, in order to decipher groups that are approximately less than 5Å apart. The peaks from 0-5ppm on the ^{1}H axis between NOEs between ^{1}H nuclei on aliphatic groups. Peaks between 6-10ppm on the ^{1}H axis show NOEs between ^{1}H nuclei on amide and aromatic groups.

The ^{1}H-^{15}N-TROSY-HSQC of NT-Als1 (Yan et al 2010) was highly similar to that of sNT-Als1, so it was possible to overlay the former spectrum onto the latter sNT-Als1 spectrum in order to gain information regarding peaks in the same or similar positions (See Figure 6.6). Although some regions of the spectra were highly similar, there is some of high variation. This was expected due to the NT-Als1 construct being longer, and also that 76% of residues had been assigned from a non-deuterated sample in which more NH groups with ^{1}H protons were present, with the remainder assigned with a deuterated NT-Als1 sample.

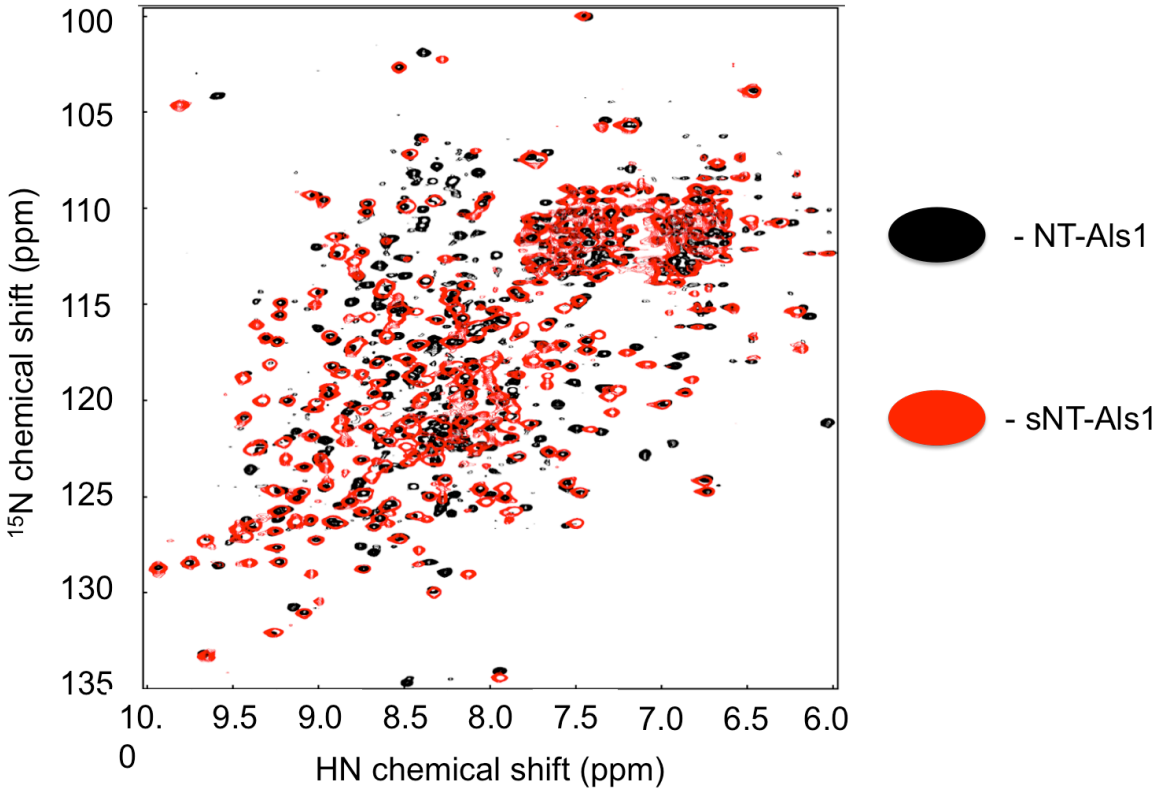


Figure 6.6: **An overlay of the ^{1}H-^{15}N TROSY-HSQC spectrum of deuterated sNT-Als1 with that of non-deuterated NT-Als1**.
The deuterated sNT-Als1 spectrum (red) is overlayed with non-deuterated NT-Als1 (black). It is noticeable that some peaks are similar in position, especially those of the protein surface. The coordinates of such similar peaks were highly useful for the manual assignment of sNT-Als1. However, some peaks in NT-Als1 have differing positions from sNT-Als1, and there is some extensive spectral overlap in similar parts of the spectra. More peaks can be seen for NT-Als1, probably due to this being a longer construct, and also due to complete back-exchange of amide protons (^{1}H-^{15}N-TROSY-HSQC from Yan et al 2010).

6.4.3. Assignment Results

From these experiments and data analysis, it was possible to assign approximately 55% of the backbone residues. Their positions are shown on the ^{1}H-^{15}N TROSY-HSQC of sNT-Als1 (Fig 6.7). In order to use NMR spectroscopy for studying the ligand binding behaviour of sNT-Als1 in the future, the important peaks to assign are those that represent residues of the PBC, especially invariant K59. This residue has been confidently assigned in the sNT-Als1 construct, and as this is the residue which forms the salt bridge interaction with peptide ligands, it is the most important to assign if ligand-binding interactions are to be studied using NMR. Nonetheless, improvements in the conditions to obtain fully exchanged amide protons for deuterated sNT-Als1 samples should allow recording of tridimensional NMR spectra for complete assignment of backbone resonances.

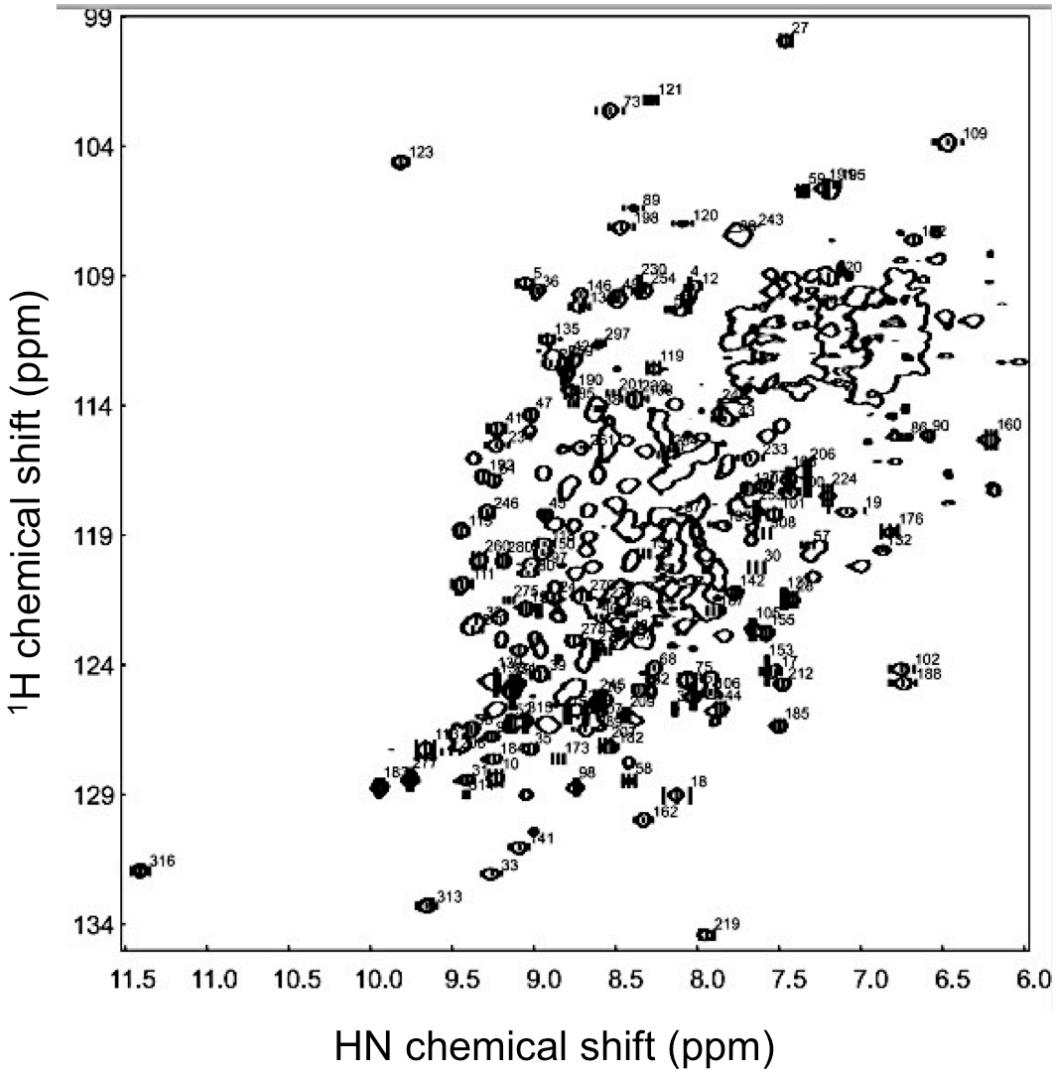


Figure 6.7: **^{1}H-^{15}N-TROSY-HSQC of sNT-Als1 with approximately 55% of assigned peaks**.
This was carried out using a combination of sequential backbone assignment experiments with MARS automated assignment, spectral overlay with the assigned ^{1}H-^{15}N-TROSY-HSQC of NT-Als1 and correspondence of sequential ^{15}N resonances in the ^{1}H-^{15}N-NOESY experiment relative to assignments for NT-Als1 (Yan et al 2010).

The residues assigned and their relation to the primary sequence and topology of sNT-Als1 is displayed in Figure 6.8. The residues assigned with medium-to-high confidence using the backbone assignment experiment and MARS correspond to 24% of the sNT-Als1 sequence, whilst those assigned from the spectral overlay of the ^{1}H-^{15}N-sNT-Als1 TROSY with that of Robert Yan's ^{1}H-^{15}N-NT-Als1 equated to 12.6% of the sNT-Als1 sequence. Assignments of medium confidence from both mine and Robert Yan's backbone assignment data equate to 17.9% of the sNT-Als1 sequence.

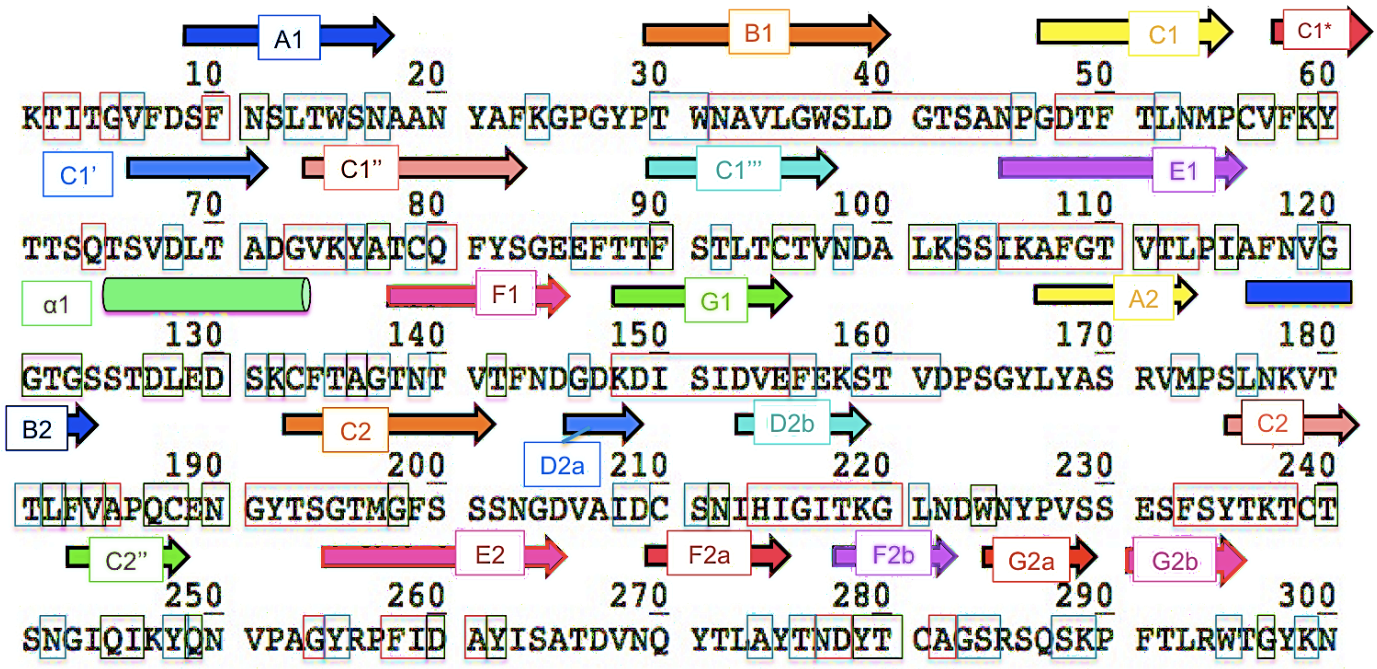


Figure 6.8: **The result of the backbone assignment experiments as displayed on the sNT-Als1 topology diagram**.
Residues boxed with red are those that were assigned with high confidence using MARS, residues boxed with green are those confidently assigned from spectral overlay with backbone assignment data courtesy of Robert Yan, and the residues coloured cyan show those that had been assigned with medium confidence from both mine and Robert Yan's backbone assignment data.

Secondary structural elements that had been completely or almost completely assigned include the B1, C1*, E1, G1, D2b and C2 β-strands. The positions of the assigned residues can be viewed with respect to the cartoon diagram of the sNT-Als1 structural model from Chapter 5 (see Figure 6.9). The conserved K59 that is crucial for the interaction of Als proteins with the C-terminal carboxyl groups of peptide ligands, was the sole amino acid of the PBC to have been assigned. It is therefore possible to use the sNT-Als1 construct in future 1H-15N-TROSY-HSQC ligand titration experiments, as chemical shift changes involving the peak assigned for K59 should indicate sNT-Als1 PBC interaction with the ligand. However, ideally, more residues of the PBC should be assigned in order to carry out more detailed ligand-binding experiments and to study the dynamics of ligand-PBC interactions.

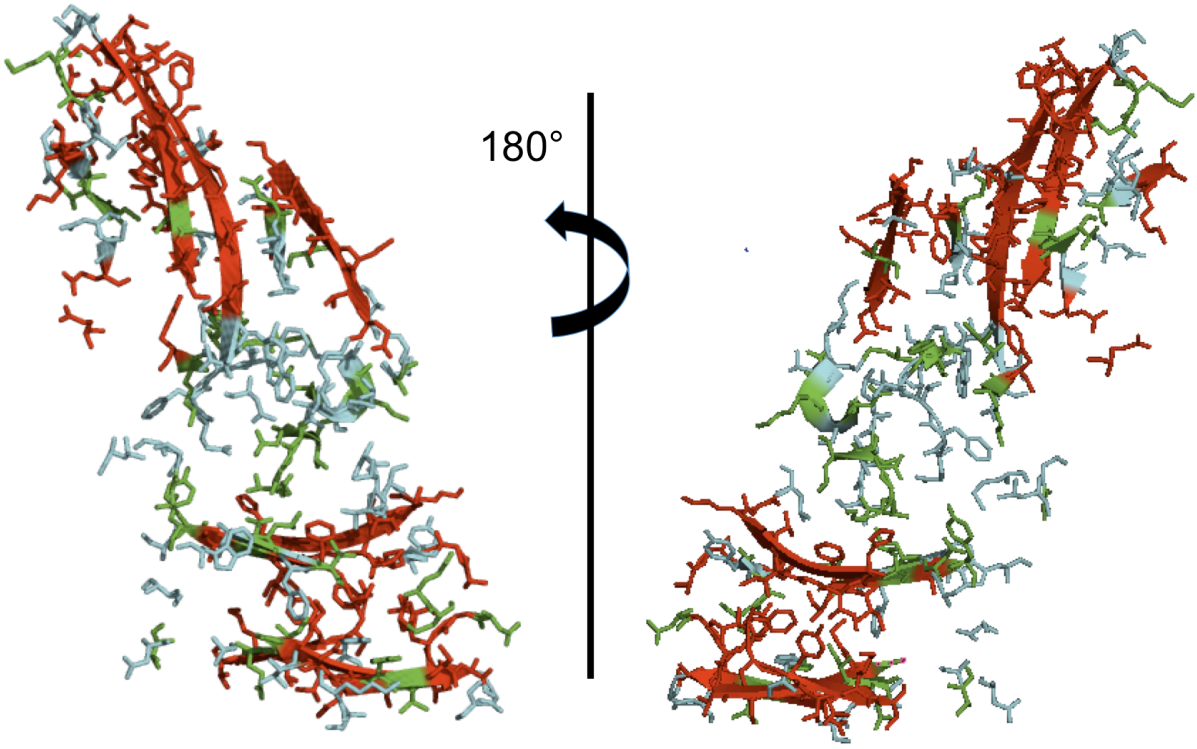


Figure 6.9: The result of the backbone assignment experiments as displayed on the sNT-Als1 cartoon diagram.
Residues coloured red show those that were assigned with high confidence using MARS, residues coloured green were those confidently assigned from spectral overlay with backbone assignment data courtesy of Robert Yan (Yan et al 2010), and the residues coloured cyan show those that had been assigned with medium confidence from both mine and Robert Yan's backbone assignment data. Figure made using PyMOL.

7. Ligand Binding Studies of sNT-Als1

Although the sNT-Als1/Hepta-T crystal structure in Chapter 4 provided information regarding the sNT-Als1 PBC interaction with Hepta-T, questions remained regarding the binding specificity of the sNT-Als1 PBC for either biologically relevant peptides or drug-like molecules. With efforts focused on optimization of co-crystallography conditions for sNT-Als1/Hepta-T, and the NMR spectroscopic backbone assignment of sNT-Als1 for future ligand-binding studies, a high-throughput screen was needed to identify target peptides and drug-like molecules likely to yield a positive interaction with sNT-Als1. Here, the technique Differential Scanning Fluorimetry (DSF) provided a valuable tool as a first-screen in order to identify such potential compounds.

7.1. Differential Scanning Fluorimetry

The technique Differential Scanning Fluorimetry (DSF) is based on the observation that the thermal stability of a protein macromolecule increases upon its interaction with a ligand (Schellman 1975). The Gibbs free energy of unfolding, ΔG_u, of a protein is a thermodynamic function, as it decreases when temperature increases alongside protein stability (Schellman 1997, Niesen et al 2007). The melting temperature of a protein (T_m) is the point of equilibrium between the concentrations of unfolded and folded protein, and is assigned the ΔG_u of 0. The binding of a ligand results in an increase in the ΔG_u for that particular protein, which can increase its T_m (Niesen et al 2007).

Protein samples mixed with the fluorescent dye SYPRO orange with and without ligand were subjected to controlled thermal denaturation in a qPCR machine. SYPRO orange has a high fluorescence intensity when in hydrophobic environments, for example the non-polar patches of unfolded proteins that are exposed as proteins are subjected to melting temperatures. Fluorescence intensity increases upon binding to these patches, and a sigmoidal curve representing the melting profile of a protein can be plotted due to the relationship between fluorescence intensity and temperature (Fig 7.1). The curve represents the transition between folded and unfolded (melted) states, with the mid-point corresponding to the T_m, calculated via Equation 7.1:

$$y = \mathrm{I_{min}} + \frac{(\mathrm{I_{max}} - \mathrm{I_{min}})}{1 + \exp\left(\frac{\mathrm{Tm} - x}{a}\right)}$$

Equation 7.1: The Equation For the Fitting of the Sigmoidal Melting Curve of a Protein According to the Relationship Between Fluorescence Intensity and Temperature.
Where I_{min} stands for minimum intensity, I_{max} stands for maximum intensity and *a* represents the slope of the curve for T_m (Niesen et al 2007).

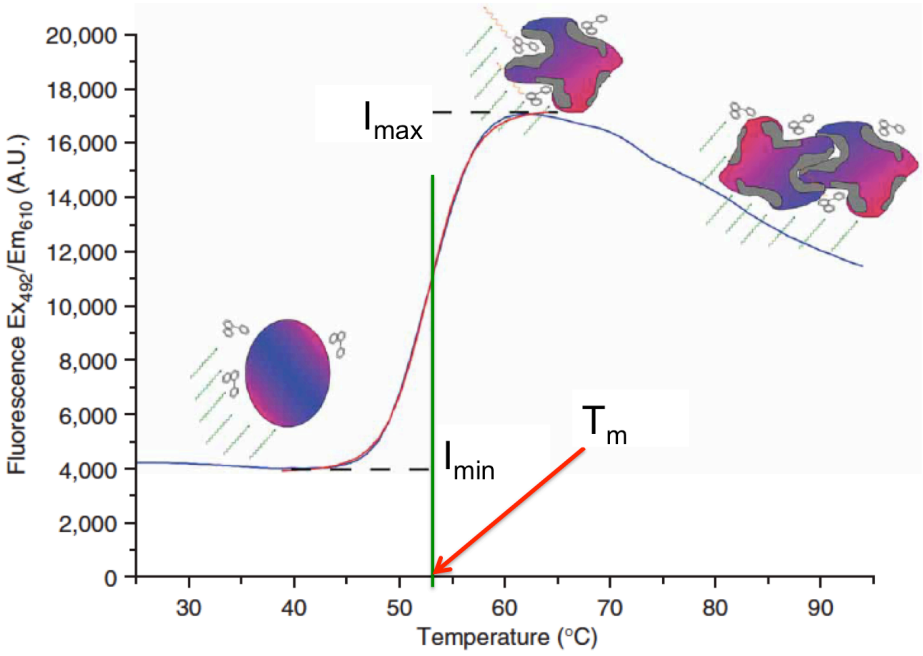


Figure 7.1: Plotting fluorescence intensity against temperature yields a sigmoidal curve that represents the melting profile of a protein.
With I_{min} and I_{max} corresponding to minimum and maximum fluorescence intensities, and also folded and unfolded protein, respectively. T_m represents point of equilibrium between folded and unfolded protein (Figure modified from Niesen et al 2007).

It is important to note that ΔT_m of a protein after the addition of a ligand does not correspond to binding affinity, and may not reflect the binding of the ligand to the native state, but perhaps an interaction with the unfolded state. Plus, large ΔT_m's may indicate ligand binding via hydrophobic interactions, with smaller ΔT_m's under-representing effects of hydrogen-bonding (Niesen et al 2007). Whilst this technique has drawbacks, the technique is a cost-effective and efficient method to screen large libraries of ligands in order to identify potential binding partners for sNT-Als1 for further study using other, more precise methods such as X-ray co-crystallography, NMR-spectroscopy and ITC.

7.2. Peptide Ligand Screening using DSF

Biologically relevant ligands for the Als proteins have been elucidated from past literature, often from studies using *C. albicans* or *S. cerevisiae* cells expressing Als proteins and test for their adhesion binding partners. For example, Als1, Als3 and Als5 expressed in *S. cerevisiae* were able to bind beads coated with fibronectin, type IV collagen and casein proteins (Gaur & Klotz 1997, Gaur et al 2002, Sheppard et al 2004). *S. cerevisiae*, when expressing Als5 was shown to bind bovine serum albumin (Gaur et al 2002), and when expressing Als3 was demonstrated to adhere to fibrinogen (Nobbs et al 2010). In a separate study, Als3 has been shown to adhere to ferritin, with important implications for a potential role in the acquisition of iron by *C. albicans* and subsequent host cell damage. However, the same study reported that Als1 was unable to bind to ferritin (Almeida et al 2008).

Als proteins not only have a role in *C. albicans* adherence to ECM and serum factors, but also to host cell surface membrane proteins that induce endocytosis, a mechanism that *C. albicans* uses to invade host tissues. Als3 has been shown to induce the endocytosis of *C. albicans* cells into host brain endothelial cells via an interaction with gp96 (Liu et al 2011). Similarly, it can induce endocytosis via adhering to EGFR, HER2 and N-cadherin in epithelial cells, and E-cadherin in epithelial cells (Phan et al 2007, Moreno-Ruiz et al 2009, Zhu et al 2012). As *C. albicans* null mutants for Als3

still retain an ability to induce endocytosis, albeit a significantly reduced ability, this suggests that other proteins may be involved in induced endocytosis (Wachtler et al 2012).

The formation of mixed-species biofilms of *C. albicans* with *S. gordonii* has been reported to be the result of the Als3 interaction with the SspB protein on *S. gordonii*. (Silverman et al 2010).

From these previous studies, it has been concluded that Als3 contributes the most prominently to the adhesion and induced endocytosis of *C. albicans* cells, with Als1 not being able to bind to ligands that Als3 can bind to, e.g. ferritin (Almeida et al 2008). Furthermore, although the N-terminal adhesion domain of the Als proteins was elucidated to be responsible for specific protein-protein interactions (Sheppard et al 2004), these past studies of Als protein binding partners did not reveal where or how these interactions occurred, with theories being put forward that the AFR region of the N-terminal domain was responsible for Als host-cell protein adhesion (Lipke et al 2012).

Fortunately, it was elucidated that the Als proteins contain a PBC that enables them to interact with the C-terminal carboxylates of a variety of peptide ligands via the invariant K59 side-chain, with ligand-binding experiments confirming that peptides with C-terminal amidation were unable to interact with both Als1 and Als9.2 (Salgado et al 2011). Mutational studies in which the function of the Als3 PBC was abolished, resulted in *C. albicans* cells with an identical reduced capacity for adhesion to buccal epithelial, human umbilical vein endothelial and FaDu cells when compared with ALS3-null mutant cells. Plus, mutations of the Als3 AFR region, in which its amyloidogenic propensity was disrupted, actually increased *C. albicans* adhesion to host cells in comparison to the Als3 Wild-type cells. This study confirmed that it is the PBC of the N-terminus, and not the AFR region, which is responsible for specific protein-peptide interactions of the Als proteins, and perhaps the amyloidogenic aggregation produced by the AFR may decrease access of the PBC to host cell surface ligands (Lin et al 2014).

Mutation of S170 at the opening of the cavity in NT-Als3 to a residue with a larger side chain (Y170) blocked the binding of it to Hepta-Threonine ligand when tested using ITC. This finding provided additional evidence that the PBC is essential for protein-peptide ligand binding interactions and that its function is independent from the AFR (Lin et al 2014).

In order to make a connection between the cell-biology and biochemical/structural studies of Als protein ligand-binding behaviour, and study the specificity of the Als1 PBC, the aim was to test the ability of sNT-Als1 to bind to C-termini of biologically relevant proteins discussed earlier. This led to the synthesis of C-terminal peptide sequences of interesting protein ligands (Cota lab) to be used in a DSF binding study.

C-terminal peptides from bovine α-s1 casein, human fibronectin and human collagen α-1 were synthesised (Cota lab) in order to test the C-termini of ECM proteins for binding to sNT-Als1. Serum factor proteins BSA, HSA and human fibrinogen-Y subunit, had their C-termini sequences synthesised and also screened. A mutant form of the BSA C-terminal peptide was also tested, in which peptide position 4 (Counting from the C-terminus) had been mutated from threonine to tyrosine.

C-terminal peptides from two ferritin chains were tested for interaction with sNT-Als1, and similarly were C-terminal peptides of Human endocytotic factors gp96, EGFR and E-cadherin plus the *S. gordonii* SspB protein C-termini. Plus, as oropharyngeal candidiasis is prevalent amongst HIV-patients (Fidel Jr 2011), attempts were made to find binding partners for sNT-Als1 that reside in the oral cavity, hence the C-terminal peptide of Human salivary amylase was also tested for ligand-binding interaction with sNT-Als1.

Peptide sequences are listed in Table 7α, and they are each comprised of 8 amino acids i.e. the last 8 amino acids of the C-termini of biologically relevant proteins. They are designed so that 5-6 of C-terminal residues can interact with the PBC, as the PBC is only capable of interacting with a maximum of 6 residues (Salgado et al 2011), and also so that the free NH_3 group of the N-terminus is away from the rest of the structure in order to prevent it from making any unnecessary interactions.

For each DSF experiment, sNT-Als1 protein at a concentration of 2μM, was mixed thoroughly with peptide at a concentration of 2mM, and Sypro Orange dye in a 1:1000 dilution to a total well

volume of 20µl within blue-white RT-PCR plates. Each plate also contained replicate negative control samples of sNT-Als1 protein without peptide ligand. Two peptide screens were carried out, the second one involved the inclusion of 0.1% DMSO in order to facilitate the peptide solubility. The first peptide screen was carried out without 0.1% DMSO.

In Table 7.1, the experimental samples are listed according to the name of the peptide used plus its peptide sequence. The table shows the T_m values of the replicates, overall average T_m values and ΔT_M values (alongside standard deviations associated with each measured value). The ΔT_M values were calculated by comparing the difference between the average T_m values of the test samples with the average T_m values of the negative control sample that had been included in the same experimental plate. Unless otherwise stated, three replicates of each peptide test condition had been carried out. The Hepta-T peptide was used as a positive control as this has been shown to bind to NT-Als1 from NMR chemical shift perturbation and ITC studies (Salgado et al 2011, Lin et al 2014), and also showed a high ΔT_m between melting curves (See Table 7.1).

In order to calculate T_m values for each replicate, a graph of relative fluorescence units against temperature was constructed, and data fitted to a sigmoidal curve according to Equation 1. Standard deviations for each T_m measurement, average T_m calculation and ΔT_m value, were calculated according to Equation 7.2. A significant ΔT_m value, indicating a highly probable interaction between peptide ligand and protein, should be larger than two standard deviations of the average T_m value for the negative control sample.

$$SD = \sqrt{\frac{\sum(X_i - M)^2}{N - 1}}$$

Equation 7.2: The Standard Deviation Equation for each T_m and ΔT_m Value.
Where SD represents standard deviation, X_i represents a replicate measurement value (e.g. T_m), M represents the average value (e.g. average T_m) and N represents the number of replicates.

Table 7.1: Peptides used in a DSF screen to test the probability of their binding to sNT-Als1.
Listed are the names of the peptides, their primary sequences, plus the T_m values of the replicates, the average T_m value for that test and the ΔT_M values (± standard deviation for each value). The ΔT_M values were calculated from the difference of the Average T_m of the test samples, from the average T_m of the negative control (sNT-Als1 protein without the addition of peptide). The table has been split depending on whether 0.1% DMSO had been included as part of the conditions and results ranked in order of descending ΔT_M values, from highest to lowest, for each screen.

Peptide Name	**Peptide Sequence**	**Tm (°C) of Replicates ± SD**			**Average Tm (°C) ± SD**	**ΔTm (°C) ± SD**
		N = 1	**N = 2**	**N = 3**		
1st Peptide Screen: 0% DMSO Used						
No peptide	N/A	45.439 ± 0.026	45.519 ±0.039	45.139 ±0.062	45.365 ± 0.200	0.000 ± 0.359
Hepta-T	TTTTTTT	51.964 ± 0.067	52.891 ± 0.065	52.442 ± 0.062	52.423 ± 0.527	7.058 ± 0.326
Bovine Serum Albumin-CT (BSA-CT)	VSTQTALA	52.065 ± 0.040	52.395 ± 0.043	52.787 ±0.056	52.416 ± 0.362	7.051 ± 0.471
Bovine α-s1 casein-CT (Bαs1-CAS-CT)	EKTTMPLW	49.933 ± 0.035	50.844 ± 0.036	50.657 ± 0.031	50.478 ± 0.481	5.112 ± 0.540
Human EGFR-CT (hEGFR-CT)	QSSEFIGA	47.626 ± 0.035	48.581 ± 0.036	48.811 ± 0.037	48.339 ± 0.628	2.973 ± 0.613
Human fibronectin-CT (hFIB-CT)	ADREDSRE	47.259 ± 0.028	48.539 ± 0.014	48.871 ± 0.028	48.223 ± 0.851	2.857 ± 0.711
Human fibrinogen-γ-subunit CT (hFIBRγ-CT)	SLYPEDDL	47.672 ± 0.047	48.149 ± 0.027	48.059 ± 0.033	47.960 ± 0.253	2.594 ± 0.399
Human E-cadherin-CT (hEcad-CT)	MYGGGEDD	46.765 ± 0.024	47.911 ± 0.044	48.040 ± 0.045	47.572 ± 0.702	2.207 ± 0.647
Human collagen-α-1-CT (hColl-α1-CT)	DGGRYYRA	46.951 ± 0.026	47.807 ± 0.082	47.477 ± 0.059	47.412 ± 0.432	2.046 ± 0.513
Human gp96-CT (hgp96-CT)	STAEKDEL	46.743 ± 0.026	47.226 ± 0.043	48.051 ± 0.105	47.340 ± 0.661	1.974 ± 0.629
Human and equine spleen ferritin light chain-CT (h&eFerritin-LC-CT	ERLTLKHD	47.094 ± 0.046	47.452 ± 0.080	47.134 ± 0.067	47.227 ± 0.196	1.861 ± 0.356
Equine spleen ferritin heavy chain-CT (eFerritin-HC-CT)	KHTLGECD	46.395 ± 0.091	47.108 ± 0.088	47.610 ± 0.192	47.038 ± 0.611	1.673 ± 0.605
Strep. Gordonii SspB-CT (SspB-CT)	GLAKRKED	46.553 ± 0.026	46.858 ± 0.041	46.895 ± 0.033	46.769 ± 0.188	1.403 ± 0.349
2nd Peptide Screen: 0.1% DMSO Used						
No peptide	N/A	52.760 ± 0.039	53.021 ± 0.028	52.800 ± 0.030	52.860 ± 0.141	0.000 ± 0.297
Hepta-T	TTTTTTT	57.877 ± 0.058	58.162 ± 0.0630	58.525 ± 0.069	58.188 ± 0.324	5.328 ± 0.440
Bovine Serum Albumin T/Y Mutant (BSA-Y-CT)	VSTQYALA	55.562 ± 0.059	55.820 ± 0.070	55.616 ± 0.067	55.666 ± 0.136	2.805 ± 0.266
Human Serum Albumin-CT (HSA-CT)	ASQAALGL	54.907 ± 0.041	55.314 ± 0.051	55.547 ± 0.048	55.256 ± 0.323	2.395 ± 0.440
Human salivary amylase-CT (hSALA-CT)	AIHAESKL	53.801 ± 0.021	53.472 ± 0.031	52.320 ± 0.039	53.501 ± 0.287	0.651 ± 0.415

The peptides that resulted in the highest ΔT_m values included Hepta-T (7.058°C ± 0.326), BSA-CT (7.051°C ± 0.326) and Bαs1-CAS-CT (5.112°C ± 0.471). These high ΔT_m values, of which are significantly more than two SDs apart from the average T_m value of the negative control (SD = 0.2), suggest a high probability of ligand-binding interaction between the peptides and sNT-Als1. DSF melting curves for the addition of these peptides to sNT-Als1, vs. the negative control are shown in Figures 7.2 and 7.3.

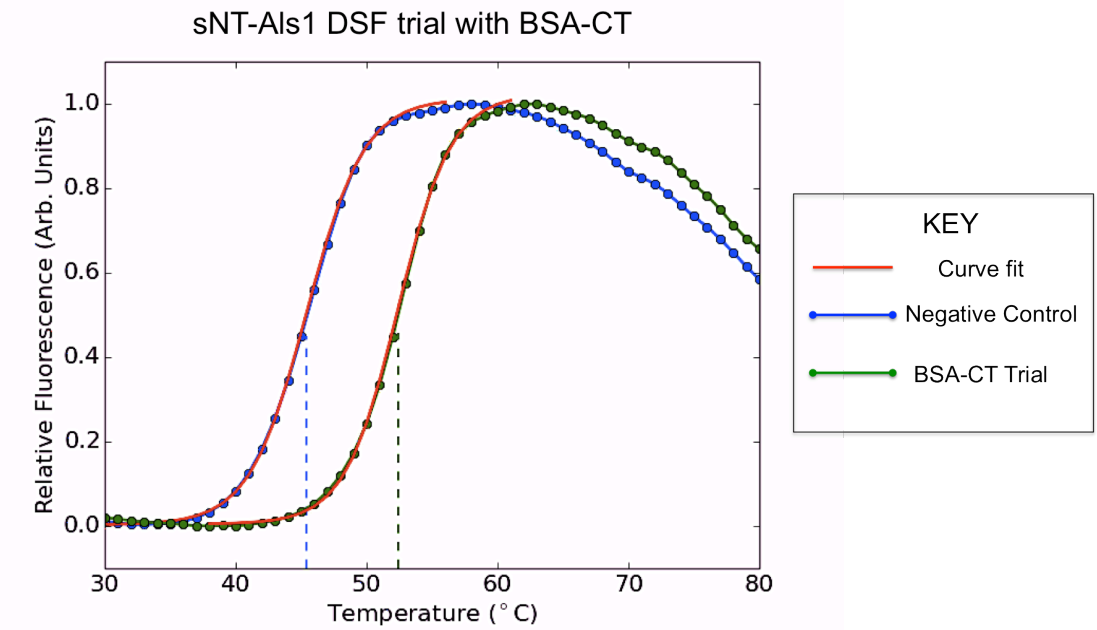


Figure 7.2: **A comparison of the sNT-Als1 negative control melting curve (N = 2) with the melting curve of the BSA-CT peptide trial (N = 2).**
Shown here is the plotted relative fluorescence data for one of the sNT-Als1 negative control replicates (N = 2, in blue) alongside the plotted relative fluorescence data for one of the sNT-Als1 plus BSA-CT peptide trial replicates (N=2, in green). In order to derive the T_m values for each trial, a curve has been fitted to the data using the equation y = Imin + (Imax – Imin)/1+ exp (T_m – x/a). The dotted lines show how midpoint has been calculated, and been used to derive the T_m values. For the negative control, a T_m of 45.519°C ± 0.039 has been calculated, and for the BSA-CT peptide trial, a T_m of 52.395°C ± 0.043 has been derived.

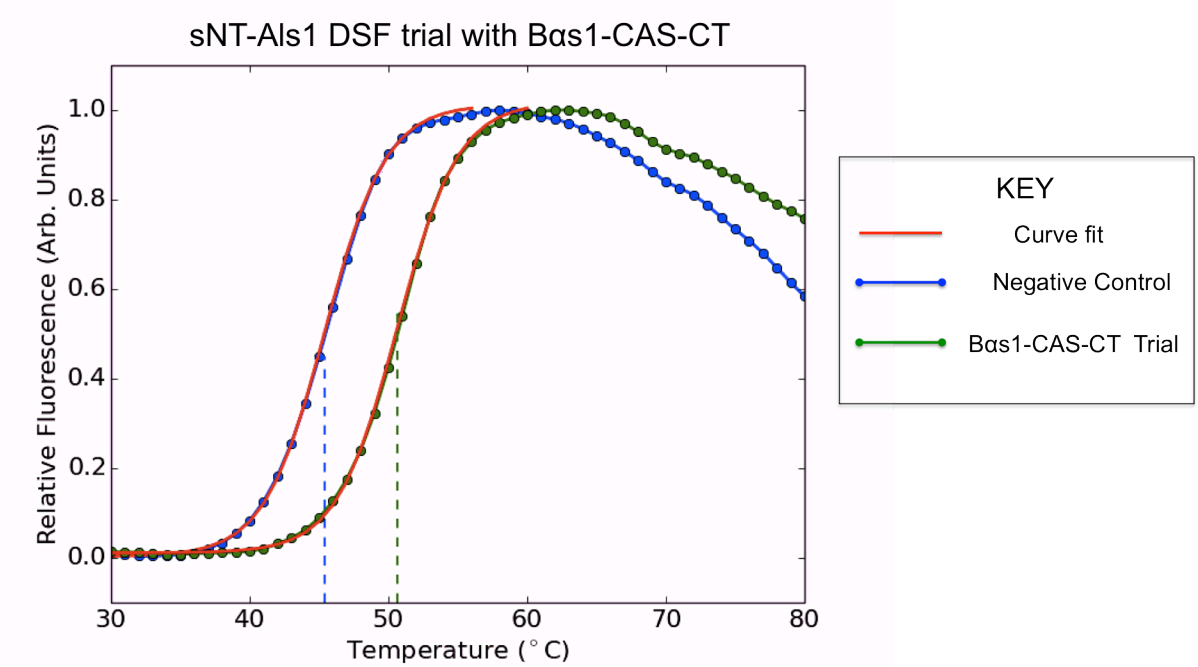


Figure 7.3: **A comparison of the sNT-Als1 negative control melting curve (N = 2) with the melting curve of the Bα1-CAS-CT peptide trial (N = 3).**
Shown here is the plotted relative fluorescence data for one of the sNT-Als1 negative control replicates (N = 2, in blue) alongside the plotted relative fluorescence data for one of the sNT-Als1 plus BSA-CT peptide trial replicates (N=2). In order to derive the T_m values for each trial, a curve has been fitted to the data using the equation y = Imin + (Imax – Imin)/1+ exp (T_m – x/a). The dotted lines show how midpoint has been calculated, and been used to derive the T_m values. For the negative control, a T_m of 45.519°C ± 0.039 has been calculated, and for the BSA-CT peptide trial, a T_m of 50.657°C ± 0.031 has been derived.

The ΔT_m for the addition of HSA-CT (2.395°C ± 0.440) did not match BSA-CT, in that there was a 4.666°C difference between their two respective ΔT_m values. However, the trial with the HSA-CT peptide had been carried out in the 2nd peptide screen plate, in which samples contained 0.1% DMSO, whilst the BSA-CT trial contained no DMSO. Additionally, the HSA-CT peptide solution showed some precipitate, indicating that a lower peptide concentration than expected may have been present in the final DSF plate.

Interestingly, the addition of 0.1% DMSO shifted the average T_m values of the sNT-Als1 negative controls from 45.365°C ± 0.200 to 52.860°C ± 0.141 – a difference of approximately 7.495°C which may indicate DMSO as a useful additive for improving the stability of the sNT-Als1 construct in future studies. However, it is possible that DMSO in the screening conditions may be present in the sNT-Als1 PBC, and so may mask the full effect of interaction with the peptides present. It is therefore not possible to make a valid comparison between BSA-CT and HSA-CT without trialling these both in the presence and absence of 0.1% DMSO, and furthermore, carrying out ligand-binding studies of higher precision such as with NMR Spectroscopy.

However, the Hepta-T peptide was used in the second screen as a positive control, resulting in a ΔT_m of 5.328°C ± 0.440. This reveals an approximately 1.73°C difference In comparison to the trial with Hepta-T in the 1st peptide screen (0% DMSO), of which a ΔT_m of 7.058°C ± 0.326 was achieved.

A ΔT_m of 2.805 ± 0.266 was seen for BSA-Y-CT peptide trial, which had been carried out in the second peptide screen. In the first peptide screen, it can be noted that the ΔT_m values for Hepta-T and BSA-CT are highly similar (both approximately 7°C). However, in the second peptide screen with 0.1% DMSO present, the ΔT_m of BSA-Y-CT has an approximately 2.523°C difference in comparison to Hepta-T. This may suggest that the addition of the tyrosine mutation at peptide position 4 of the BSA-CT peptide reduced the propensity for ligand-binding with sNT-Als1 in comparison to the original BSA-CT. However, again, it is not possible to make a precisely valid comparison between the two peptides without further screens both with and without 0.1% DMSO if possible.

The peptides hEGFR-CT (ΔT_m 2.973°C ± 0.613), hFIB-Ct (ΔT_m 2.857°C ± 0.711), hFIBRγ-CT (ΔT_m 2.594°C ± 0.399), hEcad-CT (ΔT_m 2.207°C ± 0.647) and hColl-α1-CT (ΔT_m 2.046°C ± 0.513) resulted in ΔT_m values that were more than two SD values apart (SD = 0.2) from the average T_m of the negative control sample. This does indicate that there are potential interactions between these peptides and sNT-Als1, but those that may have a weaker binding mechanism in comparison to the Hepta-T, BSA-CT and Bαs1-CAS-CT peptides.

The peptides that yielded the lowest ΔT_m values were hgp96-CT (ΔT_m 1.974 ± 0.629), h&eFerritin-LC-CT (ΔT_m 1.861 ± 0.356) eFerritin-HC-CT (ΔT_m 1.673 ± 0.605), SspB-CT (ΔT_m 1.403 ± 0.345) and hSALA-CT (ΔT_m 0.651 ± 0.415). Apart from the hSALA-CT peptide, however, the others resulted in ΔT_m values that were more than two SD values apart (SD = 0.2) from the average T_m of the negative control sample. Potential interactions may occur between such peptides, but their binding mechanism may have been much weaker in comparison to the rest of the peptides.

The trial with the hSALA-CT (ΔT_m 0.651 ± 0.415) was carried out in the presence of 0.1% DMSO, so it is possible that any weak binding of this peptide to sNT-Als1 may have been masked by DMSO. However, any binding is likely to have occurred non-specifically.

Again, it is important to highlight that ΔT_M results from these DSF technique are not directly related to binding affinity. DSF was used primarily to identify potential ligands to be validated with techniques that can measure the affinity constant, K_D, of a ligand-binding interaction, for example surface-plasmon-resonance (SPR) or isothermal titration calorimetry (ITC). Due to the imprecise nature of the DSF assay, it is also not possible to make solid conclusions about binding specificity without using X-ray co-crystallisation and NMR spectroscopy studies. However, some theoretical speculation in section 7.3, can provide clues and suggest ligands fur future structural studies.

7.3. The DSF Peptide Ligand-Binding Study in a Structural Context

The peptides that resulted in the highest ΔT_m's upon addition to sNT-Als1 include Bαs1-CAS-CT (peptide sequence EKTTMPLW), BSA-CT (peptide sequence VSTQTALA) and Hepta-T (peptide sequence TTTTTTT). The latter two peptides have small, hydrophobic residues in peptide positions 1-4 (numbering from the C-terminus). The Hepta-T peptide can be seen to interact with the sNT-Als1 PBC in the crystal structure (Chapter 5), with its side chain positions sterically favourable with regards to residues that line the PBC. Using the Mutagenesis wizard on the program PyMOL, it is possible to 'mutate' each residue of the Hepta-T peptide in order to predict the shape and structure of each of the peptides tested. When 'mutating' Hepta-T to BST-CT sequence VSTQTALA, residues at positions 1-4 should not clash sterically with sNT-Als1 PBC residues of the A1-B1 loop, the Y271 side-chain of the F2a strand and the A2 Y168 side-chain (Fig 7.4).

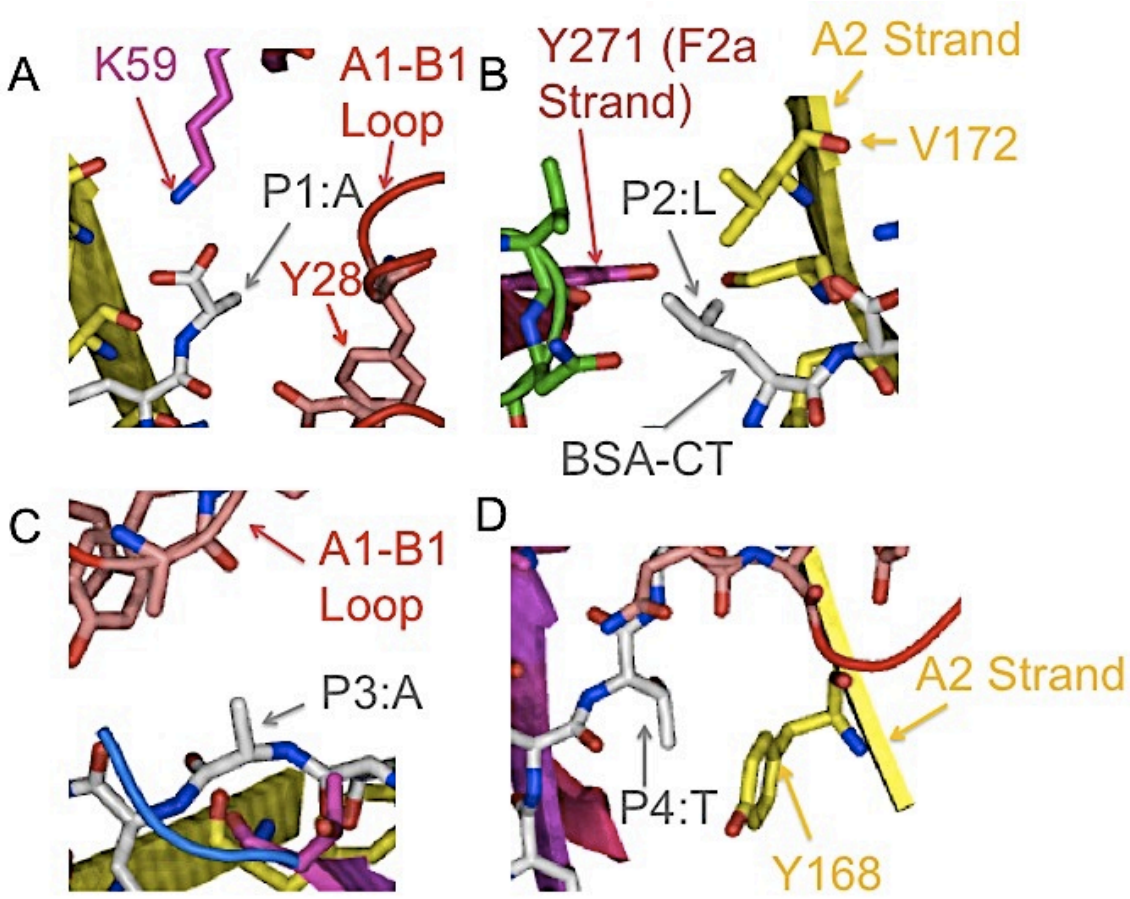


Figure 7.4: Theoretical representation of the BSA-CT in the sNT-Als1 PBC.
Carried out using the Mutagenesis wizard in PyMOL. BSA-CT (grey) has at position P1 an alanine, P2 a leucine, P3 an alanine and P4 a threonine. A) The small side-chain of alanine at P1 would not clash sterically with residues of the A1-B1 loop, and similarly so for B) the side-chains of P2 threonine with nearby Y271, and C) P3 alanine with the A1-B1 loop and D) P4 with the Y168 side-chain of the A2 Strand (PyMOL).

The sequence of the α-s1-casein-CT (EKTTMPLW) peptide, which also resulted in a relatively high ΔT_m value, contains a tryptophan at position 1. Although this is has a bulky side chain, it may be able to form favourable hydrophobic interactions with Y21 and Y28 of the sNT-Als1 A1-B1 loop. At position 4, there is a methionine residue, which may avoid clashing with Y168 and may also be able to form favourable cation-π interactions with this tyrosine as such interactions have been documented in previous literature (Tatko & Waters 2004).

The peptides hSALA-CT (peptide sequence AIHAESKL), SspB-CT (peptide sequence GLAKRKED), eFerritin-HC-CT (peptide sequence KHTLGECD), h&eFerritin-LC-CT (peptide sequence ERLTLKHD) and hgp96-CT (peptide sequence STAEKDEL) resulted in the lowest ΔT_m values, indicative of potentially

weak or non-specific interactions of such peptides with sNT-Als1. The lowest ΔT_m value was obtained from the trial with hSALA-CT peptide. In a structural context, weaker interactions may have occurred due to peptide side-chains being incompatible due to steric clashes with the residue side chains present in the sNT-Als1 PBC.

The peptides SspB-CT, eFerritin-HC-CT and h&EFerritin-LC-CT all have an aspartic acid at peptide position 1 which could be both sterically or chemically unfavourable next to the hydrophobic Y21 and Y28 of the PBC A1-B1 loop (see Figure 7.5).

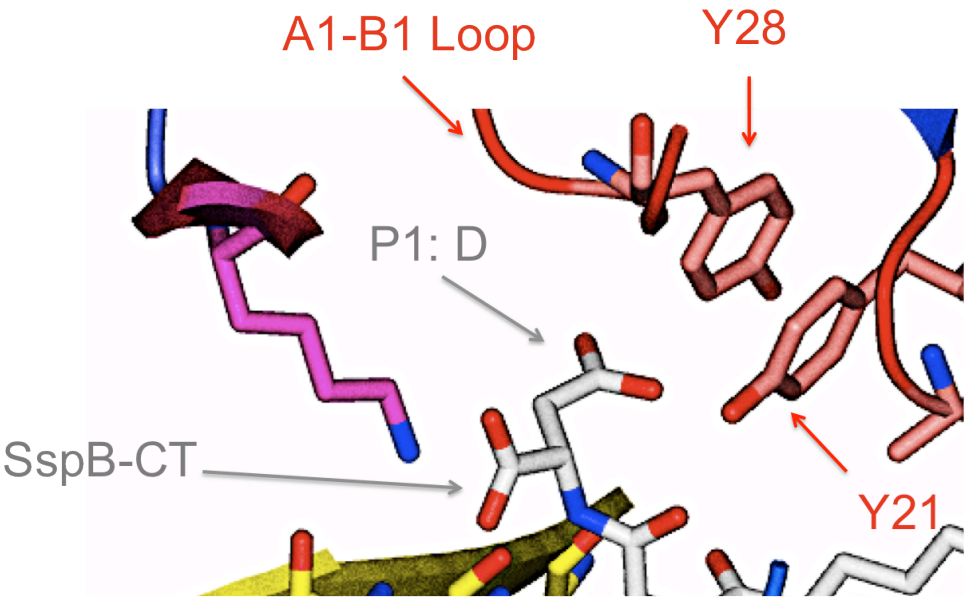


Figure 7.5: Theoretical representation of a potential steric clash between the aspartate residue at peptide position 1 of SspB-CT peptide in the sNT-Als1 PBC.
Carried out using the Mutagenesis wizard in PyMOL. SspB-CT peptide (coloured grey) has at position P1 an aspartate residue. This has the potential to clash sterically and perhaps chemically with the hydrophobic Y28 residue of the A1-B1 loop of the PBC.

The peptide hSALA-CT has a bulky lysine residue at peptide position 2, which may be sterically clash with PBC side chains Y271 of the F2a strand and N269 of the E2-F2a loop. Likewise, the SspB-CT and hgp96-CT peptides have a bulky glutamate residue in this position, which may also clash with these PBC residues (See Figure 7.6). Such large side chains at this position may mean more difficult access to the K59 residue for the carboxylic acid group of the peptide.

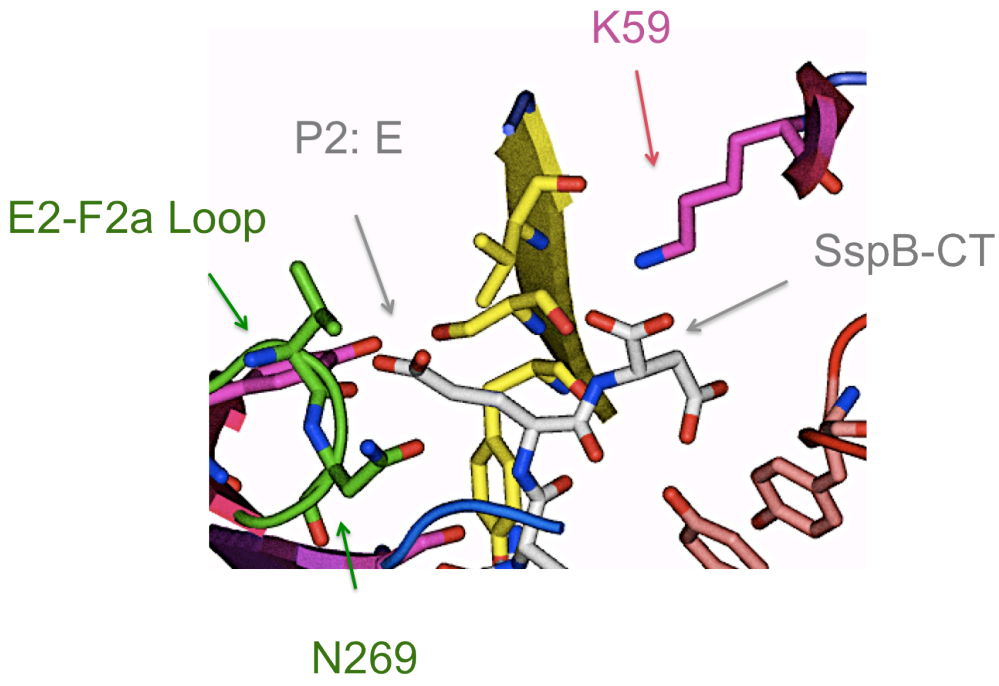


Figure 7.6: Theoretical representation of a potential steric clash between the glutamate residue at peptide position 2 of SspB-CT peptide in the sNT-Als1 PBC.
Carried out using the Mutagenesis wizard in PyMOL. SspB-CT peptide (coloured grey) has at position P2 a glutamate residue. This has the potential to clash sterically with the N269 residue of the E2-F2a loop of the PBC (coloured green). This clash may hinder the ability of the peptide C-terminal carboxylic acid from forming a salt-bridge interaction with K59.

Additionally, both SspB-CT and h&eFerritin-LC-CT peptides have a lysine at peptide position 3. The bulky lysine side-chain could result in a steric clash with the backbone and side-chain atoms of the sNT-Als1 N20 residue on the A1-B1 loop (see Figure 7.7).

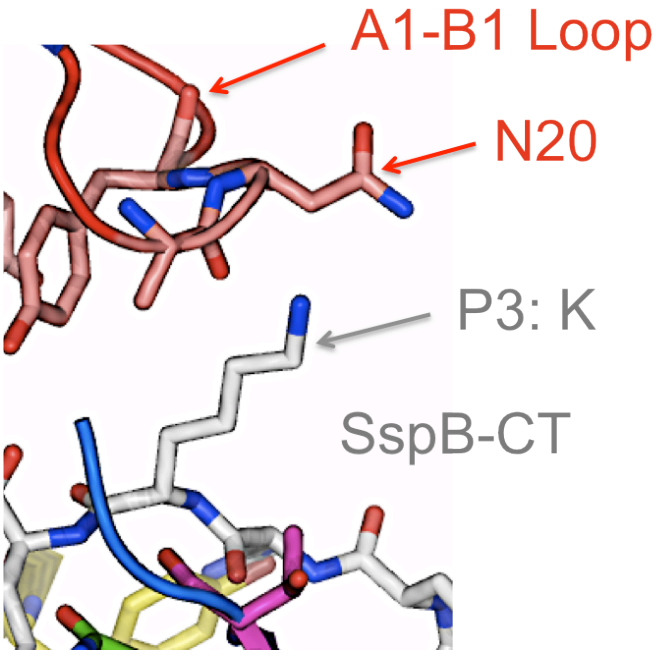


Figure 7.7: Theoretical representation of a potential steric clash between the lysine residue at peptide position 3 of SspB-CT peptide in the sNT-Als1 PBC.
Carried out using the Mutagenesis wizard in PyMOL. SspB-CT peptide (coloured grey) has at position P3 a lysine residue. This has the potential to clash sterically with atoms of the N20 residue of PBC A1-B1 loop (coloured red).

Finally, the peptides SspB-CT and hgp96-CT have bulky arginine and lysine residues, respectively, at peptide position 4. The side chains of these residues have the potential to cause a steric clash with Y168 of the A2 strand (See Figure 7.8). Similarly, for the hSALA-CT peptide, the bulky glutamate at position 4 could result in the same steric clash effect.

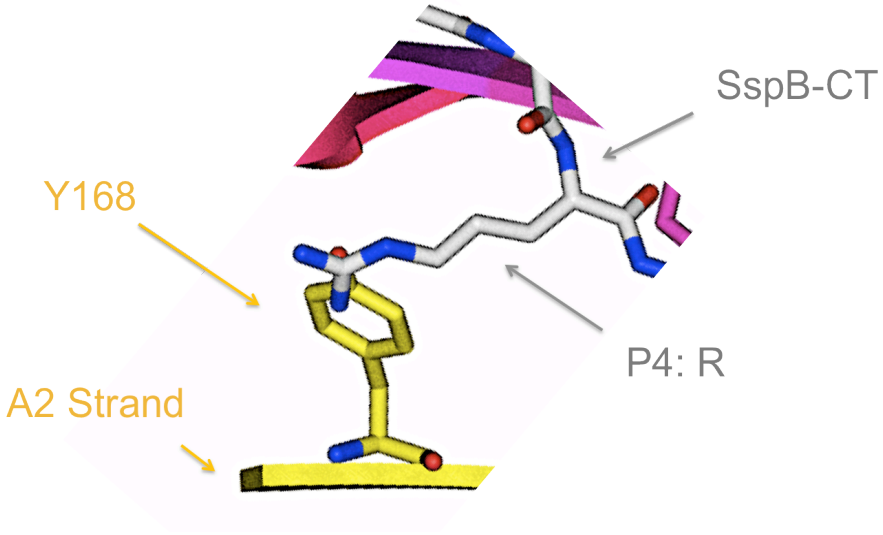


Figure 7.8: Theoretical representation of a potential steric clash between the arginine residue at peptide position 4 of SspB-CT peptide in the sNT-Als1 PBC.
Carried out using the Mutagenesis wizard in PyMOL. SspB-CT peptide (coloured grey) has at position P4 an arginine residue. This has the potential to clash sterically with the side chain of Y168 residue of the A2 strand (yellow) of the PBC.

However, it is important to state again that this theoretical speculation must be tested further as solid conclusions regarding steric clashes cannot be yielded from either DSF studies or from using PyMOL mutagenesis wizard. Moreover, the results that supported the theoretical speculation were from a screen using DSF; a technique that underestimates the contribution of non-hydrophobic interactions with regards to ligand binding. The peptides that resulted in low ΔT_m values may bind to sNT-Als1, but via less entropically driven binding mechanisms in comparison to the peptides with hydrophobic and/or small residues at peptide positions 1, 2, 3 and 4, that which yielded the highest ΔT_m values. Optimisation of the DSF technique with the inclusion of a larger library of peptides, or systematic testing of peptides with and without bulky groups at each peptide position, would be useful in order to make solid conclusions regarding peptide-binding specificity with sNT-Als1, and perhaps other sNT-Als1 constructs.

7.4. Small Drug-Like Molecules Ligand-Binding Screens

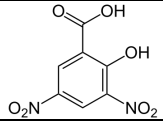
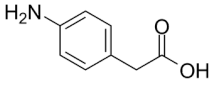
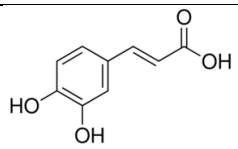
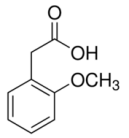
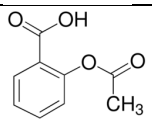
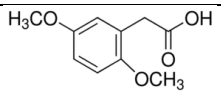
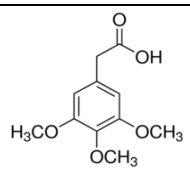
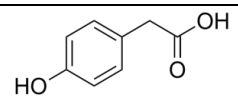
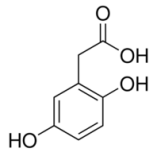
The technique DSF was not only used to test for protein-peptide ligand-binding interactions for sNT-Als1, but was also used to screen small, established drug compounds, and also drug-like compounds for interaction with sNT-Als1. Therapeutic molecules that are already in public use, and that have also been reported to decrease the production of *C. albicans* biofilms in a catheter model include Aspirin, Diclofenac and Ibuprofen (Alem & Douglas 2004). Plus, the Aspirin derivative Thimerosal has anti-*Candida* activity via its ability to kill hyphae (Wachtler et al 2012). As both Aspirin and Ibuprofen contain carboxylic acid groups, it was theorized that these might be able to bind to the K59 of the sNT-Als1 PBC. Therefore, DSF was carried out to test Aspirin and derivative drugs for potential interactions with sNT-Als1.

Additionally, a compound library of 22 phenylacetic acid derivatives (courtesy of Dr Jim Brannigan, University of York) were used in a DSF ligand-binding screen with sNT-Als1 in order to identify any lead compounds to potentially guide development of therapeutic molecules. Phenylacetic acid (PAA) was chosen as a starting point again due to the idea that the carboxylic acid group could interact with the K59 residue of the sNT-Als1 PBC, and also because phenylacetic acid has been documented to have antimicrobial activity against *C. albicans* via an unknown mechanism (Kim et al 2004).

Similarly to the peptide screen, sNT-Als1 protein at a concentration of 2µM was combined with the experimental compound at a concentration of 2mM, plus Sypro Orange dye in a 1:1000 dilution to a total well volume of 20µl. Replicate negative control samples of sNT-Als1 protein without ligand were included in each experimental RT-PCR plates. Table 7.2 represents a small summary of the results obtained from these screens, and includes the names of the drug molecules used, plus a 2D diagram of its chemical structure alongside T_m values of the replicates, average T_m values and ΔT_M values (plus standard deviations of each measurement). For separate tables featuring each plate of the screen with their respective negative controls, see the Appendix.

The ΔT_M values were again calculated by comparing the difference between the average T_m values of the test samples with the average T_m values of the negative control sample that had been included in the same experimental plate. In order to calculate T_m values for each replicate, a graph of relative fluorescence units against temperature was constructed, and data fitted to a sigmoidal curve according to Equation 1. Standard deviations for each T_m measurement, average T_m calculation and ΔT_m value, were calculated according to Equation 2. A significant ΔT_m value, indicating a highly probable interaction between peptide ligand and protein, should be larger than two standard deviations of the average T_m value for the negative control sample.

Table 7.2: Summary of drug molecules used in a DSF screen to test the probability of their interaction with sNT-Als1.
Listed are the names of the drug compounds, 2D diagrams of their chemical structures, plus the T_m values of the replicates, the average T_m values for that test and the ΔT_M values (± standard deviation for each value). The ΔT_M values were calculated from the difference of the Average T_m of the test samples, from the average T_m of the negative control (sNT-Als1 protein without the addition of drug molecules). They are listed in order of descending ΔT_M, from highest to lowest.

Drug Molecule Name	Chemistry	T_m of Replicates (°C) ± SD			Average T_m (°C) ± SD	ΔTm(°C) ± SD
		N=1	N=2	N=3		
3,5-dinitrosalicylic acid (3,5-SAL)		48.599 ± 0.036	48.930 ± 0.042	49.067 ± 0.019	48.865 ± 0.241	2.950 ± 0.442
4-amino-phenylacetic acid (4-NH2-PAA)		47.338 ± 0.020	47.147 ± 0.024	47.247 ± 0.017	47.244 ± 0.096	2.098 ± 0.243
Caffeic acid (CAF)		48.0317 ± 0.046	48.010 ± 0.058	47.674 ± 0.072	47.905 ± 0.201	1.990 ± 0.416
2-methoxy-phenylacetic acid (2-Me-PAA)		47.081 ± 0.043	47.092 ± 0.039	46.679 ± 0.034	46.951 ± 0.235	1.805 ± 0.373
Acetyl-salicylic acid (Aspirin)		46.581 ± 0.0338	46.348 ± 0.0449	46.546 ± 0.047	46.491 ± 0.126	1.201 ± 0.287
2,5-dimethoxy-phenylacetic acid (2,5-Me-PAA)		46.294 ± 0.032	46.991 ± 0.045	46.863 ± 0.028	46.716 ± 0.371	1.570 ± 0.467
3,4,5-trimethoxy-phenylacetic acid (3,4,5-Me-PAA)		46.646 ± 0.036	46.561 ± 0.036	46.818 ± 0.016	46.675 ± 0.131	1.529 ± 0.281
Para-hydroxy-phenylacetic acid (p-OH-PAA)		46.819 ± 0.035	46.582 ± 0.030	46.671 ± 0.037	46.691 ± 0.120	1.545 ± 0.270
2,5-dihydroxy-phenylacetic acid (2,5-OH-PAA)		46.745 ± 0.082	46.628 ± 0.091	46.624 ± 0.144	46.666 ± 0.069	1.520 ± 0.208

The small drug-molecules that resulted in the highest ΔT_m values and that were also more than two SD values from the average T_m value of the negative control (sNT-Als1 without drug-like molecule, SD = ~0.4) are shown in Table 7.2. The compound 3,5-dinitrosalicylic acid (3,5-SAL), a derivative of acetyl-salicylic acid (Aspirin), resulted in a ΔT_m value of 2.940°C ± 0.442. See Figure 7.9 for a representation of the shift in melting curves after the addition of 3-5-SAL to sNT-Als1.

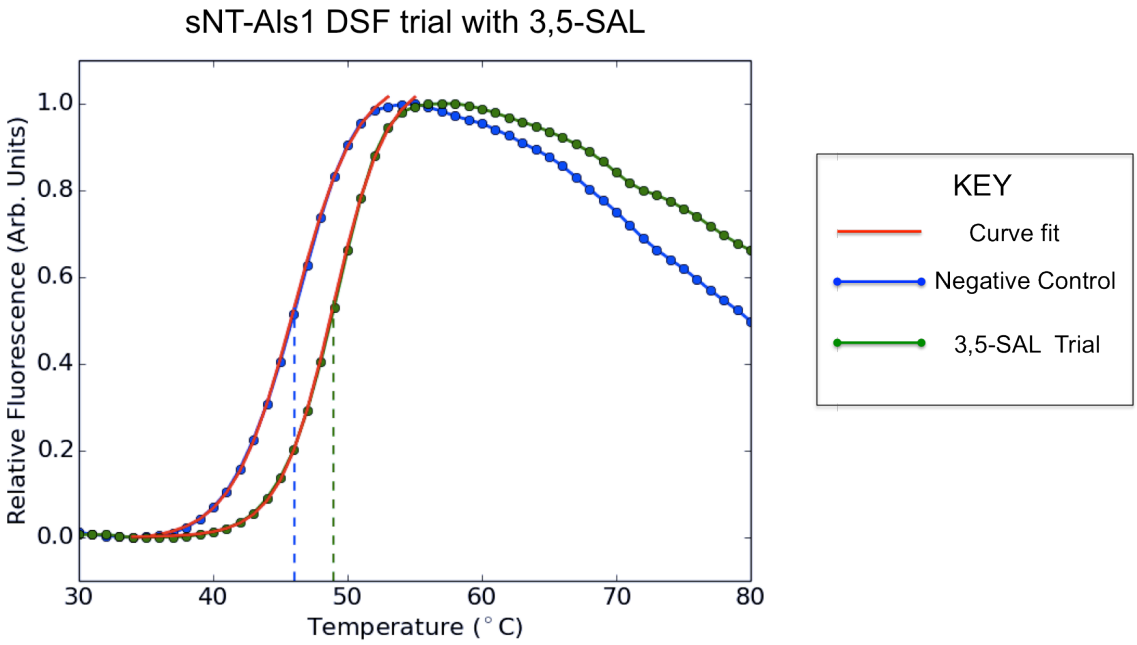
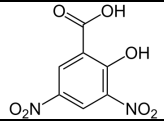


Figure 7.9: A comparison of the sNT-Als1 negative control melting curve (N = 3) with the melting curve of the 3,5-SAL drug molecule trial (N = 1).
Shown here is the plotted relative fluorescence data for one of the sNT-Als1 negative control replicates (N = 3, in blue) alongside the plotted relative fluorescence data for one of the sNT-Als1 plus 3,5-SAL drug molecule trial replicates (N=1, in green). In order to derive the T_m values for each trial, a curve has been fitted to the data using the equation y = Imin + (Imax – Imin)/1+ exp (T_m – x/a). The dotted lines show how midpoint has been calculated, and been used to derive the T_m values. For the negative control, a T_m of 46.041°C ± 0.047 has been calculated, and for the BSA-CT peptide trial, a T_m of 48.599°C ± 0.036 has been derived.

The compound was trailed again in a separate DSF screen, wherein both negative control and replicate samples included 0.1% DMSO. In this study, a ΔT_m shift value of 1.791°C ± 0.266 was obtained for the addition of 3,5-SAL to sNT-Als1. The value is lower in this second screen, possibly due to the additional effects of the DMSO. However, it was still more than two SD's of the average T_m value for the negative control (See Table 7.3).

Table 7.3: A repeated DSF trial of 3,5-SAL with sNT-Als1.
Here the T_m values of both negative control and 3,5-SAL DSF trials are shown, as well as their average T_m values and the ΔT_M values (± standard deviation for each value). The ΔT_M values were calculated from the difference of the Average T_m of the test samples, from the average T_m of the negative control (sNT-Als1 protein without the addition of drug molecules).

Drug Molecule Name	Chemistry	T_m of Replicates (°C) ± SD			Average T_m (°C) ± SD	ΔTm(°C) ± SD
Negative Control (0.1% DMSO)	N/A	52.760 ± 0.039	53.021 ± 0.028	52.800 ± 0.030	52.860 ± 0.141	0.000 ± 0.297
3,5-SAL (0.1% DMSO)		54.708 ± 0.010	54.524 ± 0.012	54.722 ± 0.019	54.651 ± 0.111	1.791 ± 0.266

These results for the 3,5-SAL compound prompted a ^{1}H-^{15}N-TROSY HSQC titration trial with ^{15}N-labelled sNT-Als1, and also a co-crystallisation screen with purified sNT-Als1. However, precipitation of the protein occurred after the addition of 3,5-SAL. In these studies, higher concentrations of 3,5-SAL (approximately 6-7mM) were used in comparison to 2mM for the DSF screen. Although strong buffers were used (100mM sodium phosphate pH 7.0 for NMR Spectroscopy and 100mM Tris-HCl pH 8.0 or crystallography), the pH effects from 3,5-SAL may have resulted in

disordered sNT-Als1 protein, of which has a pI of approximately 6.1. However, this may also suggest that pH effects were responsible for melting curve shifts, or that 3,5-SAL interacted with the unfolded state of sNT-Als1.

It is definite that other buffer conditions should be tested for optimizing further ligand-binding experiments with 3,5-SAL, before solid conclusions can be made about the interaction of 3,5-SAL with sNT-Als1. Further studies should be carried out to test this molecule, however, as salicylic acid as well as acetyl-salicylic acid (Aspirin) has been shown to be effective against the adhesion of *C. albicans* (and other micro-organisms) to catheters and subsequent biofilm formation (Farber & Wolff 1993, Alem & Douglas 2004).

7.5. Summary

The C-terminal peptides of BSA and bovine α-s1 casein resulted in high ΔT_M values when tested against sNT-Als1 using the DSF screen. This indicates that they are highly likely to bind to the sNT-Als1 PBC. As this short construct of Als1 also does not contain the AFR region, it is further evidence that the PBC only is responsible for these interactions with these C-terminal residues.

However, Als1 and other Als proteins of *C. albicans* may interact with host cell proteins by an additional mechanism, other than interacting with their C-termini. For example, free C-termini may be created in vivo due to the action of secreted aspartyl proteinases (SAPs), which can partially cleave cell surface and ECM proteins (Naglik et al 2003). Free C-terminal fragments produced in this way may remain attached to the host cell membrane, and become anchoring points for the Als-mediated adhesion of *C. albicans*. As the expression of SAPs correlates positively with *C. albicans'* propensity for adhesion, this mechanism may be an explanation, and should be studied further.

With regards to the small compounds DSF screen, the 3,5-dinitro-salicylic acid molecule was highlighted as having the potential to interact with of sNT-Als1. It may be used as a starting point in the design of drug molecule ligands that can inhibit *C. albicans* adhesion to host cell surfaces.

8. Discussion

8.1. Differences between the Adhesion PBCs of Als1, Als3 and Als9.2.

Als1 and Als3 are cited in the literature as the most biologically relevant proteins in the Als family, as they are prominent virulence factors involved in adhesion, invasion and biofilm formation in *C. albicans*. Interestingly, Als1 and Als3 also have the highest sequence identity in the family (82%) and an overlapping binding specificity to ligands, as judged by a number of cell based assays in which these proteins were expressed in *S. cerevisiae* (Sheppard et al 2004). However, differences in affinity of Als adhesins for specific proteins have been observed, for example Als3 can bind ferritin whilst Als1 cannot (Almeida et al 2008). The fact that the deletion of ALS3 results in a greater decrease of *C. albicans* adhesion to various substrates than the deletion of ALS1, this suggests that Als3 has a larger propensity for adhesion than Als1 and may have more substrate ligands (Zhao et al 2004, Oh et al 2005, Hoyer et al 2008). Importantly, mutation of residues in the PBC abolishes the contribution of Als3 to *C. albicans'* adhesion to host cells (Lin et al 2014), and a comparable effect would be expected from mutation of the same positions in the PBC of Als1. However, despite the large number of ligands identified, it is still unclear whether the observed binding is due solely to the recognition of the flexible C-termini from host proteins.

The precise role of Als9.2 during adhesion and virulence in *C. albicans* is unclear. The ALS9-2 gene is frequently expressed during candidiasis infections (Cheng et al 2005), but in one study its expression was recorded during the later stages of disease (Green et al 2006).

An explanation for differences in ligand-binding specificity between the Als proteins, especially for Als1 and Als3, may lie in differences in overall PBC volume and the shape of the cavity opening. Specificity differences may be due to the simple reasoning that peptides with bulky side chains in particular peptide residue positions may have problems with reaching the salt-bridge contact between their carboxyl terminus and the invariant K59 of the PBC.

In fact, from the peptide ligand-binding screen carried out, the peptides that resulted in the highest ΔT_m values when mixed with sNT-Als1 were those that included small, hydrophobic residues at peptide positions 1-4, and those that resulted in the lowest ΔT_m values were those with sterically bulky residues (for instance tryptophan, tyrosine or glutamate) at peptide positions 1-4. However, these assumptions will have to be tested in the future with more precise ligand-binding studies, and comparing Als1 constructs with Als3 and Als9.2

8.2. A functional relationship between Als proteins and secreted aspartyl proteinases (SAPs)?

The peptides screened against sNT-Als1 in the DSF assay were selected from the C-termini of host proteins identified as ligands of Als adhesins in different cell based assays. For extracellular proteins that are secreted, for example casein and BSA/HSA, or attached to the extracellular matrix, such as fibrinogen, it is easy to envisage recognition of their free, flexible C-termini via the PBC of Als adhesins.

A proposed mechanism to enhance the activity of Als proteins is their co-expression with secreted aspartyl proteases (SAPs) by *C. albicans* (Salgado et al 2011). SAPs are known to hydrolyse ECM components, including collagen, laminin and fibronectin, as well as immune system factors such as proinflammatory cytokines and complement proteins and interestingly, they have also been reported to enhance *C. albicans* adhesion to host cell surfaces (Naglik et al 2003, Schaller et al 2005, Dos Santos et al 2006, Gropp et al 2009), but the mechanism for this activity has not been described.

However, partial digestion of host proteins could create fragments with C-termini anchored to cell surfaces or the ECM, providing additional ligands of Als mediated adhesion.

Intriguingly, the ALS and SAP proteins are housed on the same chromosomes in *C. albicans*, *C. dubliniensis* and *C. tropicalis* (Hube 1994, Hoyer 2001, Hoyer et al 2001), and are regulated by the same transcription factors (Murad et al 2001), suggesting that their expression is coordinated during colonization. Furthermore, Als1 and Sap4 co-expression has been observed in *C. albicans* isolates from women with vaginitis (Nas et al 2008).

8.3. Recognition of cytoplasmic C-termini by Als proteins

In addition, a number of ligands for Als proteins are anchored to the host cell membrane and their C-termini are cytoplasmic (for example cadherins and fibronectin), or are soluble in the cytoplasm, as is the case of ferritin. Previously, it has been proposed that Als3 can act as an invasin, as Als3-coated latex beads are readily internalized to the cytoplasm of endothelial cells, in a mechanism mediated by the interaction of Als3 with endothelial cadherin (Phan et al 2007). Although a direct association of Als3 with the C-terminus of E-cadherin has not been demonstrated, mutations in the PBC eliminate Als3-mediated adhesion to *C. albicans* to FaDu endothelial cells (Lin et al 2014), the same cell line used to probe the invasive properties of Als3 using latex beads. Collectively, these observations support the theory that recognition of C-termini is the sole mechanism for interaction with host cell ligands.

Als adhesins share high sequence identity, and their variability is mainly concentrated in residues close to the PBC (Salgado et al 2011). If the residues involved in invasion are located on the protein surface, it will be interesting to determine whether other Als proteins are also invasins and the extent of this activity in the overall adhesive properties of *C. albicans*.

8.4. Future studies and Drug Design

In the context of drug design, Als proteins are potentially a good therapeutic target. Inactivation of their adhesive properties may impair the initial contacts, either with host cell surfaces or medical devices and implants, which lead to the formation of *C. albicans* biofilms. Constructs of NT-Als adhesins can be produced in milligram quantities and with a high level of purity to facilitate structural and functional studies. They are also amenable for the screening of different peptides and small chemical fragments using high-throughput techniques such as DSF and NMR, and in principle others such as isothermal titration calorimetry, microscale thermophoresis and surface plasmon resonance.

With NMR backbone assignment studies, it has been possible to assign more than 55% of the sNT-Als1 backbone and successfully assign residues of the PBC. With this information it will be possible to screen drug compounds against sNT-Als1 using chemical shift perturbation experiments and other ligand-binding observation NMR experiments, such as waterLOGSY and saturation transfer difference, STD. Similarly, optimization of the crystallization conditions will facilitate co-crystallisation of sNT-Als1 with drug-like molecules.

The DSF screen with small molecules in Chapter 7 pin-pointed the compound dinitro-salicylic acid as an example of a drug-like molecule able to interact with sNT-Als1. Although co-crystallization trials for this complex have been so far unsuccessful, completion of backbone NMR assignments will enable chemical shift perturbation assays to determine whether this compound interacts with K59, the key amino acid involved in ligand binding within the sNT-Als1 PBC.

Als1 and Als3, are prominent virulence factors in *C. albicans*. Their ability to recognize the model peptide Hepta-T suggests a starting point in the design of peptidomimetics with narrow

specificity to a single Als protein, or cross reactivity to all Als proteins. Although there are still a number of obstacles in the development of peptidomimetics as microbial molecules (Liskamp et al 2011, Lohan & Bisht 2013), the design of a peptide or group of peptides capable of inactivating the adhesive properties of all Als protein would be an extremely useful experimental tool to determining the contribution of these proteins to the virulence of *C. albicans*.

Similarly, peptides can provide important structural information for the design of drug-like ligands that inactivate the function of Als adhesins. Such compounds may be incorporated into a medical device (e.g. catheter or heart implant), in order to prevent the attachment of *C. albicans* and ultimately the formation of biofilms. Otherwise, anti-Als compounds could be used in conjunction with other antifungals currently in use, to provide a patient with the best chance of treatment against both superficial and invasive candidiasis.

9. References

Achkar, J. M., & Fries, B. C. (2010). *Candida* infections of the genitourinary tract. *Clinical Microbiology Reviews*, *23*(2), 253–73.

Adam, B., Baillie, G. S., & Douglas, L. J. (2002). Mixed species biofilms of *Candida albicans* and *Staphylococcus epidermidis*. *Journal of Medical Microbiology*, *51*(4), 344–9.

Afonine, P. V, Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., *et al*. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica. Section D, Biological Crystallography*, *68*(Pt 4), 352–67.

Alem, M. A. S., & Douglas, L. J. (2004). Effects of Aspirin and Other Nonsteroidal Anti-Inflammatory Drugs on Biofilms and Planktonic Cells of *Candida albicans*. *Antimicrobial Agents and Chemotherapy*, *48*(1), 41–47. doi:10.1128/AAC.48.1.41

Al-Fattani, M. A., & Douglas, L. J. (2004). Penetration of *Candida* biofilms by antifungal agents. *Antimicrobial Agents and Chemotherapy*, *48*(9), 3291–7.

Al-Fattani, M. A., & Douglas, L. J. (2006). Biofilm matrix of *Candida albicans* and *Candida tropicalis*: chemical composition and role in drug resistance. *Journal of Medical Microbiology*, *55*(Pt 8), 999–1008.

Almeida, R. S., Brunke, S., Albrecht, A., Thewes, S., Laue, M., Edwards, J. E., … Hube, B. (2008). The hyphal-associated adhesin and invasin Als3 of *Candida albicans* mediates iron acquisition from host ferritin. *PLoS Pathogens*, *4*(11), e1000217.

Almeida, R. S., Wilson, D., & Hube, B. (2009). *Candida albicans* iron acquisition within the host. *FEMS Yeast Research*, *9*(7), 1000–12.

Alsteens, D., Van Dijck, P., Lipke, P. N., & Dufrêne, Y. F. (2013). Quantifying the forces driving cell-cell adhesion in a fungal pathogen. *Langmuir : The ACS Journal of Surfaces and Colloids*, *29*(44), 13473–80.

Ancalle, I. M., Rivera, J. A., García, I., García, L., & Valcárcel, M. (2010). *Candida albicans* meningitis and brain abscesses in a neonate: a case report. *Boletín de La Asociación Médica de Puerto Rico*, *102*(1), 45–8.

Anderson, W. a., & Freeman, R. (1962). Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra. *The Journal of Chemical Physics*, *37*(1), 85.

Andes, D., Nett, J., Oschel, P., Albrecht, R., Marchillo, K., & Pitula, A. (2004). Development and characterization of an in vivo central venous catheter *Candida albicans* biofilm model. *Infection and Immunity*, *72*(10), 6023–31.

Aoki, W., Kitahara, N., Miura, N., Morisaka, H., Kuroda, K., & Ueda, M. (2012). Profiling of adhesive properties of the agglutinin-like sequence (ALS) protein family, a virulent attribute of *Candida albicans*. *FEMS Immunology and Medical Microbiology*, *65*(1), 121–4.

Asleson, C. M., Bensen, E. S., Gale, C. A., Melms, A. S., Kurischko, C., & Berman, J. (2001). *Candida albicans* INT1-induced filamentation in *Saccharomyces cerevisiae* depends on Sla2p. *Molecular and Cellular Biology*, *21*(4), 1272–84.

Azoulay, E., Timsit, J.-F., Tafflet, M., de Lassence, A., Darmon, M., Zahar, J.-R., *et al*. (2006). *Candida* colonization of the respiratory tract and subsequent *pseudomonas* ventilator-associated pneumonia. *Chest*, *129*(1), 110–7.

Baddley, J. W., Benjamin, D. K., Patel, M., Miró, J., Athan, E., Barsic, B., *et al*. (2008). *Candida* infective endocarditis. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *27*(7), 519–29.

Bailey, D. A., Feldmann, P. J., Bovey, M., Gow, N. A., & Brown, A. J. (1996). The *Candida albicans* HYR1 gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins. *Journal of Bacteriology*, *178*(18), 5353–60.

Baillie, G. S., & Douglas, L. J. (1999). Role of dimorphism in the development of *Candida albicans* biofilms. *Journal of Medical Microbiology*, *48*(7), 671–9.

Balashov, S. V, Park, S., & Perlin, D. S. (2006). Assessing resistance to the echinocandin antifungal drug caspofungin in *Candida albicans* by profiling mutations in FKS1. *Antimicrobial Agents and Chemotherapy*, *50*(6), 2058–63.

Bamford, C. V, d'Mello, A., Nobbs, A. H., Dutton, L. C., Vickerman, M. M., & Jenkinson, H. F. (2009). *Streptococcus gordonii* modulates *Candida albicans* biofilm formation through intergeneric communication. *Infection and Immunity*, *77*(9), 3696–704.

Bandara, H. M. H. N., Cheung, B. P. K., Watt, R. M., Jin, L. J., & Samaranayake, L. P. (2013a). Secretory products of *Escherichia coli* biofilm modulate *Candida* biofilm formation and hyphal development. *Journal of Investigative and Clinical Dentistry*, *4*(3), 186–99.

Bandara, H. M. H. N., K Cheung, B. P., Watt, R. M., Jin, L. J., & Samaranayake, L. P. (2013b). *Pseudomonas aeruginosa* lipopolysaccharide inhibits *Candida albicans* hyphae formation and alters gene expression during biofilm development. *Molecular Oral Microbiology*, *28*(1), 54–69.

Bearden, J.A. (1967) "X-Ray Wavelengths," *Rev. Mod. Phys.* **39**, 78.

Beaussart, A., Alsteens, D., El-Kirat-Chatel, S., Lipke, P. N., Kucharíková, S., Van Dijck, P., & Dufrêne, Y. F. (2012). Single-molecule imaging and functional analysis of Als adhesins and mannans during *Candida albicans* morphogenesis. *ACS Nano*, *6*(12), 10950–64.

Becker, D. J., & Lowe, J. B. (2003). Fucose: biosynthesis and biological function in mammals. *Glycobiology*, *13*(7), 41R–53R.

Bennett, R. J., & Johnson, A. D. (2005). Mating in *Candida albicans* and the search for a sexual cycle. *Annual Review of Microbiology*, *59*, 233–55.

Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, *10*(12), 980.

Berman, J., & Sudbery, P. E. (2002). *Candida albicans*: a molecular revolution built on lessons from budding yeast. *Nature Reviews. Genetics*, *3*(12), 918–30.

Beucher, B., Marot-Leblond, A., Billaud-Nail, S., Oh, S.-H., Hoyer, L. L., & Robert, R. (2009). Recognition of *Candida albicans* Als3 by the germ tube-specific monoclonal antibody 3D9.3. *FEMS Immunology and Medical Microbiology*, *55*(3), 314–23.

Bink, A., Kucharíková, S., Neirinck, B., Vleugels, J., Van Dijck, P., Cammue, B. P. A., & Thevissen, K. (2012). The nonsteroidal antiinflammatory drug diclofenac potentiates the in vivo activity of caspofungin against *Candida albicans* biofilms. *The Journal of Infectious Diseases*, *206*(11), 1790–7.

Boris, S., & Barbés, C. (2000). Role played by *lactobacilli* in controlling the population of vaginal pathogens. *Microbes and Infection / Institut Pasteur*, *2*(5), 543–6.

Borg-von Zepelin, M., Meyer, I., Thomssen, R., Würzner, R., Sanglard, D., Telenti, A., & Monod, M. (1999). HIV-Protease inhibitors reduce cell adherence of *Candida albicans* strains by inhibition of yeast secreted aspartic proteases. *The Journal of Investigative Dermatology*, *113*(5), 747–51.

Bowden, M. G., Heuck, A. P., Ponnuraj, K., Kolosova, E., Choe, D., Gurusiddappa, S., *et al*. (2008). Evidence for the "dock, lock, and latch" ligand binding mechanism of the staphylococcal microbial surface component recognizing adhesive matrix molecules (MSCRAMM) SdrG. *The Journal of Biological Chemistry*, *283*(1), 638–47.

Braun, B. R., & Johnson, A. D. (2000). TUP1, CPH1 and EFG1 Make Independent Contributions to Filamentation in *Candida albicans. Genetics*, *155*(1), 57–67.

Braun, B. R., van Het Hoog, M., d'Enfert, C., Martchenko, M., Dungan, J., Kuo, A., … Nantel, A. (2005). A human-curated annotation of the *Candida albicans* genome. *PLoS Genetics*, *1*(1), 36–57.

Brena, S., Omaetxebarría, M. J., Elguezabal, N., Cabezas, J., Moragues, M. D., & Pontón, J. (2007). Fungicidal monoclonal antibody C7 binds to *Candida albicans* Als3. *Infection and Immunity*, *75*(7), 3680–2.

Brena, S., Cabezas-Olcoz, J., Moragues, M. D., Fernández de Larrinoa, I., Domínguez, A., Quindós, G., & Pontón, J. (2011). Fungicidal monoclonal antibody C7 interferes with iron acquisition in *Candida albicans*. *Antimicrobial Agents and Chemotherapy*, *55*(7), 3156–63.

Brown, A. J., & Gow, N. A.. (1999). Regulatory networks controlling *Candida albicans* morphogenesis. *Trends in Microbiology*, *7*(8), 333–338.

Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*, 472–475.

Bujdáková, H., Kulková, N., & Cernáková, L. (2012). Susceptibility to caspofungin and fluconazole and als1/als3 gene expression in biofilm and dispersal cells of *Candida albicans*. *Archives of Industrial Hygiene and Toxicology*, *63*(4), 497–503.

Calderone, R. a, & Braun, P. C. (1991). Adherence and receptor relationships of *Candida albicans*. *Microbiological Reviews*, *55*(1), 1–20.

Candiracci, M., Citterio, B., & Piatti, E. (2012). Antifungal activity of the honey flavonoid extract against *Candida albicans*. *Food Chemistry*, *131*(2), 493–499.

Cassone, A. (2013). Development of vaccines for *Candida albicans*: fighting a skilled transformer. *Nature Reviews. Microbiology*, *11*(12), 884–91.

Cassone, A., & Cauda, R. (2012). *Candida* and candidiasis in HIV-infected patients: where commensalism, opportunistic behavior and frank pathogenicity lose their borders. *AIDS (London, England)*, *26*(12), 1457–72.

Chaffin, W. L. (2008). *Candida albicans* cell wall proteins. *Microbiology and Molecular Biology Reviews : MMBR*, *72*(3), 495–544.

Chai, L. Y. a, Netea, M. G., Vonk, A. G., & Kullberg, B.-J. (2009). Fungal strategies for overcoming host innate immune response. *Medical Mycology*, *47*(3), 227–36.

Chandra, J., Kuhn, D. M., Mukherjee, P. K., Hoyer, L. L., McCormick, T., & Ghannoum, M. A. (2001). Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *Journal of Bacteriology*, *183*(18), 5385–94.

Chen, H., Fujita, M., Feng, Q., Clardy, J., & Fink, G. R. (2004). Tyrosol is a quorum-sensing molecule in *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(14), 5048–52.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. a, Immormino, R. M., Kapral, G. J., *et al*. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 1), 12–21.

Cheng, G., Wozniak, K., Wallig, M. A., Fidel, P. L., Trupin, S. R., & Hoyer, L. L. (2005). Comparison between *Candida albicans* Agglutinin-Like Sequence Gene Expression Patterns in Human Clinical Specimens and Models of Vaginal Candidiasis, *73*(3), 1656–1663.

Cheng, S.-C., Joosten, L. A. B., Kullberg, B.-J., & Netea, M. G. (2012). Interplay between *Candida albicans* and the mammalian innate host defense. *Infection and Immunity*, *80*(4), 1304–13.

Cleary, I. a, Reinhard, S. M., Miller, C. L., Murdoch, C., Thornhill, M. H., Lazzell, A. L., *et al*. (2011). *Candida albicans* adhesin Als3p is dispensable for virulence in the mouse model of disseminated candidiasis. *Microbiology*, *157*(6), 1806–15.

Coleman, D. A., Oh, S.-H., Zhao, X., Zhao, H., Hutchins, J. T., Vernachio, J. H., *et al*. (2009). Monoclonal antibodies specific for *Candida albicans* Als3 that immunolabel fungal cells in vitro and in vivo and block adhesion to host surfaces. *Journal of Microbiological Methods*, *78*(1), 71–8.

Coleman, D. A., Oh, S.-H., Zhao, X., & Hoyer, L. L. (2010). Heterogeneous distribution of *Candida albicans* cell-surface antigens demonstrated with an Als1-specific monoclonal antibody. *Microbiology (Reading, England)*, *156*(Pt 12), 3645–59.

Coleman, D. a, Oh, S.-H., Manfra-Maretta, S. L., & Hoyer, L. L. (2012). A monoclonal antibody specific for *Candida albicans* Als4 demonstrates overlapping localization of Als family proteins on the fungal cell surface and highlights differences between Als localization in vitro and in vivo. *FEMS Immunology and Medical Microbiology*, *64*(3), 321–33.

Cornely, O. A., Bassetti, M., Calandra, T., Garbino, J., Kullberg, B. J., Lortholary, O., *et al*. (2012). ESCMID* guideline for the diagnosis and management of *Candida* diseases 2012: non-neutropenic

adult patients. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *18 Suppl 7*, 19–37.

Cruciani, M., de Lalla, F., & Mengoli, C. (2005). Prophylaxis of *Candida* infections in adult trauma and surgical intensive care patients: a systematic review and meta-analysis. *Intensive Care Medicine*, *31*(11), 1479–87.

Dalle, F., Wächtler, B., L'Ollivier, C., Holland, G., Bannert, N., Wilson, D., *et al*. (2010). Cellular interactions of *Candida albicans* with human oral epithelial cells and enterocytes. *Cellular Microbiology*, *12*(2), 248–71.

Deivanayagam, C. C. S., Wann, E. R., Chen, W., Carson, M., Rajashankar, K. R., Höök, M., & Narayana, S. V. L. (2002). A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. *The EMBO Journal*, *21*(24), 6660–72.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3), 277–93.

Delaloye, J., & Calandra, T. (2014). Invasive candidiasis as a cause of sepsis in the critically ill patient. *Virulence*, *5*(1), 161–9.

De Vita, D., Friggeri, L., D'Auria, F. D., Pandolfi, F., Piccoli, F., Panella, S., *et al*. (2014). Activity of caffeic acid derivatives against *Candida albicans* biofilm. *Bioorganic & Medicinal Chemistry Letters*, *24*(6), 1502–5.

Dhamgaye, S., Bernard, M., Lelandais, G., Sismeiro, O., Lemoine, S., Coppée, J.-Y., *et al*. (2012). RNA sequencing revealed novel actors of the acquisition of drug resistance in *Candida albicans*. *BMC Genomics*, *13*(1), 396.

Dick, J. D., Merz, W. G., & Saral, R. (1980). Incidence of polyene-resistant yeasts recovered from clinical specimens. *Antimicrobial Agents and Chemotherapy*, *18*(1), 158–63.

Ding, X., Liu, Z., Su, J., & Yan, D. (2014). Human serum inhibits adhesion and biofilm formation in *Candida albicans*. *BMC Microbiology*, *14*(1), 80. doi:10.1186/1471-2180-14-80

Dongari-Bagtzoglou, A., Kashleva, H., Dwivedi, P., Diaz, P., & Vasilakos, J. (2009). Characterization of mucosal *Candida albicans* biofilms. *PloS One*, *4*(11), e7967.

Donlan, R. M., & Costerton, J. W. (2002). Biofilms: survival mechanisms of clinically relevant microorganisms. *Clinical Microbiology Reviews*, *15*(2), 167–93.

Donohue, D. S., Ielasi, F. S., Goossens, K. V. Y., & Willaert, R. G. (2011). The N-terminal part of Als1 protein from *Candida albicans* specifically binds fucose-containing glycans. *Molecular Microbiology*, *80*(6), 1667–79.

Dos Santos, A. L. S. (2010). HIV aspartyl protease inhibitors as promising compounds against *Candida albicans*. *World Journal of Biological Chemistry*, *1*(2), 21–30.

Douglas, L. J. (2003). *Candida* biofilms and their role in infection. *Trends in Microbiology*, *11*(1), 30–6.

Drag, M., & Salvesen, G. S. (2010). Emerging principles in protease-based drug discovery. *Nature Reviews. Drug Discovery*, *9*(9), 690–701.

Dunyach, C., Drakulovski, P., Bertout, S., Jouvert, S., Reynes, J., & Mallié, M. (2011). Fungicidal activity and morphological alterations of *Candida albicans* induced by echinocandins: study of strains with reduced caspofungin susceptibility. *Mycoses*, *54*(4), e62–8.

Dwivedi, P., Thompson, A., Xie, Z., Kashleva, H., Ganguly, S., Mitchell, A. P., & Dongari-Bagtzoglou, A. (2011). Role of Bcr1-activated genes Hwp1 and Hyr1 in *Candida albicans* oral mucosal biofilms and neutrophil evasion. *PloS One*, *6*(1), e16218.

Edwards, J.E. Jr. (2012) Fungal cell wall vaccines: an update. *Journal of Medical Microbiology.* **Vol 61** (7):895-903.

El-Kirat-Chatel S, *et al.* (2013) *Nanoscale* analysis of caspofungin-induced cell surface remodelling in *Candida albicans*. *Nanoscale*. **Vol 5**(3):1105-15.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(Pt 1), 72–82.

Faix, R. G., & Chapman, R. L. (2003). Central nervous system candidiasis in the high-risk neonate. *Seminars in Perinatology*, *27*(5), 384–92.

Falagas, M. E., Roussos, N., & Vardakas, K. Z. (2010). Relative frequency of *albicans* and the various non-*albicans Candida spp* among candidemia isolates from inpatients in various parts of the world: a systematic review. *International Journal of Infectious Diseases : IJID : Official Publication of the International Society for Infectious Diseases*, *14*(11), e954–66.

Fanning, S., Xu, W., Solis, N., Woolford, C. a, Filler, S. G., & Mitchell, A. P. (2012). Divergent targets of *Candida albicans* biofilm regulator Bcr1 in vitro and in vivo. *Eukaryotic Cell*, *11*(7), 896–904.

Farah, C. S., Elahi, S., Drysdale, K., Pang, G., Gotjamanos, T., Seymour, G. J., *et al*. (2002). Primary Role for CD4 ε T Lymphocytes in Recovery from Oropharyngeal Candidiasis, *70*(2), 724–731.

Farber, B. F., & Wolff, A. G. (1993). Salicylic acid prevents the adherence of bacteria and yeast to silastic catheters. *Journal of Biomedical Materials Research*, *27*(5), 599–602.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology*, *22*(10), 1302–6.

Fidel, P. L. (2004). History and new insights into host defense against vaginal candidiasis. *Trends in Microbiology*, *12*(5), 220–7.

Fidel, P. L. (2011). *Candida*-host interactions in HIV disease: implications for oropharyngeal candidiasis. *Advances in Dental Research*, *23*(1), 45–9.

Filler, S. G., Swerdloff, J. N., Hobbs, C., & Luckett, P. M. (1995). Penetration and damage of endothelial cells by *Candida albicans*. *Infection and Immunity*, *63*(3), 976–83.

Finkel, J. S., & Mitchell, A. P. (2011). Genetic control of *Candida albicans* biofilm development. *Nature Reviews. Microbiology*, *9*(2), 109–18.

Fradin, C., De Groot, P., MacCallum, D., Schaller, M., Klis, F., Odds, F. C., & Hube, B. (2005). Granulocytes govern the transcriptional response, morphology and proliferation of *Candida albicans* in human blood. *Molecular Microbiology*, *56*(2), 397–415.

Frank, C. F., & Hostetter, M. K. (2007). Cleavage of E-cadherin: a mechanism for disruption of the intestinal epithelial barrier by *Candida albicans*. *Translational Research : The Journal of Laboratory and Clinical Medicine*, *149*(4), 211–22.

Frank, A. T., Ramsook, C. B., Otoo, H. N., Tan, C., Soybelman, G., Rauceo, J. M., *et al.* (2010). Structure and function of glycosylated tandem repeats from *Candida albicans* Als adhesins. *Eukaryotic Cell*, *9*(3), 405–14.

Frankel, R. I. (1996). Centennial of Röntgen's discovery of x-rays. *The Western Journal of Medicine*, *164*(6), 497–501.

French, B. Y. S., & Wilson, K. (1978). On the Treatment of Negative Intensity Observations. *Acta Crystallographica, A34*, 517–525.

Fu, Y., Rieg, G., Fonzi, W. A., Belanger, P. H., John, E., Jr, E., *et al.* (1998). Expression of the *Candida albicans* Gene ALS1 in *Saccharomyces cerevisiae* Induces Adherence to Endothelial and Epithelial Cells Expression of the *Candida albicans* Gene ALS1 in *Saccharomyces cerevisiae* Induces Adherence to Endothelial and Epithelial Cells. *Infection and Immunity*, *66*(4), 1783–1786.

Fu, Y., Ibrahim, A. S., Sheppard, D. C., Chen, Y.-C., French, S. W., Cutler, J. E., *et al*. (2002). *Candida albicans* Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway. *Molecular Microbiology*, *44*(1), 61–72.

Fu, Y., Luo, G., Spellberg, B. J., Edwards, J. E., & Ibrahim, A. S. (2008). Gene overexpression/suppression analysis of candidate virulence factors of *Candida albicans*. *Eukaryotic Cell*, *7*(3), 483–92.

Fu, Y., Phan, Q. T., Luo, G., Solis, N. V, Liu, Y., Cormack, B. P., *et al*. (2013). Investigation of the function of *Candida albicans* Als3 by heterologous expression in *Candida glabrata. Infection and Immunity*, *81*(7), 2528–35.

Gale, C., Bendel, C., & Hostetter, M. (1997). Induction of Epithelial Adhesion and Morphologic Switching by a Single Gene from *Candida albicans. Pediatric Research*, *41*(S4), 119–119.

Gale, C. A., Bendel, C. M., McClellan, M., Hauser, M., Becker, J. M., Berman, J., & Hostetter, M. K. (1998). Linkage of adhesion, filamentous growth, and virulence in *Candida albicans* to a single gene, INT1. *Science (New York, N.Y.)*, *279*(5355), 1355–8.

Gamarra, S., Rocha, E. M. F., Zhang, Y.-Q., Park, S., Rao, R., & Perlin, D. S. (2010). Mechanism of the synergistic effect of amiodarone and fluconazole in *Candida albicans*. *Antimicrobial Agents and Chemotherapy*, *54*(5), 1753–61.

Ganesh, V. K., Rivera, J. J., Smeds, E., Ko, Y.-P., Bowden, M. G., Wann, E. R., … Höök, M. (2008). A structural model of the *Staphylococcus aureus* ClfA-fibrinogen interaction opens new avenues for the design of anti-*staphylococcal* therapeutics. *PLoS Pathogens*, *4*(11), e1000226. doi:10.1371/journal.ppat.1000226

García-Sánchez, S., Aubert, S., Iraqui, I., Janbon, G., Ghigo, J.-M., & d'Enfert, C. (2004). Candida albicans biofilms: a developmental state associated with specific and stable gene expression patterns. *Eukaryotic Cell*, *3*(2), 536–45.

Gardner, K. H., & Kay, L. E. (1998). The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure*, *27*, 357–406.

Gaur, N. K., & Klotz, S. A. (1997). Expression, cloning, and characterization of a *Candida albicans* gene , ALA1 , that confers adherence properties upon *Saccharomyces cerevisiae* for extracellular matrix proteins. *Infection and Immunity*, *65*(12), 5289–5294.

Gaur, N. K., Klotz, S. A., & Henderson, R. L. (1999). Overexpression of the *Candida albicans* ALA1 gene in *Saccharomyces cerevisiae* results in aggregation following attachment of yeast cells to extracellular matrix proteins, adherence properties similar to those of *Candida albicans*. *Infection and Immunity*, *67*(11), 6040–7.

Gaur, N. K., Smith, R. L., & Klotz, S. A. (2002). *Candida albicans* and *Saccharomyces cerevisiae* expressing ALA1/ALS5 adhere to accessible threonine, serine, or alanine patches. *Cell Communication & Adhesion*, *9*(1), 45–57.

Ghalehnoo, Z. R., Rashki, A., Najimi, M., & Dominguez, A. (2010). The role of diclofenac sodium in the dimorphic transition in *Candida albicans*. *Microbial Pathogenesis*, *48*(3-4), 110–5.

Ghannoum, M. A., & Rice, L. B. (1999). Antifungal Agents : Mode of Action , Mechanisms of Resistance , and Correlation of These Mechanisms with Bacterial Resistance, *12*(4), 501–517.

Gomes, A. C., Miranda, I., Silva, R. M., Moura, G. R., Thomas, B., Akoulitchev, A., & Santos, M. A. S. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. *Genome Biology*, *8*(10), R206.

Gow, N. a R., Brown, A. J. P., & Odds, F. C. (2002). Fungal morphogenesis and host invasion. *Current Opinion in Microbiology*, *5*(4), 366–71.

Gow, N. a R., & Hube, B. (2012). Importance of the *Candida albicans* cell wall during commensalism and infection. *Current Opinion in Microbiology*, *15*(4), 406–12.

Green, C. B., Cheng, G., Chandra, J., Mukherjee, P., Ghannoum, M. A., & Hoyer, L. L. (2004). RT-PCR detection of *Candida albicans* ALS gene expression in the reconstituted human epithelium (RHE) model of oral candidiasis and in model biofilms. *Microbiology*, *150*, 267–275.

Green, C. B., Zhao, X., Yeater, K. M., & Hoyer, L. L. (2005a). Construction and real-time RT-PCR validation of *Candida albicans* PALS-GFP reporter strains and their use in flow cytometry analysis of ALS gene expression in budding and filamenting cells. *Microbiology*, *151*(4), 1051–60.

Green, C.B., Marretta, S.M., Cheng, G., Faddoul, F.F., Hoyer, L.L. (2006). RT-PCR analysis of *Candida albicans* ALS gene expression in a hyposalivatory rat model of oral candidiasis and HIV-positive human patients. *Medical Mycology*, *44*(2), 103–111.

Gregori, C., Glaser, W., Frohner, I. E., Reinoso-Martín, C., Rupp, S., Schüller, C., & Kuchler, K. (2011). Efg1 Controls caspofungin-induced cell aggregation of *Candida albicans* through the adhesin Als1. *Eukaryotic Cell*, *10*(12), 1694–704.

De Groot, P. W. J., Bader, O., de Boer, A. D., Weig, M., & Chauhan, N. (2013). Adhesins in human fungal pathogens: glue with plenty of stick. *Eukaryotic Cell*, *12*(4), 470–81.

Gropp, K., Schild, L., Schindler, S., Hube, B., Zipfel, P. F., & Skerka, C. (2009). The yeast *Candida albicans* evades human complement attack by secretion of aspartic proteases. *Molecular Immunology*, *47*(2-3), 465–75.

Grover, N. D. (2010). Echinocandins: A ray of hope in antifungal drug therapy. *Indian Journal of Pharmacology*, *42*(1), 9–11.

Grubb, S. E. W., Murdoch, C., Sudbery, P. E., Saville, S. P., Lopez-Ribot, J. L., & Thornhill, M. H. (2008). *Candida albicans*-endothelial cell interactions: a key step in the pathogenesis of systemic candidiasis. *Infection and Immunity*, *76*(10), 4370–7.

Grubb, S. E. W., Murdoch, C., Sudbery, P. E., Saville, S. P., Lopez-Ribot, J. L., & Thornhill, M. H. (2009). Adhesion of *Candida albicans* to endothelial cells under physiological conditions of flow. *Infection and Immunity*, *77*(9), 3872–8.

Harriott, M. M., & Noverr, M. C. (2009). *Candida albicans* and *Staphylococcus aureus* form polymicrobial biofilms: effects on antimicrobial resistance. *Antimicrobial Agents and Chemotherapy*, *53*(9), 3914–22.

Harriott, M. M., & Noverr, M. C. (2010). Ability of *Candida albicans* mutants to induce *Staphylococcus aureus* vancomycin resistance during polymicrobial biofilm formation. *Antimicrobial Agents and Chemotherapy*, *54*(9), 3746–55.

Harriott, M. M., & Noverr, M. C. (2011). Importance of *Candida*-bacterial polymicrobial biofilms in disease. *Trends in Microbiology*, *19*(11), 557–63.

Heilmann, C. J., Sorgo, A. G., Siliakus, A. R., Dekker, H. L., Brul, S., de Koster, C. G., *et al*. (2011). Hyphal induction in the human fungal pathogen *Candida albicans* reveals a characteristic wall protein profile. *Microbiology (Reading, England)*, *157*(Pt 8), 2297–307.

Heinsbroek, S. E. M., Brown, G. D., & Gordon, S. (2005). Dectin-1 escape by fungal dimorphism. *Trends in Immunology*, *26*(7), 352–4.

Heinz, W. J., Kurzai, O., Brakhage, A. A, Fonzi, W. A., Korting, H. C., Frosch, M., & Mühlschlegel, F. A. (2000). Molecular responses to changes in the environmental pH are conserved between the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *International Journal of Medical Microbiology : IJMM*, *290*(3), 231–8.

Hogan, D. A., & Kolter, R. (2002). *Pseudomonas-Candida* interactions: an ecological role for virulence factors. *Science (New York, N.Y.)*, *296*(5576), 2229–32.

Hogan, D. A., Vik, A., & Kolter, R. (2004). A *Pseudomonas aeruginosa* quorum-sensing molecule influences *Candida albicans* morphology. *Molecular Microbiology*, *54*(5), 1212–23.

Holcombe, L. J., McAlester, G., Munro, C. A., Enjalbert, B., Brown, A. J. P., Gow, N. A. R., *et al*. (2010). *Pseudomonas aeruginosa* secreted factors impair biofilm development in *Candida albicans*. *Microbiology (Reading, England)*, *156*(Pt 5), 1476–86.

Holmes, A. R., Gopal, P. K., & Jenkinson, H. F. (1995). Adherence of *Candida albicans* to a cell surface polysaccharide receptor on Streptococcus gordonii. *Infection and Immunity*, *63*(5), 1827–34.

Hornby, J. M., Jensen, E. C., Lisec, A. D., Tasto, J. J., Jahnke, B., Shoemaker, R., *et al.* (2001). Quorum sensing in the dimorphic fungus *Candida albicans* is mediated by farnesol. *Applied and Environmental Microbiology*, *67*(7), 2982–92.

Hoyer, L. L., Scherer, S., Shatzman, A. R., & Livi, G. P. (1995). *Candida albicans* ALS1: domains related to a Saccharomyces cerevisiae sexual agglutinin separated by a repeating motif. *Molecular **Microbiology**, **15*(1), 39–54.**

Hoyer, L. L., Payne, T. L., Bell, M., Myers, a M., & Scherer, S. (1998a). *Candida albicans* ALS3 and insights into the nature of the ALS gene family. *Current Genetics*, *33*(6), 451–9.

Hoyer, L. L., Payne, T. L., & Hecht, J. E. (1998b). Identification of *Candida albicans* ALS2 and ALS4 and Localization of Als Proteins to the Fungal Cell Surface. *Journal of Bacteriology*, *180*(20), 5334–5343.

Hoyer, L. L., Clevenger, J., Hecht, J. E., Ehrhart, E. J., & Poulet, F. M. (1999). Detection of Als proteins on the cell wall of *Candida albicans* in murine tissues. *Infection and Immunity*, *67*(8), 4251–5.

Hoyer, L. L., & Hecht, J. E. (2000). The ALS6 and ALS7 genes of *Candida albicans*. *Yeast (Chichester, England)*, *16*(9), 847–55.

Hoyer, L. L. (2001). The ALS gene family of *Candida albicans*. *Trends in Microbiology*, *9*(4), 176–180.

Hoyer, L. L., & Hecht, J. E. (2001). The ALS5 gene of *Candida albicans* and analysis of the Als5p N-terminal domain. *Yeast (Chichester, England)*, *18*(1), 49–60.

Hoyer, L. L., Fundyga, R., Hecht, J. E., Kapteyn, J. C., Klis, F. M., & Arnold, J. (2001). Characterization of agglutinin-like sequence genes from non-*albicans Candida* and phylogenetic analysis of the ALS family. *Genetics*, *157*(4), 1555–67.

Hoyer, L.L., Green, C.B., Oh, S-H., Ziaomin, X. (2008). Discovering the Secrets of the *Candida albicans* Agglutinin-Like Sequence (ALS) Gene Family - A Sticky Pursuit. *Medical Mycology*, *46*(1), 1–15.

Hube, B., Monod, M., Schofield, D. A., Brown, A. J., & Gow, N. A. (1994). Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. *Molecular Microbiology*, *14*(1), 87–99.

Hwang, I., Lee, J., Hwang, J. H., Kim, K.-J., & Lee, D. G. (2012). Silver nanoparticles induce apoptotic cell death in *Candida albicans* through the increase of hydroxyl radicals. *The FEBS Journal*, *279*(7), 1327–38.

Ibrahim, A. S., Spellberg, B. J., Avenissian, V., Fu, Y., Filler, S. G., & Edwards, J. E. (2005). Vaccination with recombinant N-terminal domain of Als1p improves survival during murine disseminated candidiasis by enhancing cell-mediated, not humoral, immunity. *Infection and Immunity*, *73*(2), 999–1005.

Ibrahim, A. S., Spellberg, B. J., Avanesian, V., Fu, Y., & Edwards, J. E. (2006). The Anti-*Candida* Vaccine Based on the Recombinant N-Terminal Domain of Als1p Is Broadly Active against Disseminated Candidiasis. *Infection and Immunity*, *74*(5), 3039–3041.

Ibrahim, A. S., Luo, G., Gebremariam, T., Lee, H., Schmidt, C. S., Hennessey, J. P., *et al*. (2013). NDV-3 protects mice from vulvovaginal candidiasis through T- and B-cell immune response. *Vaccine*, *31*(47), 5549–56.

Jackson, A. P., Gamble, J. A., Yeomans, T., Moran, G. P., Saunders, D., Harris, D., *et al*. (2009). Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Research*, *19*, 2231–2244.

Jacobsen, I. D., Wilson, D., Wächtler, B., Brunke, S., Naglik, J. R., & Hube, B. (2012). **Candida albicans** dimorphism as a therapeutic target. *Expert Review of Anti-Infective Therapy.*, *10*(1), 85–93.

James, H., Ghannoum, M., & Jurevic, R. (2011). The story of biofilms. *Journal of Invasive Fungal Infections*, 5(2), 37-42.

Johnson, B. A., & Blevins, R. A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR*, *4*(5), 603–14.

Jones, T., Federspiel, N. A., Chibana, H., Dungan, J., Kalman, S., Magee, B. B., *et al*. (2004). The diploid genome sequence of *Candida albicans. Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7329–34.

Jong, A. Y., Stins, M. F., Huang, S., Chen, S. H. M., Kim, K. S. I. K., Penetration, C. A., & The, O. F. (2001). Traversal of *Candida albicans* across Human Blood-Brain Barrier In Vitro, *69*(7), 4536–4544.

Jung, Y.-S., & Zweckstetter, M. (2004). Backbone assignment of proteins with known structure using residual dipolar couplings. *Journal of Biomolecular NMR*, *30*(1), 25–35.

Kamai, Y., Kubota, M., Kamai, Y., Fukuoka, T., & Filler, S. G. (2002). Contribution of *Candida albicans* ALS1 to the Pathogenesis of Experimental Oropharyngeal Candidiasis. *Infection and Immunity*, *70*(9), 5256–5258.

Kamai, Y., Kakuta, M., Shibayama, T., Fukuoka, T., & Kuwahara, S. (2005). Antifungal activities of R-135853, a sordarin derivative, in experimental candidiasis in mice. *Antimicrobial Agents and Chemotherapy*, *49*(1), 52–6.

Kanafani, Z. A., & Perfect, J. R. (2008). Antimicrobial resistance: resistance to antifungal agents: mechanisms and clinical impact. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *46*(1), 120–8.

Kantardjieff, K. A. (2003). Matthews coefficient probabilities : Improved estimates for unit cell contents of proteins, DNA, and protein – nucleic acid complex crystals. *Protein Science*, *12*, 1865–1871.

Kapteyn, J. C., Hoyer, L. L., Hecht, J. E., Andel, A., Verkleij, A. J., Makarow, M., *et al*. (2000). The cell wall architecture of *Candida albicans* wild-type cells and cell wall-defective mutants, *35*, 601–611.

Karkowska-Kuleta, J., Rapala-Kozik, M., & Kozik, A. (2009). Fungi pathogenic to humans: molecular bases of virulence of *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*. *Acta Biochimica Polonica*, *56*(2), 211–24.

Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*, *72*(4), 212–213.

Kay, L. E., & Gardner, K. H. (1997). Solution NMR spectroscopy beyond 25 kDa. *Current Opinion in Structural Biology*, ***7*(5), 722–31.**

Keeler, J. (2005). Understanding NMR Spectroscopy (1st Ed). Wiley-Blackwell. ISBN: 978-0470746080

Kim, Y., Cho, J.-Y., Kuk, J.-H., Moon, J.-H., Cho, J.-I., Kim, Y.-C., & Park, K.-H. (2004). Identification and antimicrobial activity of phenylacetic acid produced by *Bacillus licheniformis* isolated from fermented soybean, Chungkook-Jang. *Current Microbiology*, *48*(4), 312–7.

Kim, Y., & Mylonakis, E. (2011). Killing of *Candida albicans* filaments by *Salmonella enterica serovar Typhimurium* is mediated by sopB effectors, parts of a type III secretion system. *Eukaryotic Cell*, *10*(6), 782–90.

Kleywegt, G. J., & Brünger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure (London, England : 1993)*, *4*(8), 897–904.

Klis, F. M., de Koster, C. G., & Brul, S. (2011). A mass spectrometric view of the fungal wall proteome. *Future Microbiology*, *6*(8), 941–51.

Klotz, S. A., & Smith, R. L. (1995). Gelatin fragments block adherence of *Candida albicans* to extracellular matrix proteins. *Microbiology (Reading, England)*, *141 (1)*, 2681–4.

Klotz, S. A., Gaur, N. K., Lake, D. F., Chan, V., Rauceo, J., & Lipke, P. N. (2004). Degenerate Peptide Recognition by *Candida albicans* Adhesins Als5p and Als1p. *Infection and Immunity*, *72*(4), 2029–2034.

Klotz, S. A., Chasin, B. S., Powell, B., Gaur, N. K., & Lipke, P. N. (2007). Polymicrobial bloodstream infections involving *Candida* species: analysis of patients and review of the literature. *Diagnostic Microbiology and Infectious Disease*, *59*(4), 401–6.

M. O. Krause and J. H. Oliver (1979) Natural Widths of Atomic *K* and *L* Levels,
Ka X-Ray Lines and Several *KLL* Auger Lines. *J. Phys. Chem. Ref. Data* **8**, 329.

Kumamoto, C. A. (2011). Inflammation and gastrointestinal *Candida* colonization. *Current Opinion in Microbiology*, *14*(4), 386–91.

Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F., & Mackay, J. P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS Journal*, *278*(5), 687–703.

Laforce-nesbitt, S. S., Sullivan, M. A., Hoyer, L. L., & Bliss, J. M. (2008). Inhibition of *Candida albicans* adhesion by recombinant human antibody single-chain variable fragment specific for Als3p. *FEMS Immunology and Medical Microbiology*, *54*, 195–202.

Laskowski, R. a, & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, *51*(10), 2778–86.

Li, F., & Palecek, S. P. (2003). EAP1, a *Candida albicans* gene involved in binding human epithelial cells. *Eukaryotic Cell*, *2*(6), 1266–73.

Li, F., Svarovsky, M. J., Karlsson, A. J., Wagner, J. P., Marchillo, K., Oshel, P., *et al*. (2007). Eap1p, an adhesin that mediates *Candida albicans* biofilm formation in vitro and in vivo. *Eukaryotic Cell*, *6*(6), 931–9.

Li, Y., Ma, Y., Zhang, L., Guo, F., Ren, L., Yang, R., *et al*. (2012). In vivo inhibitory effect on the biofilm formation of *Candida albicans* by liverwort derived riccardin D. *PloS One*, *7*(4), e35543.

Lim, C. S.-Y., Rosli, R., Seow, H. F., & Chong, P. P. (2012). *Candida* and invasive candidiasis: back to basics. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *31*(1), 21–31.

Lin, J., Oh, S.-H., Jones, R., Garnett, J. a, Salgado, P. S., Rusnakova, S., Matthews, S.J., Hoyer, L.L., and Cota, E. (2014). The Peptide-binding Cavity Is Essential for Als3-mediated Adhesion of *Candida albicans* to Human Cells. *The Journal of Biological Chemistry*, *289*(26), 18401–18412.

Lin, L., Ibrahim, A. S., Xu, X., Farber, J. M., Avanesian, V., Baquir, B., *et al*. (2009). Th1-Th17 cells mediate protective adaptive immunity against *Staphylococcus aureus* and *Candida albicans* infection in mice. *PLoS Pathogens*, *5*(12), e1000703.

Liskamp, R. M. J., Rijkers, D. T. S., Kruijtzer, J. A. W., & Kemmink, J. (2011). Peptides and proteins as a continuing exciting source of inspiration for peptidomimetics. *Chembiochem : A European Journal of Chemical Biology*, *12*(11), 1626–53.

Liu, Y., Mittal, R., Solis, N. V, Prasadarao, N. V, & Filler, S. G. (2011). Mechanisms of *Candida albicans* trafficking to the brain. *PLoS Pathogens*, *7*(10), e1002305.

Lipke, P. N., Garcia, M. C., Alsteens, D., Ramsook, C. B., Klotz, S. A., & Dufrêne, Y. F. (2012). Strengthening relationships: amyloids create adhesion nanodomains in yeasts. *Trends in Microbiology*, *20*(2), 59–65.

Lohan, S., & Bisht, G. S. (2013). Recent approaches in design of peptidomimetics for antimicrobial drug discovery research. *Mini Reviews in Medicinal Chemistry*, *13*(7), 1073–88.

Lorenz, M. C., Bender, J. A., Fink, G. R., Lorenz, M. C., Bender, J. A., & Fink, G. R. (2004). Transcriptional Response of *Candida albicans* upon Internalization by Macrophages. *Eukaryotic Cell*, *3*(5).

Lott, T.J., Frade, J.P., Lockhart, S.R. (2010) Multilocus sequence type analysis reveals both clonality and recombination in populations of *Candida glabrata* bloodstream isolates from U.S. surveillance studies. *Eukaryotic Cell*. Vol 9: 619-625.

Loza, L., Fu, Y., Ibrahim, A. S., Sheppard, D. C., Filler, S. G., & Edwards, J. E. (2004). Functional analysis of the *Candida albicans* ALS1 gene product. *Yeast (Chichester, England)*, *21*(6), 473–82.

Luo, S., Skerka, C., Kurzai, O., & Zipfel, P. F. (2013). Complement and innate immune evasion strategies of the human pathogenic fungus *Candida albicans*. *Molecular Immunology*, *56*(3), 161–9.

Macherla, C., Sanchez, D. a, Ahmadi, M. S., Vellozzi, E. M., Friedman, A. J., Nosanchuk, J. D., & Martinez, L. R. (2012). Nitric oxide releasing nanoparticles for treatment of *Candida albicans* burn infections. *Frontiers in Microbiology*, *3*(June), 193.

Macomber, R.S. (1998) A complete introduction to modern NMR Spectroscopy. John Wiley & Sons, INC: New York. ISBN: 0-471-15736-8

Magee, B. B., Legrand, M., Alarco, A.-M., Raymond, M., & Magee, P. T. (2002). Many of the genes required for mating in *Saccharomyces cerevisiae* are also required for mating in *Candida albicans*. *Molecular Microbiology*, *46*(5), 1345–51.

Marchant, J., Sawmynaden, K., Saouros, S., Simpson, P., & Matthews, S. (2008). Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. *Biomolecular NMR Assignments*, *2*(2), 119–21.

Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, *33*(2), 491–7

Mayer, F. L., Wilson, D., & Hube, B. (2013). Candida albicans pathogenicity mechanisms. *Virulence*, *4*(2), 119–28.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–674.

Mendes, A., Mores, A. U., Carvalho, A. P., Rosa, R. T., Samaranayake, L. P., & Rosa, E. A. R. (2007). *Candida* albicans biofilms produce more secreted aspartyl protease than the planktonic cells. *Biological & Pharmaceutical Bulletin*, *30*(9), 1813–5.

Mikulska, M., Del Bono, V., Ratto, S., & Viscoli, C. (2012). Occurrence, presentation and treatment of candidemia. *Expert Review of Clinical Immunology*, *8*(8), 755–65.

Miranda, I., Silva-Dias, A., Rocha, R., Teixeira-Santos, R., Coelho, C., Gonçalves, T., *et al*. (2013). *Candida albicans* CUG mistranslation is a mechanism to create cell surface variation. *mBio*, *4*(4), 1–9.

Mitchell, A. P. (1998). Dimorphism and virulence in *Candida albicans*. *Current Opinion in Microbiology*, *1*(6), 687–92.

Mochon, A. B., Jin, Y., Ye, J., Kayala, M. A., Wingard, J. R., Clancy, C. J., *et al*. (2010). Serological profiling of a *Candida albicans* protein microarray reveals permanent host-pathogen interplay and stage-specific responses during candidemia. *PLoS Pathogens*, *6*(3), e1000827.

Monroy-Pérez, E., Sáinz-Espuñes, T., Paniagua-Contreras, G., Negrete-Abascal, E., Rodríguez-Moctezuma, J. R., & Vaca, S. (2012). Frequency and expression of ALS and HWP1 genotypes in

Candida albicans strains isolated from Mexican patients suffering from vaginal candidosis. *Mycoses*, *55*(3), e151–7.

Morales, D. K., & Hogan, D. A. (2010). *Candida albicans* interactions with bacteria in the context of human health and disease. *PLoS Pathogens*, *6*(4), e1000886.

Morales, D. K., Jacobs, N. J., Rajamani, S., Krishnamurthy, M., Cubillos-Ruiz, J. R., & Hogan, D. A. (2010). Antifungal mechanisms by which a novel *Pseudomonas aeruginosa* phenazine toxin kills *Candida albicans* in biofilms. *Molecular Microbiology*, *78*(6), 1379–92.

Moran, A. P. (2008). Relevance of fucosylation and Lewis antigen expression in the bacterial gastroduodenal pathogen *Helicobacter pylori*. *Carbohydrate Research*, *343*(12), 1952–65.

Moreno-Ruiz, E., Galán-Díez, M., Zhu, W., Fernández-Ruiz, E., d'Enfert, C., Filler, S. G., *et al*. (2009). *Candida albicans* internalization by host cells is mediated by a clathrin-dependent mechanism. *Cellular Microbiology*, *11*(8), 1179–89.

Mukherjee, P. K., & Chandra, J. (2004). *Candida* biofilm resistance. *Drug Resistance Updates : Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy*, *7*(4-5), 301–9.

Murciano C, *et al.* (2012) Evaluation of the role of *Candida albicans* agglutinin-like sequence (Als) proteins in human oral epithelial cell interactions. *PLoS One, 7*(3), e33362

Murillo, L. A., Newport, G., Lan, C.-Y., Habelitz, S., Dungan, J., & Agabian, N. M. (2005). Genome-wide transcription profiling of the early phase of biofilm formation by *Candida albicans. Eukaryotic Cell*, *4*(9), 1562–73.

Murshudov, G. N., Vagin, a a, & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica. Section D, Biological Crystallography*, *53*(Pt 3), 240–55.

Naglik, J. R., Challacombe, S. J., & Hube, B. (2003). *Candida albicans* Secreted Aspartyl Proteinases in Virulence and Pathogenesis. *Microbiology and Molecular Biology Reviews*, *67*(3), 400–428.
Naglik, J. R., Moyes, D. L., Wächtler, B., & Hube, B. (2011). *Candida albicans* interactions with epithelial cells and mucosal immunity. *Microbes and Infection / Institut Pasteur*, *13*(12-13), 963–76.

Nailis, H., Coenye, T., Van Nieuwerburgh, F., Deforce, D., & Nelis, H. J. (2006). Development and evaluation of different normalization strategies for gene expression studies in *Candida albicans* biofilms by real-time PCR. *BMC Molecular Biology*, *7*, 25.

Nailis, H., Vandenbroucke, R., Tilleman, K., Deforce, D., Nelis, H., & Coenye, T. (2009). Monitoring ALS1 and ALS3 gene expression during in vitro *Candida albicans* biofilm formation under continuous flow conditions. *Mycopathologia*, *167*(1), 9–17

Nailis, H., Kucharíková, S., Ricicová, M., Van Dijck, P., Deforce, D., Nelis, H., & Coenye, T. (2010). Real-time PCR expression profiling of genes encoding potential virulence factors in *Candida albicans* biofilms: identification of model-dependent and -independent gene expression. *BMC Microbiology*, *10*, 114.

Nas, T., Kalkanci, A., Fidan, I., Hizel, K., Bolat, S., Yolbakan, S., *et al*. (2008). Expression of ALS1, HWP1 and SAP4 genes in *Candida albicans* strains isolated from women with vaginitis. *Folia Microbiologica*, *53*(2), 179–83.

Nett, J. E., Lepak, A. J., Marchillo, K., & Andes, D. R. (2009). Time course global gene expression analysis of an in vivo *Candida* biofilm. *The Journal of Infectious Diseases*, *200*(2), 307–13.

Nett, J. E., Marchillo, K., Spiegel, C. A., & Andes, D. R. (2010). Development and validation of an in vivo *Candida albicans* biofilm denture model. *Infection and Immunity*, *78*(9), 3650–9.

Neuhaus, D., Williamson, M.P. (2000) The Nuclear Overhauser Effect in Structural and Conformational Analysis. Wiley-VCH, New York, ISBN: 0-471-24675-1.

Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, *2*(9), 2212-21.

Nobbs, A. H., Vickerman, M. M., & Jenkinson, H. F. (2010). Heterologous expression of *Candida albicans* cell wall-associated adhesins in *Saccharomyces cerevisiae* Reveals differential specificities in adherence and biofilm formation and in binding oral *Streptococcus gordonii*. *Eukaryotic Cell*, *9*(10), 1622–34.

Nobile, C. J., & Mitchell, A. P. (2005). Regulation of cell-surface genes and biofilm formation by the *C. albicans* transcription factor Bcr1p. *Current Biology : CB*, *15*(12), 1150–5.

Nobile, C. J., Andes, D. R., Nett, J. E., Smith, F. J., Yue, F., Phan, Q.-T., *et al*. (2006a). Critical role of Bcr1-dependent adhesins in *C. albicans* biofilm formation in vitro and in vivo. *PLoS Pathogens*, *2*(7), e63.

Nobile, C. J., Nett, J. E., Andes, D. R., & Mitchell, A. P. (2006b). Function of *Candida albicans* adhesin Hwp1 in biofilm formation. *Eukaryotic Cell*, *5*(10), 1604–10.

Nobile, C. J., Fox, E. P., Nett, J. E., Sorrells, T. R., Mitrovich, Q. M., Hernday, A. D., *et al*. (2012). A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell*, *148*(1-2), 126–38.

Nobile, C. J., Schneider, H. A., Nett, J. E., Sheppard, D. C., Scott, G., Andes, D. R., & Mitchell, A. P. (2008). Complementary adhesin function in *C. albicans* biofilm formation. *Current Biology : CB*, *18*(14), 1017–1024

Nucci, M., Colombo, A. L., Silveira, F., Richtmann, R., Salomão, R., Branchini, M. L., & Spector, N. (1998). Risk factors for death in patients with candidemia. *Infection Control and Hospital Epidemiology : The Official Journal of the Society of Hospital Epidemiologists of America*, *19*(11), 846–50.

Nucci, M., & Anaissie, E. (2001). Revisiting the source of candidemia: skin or gut? *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *33*(12), 1959–67.

O'Connor, L., Lahiff, S., Casey, F., Glennon, M., Cormican, M., & Maher, M. (2005). Quantification of ALS1 gene expression in *Candida albicans* biofilms by RT-PCR using hybridisation probes on the LightCycler. *Molecular and Cellular Probes*, *19*(3), 153–62.

Oh, S.-H., Cheng, G., Nuessen, J. a, Jajko, R., Yeater, K. M., Zhao, X., *et al*. (2005). Functional specificity of *Candida albicans* Als3p proteins and clade specificity of ALS3 alleles discriminated by the

number of copies of the tandem repeat sequence in the central domain. *Microbiology*, *151*(3), 673–81.

Ooi, L. (2009). Principles of X-ray Crystallography. OUP Oxford. ISBN: 978-0199569045

Otoo, H. N., Lee, K. G., Qiu, W., & Lipke, P. N. (2008). *Candida albicans* Als adhesins have conserved amyloid-forming sequences. *Eukaryotic Cell*, *7*(5), 776–82.

Pannu, N. S., & Read, R. J. (1996). Improved Structure Refinement Through Maximum Likelihood. *Acta Crystallographica Section A Foundations of Crystallography*, *52*(5), 659–668.

Park, K. D., Lee, J. H., Kim, S. H., Kang, T. H., Moon, J. S., & Kim, S. U. (2006). Synthesis of 13(substituted benzyl) berberine and berberrubine derivatives as antifungal agents. *Bioorganic & Medicinal Chemistry Letters*, *16*(15), 3913–6.

Perlroth, J., Choi, B., & Spellberg, B. (2007). Nosocomial fungal infections: epidemiology, diagnosis, and treatment. *Medical Mycology*, *45*(4), 321–46.

Pendrak, M. L., Yan, S. S., & Roberts, D. D. (2004). Sensing the host environment: recognition of hemoglobin by the pathogenic yeast *Candida albicans*. *Archives of Biochemistry and Biophysics*, *426*(2), 148–56.

Pervushin, K. (2000). Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Quarterly Reviews of Biophysics*, *33*(2), 161–97.

Peters, B. M., Jabra-Rizk, M. A., Scheper, M. A., Leid, J. G., Costerton, J. W., & Shirtliff, M. E. (2010). Microbial interactions and differential protein expression in *Staphylococcus aureus -Candida albicans* dual-species biofilms. *FEMS Immunology and Medical Microbiology*, *59*(3), 493–503.

Peters, B. M., Ovchinnikova, E. S., Krom, B. P., Schlecht, L. M., Zhou, H., Hoyer, L. L., *et al*. (2012). *Staphylococcus aureus* adherence to *Candida albicans* hyphae is mediated by the hyphal adhesin Als3p. *Microbiology*, *158*(12), 2975–86.

Pfaller, M. A, & Diekema, D. J. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. *Clinical Microbiology Reviews*, *20*(1), 133–63.

Pfaller, M. A. (2012). Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. *The American Journal of Medicine*, *125*(1 Suppl), S3–13.

Phan, Q. T., Belanger, P. H., & Filler, S. G. (2000). Role of hyphal formation in interactions of *Candida albicans* with endothelial cells. *Infection and Immunity*, *68*(6), 3485–90.

Phan, Q. T., Fratti, R. a, Prasadarao, N. V, Edwards, J. E., & Filler, S. G. (2005). N-cadherin mediates endocytosis of *Candida albicans* by endothelial cells. *The Journal of Biological Chemistry*, *280*(11), 10455–61.

Phan, Q. T., Myers, C. L., Fu, Y., Sheppard, D. C., Yeaman, M. R., Welch, W. H., *et al*. (2007). Als3 is a *Candida albicans* invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biology*, *5*(3), e64.

Phan, Q. T., Eng, D. K., Mostowy, S., Park, H., Cossart, P., & Filler, S. G. (2013). Role of endothelial cell septin 7 in the endocytosis of *Candida albicans*. *mBio*, *4*(6), e00542–13.

Playford, E. G., Webster, A. C., Sorrell, T. C., & Craig, J. C. (2006). Antifungal agents for preventing fungal infections in non-neutropenic critically ill and surgical patients: systematic review and meta-analysis of randomized clinical trials. *The Journal of Antimicrobial Chemotherapy*, *57*(4), 628–38.

Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., *et al.* (2003). A "dock, lock, and latch" structural model for a staphylococcal adhesin binding to fibrinogen. *Cell*, *115*(2), 217–28.

Purschke, F. G., Hiller, E., Trick, I., & Rupp, S. (2012). Flexible survival strategies of Pseudomonas aeruginosa in biofilms result in increased fitness compared with Candida albicans. *Molecular & Cellular Proteomics*, *11*(12), 1652–69.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*(1), 95–99.

Ramage, G., Vandewalle, K., Wickes, B. L., & López-Ribot, J. L. (2001). Characteristics of biofilm formation by *Candida albicans*. *Revista Iberoamericana de Micología*, *18*(4), 163–70.

Ramage, G., Bachmann, S., Patterson, T. F., Wickes, B. L., & López-Ribot, J. L. (2002). Investigation of multidrug efflux pumps in relation to fluconazole resistance in *Candida albicans* biofilms. *The Journal of Antimicrobial Chemotherapy*, *49*(6), 973–80.

Ramsook, C. B., Tan, C., Garcia, M. C., Fung, R., Soybelman, G., Henry, R., *et al*. (2010). Yeast cell adhesion molecules have functional amyloid-forming sequences. *Eukaryotic Cell*, *9*(3), 393–404.

Ramage, G., Tomsett, K., Wickes, B. L., López-Ribot, J. L., & Redding, S. W. (2004). Denture stomatitis: a role for *Candida* biofilms. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, *98*(1), 53–59.

Rauceo, J. M., Gaur, N. K., Lee, K., John, E., Klotz, S. A., Lipke, P. N., & Edwards, J. E. (2004). Global Cell Surface Conformational Shift Mediated by a *Candida albicans* Adhesin. *Infection and Immunity*, *72*(9), 4948–4955.

Rauceo, J. M., De Armond, R., Otoo, H., Kahn, P. C., Klotz, S. a, Gaur, N. K., & Lipke, P. N. (2006). Threonine-rich repeats increase fibronectin binding in the *Candida albicans* adhesin Als5p. *Eukaryotic Cell*, *5*(10), 1664–73.

Rautela, R., Singh, A. K., Shukla, A., & Cameotra, S. S. (2014). Lipopeptides from *Bacillus* strain AR2 inhibits biofilm formation by *Candida albicans*. *Antonie van Leeuwenhoek*, *105*(5), 809–21.

Rhodes, G. (2000) Crystallography made crystal clear (2nd Ed). London: Academic Press. ISBN: 012-587072-8.

Richard, M. L., & Plaine, A. (2007). Comprehensive analysis of glycosylphosphatidylinositol-anchored proteins in *Candida albicans. Eukaryotic Cell*, *6*(2), 119–33.

Ricicová, M., Kucharíková, S., Tournu, H., Hendrix, J., Bujdáková, H., Van Eldere, J., *et al.* (2010). *Candida albicans* biofilm formation in a new in vivo rat model. *Microbiology (Reading, England)*, *156*(Pt 3), 909–19.

Rosenbach, A., Dignard, D., Pierce, J. V, Whiteway, M., & Kumamoto, C. A. (2010). Adaptations of *Candida albicans* for growth in the mammalian intestinal tract. *Eukaryotic Cell*, *9*(7), 1075–86.

Rossmann, M. G. (1990). The Molecular Replacement Method. *Acta Crystallographica*, *A46*, 73–82.

Rossmann, M. G., & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, *15*(1), 24–31.

Rupp, B. (2009) Biomolecular Crystallography: Principles, Practise and Application to Structural Biology (1st Ed). New York: Garland Science. ISBN: 978-0815340812

Salgado, P. S., Yan, R., Taylor, J. D., Burchell, L., Jones, R., Hoyer, L. L., Matthews, S.J., Simpson, P.J., Cota, E. (2011a). Structural basis for the broad specificity to host-cell ligands by the pathogenic fungus *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(38), 15775–9.

Salgado, P. S., Yan, R., Rowan, F., & Cota, E. (2011). Expression, crystallization and preliminary X-ray data analysis of NT-Als9-2, a fungal adhesin from *Candida albicans*. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *67*(Pt 4), 467–70.

Samaranayake, Y.H., Cheung, B.P.K., Yau, J.Y.Y., Yeung, S.K.W., & Samaranayake, L.P. (2013). Human serum promotes *Candida albicans* biofilm growth and virulence gene expression on silicone biomaterial. *PloS One*, *8*(5), e62902.

Santos, M. A, Cheesman, C., Costa, V., Moradas-Ferreira, P., & Tuite, M. F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida spp*. *Molecular Microbiology*, *31*(3), 937–47.

Saville, S. P., Lazzell, A. L., Monteagudo, C., & Lopez-Ribot, J. L. (2003). Engineered control of cell morphology in vivo reveals distinct roles for yeast and filamentous forms of *Candida albicans* during infection. *Eukaryotic Cell*, *2*(5), 1053–60.

Sellam, A., Al-Niemi, T., McInnerney, K., Brumfield, S., Nantel, A., & Suci, P. A. (2009). A *Candida albicans* early stage biofilm detachment event in rich medium. *BMC Microbiology*, *9*(1), 25.

Schaller, M., Korting, H. C., Borelli, C., Hamm, G., & Hube, B. (2005). *Candida albicans*-secreted aspartic proteinases modify the epithelial cytokine response in an in vitro model of vaginal candidiasis. *Infection and Immunity*, *73*(5), 2758–65.

Schellman, J. A. (1975). Macromolecular binding. *Biopolymers*, *14*(5), 999–1018.

Schellman, J. A. (1997). Temperature, stability, and the hydrophobic interaction. *Biophysical Journal*, *73*(6), 2960–4.

Sheppard, D. C., Yeaman, M. R., Welch, W. H., Phan, Q. T., Fu, Y., Ibrahim, A. S., … Edwards, J. E. (2004). Functional and structural diversity in the Als protein family of *Candida albicans. The Journal of Biological Chemistry*, *279*(29), 30480–9.

Shirwaikar, A. A., Thomas, T., Shirwaikar, A., Lobo, R., & Prabhu, K. S. (2008). Treatment of onychomycosis: an update. *Indian Journal of Pharmaceutical Sciences*, *70*(6), 710–4.

Shorr, A. F., Chung, K., Jackson, W. L., Waterman, P. E., & Kollef, M. H. (2005). Fluconazole prophylaxis in critically ill surgical patients: a meta-analysis. *Critical Care Medicine*, *33*(9), 1928–35

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., Remmert, M., Söding, J., Thompson, J.D. & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7, 539.

Silverman, R. J., Nobbs, A. H., Vickerman, M. M., Barbour, M. E., & Jenkinson, H. F. (2010). Interaction of *Candida albicans* cell wall Als3 protein with *Streptococcus gordonii* SspB adhesin promotes development of mixed-species communities. *Infection and Immunity*, *78*(11), 4644–52.

Singleton, D. R., Masuoka, J., & Hazen, K. C. (2001). Cloning and analysis of a *Candida albicans* gene that affects cell surface hydrophobicity. *Journal of Bacteriology*, *183*(12), 3582–8.

Sobel, J. D. (2007). Vulvovaginal candidosis. *Lancet*, *369*(9577), 1961–71.

Soloviev, D. A., Fonzi, W. A., Sentandreu, R., Pluskota, E., Forsyth, C. B., Yadav, S., & Plow, E. F. (2007). Identification of pH-regulated antigen 1 released from *Candida albicans* as the major ligand for leukocyte integrin alphaMbeta2. *Journal of Immunology (Baltimore, Md. : 1950)*, *178*(4), 2038–46.

Sosinska, G. J., de Groot, P. W. J., Teixeira de Mattos, M. J., Dekker, H. L., de Koster, C. G., Hellingwerf, K. J., & Klis, F. M. (2008). Hypoxic conditions and iron restriction affect the cell-wall proteome of *Candida albicans* grown under vagina-simulative conditions. *Microbiology (Reading, England)*, *154*(Pt 2), 510–20.

Spellberg, B. J., Ibrahim, A. S., Avenissian, V., Filler, S. G., Myers, C. L., Fu, Y., & Edwards, J. E. (2005). The anti-*Candida albicans* vaccine composed of the recombinant N terminus of Als1p reduces fungal burden and improves survival in both immunocompetent and immunocompromised mice. *Infection and Immunity*, *73*(9), 6191–3.

Spellberg, B. J., Ibrahim, A. S., Avanesian, V., Fu, Y., Myers, C., Phan, Q. T., *et al*. (2006). Efficacy of the anti-*Candida* rAls3p-N or rAls1p-N vaccines against disseminated and mucosal candidiasis. *The Journal of Infectious Diseases*, *194*(2), 256–60.

Spellberg, B., Ibrahim, A. S., Lin, L., Avanesian, V., Fu, Y., Lipke, P., *et al*. (2008a). Antibody titer threshold predicts anti-candidal vaccine efficacy even though the mechanism of protection is induction of cell-mediated immunity. *The Journal of Infectious Diseases*, *197*(7), 967–71.

Spellberg, B., Ibrahim, A. S., Yeaman, M. R., Lin, L., Fu, Y., Avanesian, V., *et al*. (2008b). The antifungal vaccine derived from the recombinant N terminus of Als3p protects mice against the bacterium *Staphylococcus aureus*. *Infection and Immunity*, *76*(10), 4574–80.

Staab, J. F., Bradway, S.D., Fidel, P.L., Sundstrom, P. (1999). Adhesive and Mammalian Transglutaminase Substrate Properties of *Candida albicans* Hwp1. *Science*, *283*(5407), 1535–1538.

Stover, K., & Farley, J. (2014). Cardiac toxicity of some echinocandin antifungals. *Expert Opinion on Drug Safety*, *Early Online (Ahead of Print)*, 1–10.

Sudbery, P., Gow, N., & Berman, J. (2004). The distinct morphogenic states of *Candida albicans. Trends in Microbiology*, *12*(7), 317–24.

Sudbery, P. E. (2011). Growth of *Candida albicans* hyphae. *Nature Reviews. Microbiology*, *9*(10), 737–48.

Sundstrom, P., Balish, E., & Allen, C. M. (2002). Essential role of the *Candida albicans* transglutaminase substrate, hyphal wall protein 1, in lethal oroesophageal candidiasis in immunodeficient mice. *The Journal of Infectious Diseases*, *185*(4), 521–30.

Tatko, C. D., & Waters, M. L. (2004). Investigation of the nature of the methionine-pi interaction in beta-hairpin peptide model systems. *Protein Science : A Publication of the Protein Society*, *13*(9), 2515–22.

Thompson, G. R., Patel, P. K., Kirkpatrick, W. R., Westbrook, S. D., Berg, D., Erlandsen, J., *et al*. (2010). Oropharyngeal candidiasis in the era of antiretroviral therapy. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, *109*(4), 488–95.

Tournu, H., & Van Dijck, P. (2012). *Candida* biofilms and the host: models and new concepts for eradication. *International Journal of Microbiology*, *2012,* Article ID: 845352.

Drag, M., & Salvesen, G. S. (2010). Emerging principles in protease-based drug discovery. *Nature Reviews. Drug Discovery*, *9*(9), 690–701.

Tsai, N. Y., Laforce-Nesbitt, S. S., Tucker, R., & Bliss, J. M. (2011). A murine model for disseminated candidiasis in neonates. *Pediatric Research*, *69*(3), 189–93.

Tsang, P. W.-K., Bandara, H. M. H. N., & Fong, W.-P. (2012). Purpurin suppresses *Candida albicans* biofilm formation and hyphal development. *PloS One*, *7*(11), e50866.

Unkel, A., Barron, A.R., Boyd, J., Hanna, Ty., McHale, M., Oaxaca, G. (2012). Basic Knowledge of Nuclear Magnetic Resonance Spectroscopy (NMR). Connexions: Texas. Online: http://cnx.org/content/col11429/1.1/

Uppuluri, P., Dinakaran, H., Thomas, D. P., Chaturvedi, A. K., & Lopez-Ribot, J. L. (2009). Characteristics of *Candida albicans* biofilms grown in a synthetic urine medium. *Journal of Clinical Microbiology*, *47*(12), 4078–83.

Uppuluri, P., Chaturvedi, A. K., Srinivasan, A., Banerjee, M., Ramasubramaniam, A. K., Köhler, J. R., *et al*. (2010). Dispersion as an important step in the *Candida albicans* biofilm developmental cycle. *PLoS Pathogens*, *6*(3), e1000828.

Uwamahoro N, *et al.* (2012) The functions of Mediator in *Candida albicans* support a role in shaping species-specific gene expression. *PLoS Genetics.* 8(4): e1002613

Vagin, A. a, Steiner, R. a, Lebedev, A. a, Potterton, L., McNicholas, S., Long, F., & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2184–95.

Vediyappan, G., Rossignol, T., & d'Enfert, C. (2010). Interaction of *Candida albicans* biofilms with antifungals: transcriptional response and binding of antifungals to beta-glucans. *Antimicrobial Agents and Chemotherapy*, *54*(5), 2096–111.

Wächtler, B., Wilson, D., Haedicke, K., Dalle, F., & Hube, B. (2011). From attachment to damage: defined genes of *Candida albicans* mediate adhesion, invasion and damage during interaction with oral epithelial cells. *PloS One*, *6*(2), e17046.

Wächtler, B., Citiulo, F., Jablonowski, N., Förster, S., Dalle, F., Schaller, M., *et al*. (2012). *Candida albicans* -epithelial interactions: dissecting the roles of active penetration, induced endocytosis and host factors on the infection process. *PloS One*, *7*(5), e36952.

Walker, L. A., Gow, N. A. R., & Munro, C. A. (2010). Fungal echinocandin resistance. *Fungal Genetics and Biology : FG & B*, *47*(2), 117–26.

Weissman, Z., & Kornitzer, D. (2004). A family of *Candida* cell surface haem-binding proteins involved in haemin and haemoglobin-iron utilization. *Molecular Microbiology*, *53*(4), 1209–20.

White, T. C. (1997). Increased mRNA levels of ERG16, CDR, and MDR1 correlate with increases in azole resistance in *Candida albicans* isolates from a patient infected with human immunodeficiency virus. *Antimicrobial Agents and Chemotherapy*, *41*(7), 1482–7.

Wider, G. (2000). Structure Determination of Biological Macromolecules in Solution Using NMR spectroscopy. *BioTechniques*, *29*, 1278–1294.

Williams, D. W., Jordan, R. P. C., Wei, X.-Q., Alves, C. T., Wise, M. P., Wilson, M. J., & Lewis, M. A. O. (2013). Interactions of *Candida albicans* with host epithelial surfaces. *Journal of Oral Microbiology*, *5*. doi:10.3402/jom.v5i0.22434

Winn, M. D., Isupov, M. N., & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica Section D Biological Crystallography*, *57*(1), 122–133.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., *et al*. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography*, *67*(Pt 4), 235–42.

Winter, G. (2010). Xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography. 43,* 186-190.

Wojciechowicz, D., Lu, C. F., Kurjan, J., & Lipke, P. N. (1993). Cell surface anchorage and ligand-binding domains of the *Saccharomyces cerevisiae* cell adhesion protein alpha-agglutinin, a member of the immunoglobulin superfamily. *Molecular and Cellular Biology*, *13*(4), 2554–63.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., & Kay, L. E. (1994). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *Journal of the American Chemical Society*, *116*(26), 11655–11666.

Yan, L., Yang, C., & Tang, J. (2013). Disruption of the intestinal mucosal barrier in *Candida albicans* infections. *Microbiological Research*, *168*(7), 389–95.

Yan, R., Simpson, P. J., Matthews, S. J., & Cota, E. (2010). Backbone 1H, 15N, 13C and Ile, Leu, Val methyl chemical shift assignments for the 33.5 kDa N-terminal domain of *Candida albicans* ALS1. *Biomolecular NMR Assignments*, *4*(2), 187–90.

Yeater, K. M., Chandra, J., Cheng, G., Mukherjee, P. K., Zhao, X., Rodriguez-Zas, S. L., *et al*. (2007). Temporal analysis of *Candida albicans* gene expression during biofilm development. *Microbiology (Reading, England)*, *153*(Pt 8), 2373–85.

Yu, Q., Ding, X., Xu, N., Cheng, X., Qian, K., Zhang, B., *et al.* (2013). In vitro activity of verapamil alone and in combination with fluconazole or tunicamycin against *Candida albicans* biofilms. *International Journal of Antimicrobial Agents*, *41*(2), 179–82.

Zadik, Y., Burnstein, S., Derazne, E., Sandler, V., Ianculovici, C., & Halperin, T. (2010). Colonization of *Candida*: prevalence among tongue-pierced and non-pierced immunocompetent adults. *Oral Diseases*, *16*(2), 172–5.

Zakikhany, K., Naglik, J. R., Schmidt-Westhausen, A., Holland, G., Schaller, M., & Hube, B. (2007). In vivo transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cellular Microbiology*, *9*(12), 2938–54.

Zhang, Y., Cai, C., Yang, Y., Weng, L., & Wang, L. (2011). Blocking of *Candida albicans* biofilm formation by cis-2-dodecenoic acid and trans-2-dodecenoic acid. *Journal of Medical Microbiology*, *60*(11), 1643–50.

Zhao, X., Oh, S.-H., Cheng, G., Green, C. B., Nuessen, J. a, Yeater, K., … Hoyer, L. L. (2004). ALS3 and ALS8 represent a single locus that encodes a *Candida albicans* adhesin; functional comparisons between Als3p and Als1p. *Microbiology*, *150*(7), 2415–28.

Zhao, X., Pujol, C., Soll, D.R., Hoyer, L. L. (2003). Allelic variation in the contiguous loci encoding *Candida albicans* ALS5, ALS1 and ALS9. *Microbiology*, *149*(10), 2947–2960.

Zhao, X., Daniels, K. J., Oh, S.-H., Green, C. B., Yeater, K. M., Soll, D. R., & Hoyer, L. L. (2006). *Candida albicans* Als3p is required for wild-type biofilm formation on silicone elastomer surfaces. *Microbiology*, *152*(8), 2287–99.

Zhao, X., Oh, S.-H., Coleman, D. A., & Hoyer, L. L. (2011). ALS51, a newly discovered gene in the *Candida albicans* ALS family, created by intergenic recombination: analysis of the gene and protein, and implications for evolution of microbial gene families. *FEMS Immunology and Medical Microbiology*, *61*(3), 245–57.

Zhao, L.-X., Li, D.-D., Hu, D.-D., Hu, G.-H., Yan, L., Wang, Y., & Jiang, Y.-Y. (2013). Effect of tetrandrine against *Candida albicans* biofilms. *PloS One*, *8*(11), e79671.

Zhu, W., & Filler, S. G. (2010). Interactions of *Candida albicans* with epithelial cells. *Cellular Microbiology*, *12*(3), 273–82.

Zhu, W., Phan, Q. T., Boontheung, P., Solis, N. V, Loo, J. A., & Filler, S. G. (2012). EGFR and HER2 receptor kinase signaling mediate epithelial cell invasion by *Candida albicans* during oropharyngeal infection. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(35), 14194–9

10. Acknowledgements

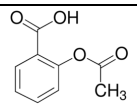
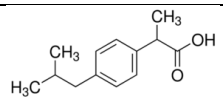
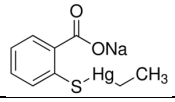
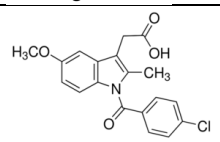
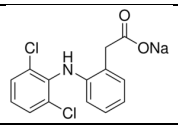
Thank you to my supervisor, Dr Ernesto Cota for providing me with a fantastic project and for reinvigorating my enthusiasm for science. Thank you also to Professor Steve Matthews for valuable advice and fascinating insights into the world of NMR Spectroscopy. Thank you to Dr James Garnett and Dr Inma Perez-Dorado for all of your help, advice and patience with X-ray crystallography. Thank you to Dr Yingqi Yu, Dr Jan Marchant and Dr Pete Simpson for all of your advice and education regarding NMR Spectroscopy. Thank you also to Dr Bernadette Byrne and Dr Rivka Isaacson for all of your advice and support.

Massive thank you to all of the past and present members of the Steve/Cota/Hare/De Simone lab for all of your advice, help and for providing such an incredible and friendly place to work, including Rhian Jones (thank you for organizing the bowling trip!), Dr Jing Lin (crystallography and karaeoki legend), Dr Inma Perez Dorado (the official Queen of electron density maps), Dr James Garnett (wise oracle of both crystallography and good music), Mostafa Jamshidiha (chemistry genius, but please don't do Walter on us), Dr Robert Yan (your NMR footsteps I couldn't quite follow, and probably also your running speed), Dr Christopher Douse (congratulations for your PhD, brilliant inspiration to complete mine), Stefi Benjamin (another great NMR pioneer), Dr Bing Liu (wise-man of crystallography and heavy water), Dr Jonathan Taylor (the cool teacher), Dr Maud Henry (merci merci beaucoup for your crystallography wisdom), Dr Masooma Rasheed (thank you for your support – I'd say more but I probably can't swear in this thesis), Dr Yi Yang (master of the machines), Dr Justin Yeoman (always on the light side of the force), Dr Stephen Hare (keep up the fantastic research), Dr Alfonso De Simone (grazie molto!), Dr Sebastian Lambert (congratulations, I hope you like sushi..), Chi-Tung Wong, Jay Patel (best of luck with your PhD – procrastinate only for worthwhile things, e.g. Honest Trailers), William Hawthorne (stay controversial) and Xeni Miliara.

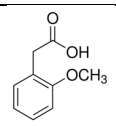
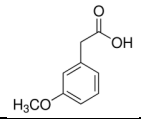
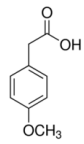
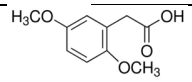
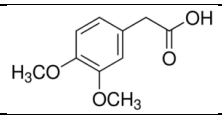
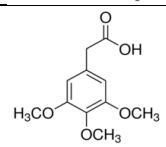
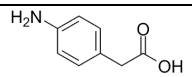
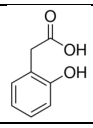
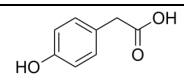
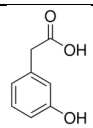
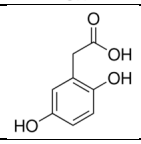
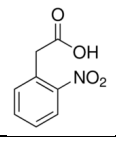
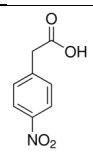
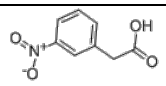
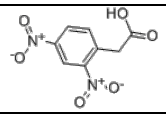
Also much gratitude for the support of my partner Carlo Vecchini, it is fantastic to be able to share a passion for science, as well as sci-fi, pasta and bourbon biscuits with you. Additionally, a special thank-you to Patricia and Kenneth Ilott for all of the laughter and support (go easy on my amateur graphic design though…).

11. Appendix

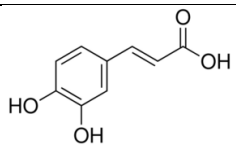
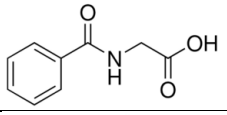
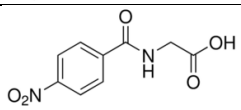
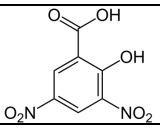
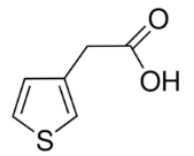
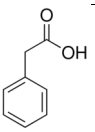
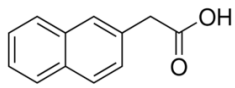
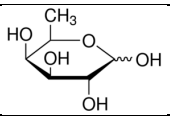
1. Established Drug molecules Screen

Drug Molecule Name	Chemistry	T_m of Replicates (°C) ± SD			Average T_m (°C) ± SD	ΔTm(°C) ± SD
		N1	N2	N3		
Control	N/A	45.249 ± 0.100	45.146 ± 0.134	45.477 ± 0.100	45.291 ± 0.169	0.000 ± 0.328
Acetyl-salicylic acid (Aspirin)		46.581 ± 0.0338	46.348 ± 0.0449	46.546 ± 0.047	46.491 ± 0.126	1.201 ± 0.287
Ibuprofen		45.921 ± 0.048	45.916 ± 0.051	46.023 ± 0.042	45.954 ± 0.060	0.662 ± 0.210
Thimerosal		45.368 ± 0.461	45.430 ± 0.045	45.397 ± 0.050	45.398 ± 0.031	0.107 ± 0.165
Indomethacin		45.739 ± 0.037	45.640 ± 0.043	45.556 ± 0.040	45.645 ± 0.092	0.354 ± 0.250
Diclofenac (Sodium Salt)		45.785 ± 0.040	45.749 ± 0.037	45.727 ± 0.042	45.754 ± 0.017	0.521 ± 0.661

2. PAA molecules screen 1. * N=3 of negative control had anomalous value so not included in calculations.

Drug Molecule Name	Chemistry	T_m of Replicates (°C) ± SD			Average T_m (°C) ± SD	ΔTm(°C) ± SD
		N = 1	N = 2	N = 3		
Negative Control	N/A	45.251 ± 0.039	45.041 ± 0.047	Anom*	45.146 ± 0.105	0.000 ± 0.254
2-methoxy-phenylacetic acid (2-Me-PAA)		47.081 ± 0.043	47.092 ± 0.039	46.679 ± 0.034	46.951 ± 0.235	1.805 ± 0.373
3-methoxy-phenylacetic acid (3-Me-PAA)		46.659 ± 0.024	46.091 ± 0.045	46.469 ± 0.031	46.409 ± 0.285	1.263 ± 0.410
4-methyoxy-phenylacetic acid (4-Me-PAA)		46.509 ± 0.037	49.694 ± 0.177	46.574 ± 0.024	47.592 ± 1.820	2.446 ± 1.027
2,5-dimethoxy-phenylacetic acid (2,5-Me-PAA)		46.294 ± 0.032	46.991 ± 0.045	46.863 ± 0.028	46.716 ± 0.371	1.570 ± 0.467
3,4-dimethoxy-phenylacetic acid (3,4-Me-PAA)		46.325 ± 0.044	45.976 ± 0.036	46.629 ± 0.021	46.310 ± 0.327	1.164 ± 0.447
3,4,5-trimethoxy-phenylacetic acid (3,4,5-Me-PAA)		46.646 ± 0.036	46.561 ± 0.036	46.818 ± 0.016	46.675 ± 0.131	1.529 ± 0.281
4-amino-phenylacetic acid (4-NH2-PAA)		47.338 ± 0.020	47.147 ± 0.024	47.247 ± 0.017	47.244 ± 0.096	2.098 ± 0.243
Ortho-hydroxy-phenylacetic acid (o-OH-PAA)		45.880 ± 0.058	46.184 ± 0.034	46.679 ± 0.033	46.248 ± 0.403	1.102 ± 0.486
Para-hydroxy-phenylacetic acid (p-OH-PAA)		46.819 ± 0.035	46.582 ± 0.030	46.671 ± 0.037	46.691 ± 0.120	1.545 ± 0.270
Meta-hydroxy-phenylacetic acid (m-OH-PAA)		46.256 ± 0.027	46.483 ± 0.032	46.700 ± 0.026	46.480 ± 0.222	1.334 ± 0.363
2,5-dihydroxy-phenylacetic acid (2,5-OH-PAA)		46.745 ± 0.082	46.628 ± 0.091	46.624 ± 0.144	46.666 ± 0.069	1.520 ± 0.208
Ortho-nitro-phenylacetic acid (o-NO2-PAA)		46.306 ± 0.043	46.245 ± 0.031	46.541 ± 0.032	46.424 ± 0.117	1.278 ± 0.267
Para-nitro-phenylacetic acid (p-NO2-PAA)		46.702 ± 0.035	46.442 ± 0.033	46.380 ± 0.032	46.508 ± 0.171	1.362 ± 0.320
Meta-nitro-phenylacetic acid (m-NO2-PAA)		45.989 ± 0.047	46.178 ± 0.038	45.965 ± 0.033	46.044 ± 0.117	0.898 ± 0.267
2,4-dinitro-phenylacetic acid (2,4-NO2-PAA)		46.298 ± 0.035	46.382 ± 0.025	46.220 ± 0.038	46.300 ± 0.081	1.154 ± 0.224

3. PAA molecules screen 2 plus fucose.

Drug Molecule Name	Chemistry	T_m of Replicates (°C) ± SD			Average T_m (°C) ± SD	ΔTm(°C) ± SD
		N1	N2	N3		
Control	N/A	45.455 ± 0.037	46.251 ± 0.039	46.041 ± 0.047	45.916 ± 0.413	0.000 ± 0.543
Caffeic acid (CAF)		48.0317 ± 0.046	48.010 ± 0.058	47.674 ± 0.072	47.905 ± 0.201	1.990 ± 0.416
Hippuric acid Na salt hydrate (HIP)		45.974 ± 0.044	45.855 ± 0.047	45.734 ± 0.046	45.854 ± 0.120	-0.061 ± 0.355
4-nitro hippuric acid (NO2-HIP)		46.466 ± 0.053	47.169 ± 0.037	46.907 ± 0.032	46.847 ± 0.355	0.921 ± 0.511
3,5-dinitrosalicylic acid (3,5-SAL)		48.599 ± 0.036	48.930 ± 0.042	49.067 ± 0.019	48.865 ± 0.241	2.950 ± 0.442
2-thiophene acetic acid (2-Thio)		46.530 ± 0.073	46.857 ± 0.042	47.157 ± 0.037	46.858 ± 0.314	0.932 ± 0.488
Phenylacetic acid (PAA)		47.068 ± 0.054	46.734 ± 0.040	47.027 ± 0.041	46.962 ± 0.147	1.047 ± 0.377
2-napthyl acetic acid (NAP)		46.796 ± 0.032	46.423 ± 0.031	46.969 ± 0.040	46.729 ± 0.279	0.814 ± 0.467
D-fucose		46.016 ± 0.048	45.959 ± 0.047	45.782 ± 0.036	45.919 ± 0.122	0.004 ± 0.357