
University of Glasgow

# Biochemical and structural characterisation of human phosphatidylinositol transfer protein Nir2 and American hookworm lipid binding protein Na-FAR-1

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

Andrei Kamenski, BSc (Hons), MRes

October 2017

# Abstract

This study focused on biochemical and structural characterisation of two lipid binding proteins: human phosphatidylinositol transfer protein (PITP) Nir2 and American hookworm fatty acid and retinol binding protein (FAR) Na-FAR-1.

Nir2 is a large multi-domain PITP that has recently been implicated in phosphoinositide signalling, where it was demonstrated to regulate phosphatidylinositol-4,5-bisphosphate [PI(4,5)$P_2$] homeostasis. Nir2 acts by reciprocally transporting phosphatidylinositol (PI) and phosphatidic acid (PA) between the plasma membrane (PM) and the endoplasmic reticulum (ER), which allows PI(4,5)$P_2$ to be re-synthesised at the PM. Upon cell stimulation, Nir2 translocates to the ER-PM contact sites and is believed to associate with the PM by binding to PA via its C-terminal LNS2 domain. Due to the proposed role of LNS2-PA binding in Nir2 targeting, a detailed investigation of the binding mechanism is desirable, which could help to reveal more details about Nir2 function in the cell. Expression screening of the difficult-to-express Nir2 LNS2 domain yielded a highly-expressed construct that was employed for characterisation of LNS2-PA binding. The data suggested that Nir2 LNS2 binds PA in a specific manner, interacts with both the polar and apolar regions of PA and might associate with the membrane via both hydrophobic and polar interactions. Although the structure of the LNS2 domain could not be determined, several assays are proposed for the identification of LNS2-PA interaction inhibitors that could be used as tool compounds in the investigation of Nir2 and its homologs.

Na-FAR-1 is a small lipid binding protein secreted by the human hookworm *Necator americanus* that infects hundreds of millions of people globally. Na-FAR-1 is known to bind a range of lipid ligands including fatty acids, retinoids and phospholipids, and was proposed to play a role in parasite-host interactions by facilitating nutrient uptake or sequestering lipid signalling molecules in the host tissues. The structure of Na-FAR-1 has been determined previously, but the molecular details of ligand binding by Na-FAR-1 remained unclear. In this study, the high-resolution structure of Na-FAR-1 in complex with its natural ligand oleic acid was determined, and the ligand binding sites were mapped. Furthermore, phospholipid binding by Na-FAR-1 was investigated, and resonance assignment of Na-FAR-1 in complex with PA was carried out, which can be used to obtain the structure of the complex. In addition, Na-FAR-1's interaction with lysophosphatidic acid was demonstrated *in vitro*. As lysophosphatidic acid is a mediator of inflammation, the interaction might have important biological implications if it also occurs *in vivo*.

# Contents

Abstract ..... 2

Contents ..... 3

List of tables ..... 10

List of figures ..... 11

Acknowledgements ..... 14

Author's declaration ..... 15

List of abbreviations ..... 16

1. Introduction ..... 21
   1.1 Lipids and their biological functions ..... 21
      1.1.1 Glycerophospholipids ..... 24
         1.1.1.1 Structure and nomenclature ..... 24
         1.1.1.2 Phosphatidylinositol and polyphosphoinositides ..... 24
         1.1.1.3 Phosphatidic acid ..... 26
      1.1.2 Lysophospholipids ..... 28
      1.1.3 Fatty acyls ..... 29
         1.1.3.1 Fatty acids ..... 29
         1.1.3.2 Eicosanoids ..... 30
      1.1.4 Retinol and retinoic acid ..... 30
   1.2 Lipid binding proteins ..... 31
      1.2.1 Phospholipid binding domains and proteins ..... 32
         1.2.1.1 PH domain ..... 32
         1.2.1.2 FYVE domain ..... 33
         1.2.1.3 PX domain ..... 33
         1.2.1.4 PA binding proteins ..... 34
      1.2.2 Nematode lipid binding proteins ..... 35
         1.2.2.1 Nematode FABPs ..... 35
         1.2.2.2 NPAs ..... 36
         1.2.2.3 FARs ..... 36
   1.3 Phosphatidylinositol transfer proteins ..... 37
      1.3.1 Class IIA PITPs ..... 38
         1.3.1.1 Nir2 ..... 41
         1.3.1.2 Nir3 ..... 44

1.3.1.3 Nir1 ..... 44
1.4 Na-FAR-1 ..... 45
1.5 Aims of this study ..... 47
1.5.1 The structure and phosphatidic acid binding mechanism of the Nir2 LNS2 domain ..... 47
1.5.2 The lipid binding mechanism of Na-FAR-1 ..... 48
2. Materials and Methods ..... 49
2.1 Materials ..... 49
2.1.1 Protein accession numbers ..... 49
2.1.2 Synthetic DNA ..... 49
2.1.3 Ligands ..... 49
2.1.4 Media recipes ..... 49
2.1.4.1 Lysogeny broth (LB) ..... 49
2.1.4.2 Autoinduction medium ..... 50
2.1.4.3 Terrific broth (TB) ..... 50
2.1.4.4 2x YT medium ..... 50
2.1.4.5 M9 minimal medium ..... 51
2.2 Protein high-throughput expression screening at OPPF-UK ..... 51
2.2.1 Molecular cloning ..... 51
2.2.2 Expression in *E. coli* ..... 52
2.2.3 Purification from *E. coli* ..... 52
2.2.4 Small-scale expression testing in insect cells ..... 53
2.2.5 Baculovirus scale-up and large-scale expression testing in insect cells ..... 54
2.3 Molecular biology, protein expression and purification ..... 54
2.3.1 Molecular cloning into pNIC28-Bsa4 vector ..... 54
2.3.2 Plasmid extraction ..... 54
2.3.3 Large-scale protein expression in *E. coli* ..... 55
2.3.3.1 Production of proteins with natural isotope abundance ..... 55
2.3.3.2 Production of isotope-labelled proteins for NMR spectroscopy ..... 55
2.3.4 Large-scale protein expression in insect cells (adherent culture) ..... 55
2.3.5 $Ni^{2+}$-affinity chromatography ..... 56
2.3.6 Size-exclusion chromatography (SEC) ..... 56
2.3.7 SEC coupled with multi-angle light scattering (SEC-MALS) ..... 56
2.3.8 SDS-PAGE ..... 57
2.3.9 Protein digestion with HRV 3C protease ..... 57

2.3.10 Western blotting ..... 57
2.3.11 Reverse-phase high pressure liquid chromatography (HPLC) ..... 58
2.3.12 Protein lyophilisation ..... 58
2.3.13 Lipid extraction ..... 58
2.3.14 Thin-layer chromatography ..... 58
2.4 Characterisation of protein-ligand binding ..... 59
2.4.1 Lipid overlay assay ..... 59
2.4.2 Liposome co-sedimentation assay (LSA) ..... 59
2.4.2.1 LSA using MLVs ..... 59
2.4.2.2 Semi-quantitative LSA using LUVs ..... 60
2.4.3 Time-resolved fluorescence lifetime measurements ..... 60
2.4.4 Preparation of large unilamellar vesicles (LUVs) by extrusion ..... 60
2.4.5 Steady-state fluorescence spectroscopy ..... 61
2.4.6 NMR spectroscopy ..... 61
2.4.6.1 Ligand-observed NMR experiments ..... 61
2.4.6.2 Chemical shift perturbation analysis ..... 62
2.4.6.3 Double- and triple-resonance experiments for structure determination ..... 62
2.4.7 Protein crystallisation, data collection and processing ..... 64
2.4.7.1 Na-FAR-1 complex with oleic acid ..... 64
2.4.7.2 Na-FAR-1 complex with DOPA ..... 64
2.4.7.3 The TF and MBP fusions of Nir2 LNS2 domain ..... 64
2.4.8 CD spectroscopy ..... 65

3. Protein expression and purification ..... 66
3.1. The LNS2 domain of Nir1-3 and RdgB ..... 67
3.1.1 Recombinant protein expression strategy ..... 67
3.1.2 Design of the LNS2 domain fragments for recombinant expression ..... 68
3.1.3 Creation of the high-throughput expression construct library ..... 69
3.1.4 High-throughput expression screening of Nir2, Nir3 and RdgB LNS2 domain in *E. coli* and insect cells ..... 73
3.1.4.1 Expression in *E. coli* ..... 73
3.1.4.2 Expression in insect cells ..... 75
3.1.5 Small-scale expression of $His_6$-tagged Nir1 and Nir2 LNS2 domain fragments in *E. coli* ..... 77
3.1.6 Large-scale expression of Nir2, Nir3 and RdgB LNS2 domain fragments in *E. coli* and insect cells ..... 78

3.1.6.1 Expression in *E. coli* ..... 79
3.1.6.2 Expression in insect cells ..... 81
3.1.7 Purification and characterisation of the TF and MBP fusions of Nir2 LNS2 domain by size-exclusion chromatography ..... 82
3.1.7.1 TF fusions of Nir2 and RdgB LNS2 domain ..... 82
3.1.7.2 MBP fusions of RdgB LNS2 domain ..... 85
3.1.8 Removal of the solubility tag of TF Nir2 LNS2-S and RdgB LNS2-S fusion ..... 86
3.2 Na-FAR-1 ..... 88
3.2.1 Expression and purification ..... 88
3.2.2 Removal of co-purifying *E. coli* lipids from recombinant Na-FAR-1 ..... 88
3.3 Conclusions ..... 89
3.3.1 The LNS2 domain of Nir1-3 and RdgB ..... 89
3.3.2 Na-FAR-1 ..... 90

4. Characterisation of the recombinant LNS2 domain of Nir2 and the interaction between the LNS2 domain and phosphatidic acid ..... 91
4.1 Introduction ..... 91
4.2 TF and MBP fusions of Nir2 LNS2 domain bind phosphatidic acid ..... 92
4.2.1 Lipid binding assays ..... 92
4.2.2 Recombinant Nir2 and RdgB LNS2 domains interact with PA ..... 93
4.2.2.1 Analysis of PA binding by lipid overlay assay ..... 94
4.2.2.2 Analysis of PA binding by liposome co-sedimentation assay ..... 95
4.3 Analysis of Nir2 LNS2 domain binding to PA by ligand-observed NMR spectroscopy ..... 98
4.3.1 Ligand-observed NMR spectroscopy ..... 98
4.3.1.1 Limitations of protein-observed NMR spectroscopy in protein-ligand binding studies ..... 99
4.3.1.2 Ligand-observed NMR spectroscopy ..... 100
4.3.1.3 WaterLOGSY ..... 101
4.3.2 Nir2 LNS2 interacts with short-chain PA and the fatty acyl chains of short-chain PC ..... 103
4.3.3 Nir2 LNS2 does not interact with free hexanoic acid nor with free glycerol 3-phosphate ..... 108
4.3.4 Nir2 LNS2 interacts with DHPA with millimolar affinity ..... 110
4.4 Analysis of Nir2 LNS2 domain binding to PA by fluorescence spectroscopy ..... 111
4.4.1 A brief overview of fluorescence spectroscopy ..... 111

4.4.1.1 Fluorescence ..... 111
4.4.1.2 Fluorescence emission, excitation and anisotropy measurements ..... 112
4.4.1.3 Fluorescence lifetime measurements ..... 114
4.4.2 Time-domain fluorescence lifetime measurements confirm DHPA binding to Nir2 LNS2 ..... 115
4.5 Analysis of the interaction between Nir2 LNS2 and PA-containing LUVs by CD spectroscopy ..... 122
4.5.1 CD spectroscopy ..... 122
4.5.2 No secondary structure change is observed in the TF fusion of Nir2 LNS2 domain upon PA binding ..... 124
4.5.3 The LNS2 domain of Nir1 has β-sheet-rich fold in solution ..... 125
4.6 Conclusions and discussion ..... 127
4.6.1 The mechanism of PA binding by the LNS2 domain of Nir2 ..... 127
4.6.2 Proposed assays for identification of Nir2 LNS2-PA interaction inhibitors ..... 131
4.6.3 Future directions ..... 132
5. Determination of the crystal structure of Na-FAR-1 complex with oleic acid ..... 134
5.1 Introduction ..... 134
5.2 A brief overview of protein X-ray crystallography ..... 135
5.2.1 The history of protein X-ray crystallography in a flash ..... 135
5.2.2 A brief introduction to the theory of protein X-ray crystallography ..... 135
5.2.2.1 Diffraction and crystals ..... 135
5.2.2.2 Bragg's law ..... 136
5.2.2.3 Structure factors and the phase problem ..... 138
5.2.2.4 Ways to solve the phase problem ..... 138
5.2.2.5 Refinement and validation of the model ..... 140
5.3 High-resolution crystal structure of Na-FAR-1 in complex with oleic acid ..... 141
5.3.1 Data collection and structure determination ..... 141
5.3.2 Structural features of the Na-FAR-1-oleic acid complex ..... 143
5.3.3 Location of the oleic acid binding sites in the Na-FAR-1 internal cavity ..... 145
5.3.3.1 Ligand fitting ..... 145
5.3.3.2 Ligand binding sites ..... 147
5.4 Conclusions and discussion ..... 150
5.4.1 Comparison of the Na-FAR-1 ligand complexes and implications for the mechanism of fatty acid binding by Na-FAR-1 ..... 150
5.4.2 Future directions ..... 152

6. Analysis of phospholipid binding by Na-FAR-1 ..... 154
6.1 Introduction ..... 154
6.2 Na-FAR-1 binds phosphatidic acid and lysophosphatidic acid ..... 155
6.3 Analysis of phosphatidic acid binding by Na-FAR-1 using protein-observed NMR spectroscopy ..... 156
6.3.1 A brief overview of protein NMR spectroscopy ..... 156
6.3.1.1 Introduction ..... 156
6.3.1.2 From nuclear spin to NMR spectrum ..... 157
6.3.1.3 Chemical shift, J-couplings and nuclear Overhauser effect (NOE) ..... 159
6.3.1.4 Multidimensional experiments in protein NMR spectroscopy ..... 160
6.3.1.5 2D HSQC and HMQC ..... 162
6.3.1.6 3D triple-resonance experiments for backbone assignment ..... 163
6.3.1.7 3D TOCSY ..... 165
6.3.1.8 2D experiments for assignment of aromatic side-chains ..... 166
6.3.1.9 3D NOESY-HSQC ..... 167
6.3.1.10 Structure calculation ..... 168
6.3.2 Chemical shift perturbation analysis of Na-FAR-1-PA binding ..... 169
6.3.2.1 Na-FAR-1 forms at least three distinct complexes with PA ..... 169
6.3.2.2 Mapping of the chemical shift changes in Na-FAR-1 induced by DOPA binding ..... 171
6.3.3 Resonance assignment of the Na-FAR-1 complex with PA ..... 175
6.3.3.1 Backbone resonance assignment ..... 175
6.3.3.2 Side-chain resonance assignment ..... 179
6.3.3.3 Assignment of DOPA resonances ..... 181
6.3.3.4 Protein resonance assignment summary ..... 182
6.3.4 Structural analysis of the Na-FAR-1-PA complex ..... 184
6.3.4.1 Structure calculation in ARIA ..... 184
6.3.4.2 Analysis of protein-ligand contacts from NOEs ..... 186
6.4 Conclusions and discussion ..... 187
6.4.1 Comparison of the Na-FAR-1 oleic acid and PA binding mechanisms ..... 188
6.4.2 Future directions ..... 190

7. Final summary ..... 191
7.1 The LNS2 domain of Nir2 ..... 191
7.2 Na-FAR-1 ..... 192

References ..... 194

Appendix A: Primers used in the expression screening of the LNS2 domains of Nir2, Nir3 and RdgB ..... 227
Nir2 ..... 227
Nir3 ..... 229
RdgB ..... 230
Appendix B: Na-FAR-1-PA complex chemical shift table ..... 232

# List of tables

Table 1-1. Lipid categories and their representative examples. 23
Table 2-1. Acquisition parameters of the protein NMR experiments carried out in this study. 63
Table 3-1. Fragments of Nir1-3 and RdgB proteins used for recombinant protein expression. 69
Table 3-2A. Expression constructs of Nir1-3 designed in this study. 71
Table 3-2B. Expression constructs of RdgB designed in this study. 72
Table 4-1. Fluorescence decay parameters of DHPA-NBD (di-6:0 PA-NBD). 118
Table 4-2. Secondary structure analysis of the TF fusion of Nir2 LNS2 domain. 125
Table 5-1. Data collection and refinement statistics for the Na-FAR-1 complex with oleic acid. 142
Table 6-1. Assignment statistics of the Na-FAR-1 residues in the Na-FAR-1-PA complex. 182
Table 6-2. Iteration parameters of the structure calculation in ARIA. 185

# List of figures

Figure 1-1. Structures of PH, FYVE and PX domains in complex with ligands. ..... 34
Figure 1-2. Solution-state structure of the repeated subunit of ABA-1, a nematode polyprotein allergen from *Ascaris suum* (PDB ID 2XV9). ..... 36
Figure 1-3. The crystal structure of human PITPα in complex with PI (PDB ID 1UW5). ..... 38
Figure 1-4. Schematic representation of the Nir1-3 and RdgB domain composition. ..... 39
Figure 1-5. Schematic depiction of Nir2 and Nir3 function in the PI(4,5)$P_2$ cycle. ..... 41
Figure 1-6. The structure of the ligand-free (apo) and *E. coli* lipid bound (holo) Na-FAR-1. ..... 46
Figure 3-1. Schematic representation of the LNS2 domain-containing Nir2 fragments used for recombinant expression. ..... 69
Figure 3-2. SDS-PAGE analysis of Nir2 and Nir3 LNS2 high-throughput expression screening in *E. coli*. ..... 74
Figure 3-3. SDS-PAGE analysis of RdgB LNS2 high-throughput expression screening in *E. coli*. ..... 76
Figure 3-4. SDS-PAGE analysis of the LNS2 domain high-throughput expression screening in insect cells. ..... 77
Figure 3-5. Small-scale expression of N-terminally $His_6$-tagged Nir1 LNS2 and Nir2 LNS2. ..... 78
Figure 3-6. Large-scale expression of trigger factor (TF) and maltose-binding protein (MBP) fusions of Nir2 and RdgB LNS2 domains in *E. coli*. ..... 80
Figure 3-7. Large-scale expression of $His_6$-tagged Nir1 LNS domain, and thioredoxin (Trx) and SUMO fusions of RdgB LNS2 in *E. coli*. ..... 81
Figure 3-8. Large-scale expression of $His_6$-tagged Nir2 LNS domain in insect cells. ..... 82
Figure 3-9. SEC trace of $Ni^{2+}$-affinity purified TF fusion of Nir2 LNS2-S. ..... 84
Figure 3-10. Tandem SEC-multi-angle light scattering (MALS) analysis of $Ni^{2+}$-affinity purified TF Nir2 LNS2-S fusion. ..... 84
Figure 3-11. SEC trace of $Ni^{2+}$-affinity purified MBP fusion of RdgB LNS2-L. ..... 86
Figure 3-12. Multi-angle light scattering (MALS) analysis of SEC-purified MBP RdgB LNS2-L fusion. ..... 86
Figure 3-13. Proteolytic cleavage of TF fusions of Nir2 and RdgB LNS2-S by HRV 3C protease. ..... 88
Figure 3-14. Purification of Na-FAR-1. ..... 89
Figure 4-1. Lipid binding assays. A. ..... 93

Figure 4-2. Lipid overlay assay with the recombinant Nir2 and RdgB LNS2 domains. ..... 97
Figure 4-3. Liposome co-sedimentation assay (LSA) with the recombinant TF fusion of Nir2 LNS2 domain. ..... 98
Figure 4-4. Analysis of DHPA (di-6:0 PA) and DHPC (di-6:0 PC) interaction with the Nir2 LNS2 domain by NMR waterLOGSY ..... 106
Figure 4-5. Analysis of DHPA (di-6:0 PA) and DHPC (di-6:0 PC) interaction with the Nir2 LNS2 domain by CPMG. ..... 107
Figure 4-6. Analysis of hexanoic acid and glycerol 3-phosphate interaction with the Nir2 LNS2 domain. ..... 109
Figure 4-7. $K_d$ calculation of the Nir2 LNS2-DHPA (di-6:0 PA) interaction using NMR waterLOGSY. ..... 111
Figure 4-8. Fluorescence lifetime measurements of DHPA-NBD (di-6:0 PA-NBD). ..... 119
Figure 4-9. Fluorescence lifetime measurements of DHPA-NBD (di-6:0 PA-NBD) displacement from TF Nir2 LNS2 by DHPA. ..... 121
Figure 4-10. Characteristic CD spectra of peptide and protein secondary structures. ..... 123
Figure 4-11. Far UV CD spectra of the TF fusion of Nir2 LNS2. ..... 125
Figure 4-12. Far UV CD spectrum of the Nir1 LNS2-S fragment. ..... 126
Figure 4-13. Multiple sequence alignment of the LNS2 domains of Nir1-3 and RdgB. ..... 130
Figure 5-1. Schematic depiction of X-ray scattering from atoms lying on two lattice planes in a crystal. ..... 137
Figure 5-2. Diffraction pattern recorded from the crystal of Na-FAR-1-oleic acid complex. ..... 142
Figure 5-3. The structure of the Na-FAR-1 oleic acid complex and its comparison with the structure of the Na-FAR-1 complex with *E. coli* lipids (PDB ID 4XCP). ..... 144
Figure 5-4. Ligand molecules inside the Na-FAR-1 complex with oleic acid. ..... 146
Figure 5-5. Oleic acid binding sites in the Na-FAR-1-oleic acid complex. ..... 149
Figure 5-6. Conformation of Na-FAR-1 K96 side-chain in the oleic acid complex and in the *E. coli* lipid complex containing palmitic acid (PDB ID 4XCP). ..... 151
Figure 6-1. DAUDA displacement assay using di-oleoyl phosphatidic acid (DOPA, di-18:1 *cis*-9 PA) and oleoyl lysophosphatidic acid (OLPA, 18:1 *cis*-9 LPA). ..... 156
Figure 6-2. Magnetisation transfer pathways in common 3D protein NMR spectroscopy experiments. ..... 165
Figure 6-3. Chemical shift perturbation analysis of the Na-FAR-1 interaction with PA. ..... 170
Figure 6-4. Backbone amide chemical shift perturbations between the $^{1}H$, $^{15}N$ HSQC spectra of apo Na-FAR-1 and Na-FAR-1 in complex with PA. ..... 174
Figure 6-5. Backbone resonance assignment using triple-resonance NMR experiments. ..... 178

Figure 6-6. Side-chain resonance assignment from HCcH-TOCSY. ..... 180
Figure 6-7. A fragment of the $^{13}$C-filtered 2D NOESY spectrum showing ligand-ligand NOEs. ..... 181
Figure 6-8. Assigned $^{1}$H, $^{15}$N HSQC spectrum of Na-FAR-1 in complex with PA. ..... 183
Figure 6-9. A region of the $^{13}$C-filtered 2D NOESY spectrum showing Na-FAR-1-PA NOEs. ..... 187

# Acknowledgements

I would like to thank everyone who has helped me on this long journey. First and foremost, my primary supervisor Brian Smith, who provided excellent guidance throughout the project, kind-heartedly shared his vast scientific knowledge with me, and reviewed the early versions of this thesis. I also thank my secondary supervisor Richard Hartley for his support and advice, Sharon Kelly for her great help in navigating the world of protein characterisation techniques, Aleksander Roszak for introducing me to X-ray crystallography, Steven Magennis and his group for kindly allowing me to use their fluorescence equipment, June Southall for the technical help with protein analysis, and all the members of the Smith group for their company, advice and help with everyday lab chores.

Special thanks go to the University of Glasgow Wellcome Trust PhD programme directors, Darren Monckton and Olwyn Byron, who organised this wonderful experience, and to my fellow PhD programme students. It was a great pleasure to spend time with you, and I have sure learned a lot of new from our research forums, retreats and meetings in the pub. I also thank my funders, the Wellcome Trust, for this opportunity and their generous support.

Finally, I thank my parents, my family, and especially my wife Alla, who kept me afloat through these years and without whom I would not be where I am today. Thank you so much for your love and your trust in me.

# Author's declaration

I declare that, except where explicit reference is made to the contributions of others, this thesis is the result of my own work and has not been submitted for any other degree, in whole or part, at the University of Glasgow or any other institution.

Andrei Kamenski

October 2017

# List of abbreviations

| | |
|---|---|
| ACN | Acetonitrile |
| ALP | Alkaline phosphatase |
| ATP | Adenosine triphosphate |
| BCIP | 5-Bromo-4-chloro-3'-indolyphosphate p-toluidine salt |
| BLAST | Basic Local Alignment Search Tool |
| BMRB | Biomolecular Magnetic Resonance Bank |
| BSA | Bovine serum albumin |
| C2 | Conserved-region 2 |
| CD | Circular dichroism |
| CDP | Cytidine diphosphate |
| CDS | Cytidine diphosphate-diacylglycerol synthase |
| CHAPS | 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate |
| cmc | Critical micelle concentration |
| CoA | Coenzyme A |
| CPMG | Carr-Purcell-Meiboom-Gill |
| CTP | Cytidine triphosphate |
| DAG | Diacylglycerol |
| DAGK | Diacylglycerol kinase |
| DAUDA | 11-(Dansylamino)undecanoic acid |
| DD-coupling | Dipole-dipole coupling |
| DHPA | Dihexanoyl phosphatidic acid |
| DHPC | Dihexanoyl phosphatidylcholine |
| DIPSI | Decoupling in the presence of scalar interactions |
| DMSO | Dimethyl sulphoxide |
| DNA | Deoxyribonucleic acid |
| DOPA | Dioleoyl phosphatidic acid |
| DOPC | Dioleoyl phosphatidylcholine |
| DPH | 1,6-Diphenyl-1,3,5-hexatriene |
| DTT | Dithiothreitol |
| EEA1 | Early endosome antigen 1 |
| ER | Endoplasmic reticulum |
| FABP | Fatty acid binding protein |

| | |
|---|---|
| FAR | Fatty acid and retinol binding protein |
| FID | Free induction decay |
| FPR2/ALX | Formyl peptide receptor 2/lipoxin A4 receptor |
| FRET | Fluorescence resonance energy transfer |
| FYVE | Fab1/YOTB/Vac1/EEA1 |
| G3P | *sn*-Glycerol 3-phosphate |
| GRAM | Glucosyltransferases, Rab-like GTPase activators and myotubularins |
| GST | Glutathione S-transferase |
| $H_6$/$His_6$ | Hexahistidine tag |
| HMQC | Heteronuclear multiple quantum coherence |
| HPLC | High pressure liquid chromatography |
| HRV | Human rhinovirus |
| HSA | Human serum albumin |
| HSQC | Heteronuclear single quantum coherence |
| $I(1,3)P_2$ | Inositol 1,3-bisphosphate |
| INEPT | Insensitive nuclei enhanced by polarization transfer |
| $IP_3$/$I(1,4,5)P_3$ | Inositol 1,4,5-trisphosphate |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| IRF | Instrument response function |
| ITC | Isothermal titration calorimetry |
| IUBMB | International Union of Biochemistry and Molecular Biology |
| IUPAC | International Union of Pure and Applied Chemistry |
| LB | Lysogeny broth |
| LNS2 | Lipin/Ned1/Smp2 |
| LPA | Lysophosphatidic acid |
| LPC | Lysophosphatidylcholine |
| LPL | Lysophospholipid |
| LSA | Liposome co-sedimentation assay |
| LUV | Large unilamellar vesicles |
| MAD | Multiple wavelength anomalous dispersion |
| MALS | Multi-angle X-ray scattering |
| MBP | Maltose-binding protein |
| MIR | Multiple isomorphous replacement |
| MLV | Multilamellar vesicle |

| | |
|---|---|
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MR | Molecular replacement |
| MW | Molecular weight |
| NaPi | Sodium phosphate |
| NBD | Nitrobenzoxadiazole |
| NBT | Nitro-blue tetrazolium chloride |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NPA | Nematode polyprotein allergen |
| NTA | Nitrilotriacetic acid |
| NUS | Non-uniform sampling |
| NusA | N-utilisation substance A |
| OD | Optical density |
| OLA | Oleic acid |
| OLPA | Oleoyl lysophosphatidic acid |
| OPPF | Oxford Protein Production Facility |
| PA | Phosphatidic acid |
| PAGE | Polyacrylamide gel electrophoresis |
| PC | Phosphatidylcholine |
| PCR | Polymerase chain reaction |
| PDB | Protein Data Bank |
| PDI | Polydispersity index |
| PE | Phosphatidylethanolamine |
| PG | Phosphatidylglycerol |
| PH | Pleckstrin homology |
| PI | Phosphatidylinositol |
| $PI(3,4,5)P_3$ | Phosphatidylinositol 3,4,5-triphosphate |
| $PI(3,4)P_2$ | Phosphatidylinositol 3,4-bisphosphate |
| $PI(3,5)P_2$ | Phosphatidylinositol 3,5-bisphosphate |
| PI(3)P | Phosphatidylinositol 3-phosphate |
| $PI(4,5)P_2$ | Phosphatidylinositol 4,5-bisphosphate |
| PI(4)P | Phosphatidylinositol 4-phosphate |
| PI(5)P | Phosphatidylinositol 5-phosphate |

| | |
|---|---|
| PIS | Phosphatidylinositol synthase |
| PKC | Protein kinase C |
| PKD | Protein kinase D |
| PLC | Phospholipase C |
| PLD | Phospholipase D |
| PM | Plasma membrane |
| PPAR | Peroxisome proliferator-activated receptor |
| PPI | Polyphosphoinositide |
| ppm | Parts per million |
| PRE | Paramagnetic relaxation enhancement |
| PS | Phosphatidylserine |
| PX | Phox homology |
| RDC | Residual dipolar coupling |
| RMSD | Root mean squared deviation |
| RNA | Ribonucleic acid |
| rpm | Rounds per minute |
| S1P | Sphingosine-1-phosphate |
| SAD | Single wavelength anomalous dispersion |
| SDS | Sodium dodecyl sulphate |
| SEC | Size-exclusion chromatography |
| SPR | Surface plasmon resonance |
| STD | Saturation transfer difference |
| SUMO | Small ubiquitin-like modifier |
| SUV | Small unilamellar vesicle |
| TB | Terrific broth |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TCSPC | Time-correlated single photon counting |
| TF | Trigger factor |
| TLC | Thin-layer chromatography |
| TOCSY | Total correlation spectroscopy |
| Tris | Tris(hydroxymethyl)aminomethane |
| TROSY | Transverse relaxation optimised spectroscopy |
| Trx | Thioredoxin |
| UDP | Uridine diphosphate |

| | |
|---|---|
| UV | Ultraviolet |
| VAP | Vesicle-associated membrane protein-associated protein |
| waterLOGSY | Water-ligand observed by gradient spectroscopy |

# 1. Introduction

## 1.1 Lipids and their biological functions

Along with proteins, carbohydrates and nuclei acids, lipids are a major class of biological molecules. The distinguishing feature of lipids is their high hydrophobicity compared to other biomolecules. Although all lipids contain hydrophobic groups, they also display great structural diversity. A systematic classification of lipids has been developed, according to which lipids are broadly divided into eight categories: 1) fatty acyls, 2) glycerolipids, 3) glycerophospholipids, 4) sterol lipids, 5) sphingolipids, 6) prenol lipids, 7) saccharolipids and 8) polyketides (Fahy *et al*., 2011, 2005). The categories are primarily characterised by the presence of certain functional groups or moieties in the lipid structure and/or the common biosynthetic pathway. For instance, glycerophospholipids are characterised by the presence of a phosphate or phosphonate group covalently linked to their glycerol backbone which distinguishes them from glycerolipids. The structures of the representative examples of the eight lipid categories are given in Table 1-1. With the information obtained from the recent large-scale lipidomics studies, several online databases containing the structures of known biological lipids have been created. To date, the most extensive lipid database is the LIPID MAPS Structure Database (http://www.lipidmaps.org/data/structure/) which is a product of the Lipid Metabolites and Pathways Strategy (LIPID MAPS), an international collaboratory project that aims to characterise all lipid species and lipid metabolic pathways in mammalian cells (Fahy *et al*., 2009; Sud *et al*., 2007).

As well having diverse structures, lipids also possess diverse functions in living organisms. The three major functions of lipids are their energetic, structural and signalling functions.

It is widely known that lipids are a major source and reservoir of energy in many organisms. In aerobes, breakdown of fatty acids in the β-oxidation process leads to generation of acetyl-CoA, which is utilised in the Krebs cycle. Total oxidation of fatty acids produces ~ 9 kcal $g^{-1}$ which is more than twice the amount of energy produced from the oxidation of sugars and proteins. Typically, triacylglycerides, which contain three fatty acyl residues attached to their glycerol backbone, are the preferred lipids for energy storage in eukaryotes.

The primary structural function of lipids is in the biological membranes that serve as the boundaries of the cell and the subcellular compartments such as the endoplasmic reticulum (ER), Golgi complex, endosomes and lysosomes. Glycerophospholipids are the major

structural units of the membranes that form the membrane bilayer. The ability of membrane glycerophospholipids to self-assemble into the bilayer structure is dictated by their physicochemical properties. Like the majority of lipids, glycerophospholipids are amphipathic molecules that possess both hydrophobic and hydrophilic groups. In an aqueous solution, it is energetically favourable for lipids to form aggregates in which their apolar regions are excluded from the interaction with the solvent (Tanford, 1980). Depending on the geometric shape of the molecule, lipids have propensity to form either bilayer or non-bilayer polymorphisms (superstructures), such as micelles and inverted cubic phases. Although biological membranes are complex mixtures of lipids that also contain non-bilayer lipids such as phosphatidylethanolamine and cardiolipin (Kruijff, 1997), the majority of glycerophospholipids in the membrane are bilayer-preferring lipids that facilitate the formation of the bilayer structure. It is widely accepted that the ability of lipids to self-assemble into superstructures has been central to the emergence of life.

The roles of lipids in cell signalling are diverse. Lipids can act as signalling molecules in their own right (Balla, 2013; Di Paolo and De Camilli, 2006; Fernandis and Wenk, 2007; Rivera and Chun, 2008), be involved in post-translational modification of proteins that can affect their signalling functions (Ahearn *et al*., 2012; Paulick and Bertozzi, 2008; Resh, 2013) or participate in formation of plasma membrane subdomains called lipid rafts which can serve as cell signalling platforms (Hancock, 2006; Lingwood and Simons, 2009; Simons and Sampaio, 2011). Lipid signalling molecules regulate numerous crucial biological processes including cell proliferation and migration, membrane trafficking, metabolism, and inflammation. In accordance with their roles in normal cellular and organismal functions, signalling lipids have also been implicated in pathogenesis of major diseases such as cancer, diabetes, neurodegenerative and infectious diseases (Wymann and Schneiter, 2008). Membrane-embedded glycerophospholipids are the major players in lipid signalling, in which they can participate by directly interacting with signalling proteins or by serving as precursors for other signalling lipids such as soluble lysophospholipids and eicosanoids.

A review of all lipid types and their functions is by far beyond the scope of this thesis. Only the functions of certain glycerophospholipids and fatty acyls that have relevance to the proteins investigated in this study will be discussed in the following sections of this chapter.

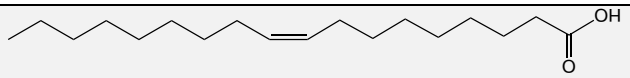
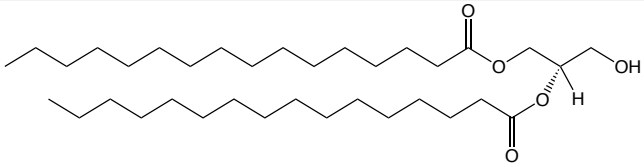
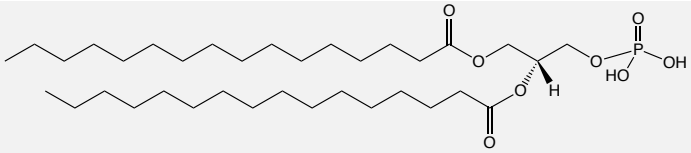
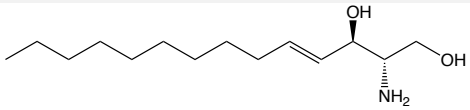
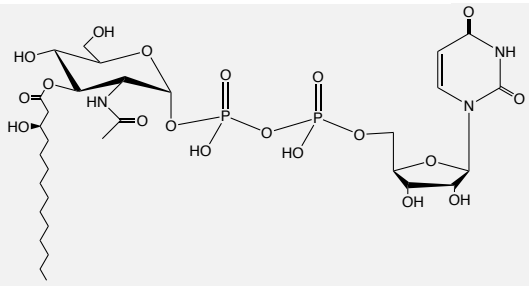
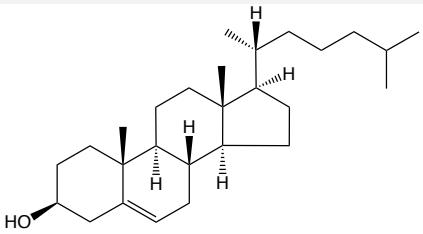
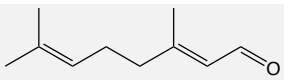
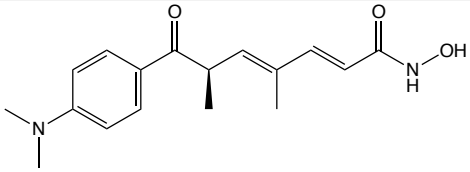
| Lipid category | Representative example |
|---|---|
| **Fatty acyls** |  Octadec-9Z-enoic acid |
| **Glycerolipids** |  1,2-Dihexadecanoyl-*sn*-glycerol |
| **Glycerophospholipids** |  1,2-Dihexadecanoyl-*sn*-glycero-3-phosphate |
| **Sphingolipids** |  Tetradecasphing-4E-enine |
| **Saccharolipids** |  UDP-3-(3R-hydroxy-tetradecanoyl)-N-acetyl-αD-glucosamine |
| **Sterol lipids** |  Cholest-5-en-3β-ol |
| **Prenol lipids** |  (Z)-3,7-dimethylocta-2,6-dienal |
| **Polyketides** |  Trichostatin A |

**Table 1-1. Lipid categories and their representative examples.**

### 1.1.1 Glycerophospholipids

#### 1.1.1.1 Structure and nomenclature

Glycerophospholipids consist of a glycerol backbone to which two fatty acyl residues and a polar head group are attached via ester linkages. The head group is attached at the *sn*-3 position in the glycerol backbone, whereas the fatty acyl tails are attached at the *sn*-1 and *sn*-2. The three *sn* positions correspond to the stereospecifically numbered carbon atoms in the glycerol molecule, according to the IUPAC-IUBMB nomenclature for lipids (International Union of Biochemistry and Molecular Biology, 1992). In the representative glycerolipid and glycerophospholipid structures shown in Table 1-1, the molecules are oriented such that their *sn*-1, *sn*-2 and *sn*-3 are shown from left to right.

Numerous glycerophospholipid species are found in nature, which can be distinguished by the type of fatty acyl residues and the head group that they contain. In biological and biochemical literature, lipids with identical head groups are typically grouped together and referred to by a single name, which is usually abbreviated. For example, if the head group of a glycerophospholipid (further, simply phospholipid) is a phosphate group, the lipid is referred to as phosphatidic acid (abbreviated as PA), if the head group is a phosphocholine moiety, the lipid is referred to as phosphatidylcholine (PC), if the head group is a phosphoinositol moiety, the lipid is referred to as phosphatidylinositol (PI), and so on. To distinguish between phospholipids that have identical head groups but different fatty acyl residues, the fatty acyl tail types can also be specified. For instance, 1,2-dioleoyl-PA (or simply dioleoyl PA or DOPA) refers to a PA species that has two oleic acid residues at *sn*-1 and *sn*-2 positions. Instead of the common fatty acid names, C:D nomenclature can be used to specify the number of carbon atoms and the number, position and stereochemistry of double bonds in the acyl chain of a fatty acid. Using C:D nomenclature, oleic acid can represented as 18:1 *cis*-9, as it contains 18 carbon atoms and a single *cis*-double bond at position ω-9 (i.e., ninth carbon-carbon bond counting from the terminal methyl group) in the acyl chain. The standard two-letter phospholipid nomenclature will be used henceforth, and the fatty acyl residues (also called fatty acyl tails) will only be specified where relevant.

#### 1.1.1.2 Phosphatidylinositol and polyphosphoinositides

PI is one of the key phospholipids in the plasma membrane that is a parent lipid of polyphosphoinositides (PPIs), a class of versatile cellular signalling lipids. PPIs are the

phosphorylated species of PI that are produced by phosphorylation of the PI *myo*-inositol ring at positions 3, 4 and 5 by the action of specific PI kinases. Depending on the pattern of the *myo*-inositol ring phosphorylation, seven different PPIs can be distinguished: PI(3)P, PI(4)P, PI(5)P, PI(3,4)$P_2$, PI(3,5)$P_2$, PI(4,5)$P_2$ and PI(3,4,5)$P_3$. The numbers in the parentheses correspond to the identifiers of the carbons in the PI inositol ring to which a phosphate group is covalently attached, not counting the phosphate group at carbon 1 via which the ring is attached to the glycerol backbone. Importantly, different PPI species can be rapidly interconverted into one another by the action of PPI kinases and phosphatases, which makes PPI signalling plastic and adaptable. In mammals, at least 19 PPI kinases and 28 PPI phosphatases have been identified, which catalyse 18 different reactions (Sasaki *et al*., 2009). The levels of PPIs vary in different cell types and membrane compartments, but are typically very low (< 1% of total cellular phospholipids).

PPIs can regulate cell signalling in at least two ways: by acting as substrates for production of signalling molecules and by engaging in direct interactions with effector proteins. A textbook example of the former function of PPIs in cell signalling is the hydrolysis of PI(4,5,)$P_2$ by phospholipase C (PLC) and the resulting generation of diacylglycerol (DAG) and inositol-1,4,5-phosphate ($IP_3$), which act as secondary messengers in the cell. The generation of $IP_3$ leads to release of $Ca^{2+}$ from the ER, which in turn activates $Ca^{2+}$-calmodulin signalling and other $Ca^{2+}$ signalling pathways (Clapham, 2007). DAG acts by binding to DAG effectors, such as the serine/threonine kinases of the protein kinase C (PKC) and protein kinase D (PKD) family, which regulate the activity of a variety of downstream signalling effectors. The role of PI(4,5,)$P_2$ and PLC in secondary messenger production was first discovered in the eighties (Berridge, 1983; Creba *et al*., 1983) and is well documented in the literature (Balla, 2013; Irvine, 2003; Katan, 2005; Rebecchi and Pentyala, 2000).

As well as being the substrates of PLCs, PPIs also regulate cell signalling processes by participating in direct protein-lipid interactions. PPI head groups are recognised by a number of phospholipid binding protein domain families that transiently interact with the membrane. Such membrane-interacting domains include the pleckstrin homology (PH) domain, PKC conserved-region 2 (C2) domain, Fab1/YOTB/Vac1/EEA1 (FYVE) domain, Phox homology (PX) domain, glucosyltransferases, Rab-like GTPase activators and myotubularins (GRAM) domain and others (Lemmon, 2008, 2003). In addition, unstructured domains containing basic and apolar residues are also known to bind certain PPIs (McLaughlin and Murray, 2005). Protein-PPI interactions can regulate signalling by modulating the activity and localisation of PPI effector proteins. Recognition of a specific

PPI by an effector allows its targeting to a specific membrane compartment, as separate membrane compartments in the cell have distinct PPI profiles. In this way, signalling partners can be brought together at a specific location to facilitate the signalling process or specific effectors can be recruited to the membrane to perform membrane-associated functions, such as the initiation of exocytosis or cytoskeleton remodelling. The dynamism of PPI generation and interconversion and the ability of PPIs to target effectors to specific subcellular compartments allows fine-tuning of cell signalling in a spatiotemporal manner.

Numerous effectors can be recruited to the membrane through interaction with PPIs. Several examples of important signalling proteins controlled by PPIs in this way include protein kinases Akt (Watton and Downward, 1999), PDK1 (Mora *et al*., 2004) and BTK (Várnai *et al*., 1999), regulators of G-proteins Cdc24 (Gulli and Peter, 2001) and Sos (Rojas *et al*., 2011), several isoforms of PLC (Williams, 1999), and SARA (Itoh *et al*., 2002), a regulator of Smad-mediated signalling. By influencing the localisation of PPI effectors, PPIs and PPI-modifying enzymes regulate a number of central pathways that control cell proliferation, survival and migration (Ooms *et al*., 2009; Vanhaesebroeck *et al*., 2005; Wymann and Pirola, 1998). Hence, it is not surprising that PPI signalling has been implicated in diseases where these crucial cellular processes are disrupted, the most prominent example of which is cancer (Bunney and Katan, 2010; Park *et al*., 2012). In addition to cancer, the role of PPIs in Type 2 diabetes, bacterial infections as well as several congenital disorders including Lowe syndrome and Charcot–Marie–Tooth disease has also been documented (Majerus and York, 2009; Pendaries *et al*., 2003).

In Section 1.2, several of the largest families of PPI binding domains will be discussed in more detail.

#### 1.1.1.3 Phosphatidic acid

Phosphatidic acid (PA) is an anionic phospholipid whose head group consists of a single phosphate group attached to the glycerol backbone via a phosphoester linkage. The basic structure of PA dictates its role as a key intermediate in *de novo* phospholipid biosynthesis. In mammals, PA is as a precursor of DAG and CDP-DAG, which are used as substrates for the synthesis of all membrane glycerophospholipids in mammalian cells (Vance, 2015). DAG is produced through PA dephosphorylation by phosphatidic acid phosphatase-1 (PAP-1), and CDP-DAG is generated from PA and CTP by CDP-DAG synthase. DAG is then used as a precursor for the synthesis of PC and PE, and CDP-DAG for the synthesis of PI and cardiolipin. In turn, PC and PE can be converted into phosphatidylserine (PS). PA

can be generated *de novo* from lysophosphatidic acid by the activity of lysophosphatidic acid acyltransferase. Alternatively, phospholipase D (PLD)-catalysed hydrolysis of PC and phosphorylation of DAG by DAG kinase also result in PA production. Due to the rapid turnover of PA, its levels in the membrane are typically low (~ 1-2% of total cellular phospholipids).

As well as functioning in phospholipid biosynthesis, PA also acts as a signalling lipid by directly interacting with PA binding proteins and facilitating their membrane recruitment. Examples of signalling proteins regulated by PA binding include Raf-1 (Ghosh *et al*., 1996), mTOR (Fang *et al*., 2001), Sos (Zhao *et al*., 2007), Rho (Kurooka *et al*., 2011) and PKCε (Jose Lopez-Andreo *et al*., 2003). It is thus not a surprise that similar to PPI signalling, PA signalling is known to regulate a breadth of cellular processes including vesicular trafficking, cell proliferation, migration and survival, and was shown to play a role in cancer (Wang *et al*., 2006).

Usually, PA effectors employ basic residues such as arginine and lysine for binding of the anionic phosphomonoester head group of PA via electrostatic interactions (Ghosh *et al*., 1996; Nakanishi *et al*., 2004; Stace and Ktistakis, 2006). The ionisation state of PA plays an important role in protein-PA binding and PA signalling. This is because the phosphomonoester group of PA (and lysophosphatidic acid) has two $pK_a$s, unlike the phosphodiester groups of other phospholipids. As one of the $pK_a$s is in the physiological pH range (Kooijman *et al*., 2005a), PA is particularly sensitive to small pH changes in the cell and has been demonstrated to have a pH sensing function in yeast (Young *et al*., 2010).

Importantly, deprotonation of PA can be facilitated by hydrogen bonding between the basic residues in the protein and the head group of PA. This effect contributes to the so-called electrostatic/hydrogen-bond switch mechanism that allows specific recognition of PA by PA binding proteins (Kooijman *et al*., 2007). In the first step of the switch mechanism, a PA binding protein is electrostatically attracted to partially-protonated PA in the membrane that carries a single negative charge. Basic residues in the ligand binding site of the protein form hydrogen bonds to the phosphate group of PA, leading to full deprotonation of the PA head group and formation of the second negative charge. As a result, the protein is able to make additional electrostatic interactions with PA, which increases its binding affinity to PA (Kooijman *et al*., 2007).

In addition to participating in direct protein-lipid interactions, PA can influence the activity of membrane-associated proteins indirectly by affecting the membrane structure. The

reason for this is the cone-shape shaped structure of PA which is unique among membrane phospholipids at physiological conditions (Kooijman *et al*., 2005b, 2003). The cone-shaped structure of PA affects the packing of the phospholipid head groups and allows the apolar layer of the membrane to be partially exposed. As a result, the presence of PA in the membrane can enhance membrane insertion of membrane-binding proteins such as dynamin (Burger *et al*., 2000).

### 1.1.2 Lysophospholipids

Lysophospholipids (LPLs) are a class of bioactive phospholipids that contain a single acyl chain in their structure. Depending on the nature of the LPL backbone, LPLs can be divided into glyceroLPLs and lysosphingolipids that belong to the glycerophospholipid and sphingolipid categories, respectively. Most prominent types of LPLs include lysophosphatidylcholine (LPC), lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P). Due to the lack of the second fatty acyl residue, LPLs are considerably more soluble in water than diacylphospholipids. GlyceroLPLs are produced as a result of diacylglycerophospholipid hydrolysis catalysed by phospholipases A (PLAs). LPA can also be synthesised *de novo* from glycerol-3-phosphate or 1-acyl-dihydroxyacetone-phosphate (Vance, 2015), or generated via lysoPLD-mediated hydrolysis of LPC. *De novo* LPA biosynthesis is of large importance in phospholipid metabolism, as LPA is a precursor of PA. S1P is produced by phosphorylation of sphingoid bases by sphingosine kinases (Vance, 2015).

LPA and S1P are well-characterised cellular signalling lipids, which act through a set of specific G-protein coupled receptors. Six distinct LPA receptors ($LPA_{1\text{-}6}$) and five distinct S1P receptors ($S1P_{1\text{-}5}$) have been characterised to date (Yung *et al*., 2014). Importantly, LPL binding to distinct LPL receptors elicits different signalling responses in the cell. Due to their solubility, LPLs can act as both extracellular and intracellular signalling mediators, and are abundant in various biological fluids including serum, saliva and follicular fluid (Rivera and Chun, 2008). Specifically, LPA signalling has been implicated in pro-, anti-inflammatory and immune responses (Choi *et al*., 2010; Gobeil *et al*., 2003; Palmetshofer *et al*., 1999; Rubenfeld *et al*., 2005; Zheng *et al*., 2000), wound healing (Khurana *et al*., 2008), the pathogenesis of cancer (Bian *et al*., 2004; Fishman *et al*., 2001; Stracke *et al*., 1992), and in inflammatory (Zhao *et al*., 2009; Zhao and Natarajan, 2013) and cardiovascular diseases (Rother *et al*., 2003; Siess, 2002; Tigyi *et al*., 1995). S1P signalling has been found to regulate processes similar to LPA including inflammation and immune cell trafficking (Huang *et al*., 2013; Pyne *et al*., 2016; Spiegel and Milstien, 2011).

### 1.1.3 Fatty acyls

Fatty acyls are a diverse class of lipids that comprises fatty acids, fatty esters, fatty amides, eicosanoids, docosanoids and other lipids. Fatty acyls are characterised by a common biosynthetic pathway, in which fatty acyl carbon chains are produced by sequential condensation of malonyl- or methylmanonyl-CoA on the acetyl-CoA primer. In this section, attention will be given primarily to polyunsaturated fatty acyls involved in cell signalling.

#### 1.1.3.1 Fatty acids

Fatty acids are the most prominent group of fatty acyls. Naturally occurring fatty acids have diverse chemical structures, and include saturated, monounsaturated, polyunsaturated, branched, carbocyclic, oxo, amino and other fatty acids. Fatty acids play a central role in metabolism, where they can act as precursors of other lipid classes, such as glycerophospholipids, sphingolipids, tri- and diacylglycerides, eicosanoids, or can be broken down in the β-oxidation pathway to produce acetyl-CoA for cellular respiration. In lipid biosynthesis, fatty acids can be esterified to the glycerol backbone by the action of acyltransferases to produce fatty acyl chains of glycerolipids and glycerophospholipids, or can be oxidised by cyclooxygenases or lipoxygenases to yield eicosanoids. Furthermore, palmitic acid is a key precursor of sphinganine, a central intermediate in the *de novo* synthesis of all sphingolipids. As lipid biosynthetic pathways differ between species, so do nutritional requirements of organisms. An example of a class of organisms with restricted fatty acid biosynthesis are parasitic nematodes, which heavily rely on acquisition of essential lipids from the host.

As well as being precursors of major signalling lipids in the cells, fatty acids can act directly as intracellular and extracellular signalling molecules (Papackova and Cahova, 2015). In the cell, fatty acids are stored in the form of triacylglycerol and phospholipid fatty acyl tails, and are released by the action of intracellular lipases. Due to their low water solubility, fatty acids are carried from their release site to their cognate receptors by the cytosolic fatty acid transport proteins including those of the fatty acid binding protein (FABP) family (Chmurzyńska, 2006; Smathers and Petersen, 2011). Intracellular fatty acid receptors include peroxisome proliferator-activated receptors (PPARs), which bind to a range of saturated and unsaturated fatty acids, and sterol-regulatory element binding protein 1 (SREBP1), which binds only polyunsaturated fatty acids. PPARs and SREBP1 are transcription factors that control the expression of genes involved in lipid transport and

metabolism and are regulated by fatty acid binding (Kersten *et al*., 2000). Importantly, extracellular fatty acids are also known to participate in cell signalling by binding to Toll-like receptors in the plasma membrane (Fessler *et al*., 2009; Guo and Friedman, 2010). Specifically, dietary saturated fatty acids have been shown to elicit pro-inflammatory responses by activating Toll-like receptors 2 and 4 (Huang *et al*., 2012).

#### 1.1.3.2 Eicosanoids

Eicosanoids are a group of bioactive lipids produced from the oxidation of arachidonic acid and other polyunsaturated fatty acids through enzymatic and non-enzymatic mechanisms. The major classes of eicosanoids include prostaglandins, leukotrienes and thromboxanes, which are generated by cyclooxygenases, lipoxygenases and cytochrome P450 enzymes, respectively. Eicosanoid production is initiated after the release of arachidonic acid from internal lipid stores by $PLA_2$, which is activated in response to external stimuli.

Eicosanoids are key regulators of inflammatory responses in the cell. They bind a variety of cognate receptors including PPARs, formyl peptide receptor 2 (FPR2, also known as lipoxin A4 receptor or ALX) and prostaglandin D, E, F receptors (Dennis and Norris, 2015). Eicosanoids can act as pro- or anti-inflammatory agents depending on the type of receptors they activate and the cell type in which the signalling occurs. Prostaglandins control classical inflammatory responses such as redness, pain, tissue swelling, heat and loss of function. Correspondingly, prostaglandins and other eicosanoids have been implicated in acute infection and injury, as well as chronic inflammatory diseases such as allergy, asthma and arthritis (Dennis and Norris, 2015; Ricciotti and FitzGerald, 2011). Interestingly, prostaglandin $E_2$ has been found to be secreted by filarial worms during infection, where it is believed to be involved in immunomodulation of the host (Liu *et al*., 1990).

### 1.1.4 Retinol and retinoic acid

Retinoic acid is a prenol lipid that is involved in a variety of cellular and organismal processes. Similar to many other lipid signalling mediators, retinoic acid is an important regulator of inflammation and immunity. It is a metabolite of retinol (vitamin A), an essential component of human diet. In the cell, retinoic acid signalling is mediated via intracellular retinoic acid receptors RAR and RXR (Altucci *et al*., 2007), which form heterodimers on ligand binding and act as transcription factors. Retinoic acid signalling has

a well-established role in gut inflammation and regulation of immune cells (Mora *et al*., 2008), and appears to be important in gastrointestinal parasitic nematode infections (Hurst and Else, 2012). Although the exact role of retinoic acid signalling in parasitic infections is unclear, it has been implicated in several immunological processes. For instance, retinoic acid signalling was found to regulate Th1, Th2 and Treg responses during pig infection by *Ascaris suum* (Dawson *et al*., 2009), and vitamin A-deficient mice have been shown to have lower *Trichinella spiralis* expulsion rates compared to non-deficient mice, as well as lower frequency of IgG1-secreting B lymphocytes (Carman *et al*., 1992). Moreover, vitamin A deficiency has been associated with an increase in Th1 responses and reduction in Th2 responses in *T. spiralis*-infected mice (Cantorna *et al*., 1994). In addition, retinoic acid signalling has been demonstrated to play a role in the maintenance of gut epithelium integrity (Osanai *et al*., 2006), which acts as a natural barrier for gut-inhabiting parasites. Interestingly, many species of parasitic nematodes secrete several distinct classes of retinoic acid binding proteins, at least some of which are believed to participate in immunomodulation of the host by sequestration of retinoic acid or its precursors from the host tissues. The lipid binding proteins of nematodes will be discussed below in Section 1.2.

## 1.2 Lipid binding proteins

In the previous section, diverse signalling functions of lipids were introduced. From the various examples presented, a common pattern can be observed where lipid signalling is mediated and regulated via the action of lipid binding proteins. Proteins can participate in lipid signalling directly by binding to signalling lipids. In turn, lipid binding can aid protein recruitment to biological membranes or lead to conformational changes in proteins that allow them to engage in interactions with their signalling partners and initiate downstream signal transduction. Proteins can also maintain lipid-dependent signalling pathways and control the levels of specific bioactive lipids available for protein-lipid interactions by acting as lipid transporters. In addition, plasma lipid binding proteins such as albumin transport nutrient lipids for energy metabolism and substrates for biosynthesis of lipids across the tissues. Intracellular counterparts of albumin are FABPs, which distribute fatty acids in the cytoplasm.

A review of all lipid binding protein types is beyond the scope of this thesis. Only the structure and functions of several phospholipid binding domains and fatty acid binding proteins will be briefly discussed below, with primary attention given to mammalian

phosphatidylinositol transfer proteins (PITPs) and nematode fatty acid and retinol binding proteins (FARs).

### 1.2.1 Phospholipid binding domains and proteins

Phospholipid signalling is mediated through the activity of phospholipid binding proteins. Many phospholipids, examples of which include several types of PPIs, bind to a number of conserved modular phospholipid binding domains shared between a large variety of phospholipid binding proteins (Cullen *et al*., 2001; Hurley *et al*., 2000). In contrast, no conserved common PA binding domain has been identified, and each family of PA binding proteins has a unique sequence typically containing basic and hydrophobic residues that interact with PA (Stace and Ktistakis, 2006). It is important to note that protein-phospholipid binding typically occurs at the membrane (binding to lysophospholipids being the exception). Hence, interactions between the phospholipid binding proteins and the bilayer lipids neighbouring the ligand often also play a role in the interaction. Furthermore, several types of lipid binding domains interact with the membrane non-specifically, recognising general membrane properties such as charge and amphiphilicity. Below, several proteins and protein domains that demonstrate specific phospholipid binding will be described.

#### 1.2.1.1 PH domain

The PH domain is one of the most widespread protein domains in the cell. Several variants of PH domain are known, which have been demonstrated to interact with $PI(3,4,5)P_3$, $PI(3,4)P_2$, $PI(4,5)P_2$, as well as PA and PS (Lemmon, 2008). Different PH domain variants have different phospholipid specificity, with several variants displaying dual specificity. Structures of several PH domains have been determined, including those of PLCδ (Ferguson *et al*., 1995), pleckstrin (Yoon *et al*., 1994), Sos1 (Zheng *et al*., 1997) and ceramide trafficking protein (Sugiki *et al*., 2012). Structural studies revealed that PH domains share a common fold consisting of a β-sandwich motif capped at one end by an α-helix (Fig 1-1A). The phospholipid-interacting face of PH domains carries a net positive charge, which is believed to facilitate the interaction with the anionic head group of its ligands (Lemmon and Ferguson, 2000). PH domains interact with PPIs via a set of basic residues in the β1–β2 loop that form hydrogen bonds to the polar head group of their ligands. Although PH domains have been shown to insert into the membrane under certain conditions (Flesch *et al*., 2005), typically, PH binding to membrane PPIs does not require membrane insertion. Indeed, it has been demonstrated that certain PH domains interact

with soluble inositol phosphates with higher affinity than with intact PPIs (Lemmon, 2008; Lemmon *et al*., 1995). In addition to phospholipid binding, PH domains have been implicated in protein-protein interactions (Lemmon, 2004; Scheffzek and Welti, 2012), suggesting than the functions of PH domains are much more diverse than simple membrane targeting of proteins.

#### 1.2.1.2 FYVE domain

FYVE domains constitute another class of modular phospholipid binding domains. Unlike PH domains, FYVE domains are only known to recognise one phospholipid species: PI(3)P (Gaullier *et al*., 1998; Patki *et al*., 1998). FYVE domains are Zinc finger proteins, and consist of two antiparallel β-sheets and an α-helix stabilised by two $Zn^{2+}$ ions (Fig 1-1B). Binding of FYVE domains to PI(3)P is mediated by the basic (R/K)(R/K)HHCR motif in their β1-strand, the residues in which interact with the phosphoinositol head group of PI(3)P via hydrogen bonding (Kutateladze and Overduin, 2001; Misra and Hurley, 1999; Saio *et al*., 2014). FYVE domains display a different mode of membrane association to PH domains. In contrast to PH domains, FYVE domains insert a loop into the lipid bilayer upon membrane binding (Kutateladze and Overduin, 2001), which allows them to enhance the affinity of PI(3)P binding. Correspondingly, FYVE domains have been demonstrated to interact with membrane-embedded PI(3)P with at least 20-fold higher affinity than with short-chain PI(3)P or the soluble $IP_2$ head group of PI(3)P (Gaullier *et al*., 2000; Kutateladze *et al*., 1999). Unlike the monomeric PH domains, FYVE domains have been shown to dimerise on membrane binding. The major function of FYVE domains is believed to be protein targeting to endosomes, which are enriched in PI(3)P (Raiborg *et al*., 2013).

#### 1.2.1.3 PX domain

PX domains were first identified in the $p40^{phox}$ and $p47^{phox}$ subunits of phagocyte NADPH oxidase, from which they take their name (Ponting, 1996). Similar to FYVE domains, PX domains are known to interact predominantly with PI(3)P, and regulate the endosomal localisation of proteins involved in membrane trafficking. PX domains comprise a subdomain consisting of three β-sheets and a subdomain consisting of three α-helices. The pocket formed between the two subdomains serves as the binding site for the PI(3)P head group (Bravo *et al*., 2001; Cheever *et al*., 2001). Like in the case of FYVE domains, PX domain binding to PI(3)P-containing membranes is achieved through a combination of

specific hydrogen bonding to PI(3)P head group and insertion of a hydrophobic loop into the membrane bilayer (Kutateladze, 2007; Lemmon, 2003).

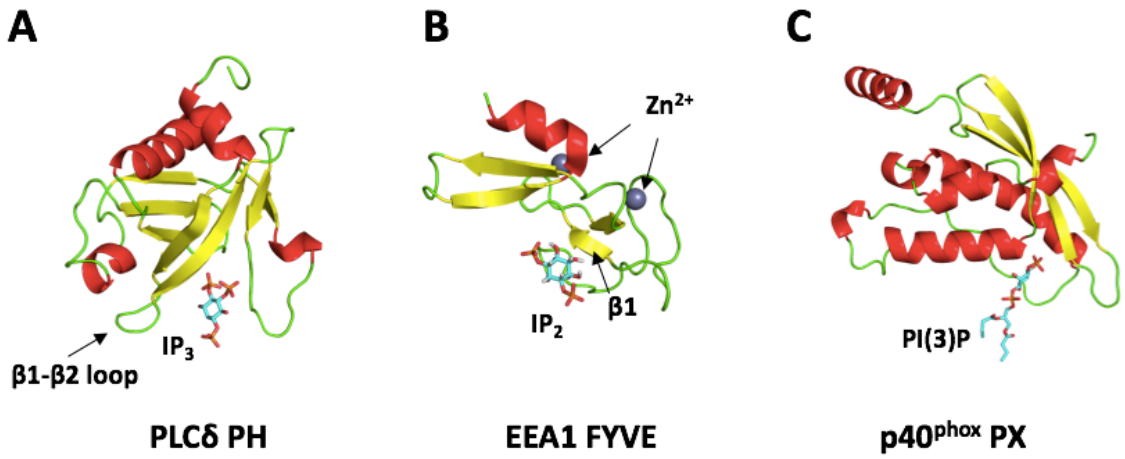



**Figure 1-1. Structures of PH, FYVE and PX domains in complex with ligands.** The structures are depicted in cartoon representation. Helices are coloured in red, β-strands in yellow and loops in green. The ligands are shown in teal. **A.** Crystal structure of the PH domain of PLCδ in complex with $I(1,4,5)P_3$, the free head group of $PI(4,5,)P_2$ (PDB ID 1MAI) (Ferguson *et al*., 1995). Protein interaction with the ligand is mediated via basic residues in the β1-β2 loop. **B.** Solution structure of the FYVE domain of early endosome antigen 1 (EEA1) in complex with $I(1,3)P_2$, the free head group of PI(3)P (PDB ID 1HYI) (Kutateladze and Overduin, 2001). $Zn^{2+}$ ions are shown as grey spheres. The protein forms polar contacts with the ligand via basic residues in its β1-strand. **C.** Crystal structure of the PX domain of $p40^{phox}$ in complex with di-butyl (di-4:0) PI(3)P (PDB ID 1H6H) (Bravo *et al*., 2001). The ligand binding pocket is formed by three α-helices and three β-strands.

### 1.2.1.4 PA binding proteins

Although several PH domains (Bullen *et al*., 2016; Zhao *et al*., 2007) and at least one PX domain (Karathanassis *et al*., 2002) were reported to bind PA, no common well-defined globular protein domain that specifically recognises PA has been identified to date. Typically, proteins interact with PA via a number of hydrophobic and basic residues which are often conserved between related proteins (Stace and Ktistakis, 2006). For instance, in mammalian Raf-1, two neighbouring arginines and a lysine in the highly conserved Raf-1 PA binding region have been shown to be directly involved in PA recognition (Rizzo *et al*., 2000). Similarly, in the yeast SNARE protein Spo21p, three lysines, an arginine and a leucine in the RLHV**KLK**SL**R**N**K**IHKQLH sequence have been directly implicated in PA binding (Nakanishi *et al*., 2004). The lack of a defined PA binding domain makes identification of PA effectors difficult. Nevertheless, the importance of PA signalling in crucial cellular processes is becoming more and more clear and novel PA binding proteins are continued to be discovered. As described in Section 1.1.1.3, electrostatic/hydrogen bond switch mechanism has been proposed to explain how PA effectors achieve PA

specificity (Kooijman *et al*., 2007). However, since few PA effectors have been well-characterised to date, the understanding of PA binding is far from complete.

### 1.2.2 Nematode lipid binding proteins

Nematodes or roundworms are a diverse class of organisms which include free-living species such as the model organism *Caenorhabditis elegans,* as well as important human, animal and plant parasites. Human parasitic nematodes, such as *Ascaris lumbricoides* and *Necator americanus,* infect more than 1.6 billion people globally (de Silva *et al*., 2003; Pullan *et al*., 2014), causing considerable human suffering and a profound deleterious impact on the society (Stephenson *et al*., 2000) in many developing regions of the world.

Parasitic nematodes have restricted lipid metabolism (Smyth and Wakelin, 1994) and rely on acquisition of lipids essential for their survival from the environment. In order to transport, store and protect lipids from degradation, nematodes produce a range of lipid binding proteins. Due to their functions in nematode survival and parasitism, lipid binding proteins have been suggested as potential therapeutic or vaccine targets in the fight against nematode infections. To date, several classes of nematode lipid binding proteins have been identified and characterised, which include nematode fatty acid binding proteins (nemFABPs), nematode polyprotein allergens (NPAs) and fatty acid and retinol binding proteins (FARs).

#### 1.2.2.1 Nematode FABPs

FABPs are ~14 kDa proteins widespread in animals, including mammals. Although the primary sequence of FABPs can vary considerably, all known FABP structures contain a conserved β-barrel fold with a water-filled internal cavity that contains the ligand binding sites (Sacchettini *et al*., 1989; Thompson *et al*., 1997; Zimmerman and Veerkamp, 2002). In mammals, FABPs are cytosolic proteins involved in cell growth (Sorof, 1994), signalling (Wolfrum *et al*., 2001), differentiation (Veerkamp and Zimmerman, 2001) and pathogenesis (Baier *et al*., 1995). Unlike mammalian FABPs, nemFABPs are extracellular proteins that are secreted by the parasite. While the definite biological function of nemFABPs is yet to be discovered, they have been implicated in the maintenance of the worm egg lipid layer (Mei *et al*., 1997; Michalski *et al*., 2002), and have been proposed to play a role in nutrient acquisition by the parasites (Franchini *et al*., 2015). No empirical structures of nemFABPs have been published to date.

### 1.2.2.2 NPAs

In contrast to FABPs, NPAs are found exclusively in nematodes. NPAs are produced as protein polymers, which are proteolytically cleaved into functional monomeric units and secreted by the parasite. NPAs have been identified from the strong IgE-type response that they elicit in the hosts (Tomlinson *et al*., 1989). They have been demonstrated to bind fatty acids and retinol and are believed to be involved in the transport or sequestration of small lipids (Kennedy, 2011; Kennedy *et al*., 1995). The structure of the repeated unit of *Ascaris suum* NPA ABA-1 has been solved, revealing an α-helical fold with two lipid binding pockets (Fig 1-2) (Meenan *et al*., 2011).

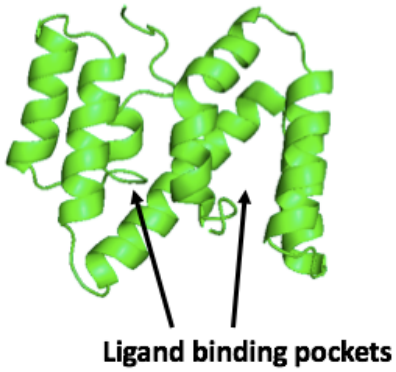



**Figure 1-2. Solution-state structure of the repeated subunit of ABA-1, a nematode polyprotein allergen from *Ascaris suum* (PDB ID 2XV9).** The structure is shown in cartoon representation. The locations of the two ligand binding pockets are shown. (Meenan *et al*., 2011).

### 1.2.2.3 FARs

FARs are another distinct family of nematode lipid binding proteins. Like FABPs and NPA monomers, FARs are small and soluble proteins, with the average molecular weight equal to 14-20 kDa. FARs have been identified in *C. elegans* (Ce-FAR-1 to -8) and in a variety of parasitic nematodes, including human parasites *Onchocerca volvulus* (Ov-FAR-1), *Brugia malayi* (Bm-FAR-1) and *Necator americanus* (Na-FAR-1), as well as several animal and plant parasites. FARs demonstrate high expression levels in the worm tissues (Basavaraju *et al*., 2003; Popeijus *et al*., 2000; Ranjit *et al*., 2006) and some FARs, including Na-FAR-1, are secreted into the host. The ligand repertoire of FARs is wide and includes fatty acids, retinoids and phospholipids (Basavaraju *et al*., 2003; Garofalo *et al*., 2003, 2002; Iberkleid *et al*., 2013; Jordanova *et al*., 2009; Rey-Burusco *et al*., 2015). In addition, Ov-FAR-1 has been shown to bind ivermectin, an antihelmintic drug (Lal and James, 1996; Sani and Vaid, 1988). Although the exact function of FARs in parasitic nematodes remains unclear, they were proposed to play a role in nutrient acquisition and/or sequestration of lipid signalling molecules to modulate the host's response to the parasitic infection (Garofalo *et al*., 2002; Rey-Burusco *et al*., 2015). To date, the only FAR that has

been directly implicated in the host immunomodulation is *Meloidogyne javanica* FAR-1 (Mj-FAR-1), which has been shown to increase tomato susceptibility to root-knot nematode infections (Iberkleid *et al*., 2015, 2013).

Na-FAR-1, which is the focus of this study and has potential biomedical importance due to its presence in a widespread human parasite, will be discussed in more detail in Section 1.4.

## 1.3 Phosphatidylinositol transfer proteins

Phosphatidylinositol transfer proteins (PITPs) are phospholipid binding proteins involved in a wide range of biological processes including cytokinesis (Giansanti *et al*., 2006; Litvak *et al*., 2004), cell proliferation (Cockcroft and Garner, 2013; Schenning *et al*., 2008; Snoek, 2004), vesicular transport (Cockcroft, 1999; Jones *et al*., 1998; Simon *et al*., 1998), neurodevelopment (Cosker *et al*., 2008; Xie *et al*., 2005) and phototransduction (Harris and Stark, 1977; Kohn *et al*., 2007; Trivedi and Padinjat, 2007). The defining biochemical characteristic of PITPs is their ability to bind and transport phospholipids between lipid membranes. Consistently, they were shown to be key players in the non-vesicular phospholipid transport in the cell (Cockcroft and Garner, 2011).

Based on their phospholipid binding properties and protein domain composition, PITPs are commonly divided into three classes: class I, class IIA and class IIB PITPs. Class I and IIB PITPs comprise a single PITP domain, whereas class IIA PITPs also contain additional domains (Cockcroft and Garner, 2011). In humans, five PITPs are present: two class I PITPs: PITPα and PITPβ; two class IIA PITPs: Nir2 (also known as RdgBαI or PITPNM1) and Nir3 (RdgBαII or PITPNM2); and one class IIB PITP: RdgBβ (PITPNC1). Furthermore, Nir1 (RdgBαIII or PITPNM3), a homolog of Nir2 and Nir3 lacking the PITP domain, is also found in the genome of humans. PITPs are highly conserved between mammals, and are also found in other classes of eukaryotes. All PITPs bind PI with high affinity and transfer it between cellular membranes. In addition, PITPα and PITPβ have been demonstrated to bind and transfer PC, and Nir2, Nir3 and RdgBβ have been shown to bind and transfer PA.

PITPα and PITPβ are small ~ 35 kDa proteins. The structures of the apo (Schouten *et al*., 2002), PI- (Tilley *et al*., 2007) and PC-bound (Yoder *et al*., 2001) PITPα, as well as the structure of PC-bound PITPβ (Vordtriede *et al*., 2005) have been determined using X-ray crystallography. Structural analysis revealed that the proteins share a common fold,

featuring a large concave β-sheet and seven α-helices (Fig 1-3). PITPα and PITPβ comprise three structural-functional units: 1) the lipid binding core, 2) the regulatory loop and 3) the C-terminal region (Cockcroft and Carvou, 2007). The lipid binding core forms the largest part of the proteins and can accommodate a single PI or PC molecule. The regulatory loop contains a PKC phosphorylation site, and has been proposed to act as an interface for interactions with protein- and lipid modifying enzymes (Yoder *et al*., 2001). The C-terminal region contains helix G and eleven C-terminal amino acids, which cap the entrance to the lipid binding core upon PI or PC binding. The phospholipid exchange happens in the open protein conformation when helix G is shifted outward and the cavity is open to the membrane interface. In the closed conformation, class I PITPs travel through the cytosol between subcellular compartments carrying PI or PC for the exchange (Cockcroft and Carvou, 2007). While the mammalian class I PITPs have been reasonably well-characterised, less is known about class II PITPs, especially in regard to the structure and function of domains absent in other PITPs. Below, the current knowledge about human class IIA PITPs will be briefly summarised.

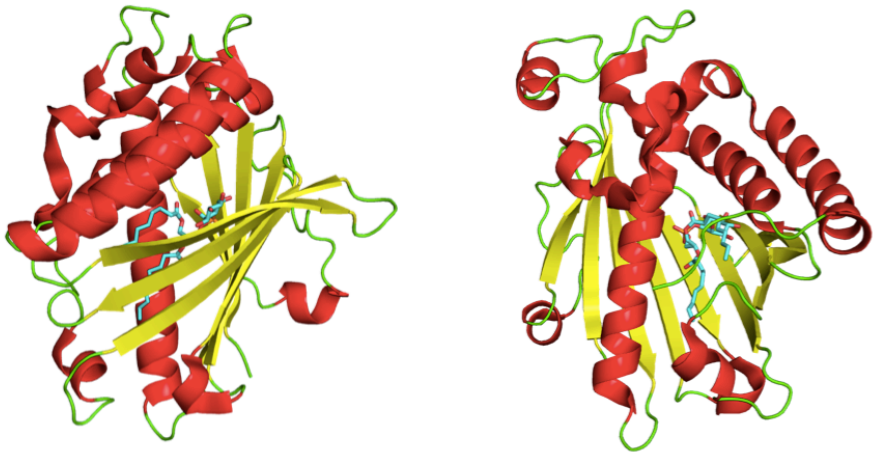


**Figure 1-3. The crystal structure of human PITPα in complex with PI (PDB ID 1UW5).** Two views of the structure are shown in cartoon representation. Protein α-helices are coloured in red, β-strands in yellow and loops in green. The ligand (teal) is shown inside the structure. (Tilley *et al*., 2007).

### 1.3.1 Class IIA PITPs

Compared to PITPα and PITPβ, class IIA are much larger proteins with molecular masses of ~ 150-170 kDa. The founding member of the IIA sub-class of PITPs is RdgB (PITPNM) protein that was identified from the rdgB (retinal degeneration B) mutation in *Drosophila* (Vihtelic *et al*., 1993). The PITP domain of class IIA PITPs is located at the N-terminal end of the proteins. In addition, class IIA PITPs contain a FFAT motif, a DDHD domain and an Lipin/Ned1/Smp2 (LNS2) domain. The schematic domain composition of class IIA

PITPs Nir2, Nir3 and RdgB is depicted in Fig 1-4. Although Nir1 does not contain a PITP domain, it was included in Fig 1-4 with class IIA PITPs due to the shared domain composition and relatively high sequence homology with the PITPs.

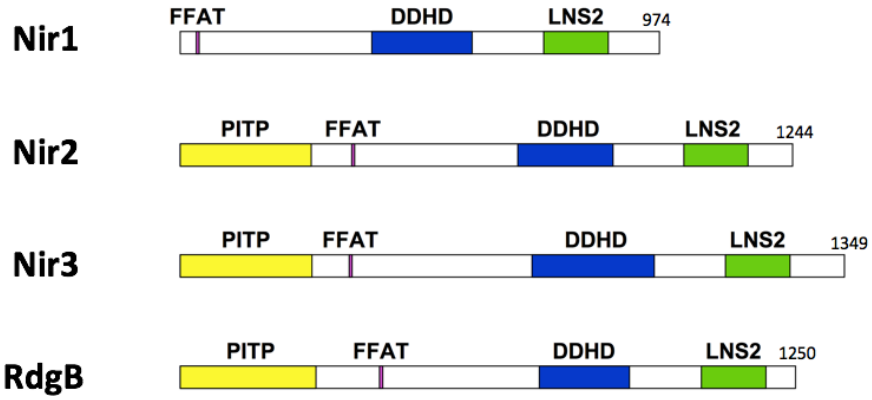



**Figure 1-4. Schematic representation of the Nir1-3 and RdgB domain composition.** Nir1-3 and RdgB comprise a VAP-interacting FFAT motif, a putative metal binding DDHD domain and a phosphatidic acid binding haloacid dehalogenase-like LNS2 domain. Nir2, Nir3 and RdgB contain a PITP domain at their N-terminus, which is absent in Nir1.

The FFAT motif is a short peptide (EFFDAxE, where x is any amino acid) that binds to VAPs [vesicle-associated membrane protein (VAMP)-associated proteins] (Loewen *et al*., 2003; Loewen and Levine, 2005), which are integral membrane proteins localised at the cytoplasmic face of the ER. Correspondingly, Nir1-3 have been shown to interact with VAP-B via their FFAT motifs (Amarilio *et al*., 2005), which is consistent with the proposed function of class IIA PITPs in the PI(4,5)$P_2$ cycle (see below). The acidic residues surrounding the FFAT motif have also been shown to bind $Ca^{2+}$ *in vitro* (Lev *et al*., 1999; Vihtelic *et al*., 1993), however, it is unclear whether this interaction occurs or plays any role in the regulation of class IIA PITPs and Nir1 *in vivo*.

The DDHD domain comprises 195 amino acids and is found in class IIA PITPs and several phospholipases (Lev, 2004). It is named after the four residues (DDHD) that are conserved between all DDHD domains. Although the function of the DDHD domain is unknown, the DDHD peptide may form a metal ion binding site (Lev, 2004). The DDHD domain is a part of the larger C-terminal region of Nir1-3 that was found to interact with PYK2 tyrosine kinase in the study by Lev *et al*. (Lev *et al*., 1999), in which Nirs (PYK2 N-terminal domain-interacting receptors) were identified. Recently, the DDHD domain was shown to interact with PI(4)P and PA *in vitro* (Klinkenberg *et al*., 2014).

The LNS2 domain is composed of ~ 130 amino acids and is proposed to belong to the haloacid dehalogenase (HAD) protein superfamily. Apart from class IIA PITPs and Nir1, the LNS2 domain is also found in mammalian lipins and yeast lipins orthologs Smp2 and Ned1. In lipins, Smp2 and Ned1, the LNS2 domain has a $Mg^{2+}$-dependent phosphatidic acid phosphatase (PAP) activity (Donkor *et al*., 2007; Han *et al*., 2006). In contrast, in class IIA PITPs, LNS2 domain lacks the catalytic function and simply binds PA (Kim *et al*., 2013). The lack of PAP function is attributed to an aspartate to alanine change in the catalytic DxDxT motif that is critical for the PAP activity in lipins (Mietkiewska *et al*., 2011; Reue and Dwyer, 2009). Crucially, PA binding by the LNS2 domain is believed to be required for the plasma membrane targeting of Nir2 and Nir3 (Chang and Liou, 2015; Kim *et al*., 2013).

Nir2, Nir3 and RdgB have been directly implicated in PLC-mediated signalling, where their PI-PA exchange function is believed to be required for the maintenance of $PI(4,5)P_2$ cycle (Chang and Liou, 2015; Cockcroft *et al*., 2016; Kim *et al*., 2013, 2015; Yadav *et al*., 2015). The $PI(4,5)P_2$ cycle is a series of biochemical reactions that lead to generation of $PI(4,5)P_2$ in the plasma membrane (PM) (Fig 1-5) (Cockcroft and Raghu, 2016). In the $PI(4,5)P_2$ cycle, $PI(4,5)P_2$ hydrolysis in response to PLC activation by external stimuli leads to production of DAG and $IP_3$, of which DAG can be phosphorylated to generate PA. In order to maintain PLC-mediated signalling, re-synthesis of $PI(4,5)P_2$ from PI is required. Since only a limited pool of PI is available at the PM for $PI(4,5)P_2$ synthesis, PI needs to be transported from its *de novo* synthesis site in the ER to replenish $PI(4,5)P_2$ levels at the PM and maintain the signalling processes. Conversely, the PA that has accumulated at the PM as a result of PLC-mediated signalling needs to be transported to the ER to maintain the supply of PI required for the production of $PI(4,5)P_2$. By reciprocally transferring PI and PA between the ER and PM, class IIA PITPs provide the substrate for $PI(4,5)P_2$ synthesis in PM and PI synthesis in ER. The reported localisation of Nir2, Nir3 and RdgB at the ER-PM junctions (Chang *et al*., 2013; Chang and Liou, 2015; Cockcroft *et al*., 2016; Kim *et al*., 2013), which are the narrow (< 40 nm) contact sites between the two membranes, is in agreement with this putative function.

Below, human class IIA PITPs Nir2 and Nir3, as well as their sister protein Nir1 will be described in more detail.

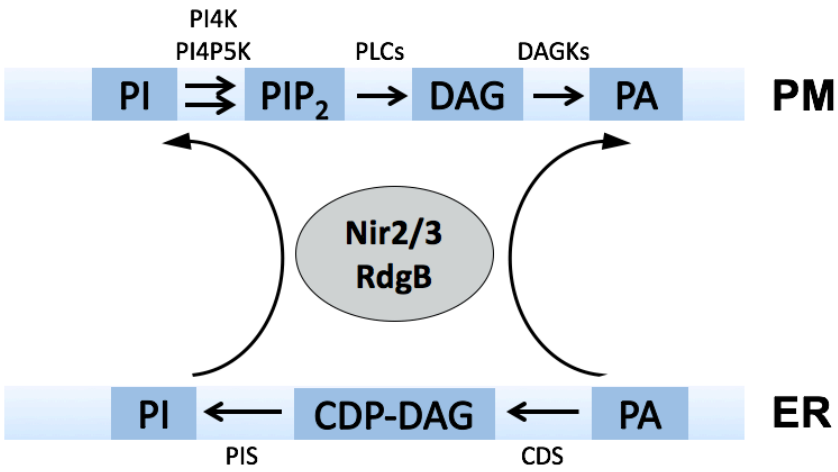



**Figure 1-5. Schematic depiction of Nir2 and Nir3 function in the PI(4,5)$P_2$ cycle.** During PLC-mediated signalling, PI(4,5)$P_2$ ($PIP_2$) is hydrolysed at the plasma membrane (PM) to produce DAG that is converted into PA. To maintain PLC-medicated signalling, PI is consumed at PM for the synthesis of $PIP_2$. As PI is synthesised from PA in the endoplasmic reticulum (ER), the ER-PM transport of PI and the PM-ER transport of PA is required to maintain the levels of PI, and, correspondingly, $PIP_2$ at the plasma membrane. The PI and PA transfer function of Nir2, Nir3 and RdgB is required to transport the lipids between the ER and PM. PI4K = PI 4-kinase, PI4P5K = PI 4-phosphate 5-kinase, PLC = phospholipase C, DAGK = DAG kinase, CDS = CDP-DAG synthase, PIS = PI synthase.

#### 1.3.1.1 Nir2

Although Nir2 is the most well-characterised human class IIA PITP to date, its cellular functions are only beginning to be elucidated. As mentioned above, Nir2 was isolated as a PYK2-binding protein together with other Nirs in 1999 (Lev *et al*., 1999). Several studies in cultured mammalian cells have followed, which revealed that Nir2 is a regulator of cellular morphogenesis (Tian *et al*., 2002) and cytokinesis (Litvak *et al*., 2004, 2002). In the study by Tian *et al*., co-expression of RhoA and Nir2 in human cerebellar medulloblastoma cells resulted in inhibition of RhoA-mediated neurite retraction, whereas injection of anti-Nir2 antibodies lead to a reduced neurite extension, suggesting that Nir2 plays a role in cytoskeleton remodelling and controls cell shape (Tian *et al*., 2002). Consistent with this function, overexpression of different truncated Nir2 mutants resulted in morphological changes in a range of mammalian cell lines (Tian *et al*., 2002). Similar results were obtained by Litvak *et al*. who demonstrated that injection of anti-Nir2 antibodies into HeLa cells resulted in aberrant cytokinesis (Litvak *et al*., 2002) and that phosphorylation of Nir2 was required for the completion of cytokinesis in mitotic cells (Litvak *et al*., 2004), confirming that Nir2 plays a role in cytoskeletal regulation.

In another study by Litvak *et al*., Nir2 was found to be involved in Golgi secretion function through regulation of DAG levels in the Golgi apparatus (Litvak *et al*., 2005). Here, downregulation of Nir2 expression by RNAi resulted in reduced DAG levels in the Golgi membrane and lead to impaired Golgi secretion function, which was rescued by inhibition of the CDP-DAG pathway (Litvak *et al*., 2005).

Even more striking is Nir2 function in $PI(4,5)P_2$ homeostasis, which has been investigated by multiple groups (Chang *et al*., 2013; Chang and Liou, 2015; Kim *et al*., 2013, 2015). The ability of Nir2 to regulate $PI(4,5)P_2$ levels at the PM was first reported by Kim *et al*. (Kim *et al*., 2013). Kim *et al*. overexpressed a Myc-tagged Nir2 in HeLa cells and observed that Nir2 translocates to PM upon EGF stimulation of the serum-starved cells. The addition of PA to the cells lead to a similar effect, whereas inhibition of PA production by 1-butanol resulted in decreased Nir2 translocation to the PM, suggesting that Nir2 translocation is mediated via PLD activation by EGF signalling (Kim *et al*., 2013). The following investigation of truncated Nir2 mutants revealed that the C-terminal portion of the protein containing the LNS2 domain was responsible for the PA-dependent PM targeting of Nir2. This was evident from the fact that the Nir2 mutant lacking the C-terminal region failed to translocate to the PM upon EGF stimulation, whereas the mutant consisting of only the C-terminal region was detected at the PM after the addition of EGF. The ability of the C-terminal domain to bind PA with high specificity was demonstrated *in vitro* (Kim *et al*., 2013). Crucially, overexpression of Nir2 was found to cause an increase in $PI(4,5)P_2$ and $PI(3,4,5)P_3$ levels at the PM and depletion of Nir2 lead to a remarkable reduction in the levels of the two phospholipids. The regulation of $PI(4,5)P_2$ appeared to be dependent on both the lipid transfer function of Nir2 and its PA-dependent PM translocation, as both PITP- and C-terminal domain truncated Nir2 mutants failed to restore the $PI(4,5)P_2$ and $PI(3,4,5)P_3$ levels in Nir2-depleted cells (Kim *et al*., 2013). Furthermore, Kim *et al*. have generated a Nir2 mutant containing the D1128A mutation in the LNS2 domain, which has demonstrated reduced PA binding *in vitro* and reduced PM translocation *in vivo* (Kim *et al*., 2013). Although Kim *et al*. claimed that the mutation site was chosen based on the analysis of conserved motifs in HAD proteins, it is unclear whether the mutation caused a specific reduction in PA binding affinity of Nir2 LNS2 or whether the PA binding was affected due to a structural change in the domain caused by the mutation, as structural analysis of the mutant was not carried out.

Further studies by Chang *et al*. and Kim *at al.* have confirmed the role of Nir2 in positive regulation of $PI(4,5)P_2$ levels in the PM. Chang *et al*. demonstrated that Nir2 translocates to ER-PM junctions in response to histamine stimulation and replenishes $PI(4,5)P_2$ levels

during PLC-mediated signalling (Chang *et al*., 2013). Consistent with the previously described results by Kim *et al*., Chang *et al*. observed that the PITP domain is not required for PM translocation of Nir2. Interestingly, Chang *et al*. also showed that receptor stimulation enhanced both PM targeting of Nir2 and co-localisation of Nir2 with VAP-A and VAP-B. Consistently, cultured mammalian cells overexpressing Nir2 with a functional mutation in the FFAT motif displayed reduced $PI(4,5)P_2$ levels compared to the cells overexpressing wild-type Nir2, suggesting that VAP binding is important for Nir2 function in $PI(4,5)P_2$ homeostasis (Chang *et al*., 2013). In another study, Chang and Liou have shown that PA plays a central role in PM targeting of Nir2 (Chang and Liou, 2015). Addition of PLD and DAGK inhibitors to HeLa cells inhibited PM translocation of Nir2, whereas addition of PA alone was able to initiate the translocation (Chang and Liou, 2015). By studying the chimeras of Nir2 and Nir3, Chang and Liou were able to show that translocation of Nir2 and Nir3 to the PM was dependent on their C-terminal domains (Chang and Liou, 2015).

A more recent study by Kim *et al*. has provided further insight into the role of Nir2 in $PI(4,5)P_2$ regulation (Kim *et al*., 2015). Kim *et al*. were able to monitor the $PI(4,5)P_2$, PA, and DAG pools at the PM by using specific phospholipid binding domains coupled to two fluorescent probes that can act as a donor and acceptor in fluorescence resonance energy transfer (FRET). Consistent with previous results, Kim *et al*. observed that PM $PI(4,5)P_2$ and DAG levels were reduced in Nir2-depleted cells. Surprisingly, an increase in PA levels was observed, which was believed to be due to decreased consumption of PA in Nir2-deficient cells (Kim *et al*., 2015). As in the previous studies, Nir2 was found to translocate from the cytoplasm to the ER-PM junctions on receptor stimulation. By using a fluorescently-labelled PA binding motif from yeast sporulation protein Spo20 and a GFP-Nir2 fusion, Kim *et al*. demonstrated that Nir2 overexpression facilitates PA removal from PM in angiotensin II-stimulated cells, and that PA transported by Nir2 is utilised in the ER membrane for PI synthesis (Kim *et al*., 2015). Interestingly, the kinetics of Nir2-GFP membrane translocation upon receptor stimulation were faster than those of the Spo20 PA binding motif, which suggests that an interaction between Nir2 and a membrane lipid other than PA might also be involved in PM targeting of Nir2. A short putative DAG binding sequence preceding the LNS2 domain was identified by Kim *et al.,* the absence of which from Nir2 prevented PM recruitment (Kim *et al*., 2015). These data suggest that DAG might also be a regulator of Nir2 localisation along with PA. However, the putative DAG binding function of Nir2 C-terminal region was not directly demonstrated by Kim *et al*. and is yet to be confirmed by other groups.

Remarkably, overexpression of Nir2 was shown to enhance epithelial-to-mesenchymal transition in cultured breast cancer cells and injection of Nir2-overexpressing cells into mice was found to promote lung metastasis in mice metastasis models (Keinan *et al*., 2014). Furthermore, high Nir2 expression levels were associated with poor prognosis in breast cancer patients (Keinan *et al*., 2014), pointing to its potential as a target for cancer therapies.

#### 1.3.1.2 Nir3

Nir3 is much less studied than Nir2. Analogously to Nir2, Nir3 was identified as a PYK2 binding protein (Lev *et al*., 1999) and was shown to bind VAP-B through its FFAT domain (Amarilio *et al*., 2005). The predominant source of information about the cellular function of Nir3 is the study by Chang and Liou referenced in the Nir2 section above (Chang and Liou, 2015). In the study, Nir3 was found to be involved in $PI(4,5)P_2$ homeostasis together with Nir2. Interestingly, Nir3 appeared to be less efficient at $PI(4,5)P_2$ replenishment than Nir2 in response to receptor stimulation, which was believed to be due to the difference in the activity of the PITP domains of Nir3 and Nir2. However, in contrast to Nir2, Nir3 was found to play a clear role in the maintenance of basal $PI(4,5)P_2$ levels and was shown to be more important than Nir2 in $PI(4,5)P_2$ homeostasis during low levels of receptor activation (Chang and Liou, 2015). These results indicate that Nir2 and Nir3 have complementary functions in $PI(4,5)P_2$ cycle. Further studies are required to better understand the role of Nir3 in cell signalling and lipid metabolism.

#### 1.3.1.3 Nir1

Nir1 is an even more mysterious RdgB-like protein. It shares the PYK2 and VAP-B binding function with Nir2 and Nir3 (Amarilio *et al*., 2005; Lev *et al*., 1999), but lacks the PITP domain and thus is not believed to transport lipids. Knowledge about the cellular function of Nir1 is scarce. A mutation in Nir1 was documented to cause autosomal dominant cone dystrophy, a rare congenital vision disorder, indicating a role for Nir1 in mammalian phototransduction (Kohn *et al*., 2007). This putative function of Nir1 is similar to the function of its *Drosophila* ortholog RdgB, which is a known player in phototransduction (Harris and Stark, 1977; Vihtelic *et al*., 1993). Recently, Nir1 was also proposed to be a receptor of chemokine (C-C motif) ligand 18 (CCL18) (Chen *et al*., 2011; Zhang *et al*., 2013). However, evidence obtained by a different group has shed doubt on this notion (Krohn *et al*., 2013). Furthermore, like Nir2 and Nir3, Nir1 is not believed to be an integral membrane protein (Cockcroft and Raghu, 2016), and hence it is unlikely to act

as a cell-surface receptor of CCL18. Further research will be required to obtain a clearer picture of Nir1 function in the cell.

## 1.4 Na-FAR-1

Na-FAR-1 is one of the most well-studied FARs to date. As it is one of the six FARs identified in an important intestinal human parasite, American hookworm *Necator americanus* (Tang *et al*., 2014), Na-FAR-1 is of interest as a potential drug or vaccine target. Importantly, Na-FAR-1 contains a secretion signal and is highly expressed in the blood-feeding, adult stage of the parasite (Tang *et al*., 2014), which suggests that it may play a role in parasitism. Among FARs, the closest relatives of Na-FAR-1 are other parasite-specific FARs, such as Ov-FAR-1 and Bm-FAR-1 from human parasites, Ace-FAR-1 and Hp-FAR-1 from animal parasites and Gp-FAR-1 from a plant parasite (Rey-Burusco *et al*., 2015), some of which also have high expression levels in the parasitic stages of the worms (Jones *et al*., 2009; Popeijus *et al*., 2000; Ranjit *et al*., 2006). Similar to many other FARs, Na-FAR-1 has been demonstrated to bind retinoic acid and different fatty acids, displaying a preference for the long-chain fatty acids (Rey-Burusco *et al*., 2015). Interestingly, Na-FAR-1 has also been shown to bind PG and PE (Rey-Burusco *et al*., 2015), which is the first documented evidence of phospholipid binding by a FAR.

The structure of Na-FAR-1 has been determined by both solution-state NMR spectroscopy and X-ray crystallography in an apo (ligand-free) and holo (ligand-bound) forms, respectively (Gabrielsen *et al*., 2012; Rey-Burusco *et al*., 2015, 2014). It was found that Na-FAR-1 has an α-helical fold, which is largely similar to the previously determined fold of Ce-FAR-7 (Jordanova *et al*., 2009), the only other FAR whose three-dimensional structure is known. The fact that Na-FAR-1 and Ce-FAR-7 share a common fold despite their low sequence similarity suggests that the α-helical fold observed in the two proteins is a common feature of the FAR family.

Na-FAR-1 comprises eleven α-helices (α1-11) of varying lengths that form a wedge-like structure (Rey-Burusco *et al*., 2015) (Fig 1-6). Helices α1-3, α6 and α9-11, of which α6, α9 and α10 are the largest, form one face of the wedge, whereas α5, α7 and α8 form the opposite face. A cavity lined with predominantly hydrophobic residues is located at the centre of Na-FAR-1. Helix α4 is out of plane with the other helices, obstructing the cavity at one side of Na-FAR-1 (Rey-Burusco *et al*., 2015). The central cavity of Na-FAR-1 was shown to act as a ligand binding pocket, which is consistent with its hydrophobicity.

The structures of the apo and holo forms of Na-FAR-1 are largely similar. The most remarkable difference between the apo and the holo Na-FAR-1 is that the volume of the cavity of the holo protein is more than double the size of the apo protein cavity, indicating Na-FAR-1 expansion upon ligand binding (Rey-Burusco *et al*., 2015).

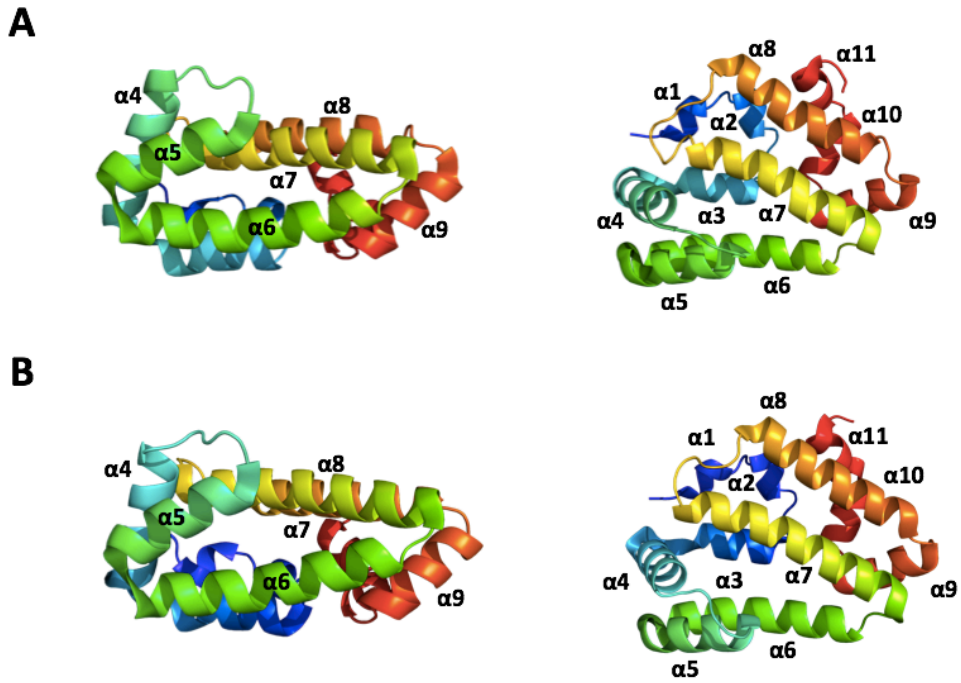



**Figure 1-6. The structure of the ligand-free (apo) and *E. coli* lipid bound (holo) Na-FAR-1.** **A.** Cartoon representation of the solution structure of apo Na-FAR-1 (PDB ID 4UET) (Rey-Burusco *et al*., 2015). **B.** Cartoon representation of the crystal structure of holo Na-FAR-1 (PDB ID 4XCP) (Rey-Burusco *et al*., 2015). No ligands are shown in the structure. The protein is coloured from blue (N-terminus) to red (C-terminus). The α-helices are numbered (α1-11).

In the study by Rey-Burusco *et al*., recombinant Na-FAR-1 expressed in *E. coli* was found to co-purify with endogenous *E. coli* lipids. In order to determine the structure of holo Na-FAR-1, Rey-Burusco *et al*. crystallised the *E. coli* lipid bound Na-FAR-1 without further purification steps. As the protein contained a heterogeneous mixture of lipids, identification of ligands in the structure was difficult, and only one ligand binding site was mapped with confidence (Rey-Burusco *et al*., 2015). The ligand was identified as palmitic acid, which appeared to interact with the side-chain amine group of K96 via its carboxylic acid head group (Rey-Burusco *et al*., 2015). In addition, the fatty acyl tail of the ligand contacted eleven hydrophobic residues in the internal cavity of Na-FAR-1 (Rey-Burusco *et al*., 2015). Importantly, other electron density peaks were observed in the central cavity of Na-FAR-1 which could not be filled with water molecules alone, suggesting the presence of unidentified ligands at alternative binding sites.

Furthermore, chemical shift perturbation analysis by NMR spectroscopy using oleic acid revealed that Na-FAR-1 forms four distinct complex with oleic acid during the ligand titration that were assumed to contain one, two, three and at least four ligand molecules.

From the chemical exchange regimes observed in the experiment, it was found that Na-FAR-1 binds three oleic acid molecules with a higher affinity than the fourth oleic acid molecule (Rey-Burusco *et al*., 2015).

## 1.5 Aims of this study

The two primary aims of this study were to: a) determine the phosphatidic acid binding mechanism and the structure of the LNS2 domain of Nir2 and b) determine the fatty acid and phospholipid binding mechanism of Na-FAR-1 and map the previously unidentified fatty acid and phospholipid binding sites of Na-FAR-1. The rationale behind the study is described below.

### 1.5.1 The structure and phosphatidic acid binding mechanism of the Nir2 LNS2 domain

The LNS2 domain of Nir2 has been proven to be a pivotal regulator of Nir2 PITP localisation, where the LNS2-PA interaction appears to play a central role. It is thus surprising that little to nothing is known about the PA binding mechanism of the LNS2 domain and its three-dimensional structure. The structural and functional analysis of the LNS2-PA interaction will shed light on the mechanism of PA binding by the LNS2, and will help to better understand the role of the LNS2 domain in Nir2 PM targeting.

Furthermore, determination of the atomic level details of the LNS2-PA binding could inform the development of an LNS2-PA interaction inhibitor tool compound, which could be used to obtain further insights into the cellular function of Nir2. Importantly, since the LNS2-PA interaction has been implicated in maintaining the signalling through MAP kinase and PI3K/Akt pathways (Kim *et al*., 2013), it is a possibility that PA recognition by Nir2 might play a role in cancer, where these key pathways are very often disrupted (Altomare and Testa, 2005; Martin, 2003; Sever and Brugge, 2015; Spiegel and Milstien, 2003). Hence, structure determination of the LNS2 domain and/or the LNS2-PA complex and characterisation of PA binding by the LNS2 domain might aid the assessment of the potential of the Nir2 LNS2 as a drug target in cancer.

To obtain the structural and functional information on LNS2-PA binding, this study aimed to: a) produce the LNS2 domain of Nir2 by recombinant expression, b) characterise the LNS2 domain binding to PA to determine what mechanisms are at play in the interaction and estimate the binding affinity, c) determine the three-dimensional structure of the LNS2 domain in the PA-free and/or PA-bound form and map the PA binding site, d) use the

structural information to guide the development of the LNS2-PA inhibitor by *in silico* drug screening.

### 1.5.2 The lipid binding mechanism of Na-FAR-1

Although the structure of Na-FAR-1 has been determined and it was demonstrated that Na-FAR-1 is able to form four distinct complexes with oleic acid, little else is known about its fatty acid binding mechanism. This includes the location of the putative long-chain fatty acid binding sites of Na-FAR-1, of which only one was identified. The recently discovered ability of Na-FAR-1 to interact with phospholipids also needs to be explored, and it is desirable to know whether Na-FAR-1 interacts with lysophospholipids, which are important extracellular mediators of inflammation. Furthermore, a better understanding of the mechanisms that underlie the relatively broad ligand specificity of Na-FAR-1 is also required. Answers to these questions may help to elucidate the biological function of Na-FAR-1 and enhance our understanding of its role in parasitism.

To obtain mechanistic details about the lipid binding by Na-FAR-1, this study aimed to: a) produce Na-FAR-1 by recombinant expression, b) confirm the phospholipid binding function of Na-FAR-1 *in vitro*, c) investigate whether Na-FAR-1 binds lysophospholipids, and d) determine the structure of Na-FAR-1 in complex with a fatty acid and in complex with a phospholipid, and compare the binding mechanisms.

The experimental approaches used to achieve the aims stated in this section and in Section 1.5.1 are described at the beginning of each relevant chapter of this thesis.

# 2. Materials and Methods

## 2.1 Materials

### 2.1.1 Protein accession numbers

Protein sequences of Nir1-3 and RdgB were obtained from GenBank. Accession numbers are given in brackets: Nir1 (AAK01446), Nir2 (AAK01444), Nir3 (AAK01445), RdgB (CAA69291). Na-FAR-1 sequence was obtained from the Nematode Transcriptome Database (NEMBASE4; sequence ID NAC00128).

### 2.1.2 Synthetic DNA

Synthetic DNA fragments encoding the C-terminal domain portions of Nir1-3 and RdgB were purchased from GenScript (China). The DNA fragments were codon optimized for expression in *E. coli*. Synthetic DNA fragment encoding Na-FAR-1 was purchased from GeneArt AG (Germany), as described previously (Gabrielsen *et al*., 2012).

### 2.1.3 Ligands

Dihexanoyl (di-6:0), dioleoyl (di-18:1 *cis*-9) and egg yolk phosphatidic acid sodium salt and phosphatidylcholine were purchased from Avanti Polar Lipids. Oleic acid sodium salt, lysophosphatidic acid sodium salt, *sn*-glycerol 3-phosphate bis(cyclohexylammonium) salt and hexanoic acid were purchased from Sigma-Aldrich.

### 2.1.4 Media recipes

#### 2.1.4.1 Lysogeny broth (LB)

For 1 L:

| | |
|---|---|
| Tryptone | 10 g |
| Yeast extract | 5 g |
| NaCl | 10 g |
| $dH_2O$ | to 1 L |

Autoclaved.

### 2.1.4.2 Autoinduction medium

For 1 L:

| | |
|---|---|
| Tryptone | 20 g |
| Yeast extract | 5 g |
| NaCl | 5 g |
| $dH_2O$ | to 1 L |

Autoclaved.

| | |
|---|---|
| 60% glycerol (v/v), f.s. | 10 ml |
| 10% glucose (w/v), f.s. | 5 ml |
| 8% lactose (w/v), f.s. | 25 ml |

Added to the autoclaved medium. f.s. = filter-sterilised.

### 2.1.4.3 Terrific broth (TB)

For 1 L:

| | |
|---|---|
| Tryptone | 12 g |
| Yeast extract | 24 g |
| 80% glycerol (v/v) | 5 ml |
| $dH_2O$ | to 900 ml |

Autoclaved.

Hundred milliliters of 0.17 M $KH_2PO_4$, 0.72 M $K_2HPO_4$ filter-sterilised solution was added to the autoclaved medium.

### 2.1.4.4 2x YT medium

For 1 L:

| | |
|---|---|
| Tryptone | 16 g |
| Yeast extract | 10 g |
| $dH_2O$ | to 1 L |

Autoclaved.

#### 2.1.4.5 M9 minimal medium

5x M9 salt solution, for 1 L:

| | |
|---|---|
| $Na_2HPO_4$ | 34 g |
| $KH_2PO_4$ | 15 g |
| NaCl | 2.5 g |
| $dH_2O$ | to 1 L |

Autoclaved.

M9 minimal medium, for 1 L:

| | |
|---|---|
| 5x M9 salt solution | 200 ml |
| 1 M $MgSO_4$[a] | 1 ml |
| 50 mM $CaCl_2$[a] | 1 ml |
| 20% (w/v) glucose[a,b] | 15 ml |
| $NH_4Cl$[c] | 1 g |
| 50 mg/ml thiamine[a] | 0.4 ml |
| $dH_2O$ | to 1 L |

a. Filter-sterilised.
b. $^{13}$C6-glucose for $^{13}$C labelling.
c. $^{15}NH_4Cl$ for $^{15}$N labelling.

## 2.2 Protein high-throughput expression screening at OPPF-UK

Protein expression screening in *E. coli* and insect cells was carried out at Oxford Protein Production Facility-UK (OPPF-UK), as per the protocols given on the OPPF-UK website (www.oppf.rc-harwell.ac.uk, accessed on 02/07/2017) (Bird, 2016; Nettleship, 2016). The process is briefly described below.

### 2.2.1 Molecular cloning

Nir2, Nir3 and RdgB LNS2 domain fragments were amplified by PCR using a set of primers designed for ligation independent cloning. Forward primers contained AAGTTCTGTTTCAGGGCCCG adaptor sequence added to their 5' end, reverse primers contained ATGGTCTAGAAAGCTTTA adaptor sequence added to their 5' end. The

PCR-amplified fragments were cloned into 12 vectors of the pOPIN vector suite (Bird, 2011) (pOPINE-3C-HALO7, pOPINEneo, pOPINF, pOPINHALO7, pOPINJ, pOPINM, pOPINTF, pOPINTRX) by using In-Fusion® Cloning Kit (Takara). Prior to InFusion® reaction, vectors were digested with KpnI and HindIII restriction endonucleases by the OPPF-UK staff. After cloning, the plasmids were transformed into OmniMaxII competent cells (Invitrogen) by heat-shock. The presence of the insert was validated by PCR and, in some cases, by Sanger sequencing (Source Bioscience, UK). For protein expression, the plasmids were transformed into Lemo21 (DE3) competent cells or Rosetta™ 2 (DE3) pLacI competent cells (NEB) or used to construct a recombinant baculovirus by using a bacmid produced by Ian Jones (University of Reading) for protein expression in insect Sf9 cells (Thermo Fisher Scientific).

### 2.2.2 Expression in *E. coli*

Expression screening was carried out in a 96-well plate format. Individual colonies were picked from Lemo21 (DE3) and Rosetta™ 2 (DE3) pLacI agar plates and each was used to inoculate 0.7 ml of PowerBroth™ medium (Molecular Dimensions). The cultures were incubated overnight at 37 °C. The next day, 150 µl of Lemo21 (DE3) and 250 µl of Rosetta™ 2 (DE3) pLacI overnight culture were used to inoculate 3 ml aliquots of PowerBroth™ for IPTG-induced expression and 3 ml aliquots of OvernightExpress™ medium (Novagen) for expression with autoinduction.

For IPTG induction, cultures in PowerBroth™ were shaken at 220-225 rpm, 37 °C for ~ 4 hours, and cooled down to 20 °C whilst shaking at 220-225 rpm for 20 min. IPTG was added to the final concentration of 1 mM, and the cultures were incubated for ~ 18 hours at 20 °C, shaking at 220-225 rpm.

For autoinduction, cultures in the OvernightExpress™ medium were shaken at 220-225 rpm, 37 °C for ~ 4 hours. The temperature of the shaker was then reduced to 25 °C and the cultures were shaken at 220-225 rpm for 24 hours.

The cells were harvested by centrifugation at 6000 x g for 15 min.

### 2.2.3 Purification from *E. coli*

Cell pellets were briefly frozen at -80 °C, thawed and resuspended in lysis buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% Tween 20) supplemented

with 1 mg/ml lysozyme and 3 units of Benzonase® nuclease (EMD Millipore). Cells were lysed by shaking on an orbital microplate shaker (30 min, ~ 1000 rpm) and lysates were centrifuged at 6000 x g for 30 min. Cleared lysates were used for $Ni^{2+}$-affinity purification, which was carried out on QIAGEN BioRobot 8000. Lysates were mixed with 20 µl of QIAGEN Ni-NTA magnetic beads and shaken on an orbital microplate shaker for 30 min. Beads were separated from the supernatant using a QIAGEN magnet and the supernatant was discarded. Beads were washed with the wash buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% Tween 20), and protein elution was carried out by the addition of 50 µl elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole, 0.05% Tween 20).

### 2.2.4 Small-scale expression testing in insect cells

Expression screening was carried out in a 24-well plate format. Sf9 cells were cultured in Sf-900 II serum-free medium (Thermo Fisher Scientific). For transfection with the bacmid and the plasmid encoding the protein constructs of interest, cells were grown to ~ 5 x $10^5$ cells/ml, and 0.5 ml of cells was added to each well. A mixture of bacmid (250 ng, 2.5 µl), recombinant plasmid (100 – 500 ng), transfection reagent FuGENE® HD (1.5 µl, Promega) and Sf-900 II medium (50 µl) was prepared and added to the cells. The plate was slowly swirled to allow the transfection mixture to disperse in the well. Cells were incubated for 6 days at 27 °C. After 6 days, the supernatant was removed from the cells and used as the P0 viral stock. Viral stocks were stored in the dark at 4 °C.

Virus amplification, expression testing and virus scale-up was carried out by Mr Valtteri Järvinen (OPPF-UK).

For production of P1 viral stock, 5 µl of P0 stock was added to 0.5 ml Sf9 cells (1.0 x $10^6$ cells/ml) in each well. The cells were incubated for 6 days at 27 °C to allow viral amplification. The supernatant was then harvested and used as P1 stock.

For expression testing, 3 or 30 µl of P1 stock was added to 3 ml of Sf9 cells (1.0 x $10^6$ cells/ml) in each well. The cells were incubated at 27 °C for 3 days, shaking at 250 rpm. After 3 days, 1 ml of cells was transferred from each well to a well in a 96-well plate. The plate was centrifuged for 15 min at 6000 x g to pellet the cells. Supernatant was discarded, and the cells were frozen at -80 °C. Protein purification was carried out as described above for *E. coli*.

### 2.2.5 Baculovirus scale-up and large-scale expression testing in insect cells

P2 viral stock used for large-scale expression screening was produced as follows. Fifty-milliliter suspension culture of Sf9 cells (1.0 x $10^6$ cells/ml) was infected with 400 µl of P1 virus. The culture was incubated at 27 °C for 6-7 days, shaking at 250 rpm, and the cells were then pelleted for 10 min at 1000 x g. The supernatant was harvested and used as P2 virus stock.

Large-scale protein expression was tested by infecting 2.5 L of Sf9 cells (1 x $10^6$ cells/ml) in a Thomson flask with either 2.5 ml (1:1000 dilution) or 25 ml (1:100 dilution) of P2 viral stock. The infected cultures were incubated at 27 °C for 3 days, shaking at 250 rpm. Samples (2 ml) were taken and used for $Ni^{2+}$-affinity purification to assess the protein yield as described above. The remaining cultures were centrifuged at 6000 x g, frozen at -80 °C and used for large-scale protein purification.

## 2.3 Molecular biology, protein expression and purification

### 2.3.1 Molecular cloning into pNIC28-Bsa4 vector

For cloning into pNIC28-Bsa4 vector, DNA fragments were amplified by PCR with a set of primers containing adaptor sequences for ligation independent cloning. Forward primers contained TACTTCCAATCCATG adaptor sequence added to their 5' end, reverse primers contained TATCCACCTTTACTG adaptor sequence added to their 5' end. Cloning was carried out using InFusion® Cloning Kit (Takara), as per manufacturer's protocol. DNA sequence was confirmed by Sanger sequencing (service was provided by Source Bioscience, UK).

### 2.3.2 Plasmid extraction

Plasmids for cloning and protein expression were purified from 10 ml LB culture of *E. coli*. The culture was inoculated with a single bacterial colony and grown at 37 °C overnight, shaking at 225 rpm. Cells were harvested by centrifugation at 3000 x g. Plasmid purification was carried out using Wizard® Plus SV Miniprep DNA Purification System (Promega), as per manufacturer's protocol. DNA was eluted by addition of 20-30 µl of nuclease free water supplied with the kit. Plasmid solutions were stored at -20 °C until required.

### 2.3.3 Large-scale protein expression in *E. coli*

#### 2.3.3.1 Production of proteins with natural isotope abundance

Tuner (DE3) pLysS cells (Novagen) or T7 Express lysY/Iq cells (New England Biolabs) were used for large-scale protein expression. For expression, inoculation cultures (10 ml) were prepared in LB the day before and grown at 37 °C, 225 rpm overnight. The cultures were used to inoculate the expression media (typically, LB or TB), with 10 ml inoculation culture used per 0.5 L media. Expression cultures were grown until $OD_{600}$ reached ~ 0.5-0.7 in LB or ~ 1.0-1.4 in TB. All media were supplemented with an appropriate antibiotic, depending on the plasmid and the cell strain used. The antibiotics were used at the following concentrations: 100 μg/ml ampicillin, 30 μg/ml kanamycin or 25 μg/ml chloramphenicol. Protein expression was induced by addition of 0.4 mM IPTG, followed by an overnight incubation at 16 °C or 20 °C for expression of Nir2 and RdgB fragments, or by a 3-hour incubation at 37 °C for expression of Na-FAR-1, shaking at 225 rpm. Cells were harvested by centrifugation at 9000 rpm in Beckman J2-21 centrifuge using JA-21 rotor. Cells were lysed by sonication (15 sec on/off pulses, 50% intensity) on a MSE Soniprep 150 (Sanyo) sonicator or by using a French press cell disruptor and purified by $Ni^{2+}$-affinity chromatography (Ni-Superflow resin, Generon, UK).

#### 2.3.3.2 Production of isotope-labelled proteins for NMR spectroscopy

For production of isotope-labelled Na-FAR-1 for NMR experiments, an alternative growth strategy was used. Expression cultures were initially grown in 2x YT medium until $OD_{600}$ reached ~ 1.5-2.0. After this, the cultures were centrifuged as described above, resuspended in M9 minimal medium (without nitrogen and carbon sources) and centrifuged again to remove any nutrients retained in the pellet after growth in 2x YT. The medium was then decanted and replaced with fresh M9 medium supplemented with 1 g/L $^{15}NH_4Cl$ and 3 g/L glucose or $^{13}C_6$-glucose (Sigma-Aldrich). The cultures were incubated for 1 hour at 20 °C, and protein expression was induced by addition of 0.4 mM IPTG, followed by overnight incubation at 20 °C. The cells were harvested the next day and lysed as described above.

### 2.3.4 Large-scale protein expression in insect cells (adherent culture)

Protein large-scale expression was carried out using P2 stock of the recombinant baculovirus produced by Mr Valtteri Järvinen at OPPF-UK. Sf9 cells (Thermo Fisher

Scientific) were grown at 27-28 °C in TC-100 Insect medium (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fisher Scientific), which was heat-inactivated by 30 min incubation at 56 °C in a water bath. Corning® 150 $cm^2$ cell culture flasks with a plug seal cap were used for cell growth. Cells were infected at ~ 90% confluency with 1:1000 dilution of the P2 stock, which was determined to be an optimal ratio of the virus to the culture in the small-scale expression trials carried out at OPPF-UK. After infection, cells were incubated for 72 hours at 27-28 °C, harvested by centrifugation at 1000 x g, and stored at -80 °C until required. Cells were lysed by a ~ 15 min incubation with the lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 1 % Tween 20), gently rocking. The lysate was used for protein purification.

### 2.3.5 $Ni^{2+}$-affinity chromatography

$Ni^{2+}$-affinity chromatography was performed using Ni-Superflow resin (Generon, UK). Typically, ~ 0.5-1 ml of resin was used for protein purification from 1 L of LB culture or 0.5 L of TB culture. Purification was carried out essentially as per manufacturer's protocol. Typical chromatography buffers contained 50 mM Tris-HCl or NaPi pH 7.5, 300 mM NaCl, 1 mM TCEP, with 10 mM, 30 mM and 300 mM imidazole used in the binding, wash and elution buffer, respectively. After protein binding, the column was washed with 10-20 CV of the wash buffer and protein was eluted with 3 CV of the elution buffer. All fractions were collected into 20 ml universal tubes and stored on ice or at 4 °C until required.

### 2.3.6 Size-exclusion chromatography (SEC)

Size-exclusion chromatography was carried out on ÄKTA explorer chromatography system (GE Healthcare) equipped with a Superdex 75 10/300 GL column (GE Healthcare) or Superdex 200 Increase 10/300 GL column (GE Healthcare) at 8 °C. Typical buffer solution contained 20 mM sodium phosphate or Tris-HCl pH 7.5, 150 mM NaCl, 1 mM TCEP and 0.01% $NaN_3$. Buffer components were varied depending on the nature and downstream applications of the protein sample. Normally, 0.5 ml of sample was loaded onto the column at a concentration of ~ 1-15 mg/ml. Protein elution was carried out with 2 CV of buffer, with the typical flow rate set to 0.4 ml/min.

### 2.3.7 SEC coupled with multi-angle light scattering (SEC-MALS)

Analysis by SEC-multi-angle light scattering (SEC-MALS) was carried out on an Alliance HPLC system (Waters) equipped with a BEH SEC 200 Å, 3.5 μm column (Waters). The

HPLC system was connected to Viscotek SEC-MALS 20 multi-angle scattering detector (Malvern) and Viscotek VE 3589 refractive index detector (Malvern). The protein sample was prepared in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM TCEP buffer, which was also used as a running buffer. The column was calibrated by using human serum albumin standards prior to each run.

### 2.3.8 SDS-PAGE

SDS-PAGE was carried out using RunBlue 4-20% SDS Precast Gels (Expedeon). Prior to electrophoresis, protein samples were mixed with NuPage® LDS Sample Buffer (Novex) with β-mercaptoethanol added to a final 5% (v/v) concentration and heated at 85 °C for 5 min. Protein bands were resolved in RunBlue running buffer (60 mM Tris, 30 mM MOPS, 0.1% SDS). Electrophoresis was performed at 180-200 V (constant voltage) for ~ 45 min. The gels were stained with InstantBlue™ (Expedeon) protein stain and images were acquired using Kodak Image Station 440CF.

### 2.3.9 Protein digestion with HRV 3C protease

Proteolytic digests were carried out using $His_6$-tagged HRV 3C produced in our home lab. The activity of the protease was confirmed by comparison with a commercial GST-tagged HRV 3C PreScission Protease (GE Healthcare). The digests were performed in a 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 10 mM EDTA buffer at 4 °C for 24-48 hours. Approximately 1 μg of the protease was used for the cleavage of 10 μg of protein.

### 2.3.10 Western blotting

Protein samples were resolved by SDS-PAGE and transferred from polyacrylamide gels onto nitrocellulose membrane (Whatman Protran BA-83, pore size 0.2 μm) in 25 mM Tris, 1.9 mM glycine, 20% methanol at 400 mA (constant current) for 45 min. The membrane was blocked by a 1 hour incubation with 5% fat-free milk solution in 20 mM Tris-HCl pH 7.6, 140 mM NaCl, 0.1% Tween-20 (TBST) at room temperature on a rocking platform. Membrane was washed three times with 5 ml TBST and incubated with a solution of a primary anti-His-tag mouse antibody (Abcam) in TBST for 2 hours at room temperature, gently rocking. After that, the membrane was washed again three times with TBST and incubated with a secondary anti-mouse ALP-conjugated antibody (Promega) in TBST for 1 hour at room temperature, gently shaking. After three final washes, the membrane was incubated with BCIP/NBT solution (Sigma-Aldrich) for 10 minutes at

room temperature to develop the protein bands.

### 2.3.11 Reverse-phase high pressure liquid chromatography (HPLC)

Reverse-phase HPLC (RP-HPLC) was carried out using a Supelco Discovery® BIO wide-pore C18 column (Sigma-Aldrich) on an ÄKTA purifier (GE Healthcare) chromatography system. The column was pre-equilibrated with 10% acetonitrile solution in $dH_2O$ (10% ACN) containing 0.01% trifluoroacetic acid. The sample was prepared in 10% ACN before loading into the injection loop. The elution program consisted of four steps: 1) washing step with 10% ACN (0.5 CV), 2) increase in ACN concentration to 50% (0.5 CV), 3) increase in ACN concentration to 60% (2 CV) and 4) increase in ACN concentration to 100% (1 CV). The flow rate was set to 4 ml/min. The fractions were collected during each elution step and stored at 4 °C, covered.

### 2.3.12 Protein lyophilisation

Protein lyophilisation was carried out on a Heto PowerDry LL1500 freeze dryer for ~ 16 hours with the cold trap cooled to -110 °C. Prior to lyophilisation, protein samples were frozen in liquid nitrogen. The samples were monitored during first 10 minutes of lyophilisation. If any thawing of the sample was observed, it was refrozen. Dry protein was stored at -20 °C, and redissolved in a buffer of interest.

### 2.3.13 Lipid extraction

Lipid extraction for TLC was carried out essentially as described previously (Obal *et al*., 2012). Three milligrams of protein were added to 5 ml chloroform and the mixture was vigorously shaken on ice for 10 min. The extract was washed with 200 µl of 3 M NaCl solution and centrifuged at 5000 x g for 15 minutes. The aqueous phase was removed by pipetting and chloroform was evaporated under a stream of dry $N_2$. The dry lipids were redissolved in 200 µl chloroform and stored at -20 °C under a layer of $N_2$ until required.

### 2.3.14 Thin-layer chromatography

Thin-layer chromatography was carried out on TLC Silica gel 60 plates (EMD Millipore). The plates were pre-washed with methanol/chloroform 1:2 mixture by volume and activated at 100 °C for 30 minutes. Approximately 5 µl lipid extract was spotted onto the plate and the lipids were allowed to dry under a stream of $N_2$. The lipids were resolved in hexane/diethyl ether/acetic acid 80:20:1 mixture by volume, sprayed with 8% (w/v) $CuSO_4$

solution in 10% (v/v) phosphoric acid and charred at 150 °C for 10 minutes until black spots appeared.

## 2.4 Characterisation of protein-ligand binding

### 2.4.1 Lipid overlay assay

Phospholipids in chloroform were spotted onto nitrocellulose membrane (Whatman Protran BA-83, pore size 0.2 μm), ~ 10 μg of lipid per dot. Lipids were dried under a stream of $N_2$ and the membrane was blocked with 1% fatty acid free BSA in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20 (TBST) buffer for 1 hour at room temperature. The membrane was incubated with ~ 20-30 nM protein overnight at 4 °C and washed three times with TBST. In the next step, the membrane was incubated with anti-His tag primary mouse antibody (Abgent) in TBST (1:1000 dilution) for 2 hours at room temperature and again washed three times with TBST. Secondary anti-mouse antibody conjugated to alkaline phosphatase (Promega) in TBST was added to the membrane (1:2500 dilution). The membrane was incubated for 1 hour at room temperature, and was washed three times with TBST for the last time. The signal was obtained by incubating the membrane with NBT-BCIP solution (Sigma-Aldrich) for 10 minutes at room temperature.

### 2.4.2 Liposome co-sedimentation assay (LSA)

#### 2.4.2.1 LSA using MLVs

The protocol was essentially as described previously (Kim *et al*., 2013). Lipid films containing either egg yolk PC or a mixture of egg yolk PC and PA in 2:1 (w/w) ratio were prepared in glass vials by mixing the phospholipid solutions in chloroform and drying them under nitrogen gas stream for 15 minutes. Tris buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl) was added to the lipid films to a total lipid concentration of 1 mg/ml. The films were allowed to hydrate at room temperature for 20 min, after which the vials were vigorously vortexed for 2 min to produce the multilamellar vesicle (MLV) suspension. The TF fusion of Nir2 LNS2 (5 μg in 50 μl Tris buffer) was mixed with 16.5 μl of MLV suspension. The mixture was incubated at room temperature for 20 minutes and centrifuged at 16000 x g, 4 °C for 30 minutes. The supernatant was separated from the pellet, and the pellet was resuspended in 21.5 μl of Tris buffer. Both fractions were used for analysis by SDS-PAGE.

#### 2.4.2.2 Semi-quantitative LSA using LUVs

Ten millimolar stock of LUVs in Tris buffer was prepared by extrusion as described in Section 2.4.4 below. Protein was prepared in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM TCEP buffer, and centrifuged at 150,000 x g, 16 °C for 10 minutes to pellet any aggregated material. Protein solution and LUVs were mixed to prepare a series of samples with varying concentration of LUVs for analysis of PA binding affinity. Protein concentration was 2 μM in all samples. The samples were incubated for 20 minutes at room temperature and centrifuged at 150,000 x g, 16 °C for 30 minutes.

### 2.4.3 Time-resolved fluorescence lifetime measurements

Fluorescence lifetime measurements were carried out on a PicoQuant FluoTime 300 fluorescence spectrometer by using time-correlated single photon counting (TCSPC) technology. The excitation source was a super continuum laser (WL-SC-400-4-PP, Fianium) with tunable wavelength filter and multimode fibre (Superchrome-Vis-FDS-MM, Fianium). The ligands and the protein were prepared in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT. The DHPA-NBD concentration was 3 μM in the fluorescence lifetime measurement in the absence of the protein, and 2 μM in the measurements in the presence of the protein (8 μM). The concentrations of the ligand and the protein were determined by UV-Vis spectrophotometry using the reported NBD molar extinction coefficient of 25,000 $M^{-1}cm^{-1}$ at 480 nm (Ladokhin *et al*., 2002) and the TF Nir2 LNS2 molar extinction coefficient of 37,550 $M^{-1}$ $cm^{-1}$ at 280 nm calculated from the protein amino acid sequence in CLC Genomics Workbench (QIAGEN). The decays were measured at 534 nm and the fluorophore was excited at 460 nm. Fitting of fluorescence decay data was carried out with FluoFit software (PiqoQuant) using the exponential model with reconvolution. The quality of the fits was assessed by $\chi^2$ values and the randomness of residuals.

### 2.4.4 Preparation of large unilamellar vesicles (LUVs) by extrusion

Dioleoyl PA and PC dissolved in chloroform were used for LUV preparation. Lipids were mixed in glass vials in 70:30 PC:PA ratio for use in the liposome co-sedimentation assay and CD spectroscopy. Lipids were dried under a stream of dry $N_2$ for 15 minutes to obtain lipid films. The lipid films were hydrated by the addition of buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 1 mM DTT or 1 mM TCEP to yield a total lipid concentration of 10 mM. The lipid films were incubated in the buffer for 1 hour. After the incubation, the liposome suspension was vortexed to dissociate the liposomes from the

glass and subjected to 4 freeze-thaw cycles using a dry ice/ethanol bath and a 37 °C water bath. The liposomes were downsized by extrusion through a nitrocellulose membrane (0.1 μm pore size, Avanti Polar Lipids) with a mini-extruder (Avanti Polar Lipids) to yield a transparent LUV suspension.

### 2.4.5 Steady-state fluorescence spectroscopy

Steady-state fluorescence spectroscopy experiments were carried out using Perkin Elmer LS 50B fluorimeter.

Stocks of 11-[5-(Dimethylamino)-1-naphthalenesulfonylamino]undecanoic acid (DAUDA, 1.4 mM), dioleoyl PA (DOPA, 10 mM) and oleoyl lysoPA (OLPA, 10 mM) were prepared in methanol for use in the assay. DAUDA stock concentration was calculated using the molar extinction coefficient of 4800 $M^{-1}$ $cm^{-1}$ at 335 nm in methanol (Haugland and Spence, 1996). DAUDA was diluted with 20 mM sodium phosphate pH 7.2 buffer in a quartz-glass cuvette to a concentration of ~ 1.5 μM. Na-FAR-1 (0.5 μM) was added to DAUDA, followed by DOPA or OLPA titration. Fluorescence emission intensity ($\lambda_{ex}$ = 345 nm) was recorded after each step to monitor DAUDA displacement. Na-FAR-1 concentration was calculated using molar extinction coefficient of 7680 $M^{-1}$ $cm^{-1}$ at 280 nm estimated from the protein primary sequence.

### 2.4.6 NMR spectroscopy

All NMR experiments were carried out on a Bruker 600 MHz Avance IIIHD NMR spectrometer fitted with a TCI cryoprobe. Protein and ligand samples were prepared in 5 mm thin-walled NMR sample tubes (Wilmad).

#### 2.4.6.1 Ligand-observed NMR experiments

For the analysis of LNS2 ligand binding, CPMG (Carr and Purcell, 1954; Meiboom and Gill, 1958), waterLOGSY (Dalvit *et al*., 2001) and standard $^{1}H$ spectra with water suppression were recorded out at 288 K. Protein samples (~ 10 μM) were prepared in 20 mM sodium phosphate pH 7.5, 5% $D_2O$. Dihexanoyl PA in d6-DMSO or $dH_2O$ were titrated into the protein samples in a series of steps. Dihexanoyl PC, hexanoic acid, oleic acid sodium salt and *sn*-glycerol 3-phosphate bis(cyclohexylammonium) salt were used at a concentration of 100 μM.

#### 2.4.6.2 Chemical shift perturbation analysis

Chemical shift perturbation caused by dioleoyl PA (DOPA) binding to $^{15}$N-labelled Na-FAR-1 was analysed in a series of $^{1}$H-$^{15}$N heternonuclear single quantum coherence (HSQC) and SOFAST heternonuclear multiple quantum coherence (HMQC) (Schanda and Brutscher, 2005) experiments. For the analysis, DOPA sodium salt was prepared in methanol-d4 at ~ 10 mM concentration. DOPA was added to the protein sample in a series of titration steps until the apparent saturation of binding was reached.

#### 2.4.6.3 Double- and triple-resonance experiments for structure determination

In order to prepare the complex of Na-FAR-1 with DOPA, DOPA in chloroform was added to a clean NMR tube, the solvent was evaporated in the stream of dry $N_2$ and the dry lipid film was hydrated by the addition of Na-FAR-1 solution (0.5 mM) in 20 mM sodium phosphate pH 7.2, 0.01% $NaN_3$ buffer to the final ligand concentration of 3.5 mM (7-fold excess of ligand to the protein). The phospholipids were mixed into the protein solution by gentle pipetting and the mixture was incubated at 35 °C for 2-4 hours until the sample became transparent and the phospholipid binding to the protein had equilibrated.

The backbone resonances of Na-FAR-1 complex with DOPA were assigned with the help of 2D $^{1}$H, $^{15}$N HSQC and 3D HNCA, HNCO, HNcoCA, HBHANH, HBHAcoNH, CBCAcoNH, CBCANH (Grzesiek and Bax, 1993, 1992a; Kay *et al*., 1993; Muhandiram and Kay, 1994) and $^{15}$N NOESY-HSQC (Marion *et al*., 1989a, 1989b; Zuiderweg and Fesik, 1989) experiments. Side-chain resonance assignment was carried out using 2D $^{1}$H, $^{13}$C HSQC, and 3D HCccoNH-TOCSY (Montelione *et al*., 1992), HCcH-, hCCH-TOCSY (Bax *et al*., 1990) and $^{13}$C NOESY-HSQC experiments. The acquisition parameters of the protein NMR experiments are summarised in Table 2-1. Resonance assignments were performed using CcpNmr Analysis software (Vranken *et al*., 2005). Distance restraints were obtained from $^{15}$N and $^{13}$C NOESY-HSQC spectra. Dihedral restraints were obtained from the chemical shifts of assigned backbone resonances using DANGLE (Cheung *et al*., 2010).

| | $F_1$ | | | | $F_2$ | | | | $F_3$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Experiment** | **N[a]** | **AT[b] (s)** | **SW[c] (ppm)** | **O[d] (ppm)** | **N** | **AT (s)** | **SW (ppm)** | **O (ppm)** | **N** | **AT (s)** | **SW (ppm)** | **O (ppm)** | **NUS[e] (%)** | **Solvent** | **Pulse seq.[f]** |
| $^{1}H$, $^{15}N$ HSQC | $^{15}N$ | 0.1065 | 21.00 | 116.0 | $^{1}H$ | 0.1003 | 16.03 | 4.702 | | | | | | $H_2O/D_2O$ | fhsqcf3gpph |
| $^{1}H$, $^{15}N$ HMQC | $^{15}N$ | 0.0702 | 30.00 | 115.0 | $^{1}H$ | 0.1024 | 16.67 | 4.698 | | | | | | $H_2O$ | sfhmqcf3gpph |
| hbCBcgcdHD | $^{13}C$ | 0.0106 | 40.00 | 39.00 | $^{1}H$ | 0.1065 | 16.03 | 4.702 | | | | | | $D_2O$ | hbcbcgcdhdgp |
| hbCBcgcdceHE | $^{13}C$ | 0.0038 | 40.00 | 39.00 | $^{1}H$ | 0.1065 | 16.03 | 4.702 | | | | | | $D_2O$ | hbcbcgcdcehegp |
| aro $^{1}H$, $^{13}C$ TROSY | $^{13}C$ | 0.0106 | 40.16 | 120.0 | $^{1}H$ | 0.1217 | 14.03 | 4.702 | | | | | | $D_2O$ | trosyargpphwg |
| $^{13}C$-filtered NOESY | $^{1}H$ | 0.0640 | 4.702 | 4.702 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | | | | | | $D_2O$ | noesygpphwgx1 |
| HNCA | $^{13}C$ | 0.0067 | 75.00 | 54.00 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 20 | $H_2O$ | hncagpwg3d |
| CBCANH | $^{13}C$ | 0.0067 | 75.00 | 39.40 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 20 | $H_2O$ | hncacbgpwg3d |
| CBCAcoNH | $^{13}C$ | 0.0067 | 75.00 | 43.00 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 20 | $H_2O$ | cbcaconhgpwg3d |
| HNCO | $^{13}C$ | 0.0193 | 22.00 | 176.0 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | hncogpwg3d |
| HNcaCO | $^{13}C$ | 0.0193 | 22.00 | 176.0 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | hncacogpwg3d |
| HBHANH | $^{1}H$ | 0.0133 | 8.00 | 4.702 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | hbhanhgpwg3d |
| HBHAcoNH | $^{1}H$ | 0.0133 | 8.00 | 4.702 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | hbhaconhgpwg3d |
| HCcH-TOCSY | $^{1}H$ | 0.0260 | 8.33 | 4.702 | $^{13}C$ | 0.0080 | 33.00 | 29.50 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 40 | $H_2O$ | hcchdigp3d |
| hCCH-TOCSY | $^{13}C$ | 0.0092 | 75.00 | 29.50 | $^{13}C$ | 0.0080 | 33.00 | 29.50 | $^{1}H$ | 0.1065 | 16.02 | 4.702 | | $D_2O$ | hcchdigp3d2 |
| hCCcoNH-TOCSY | $^{13}C$ | 0.0113 | 75.00 | 39.00 | $^{15}N$ | 0.0235 | 21.00 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 33 | $H_2O$ | hccconhgpwg3d3 |
| $^{13}C$ NOESY-HSQC | $^{1}H$ | 0.0267 | 12.50 | 4.702 | $^{13}C$ | 0.0080 | 33.00 | 62.50 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | noesyhsqcetgp3d |
| $^{15}N$ NOESY-HSQC | $^{1}H$ | 0.0267 | 12.50 | 4.702 | $^{15}N$ | 0.0256 | 20.50 | 116.0 | $^{1}H$ | 0.1024 | 16.67 | 4.702 | 25 | $H_2O$ | noesyhsqcf3gp193d |

a. Nucleus.
b. Acquisition time. In NOESY-HSQC, acquisition in $F_2$ was carried in the Echo-Antiecho mode. In all other experiments, acquisition in the indirect dimensions was carried in the States-TPPI mode.
c. Spectral width.
d. Transmitter frequency offset.
e. Amount of non-uniform sampling.
f. The noesyhsqcf3gp193d pulse sequence was modified to improve water suppression. In all other experiments, standard Bruker pulse sequences were used.

**Table 2-1. Acquisition parameters of the protein NMR experiments carried out in this study.**

## 2.4.7 Protein crystallisation, data collection and processing

### 2.4.7.1 Na-FAR-1 complex with oleic acid

RP-HPLC purified protein was prepared in pure $dH_2O$ at 5 mg/ml or 10 mg/ml concentration. Oleic acid sodium salt was added in a 4-fold molar excess to the protein for co-crystallisation. Crystallisation was carried out using JCSG+ and PACT Premier screens (Molecular Dimensions) by sitting drop vapour diffusion method. For crystallisation, 0.5 µl protein was mixed with reservoir solution in 1:1 ratio by Cartesian Honeybee robot (Genomic Solutions) at room temperature. The plates were incubated at 16 °C.

Cubic crystals were observed after 24 hours in JCSG+ condition C6 (0.1 M phosphate-citrate pH 4.2, 40% PEG-300) and were grown further for 10 days. Largest crystal (approximate dimensions = 150 x 150 x 150 µm) was frozen without any further cryoprotection in liquid $N_2$ and used for remote data collection at Diamond Light Source (Research Complex at Harwell, Oxfordshire, UK) beamline I03. Two data sets were merged and processed by xia2 (Winter, 2010) using the flag -3dii for XDS (Kabsch, 2010) pipeline to produce the working data set.

### 2.4.7.2 Na-FAR-1 complex with DOPA

RP-HPLC purified protein was prepared in $dH_2O$ at 5 mg/ml or 10 mg/ml concentration. Dioleoyl PA (DOPA) sodium salt was added to the protein in 7-fold molar excess and the mixture was incubated at 35 °C for ~ 2 hours to allow ligand binding to the protein. The sample was centrifuged at 11000 x g table-top microcentrifuge to pellet the unbound DOPA. The supernatant was transferred to a clean Eppendorf tube and used for crystallisation screening. Crystallisation screening was carried out using JCSG+ and PACT Premier screens (Molecular Dimensions). The screens were prepared as described for Na-FAR-1 complex with oleic acid.

No protein crystals were obtained.

### 2.4.7.3 The TF and MBP fusions of Nir2 LNS2 domain

TF Nir2 LNS2 samples for crystallisation screening were prepared in 20 mM Tris-HCl pH 7.5, 0.5 mM TCEP at 4 mg/ml and 10 mg/ml concentrations. MBP RdgB LNS2 samples were prepared in 10 mM Tris-HCl pH 7.5, 5 mM maltose. Crystallisation screening was carried out using JCSG+ and PACT Premier screens (Molecular

Dimensions). The crystallisation screening trays were prepared as described for Na-FAR-1 complex with oleic acid.

No protein crystals were obtained.

### 2.4.8 CD spectroscopy

CD spectroscopy was carried out on a Jasco J-810 spectrapolarimeter. Quartz cuvettes with 0.02 cm path-length were used for measurements. TF Nir2 LNS2-S samples with or without LUVs (1 mM) were prepared in 20 mM Tris-HCl pH 7.5, 150 mM NaCl. The protein concentration was 0.71 mg/ml. Nir1 LNS2-S (0.12 mg/ml) was prepared in 50 mM Tris-HCl pH 7.5, 300 mM NaCl. Spectra recorded from the buffer-only or liposome-only samples were subtracted from the spectra recorded from the protein-containing samples to correct for background light scattering. Measured ellipticity values were converted into the molar ellipticity values by correcting for protein concentration in the sample. TF Nir2 LNS2-S spectra were analysed using K2d algorithm developed by Andrade *et al.* (Andrade *et al*., 1993). The spectrum of Nir1 LNS2-S was analysed using the CDSSTR algorithm (Sreerama and Woody, 2000), a modification of VARSLC algorithm developed by Johnson (Compton and Johnson, 1986; Manavalan and Johnson, 1987), using reference data set 7. The analysis was carried out on DICHROWEB (Whitmore and Wallace, 2004).

# 3. Protein expression and purification

Biochemical and structural characterisation of proteins and protein-ligand interactions typically requires milligram amounts of pure, homogenous protein. There are two ways of obtaining protein for characterisation: purification from natural source or recombinant expression. Protein purification from natural source is ridden with technical difficulties, costly, time-consuming and is often limited to proteins that are expressed at high levels in the organism of interest. Conversely, recombinant protein expression, which involves introduction of an artificial gene encoding the protein of interest into a heterologous host, is a cost-effective, scalable and often quick method of obtaining large amounts of pure protein. Hence, recombinant expression was used as the method-of-choice for producing Na-FAR-1 and the LNS2 domain-containing C-terminal fragments of the human Nirs and *Drosophila* RdgB proteins for their characterisation.

In previous studies, Na-FAR-1 was successfully produced and purified for structure determination by X-ray crystallography and NMR spectroscopy. It was expressed to high levels from a pET-based expression vector and proved to be stable during manipulations required for structural investigations. Hence, existing protocols were employed for expression and purification of Na-FAR-1 with a few modifications allowing a higher yield of the $^{15}$N-labelled protein to be achieved for NMR spectroscopy.

As opposed to Na-FAR-1, the C-terminal fragments of Nirs and RdgB have been previously produced only in small quantities for biochemical lipid binding assays (Chang and Liou, 2015; Kim *et al*., 2013). Because of the difficulty of producing recombinant Nir1-3 and RdgB fragments that was observed early in this project, a significant amount of work had to be carried out to identify the constructs and conditions that enabled sufficient yield of recombinant proteins to be obtained for characterisation.

Below, the detailed process of designing, cloning, expressing and purifying of the LNS2 domain-containing fragments of Nir1-3 and RdgB will be described. The process of Na-FAR-1 expression and purification will be touched upon in less detail owing to the ease of Na-FAR1 production and the small number of optimisations that were carried out.

## 3.1. The LNS2 domain of Nir1-3 and RdgB

### 3.1.1 Recombinant protein expression strategy

Many recombinant expression systems have been developed over the years to suit different research needs. These can be divided into prokaryotic [e.g, *Escherichia coli* (Baneyx, 1999; Makrides, 1996), *Brevibacillus chosinensis* (Mizukami and Miyauchi, 2010)] eukaryotic [e.g., insect cells (Jarvis, 2009), mammalian cells (Jäger *et al*., 2015), yeast (Mattanovich *et al*., 2012)] and cell-free systems (Carlson *et al*., 2012). Important considerations in choosing the suitable expression systems are the cost-effectiveness of the system, the source organism of the protein, ease of manipulation, the time required for protein expression and scalability of expression.

The most commonly used expression system is *E. coli*, which has many advantageous properties for protein production. These include fast growth rates, high recombinant protein yields, ease of cell transformation, availability of a range of expression vectors, growth media and cell strains, and good understanding of *E. coli* genetics and protein expression machinery. In addition, the relative ease of isotope-labelled protein production in *E. coli* offers a specific advantage for structural investigations using biomolecular NMR spectroscopy, which relies heavily on the use of isotope-labelled proteins.

Nevertheless, there are also a number of disadvantages associated with the use of *E. coli* for expression of eukaryotic proteins such as Nir2, our protein of interest. Eukaryotic proteins often fail to express and/or fold in the prokaryotic expression systems due to lack of the eukaryotic protein expression and folding machinery, the absence of the post-translational modifications, inadequate formation of disulfide bridges or the high rate of translation in prokaryotic cells. However, many strategies exist to circumvent these problems, including protein secretion to the periplasm for disulfide-bond formation, co-expression with molecular chaperones such as trigger factor (TF) (Nishihara *et al*., 2000) for enhanced protein folding or the use of solubility-enhancing fusion partners such as maltose-binding protein (MBP) (Kapust and Waugh, 1999) or thioredoxin (Trx) (LaVallie *et al*., 2000) for increased stability and solubility.

In this study, *E. coli* was chosen as the primary expression system due to the advantages described above and the presence of the technical facilities suitable for *E. coli* cultivation. Additionally, LNS2 domain expression was also tested in *Spodoptera frugiperda* Sf9 cells

to determine whether expression in eukaryotic system is preferred for the protein of interest.

### 3.1.2 Design of the LNS2 domain fragments for recombinant expression

The LNS2 domain of Nir2 was chosen as the primary target for the investigation due to its regulatory function in PI(4,5)$P_2$ signalling, as described in Chapter 1. However, expression of the LNS2 domains of three close Nir2 homologs (Nir1, Nir3 and RdgB) was also attempted. This was done to increase the probability of obtaining soluble protein, as even small variations in the amino acid sequence of homologous domain are known to influence protein yields in recombinant expression.

It also is often the case that the yield and solubility of a recombinant protein domain depends on the choice of expression boundaries when designing the construct (Edavettal *et al*., 2012). Hence, two different fragments containing the LNS2 domain were designed for each of the four proteins, the 'short' fragment (LNS2-S) and the 'long' fragment (LNS2-L), which differed in the length of the sequence flanking the LNS2 domain. The LNS2-S fragments contained the LNS2 domain flanked by a few neighbouring residues, whereas LNS2-L comprised the LNS2 domain as well as the larger portion of the C-terminal part of the protein. To give an example, the LNS-S and the LNS2-L fragments of Nir2 comprised amino acids 995-1221 and 911-1244, respectively (Fig 3-1). The full list of the fragments is given in Table 3-1.

The boundaries of the LNS2 domain were determined by analyzing the C-terminal portions of Nir1-3 and RdgB using Simple Modular Architecture Research Tool (SMART, www.smart.embl-heidelberg.de) (Letunic *et al*., 2015; Schultz *et al*., 1998) and Protein BLAST (www.blast.ncbi.nlm.nih.gov/) (Altschul *et al*., 1997) to identify conserved sequence that might delineate the extent of the domain. Additionally, the C-terminal fragments of the proteins were subjected to analysis by the disorder predicting algorithm VL-XT (Li *et al*., 1999) using Predictor of Naturally Disordered Regions (PONDR, www.pondr.com) and the secondary structure predicting algorithm JNet (Drozdetskiy *et al*., 2015) using JPred 4 (http://www.compbio.dundee.ac.uk/jpred/) to identify the likely ordered and structured regions in the C-terminal parts of the protein containing the LNS2 domain. Based on SMART, protein BLAST, JPred 4 and PONDR analyses, the expression fragments were chosen in such a way that they contained a complete LNS2 homology region and an additional stretch of residues at the both termini of LNS2-S and the

N-terminus of LNS2-L that are predicted to be ordered. The sequences were chosen such that they started and ended with hydrophilic, polar residues. This was assumed to increase the likelihood of production of a stable and soluble protein by ensuring the lack of exposed hydrophobic residues that may interact in an inter-molecular fashion and cause protein aggregation. The intention was also to improve the chance of crystallising each protein, as the presence of the disordered regions at the protein termini may lead to sample inhomogeneity and inhibit crystal formation. The LNS2-L fragment of Nir2 was chosen to be identical to the Nir2 fragment expressed in *E. coli* in two previous studies (Chang and Liou, 2015; Kim *et al*., 2013).

| Protein | Fragment | Residues* |
|---|---|---|
| Nir1 | LNS2-S | 726-895 |
| Nir1 | LNS2-L | 726-971 |
| Nir2 | LNS2-S | 995-1171 |
| Nir2 | LNS2-L | 911-1244 |
| Nir3 | LNS2-S | 1057-1261 |
| Nir3 | LNS2-L | 1057-1341 |
| RdgB | LNS2-S | 1046-1198 |
| RdgB | LNS2-L | 1001-1250 |

* Residue numbers in the full-length protein.

**Table 3-1. Fragments of Nir1-3 and RdgB proteins used for recombinant protein expression.**

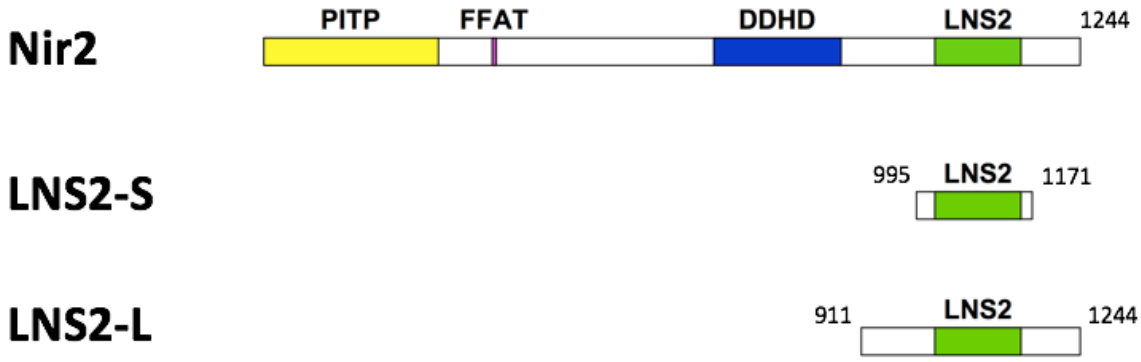



**Figure 3-1. Schematic representation of the LNS2 domain-containing Nir2 fragments used for recombinant expression.** Full-length Nir2 is shown for comparison.

### 3.1.3 Creation of the high-throughput expression construct library

After selecting the C-terminal fragments of Nir proteins and RdgB to produce, a library of recombinant constructs for high-throughput expression screening of the protein fragments

was generated using the vectors of the pOPIN vector suite (Bird, 2011). The pOPIN suite allows rapid production of constructs for protein expression with a variety of solubility-enhancing fusion partners and purification tags, and is suitable for expression screening in *E. coli*, insect and mammalian cells, which aids the identification of optimally expressed constructs in a relatively short timeframe. The library creation and screening was carried at the Oxford Protein Production Facility (OPPF-UK) located at the Research Complex at Harwell in Oxfordshire, UK under the guidance of Dr Louise Bird and with help from Mr Valtteri Järvinen.

Due to the limited time available at the OPPF-UK for creation and testing of the expression constructs, only the fragments of Nir2, Nir3 and RdgB proteins were chosen for the high-throughput screening. Of these three, the majority of the constructs were created with the LNS2 domains of Nir2 and RdgB owing to the relatively large evolutionary distance between Nir2 and RdgB compared to Nir2 and Nir3, and hence the higher possibility of observing differences in the expression yield and stability of the protein fragments.

The cloning was carried out using In-Fusion® ligation independent cloning kit (Raman and Martin, 2014). For cloning, each Nir and RdgB fragment was amplified by PCR from a synthetic DNA template, which was codon-optimised for expression in *E. coli* (See Appendix A for the list of the PCR primers). The fragments were cloned into a set of pOPIN vectors pre-digested with HindIII and KpnI restriction endonucleases. The presence of the insert in the reconstituted vectors was confirmed by PCR and, in some instances, by Sanger sequencing. Subsequently, the constructs were transformed into *E. coli* or used for creation of recombinant baculovirus for insect cell expression.

Additionally, a number of polyhistidine-tagged Nir and RdgB constructs were produced by ligation independent cloning in our home laboratory using a pNIC28-Bsa4 vector. The full list of the constructs created in this study for expression screening of Nir and RdgB LNS2 domains is given in Tables 3-2A-B.

| Protein | Fragment | Plasmid | Tag |
|---|---|---|---|
| Nir1 | LNS2-S | pNIC28-Bsa4 | N-H6 |
| Nir1 | LNS2-L | pNIC28-Bsa4 | N-H6 |
| Nir2 | LNS2-S | pOPINF | N-H6 |
| Nir2 | LNS2-L | pOPINF | N-H6 |
| Nir2 | LNS2-S | pOPINE-3C-HALO7 | HALO7-H6-C |
| Nir2 | LNS2-L | pOPINE-3C-HALO7 | HALO7-H6-C |
| Nir2 | LNS2-S | pOPINEneo | H6-C |
| Nir2 | LNS2-L | pOPINEneo | H6-C |
| Nir2 | LNS2-S | pOPINF | N-H6 |
| Nir2 | LNS2-L | pOPINF | N-H6 |
| Nir2 | LNS2-S | pOPINHALO7 | N-H6-HALO7 |
| Nir2 | LNS2-L | pOPINHALO7 | N-H6-HALO7 |
| Nir2 | LNS2-S | pOPINJ | N-H6-GST |
| Nir2 | LNS2-L | pOPINJ | N-H6-GST |
| Nir2 | LNS2-S | pOPINM | N-H6-MBP |
| Nir2 | LNS2-L | pOPINM | N-H6-MBP |
| Nir2 | LNS2-S | pOPINNUSA | N-H6-NusA |
| Nir2 | LNS2-L | pOPINNUSA | N-H6-NusA |
| Nir2 | LNS2-S | pOPINS3C | N-H6-SUMO |
| Nir2 | LNS2-L | pOPINS3C | N-H6-SUMO |
| Nir2 | LNS2-S | pOPINTF | N-H6-TF |
| Nir2 | LNS2-L | pOPINTF | N-H6-TF |
| Nir2 | LNS2-S | pOPINTRX | N-H6-Trx |
| Nir2 | LNS2-L | pOPINTRX | N-H6-Trx |
| Nir2 | LNS2-S | pNIC28-Bsa4 | N-H6 |
| Nir2 | LNS2-L | pNIC28-Bsa4 | N-H6 |
| Nir3 | LNS2-S | pOPINEneo | H6-C |
| Nir3 | LNS2-S | pOPINF | N-H6 |
| Nir3 | LNS2-S | pOPINHALO7 | N-H6-HALO7 |
| Nir3 | LNS2-S | pOPINJ | N-H6-GST |
| Nir3 | LNS2-S | pOPINM | N-H6-MBP |
| Nir3 | LNS2-S | pOPINS3C | N-H6-SUMO |
| Nir3 | LNS2-S | pOPINTRX | N-H6-Trx |
| Nir3 | LNS2-S | pOPINEneo | H6-C |
| Nir3 | LNS2-S | pNIC28-Bsa4 | N-H6 |
| Nir3 | LNS2-L | pNIC28-Bsa4 | N-H6 |

**Table 3-2A. Expression constructs of Nir1-3 designed in this study.** The names and positions of the solubility and purification tags are given in the last column. N- or -C = N- or C-terminal tag, H6 = hexahistidine tag, GST = glutathione transferase, MBP = maltose-binding protein, TF = trigger factor, Trx = thioredoxin.

| Protein | Fragment | Plasmid | Tag |
|---|---|---|---|
| RdgB | LNS2-S | pOPINE-3C-HALO7 | HALO7-H6-C |
| RdgB | LNS2-L | pOPINE-3C-HALO7 | HALO7-H6-C |
| RdgB | LNS2-S | pOPINEneo | H6-C |
| RdgB | LNS2-L | pOPINEneo | H6-C |
| RdgB | LNS2-S | pOPINF | N-H6 |
| RdgB | LNS2-L | pOPINF | N-H6 |
| RdgB | LNS2-S | pOPINHALO7 | N-H6-HALO7 |
| RdgB | LNS2-L | pOPINHALO7 | N-H6-HALO7 |
| RdgB | LNS2-S | pOPINJ | N-H6-GST |
| RdgB | LNS2-L | pOPINJ | N-H6-GST |
| RdgB | LNS2-S | pOPINM | N-H6-MBP |
| RdgB | LNS2-L | pOPINM | N-H6-MBP |
| RdgB | LNS2-S | pOPINNUSA | N-H6-NusA |
| RdgB | LNS2-L | pOPINNUSA | N-H6-NusA |
| RdgB | LNS2-S | pOPINS3C | N-H6-SUMO |
| RdgB | LNS2-L | pOPINS3C | N-H6-SUMO |
| RdgB | LNS2-S | pOPINTF | N-H6-TF |
| RdgB | LNS2-L | pOPINTF | N-H6-TF |
| RdgB | LNS2-S | pOPINTRX | N-H6-Trx |
| RdgB | LNS2-L | pOPINTRX | N-H6-Trx |
| RdgB | LNS2-S | pNIC28-Bsa4 | N-H6 |

**Table 3-2B. Expression constructs of RdgB designed in this study.** The names and the locations of solubility and purification tags are given in the last column. N- or -C = N- or C-terminal tag, H6 = hexahistidine tag, GST = glutathione transferase, MBP = maltose-binding protein, TF = trigger factor, Trx = thioredoxin.

### 3.1.4 High-throughput expression screening of Nir2, Nir3 and RdgB LNS2 domain in *E. coli* and insect cells

#### 3.1.4.1 Expression in *E. coli*

The high-throughput expression screening was carried out in Lemo21 (DE3) and Rosetta cells either in Overnight Express™ autoinduction medium or TB medium with IPTG induction. The recombinant proteins were purified from 2 ml of lysate by $Ni^{2+}$-affinity chromatography, and protein elution fractions were analysed by SDS-PAGE to identify the expression constructs exhibiting the highest expression levels of Nir2-3 and RdgB LNS2 domain fragments.

The LNS2 domains from the different proteins displayed somewhat different protein yields. Namely, RdgB LNS2 fragments appeared to be expressed slightly better than Nir2 and Nir3 fragments, with expression of 7 out of 20 constructs detected for RdgB, 5 out of 20 for Nir2, and 2 out of 10 for Nir3 (Fig 3-2, 3-3). The best overall protein yield was observed in Lemo21 cells grown in Overnight Express™ medium, whereas protein yields obtained with IPTG induction were lower, as judged from the intensity of the recombinant bands. This result was not unexpected, as Overnight Express™ medium supports higher cell densities than the rich medium used with IPTG induction.

Expression screening revealed that the LNS2 domain is a difficult protein to produce. It was observed that most LNS2 domain constructs containing short purification tags such as $His_6$-tag, GST (Harper and Speicher, 2001) or SUMO (Malakhov *et. al*, 2004; Marblestone *et al*., 2006) either failed to express or showed very low protein yield, as the corresponding bands were not clearly visible on the polyacrylamide gels. An exception is the thioredoxin (Trx) fusion of RdgB LNS2-L, which notably was present in the cell lysate only after expression in the autoinduction medium, and not after expression in the rich medium with IPTG induction.

Generally, only bands containing high molecular weight (MW) (> 65 kDa) fusions of the LNS2 domain, i.e. Trigger factor (TF), N-utilisation substance A (NusA) (Davis *et al*., 1999) and maltose-binding protein (MBP) fusions, were detectable on the gels. Of these, the intensity of the bands corresponding to the MBP fusion of Nir2 LNS2-S and LNS2-L, the NusA fusion of Nir2 LNS2-S, the TF fusion of Nir2 LNS2-L, the SUMO fusion of Nir3 LNS2-S, the Trx fusion of Nir3 LNS2-S, the NusA fusion of RdgB LNS2-L and the Trx fusion of RdgB LNS2-L was weak and comparable to the intensity of the endogenous

*E. coli* protein bands observed in the gel. In contrast, the bands corresponding to the TF fusion of Nir2 LNS2-S, the MBP fusion of RdgB LNS2-L, and the TF fusion of RdgB LNS2-S and LNS2-L were intense and clearly stood out of the background, indicating that the expression yield of those constructs was relatively high. Here, the TF fusion of Nir2 LNS-S fusion demonstrated the highest expression yield compared to other recombinant proteins, and hence appeared to be the most promising candidate for expression scale-up.

It is unclear why the majority of LNS2 fusion proteins exhibited very low expression levels in *E. coli*, but the possible explanations include the lack machinery necessary for successful folding of mammalian proteins or post-translational modifications in bacterial cells, general poor solubility of LNS2 domain or the toxic effects it may have on the cells. Not entirely unexpected, TF and MBP appeared to be most successful at solubilising the fragments, which is consistent with the previous evidence demonstrating their effectiveness as solubility tags (Kapust and Waugh, 1999; Lebendiker and Danieli, 2014).

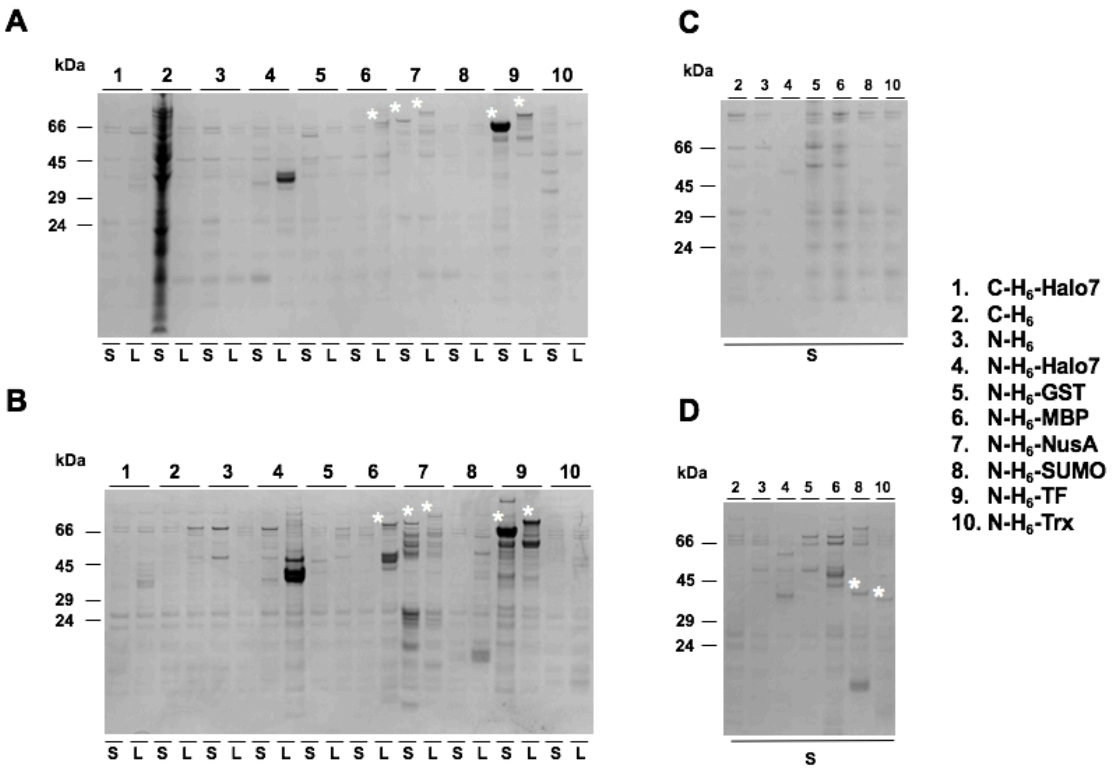



**Figure 3-2. SDS-PAGE analysis of Nir2 and Nir3 LNS2 high-throughput expression screening in *E. coli*.** $Ni^{2+}$-affinity purified lysates are shown. Coomassie® protein staining. **A**. Nir2 LNS2 expression with IPTG induction in rich medium. **B**. Nir2 LNS2 expression with autoinduction in Overnight Express™ medium. **C**. Nir3 LNS2 expression with IPTG induction in rich medium. **D.** Nir3 LNS2 expression with autoinduction in Overnight Express ™ medium. Recombinant protein bands are marked by white asterisks. The intense band in the N-$H_6$-Halo7 L lanes corresponds to Halo7 expressed without the recombinant LNS2 fragment or a degradation product of the fusion protein. The names of the LNS2 domain constructs tested are specified below the lanes (S = LNS2-S, L = LNS2-L). Types and positions of the recombinant expression tags are specified above the lanes. N = N-terminal tag, C = C-terminal tag. $H_6$ = hexahistidine tag, GST = glutathione S-transferase, MBP = maltose binding protein, SUMO = small ubiquitin-like modifier, TF = trigger factor, Trx = thioredoxin.

### 3.1.4.2 Expression in insect cells

In contrast to *E. coli,* insect cells (Sf9) appeared more efficient at producing recombinant LNS2. As judged by SDS-PAGE analysis (Fig 3-4), a variety of lower MW (< 65 kDa) LNS2 fusion proteins appeared to be expressed at markedly higher levels than in *E. coli* allowing their detection by Coomassie® staining. This included a 37 kDa C-terminally His-tagged Nir2 LNS2-L construct, which was thought to be suitable for analysis by protein NMR and X-ray crystallography due its small size and the absence of large fusion partners that have been reported to inhibit protein crystallisation. Interestingly, the analogous N-terminally $His_6$-tagged LNS2-L construct was expressed to much lower yield, suggesting that the position of the tag can influence the efficiency of LNS2 domain expression in insect cells.

In contrast to screening in *E. coli*, Nir2 constructs were generally better expressed than RdgB constructs in insect cells. Expression of eleven Nir2 constructs was detected, compared to only nine RdgB constructs. Interestingly, expression of $His_6$-tagged RdgB fragments was not detected at all. The LNS2 domain of Nir3 has demonstrated the lowest protein yield compared to Nir2 and RdgB. The reason for the observed differences in expression is unknown, but has probably to do with the differences in the physiochemical properties of the constructs arising from the variations in the primary sequence.

Furthermore, in insect cell, the longer LNS2-L fragments demonstrated higher expression levels that the shorter LNS-S fragments. This was also in contrast with screening in *E. coli*, where fusions with the short LNS2 fragment were expressed to higher yield. This might suggest that the extra protein sequence flanking the LNS2 domain in LNS2-L fragments aids its folding in eukaryotic cells.

Based on the satisfactory results of expressing screening in insect cells, large-scale overexpression of C-terminally $His_6$-tagged Nir2 LNS2-L construct was attempted, as described below. Expression of other constructs was not taken further due to time constraints and the fact that some of the large LNS2 domain fusions had already demonstrated reasonable expression levels in *E. coli*.

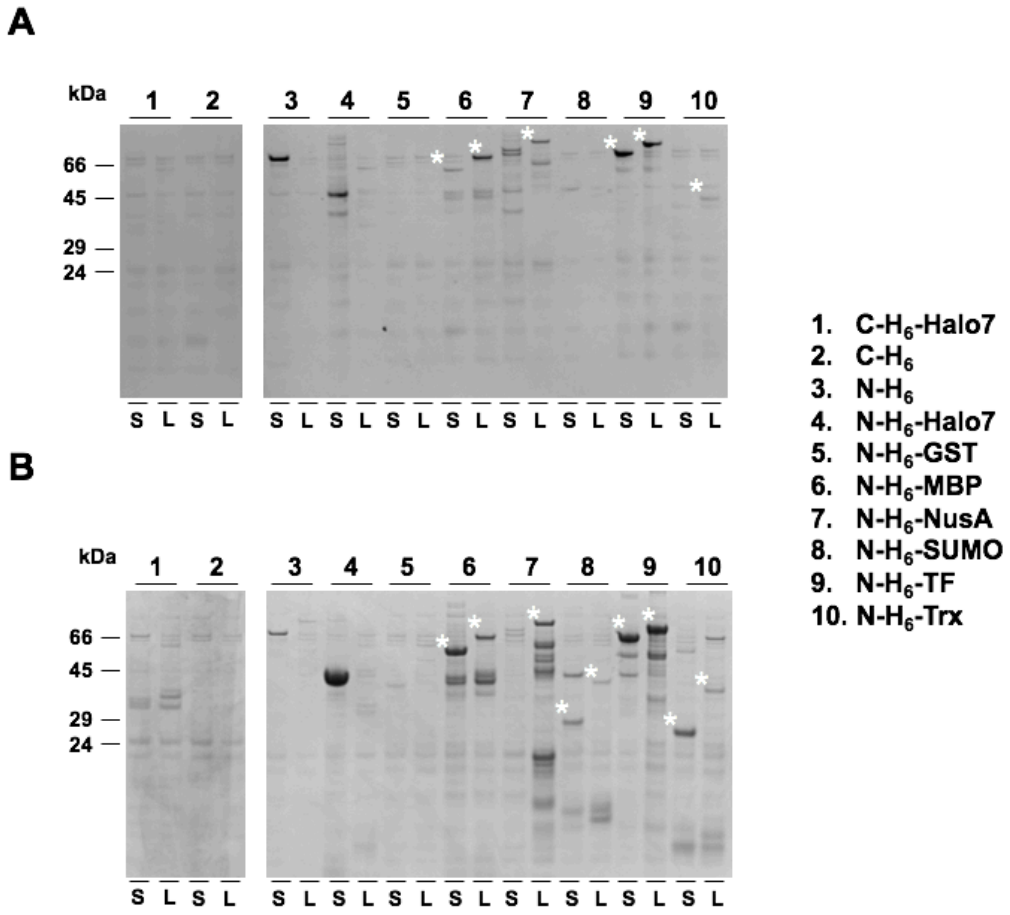



**Figure 3-3. SDS-PAGE analysis of RdgB LNS2 high-throughput expression screening in *E. coli*.** $Ni^{2+}$-affinity purified lysates are shown. Coomassie® protein staining. Recombinant protein bands are marked by white asterisks. Type and position of recombinant tags are specified above the lanes. The intense band in the N-$H_6$-Halo7 S lanes corresponds to Halo7 expressed without the recombinant LNS2 fragment or a degradation product of the fusion protein. **A**. IPTG induction in rich medium. **B**. Autoinduction in Overnight Express™ medium. The names of the LNS2 domain constructs tested are specified below the lanes (S = LNS2-S, L = LNS2-L). Types and positions of the recombinant expression tags are specified above the lanes. N = N-terminal tag, C = C-terminal tag. $H_6$ = hexahistidine tag, GST = glutathione S-transferase, MBP = maltose binding protein, SUMO = small ubiquitin-like modifier, Trx = thioredoxin.

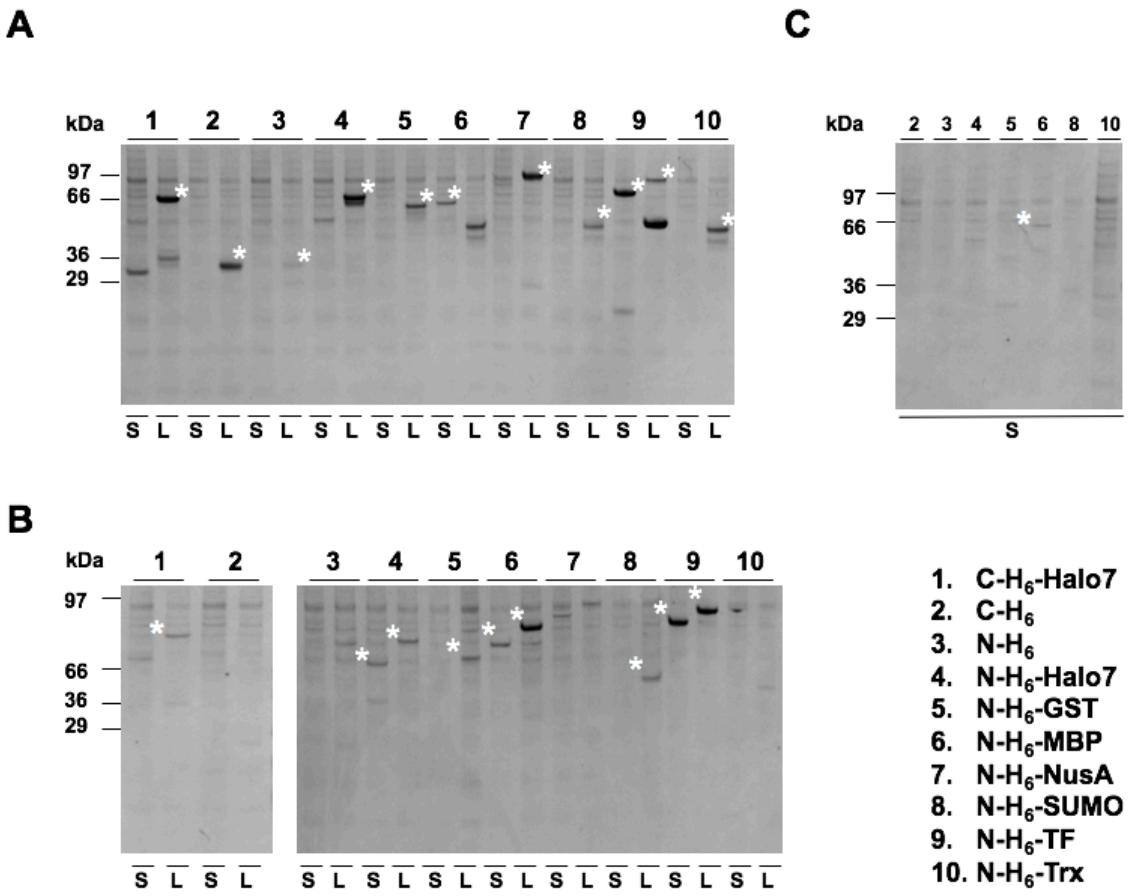



**Figure 3-4. SDS-PAGE analysis of the LNS2 domain high-throughput expression screening in insect cells.** $Ni^{2+}$-affinity purified lysates are shown. Coomassie® protein staining. Recombinant protein bands are marked by white asterisks. **A**. Nir2 fragments. **B**. RdgB fragments. **C**. Nir3 fragments. The names of the LNS2 domain constructs tested are specified below the lanes (S = LNS2-S, L = LNS2-L). Types and positions of the recombinant expression tags are specified above the lanes. N = N-terminal tag, C = C-terminal tag. $H_6$ = hexahistidine tag, GST = glutathione S-transferase, MBP = maltose binding protein, SUMO = small ubiquitin-like modifier, TF = trigger factor, Trx = thioredoxin.

### 3.1.5 Small-scale expression of $His_6$-tagged Nir1 and Nir2 LNS2 domain fragments in *E. coli*

As no soluble $His_6$-only tagged LNS2 domain fragments of Nir2, Nir3 and RdgB were obtained in the high-throughput screen in *E. coli* from pOPIN vectors, expression of Nir1 LNS2 domain was attempted from pNIC28-Bsa4 vector. For comparison, expression of $His_6$-only tagged Nir2 LNS2 domain was also carried out using the same protocol.

Small-scale (1 ml) expression test were performed using N-terminally $His_6$-tagged Nir1 LNS2-S (21 kDa) and Nir2 LNS2-S (22 kDa) fragments. In order to determine whether there is any recombinant protein accumulation in *E. coli* inclusion bodies, the lysate was separated by centrifugation into two fractions: a fraction containing soluble proteins and a

fraction containing insoluble cell debris and inclusion bodies. For reference, the lysate from uninduced cells and unseparated post-induction lysate were also analysed.

From the SDS-PAGE analysis (Fig 3-5), it became evident that the expression level of Nir1 LNS2 was higher than that of Nir2 LNS2. This is suggested by the fact that only the band corresponding to Nir1 LNS2 construct was observed, albeit rather weakly, in all post-induction fractions, at slightly below 25 kDa, whereas a band corresponding to Nir2 LNS2 could not be clearly distinguished on the gel.

Nevertheless, it should be noted that the expression levels of both proteins were likely very low, as the intensity of all recombinant protein bands is lower compared to the *E. coli* protein bands. Also, $His_6$ Nir1 LNS2-S appeared to be mostly found in the insoluble fraction, which likely indicates problems with protein folding in the cell, and could explain why LNS2 domain fusions with large solubility tags were more successfully expressed in the screens.

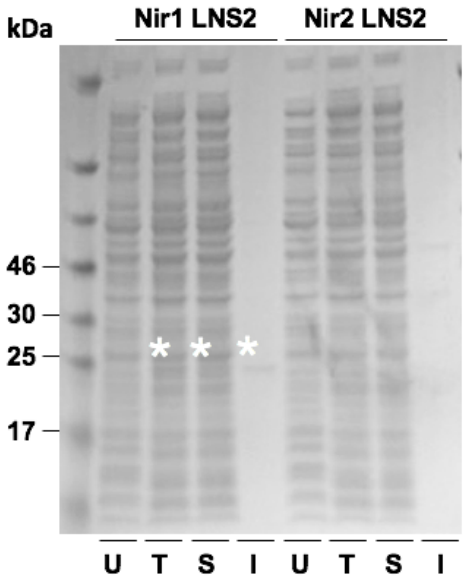



**Figure 3-5. Small-scale expression of N-terminally $His_6$-tagged Nir1 LNS2 and Nir2 LNS2.** Coomassie® protein staining. Nir1 LNS2 bands are marked by white asterisks. Nir2 LNS2 bands are not distinguishable from the background. U = total lysate from uninduced cells, T = total lysate from induced cells, S = soluble fraction of induced cell lysate, I = insoluble fraction of induced cell lysate.

### 3.1.6 Large-scale expression of Nir2, Nir3 and RdgB LNS2 domain fragments in *E. coli* and insect cells

Next, it was attempted to scale-up the expression of recombinant LNS2 domain fragments identified in the expression tests and screens. The scale-up of all constructs, except C-$His_6$-tagged Nir2 LNS2-L, was carried out in *E. coli.* Expression of C-$His_6$-tagged Nir2 LNS2-L was carried out in insect cells.

### 3.1.6.1 Expression in *E. coli*

In *E. coli*, large-scale expression was carried out in LB medium or rich phosphate-buffered terrific broth (TB) medium using IPTG induction. Large TF Nir2 LNS2-S (70 kDa), TF RdgB LNS2-S (67 kDa), MBP RdgB LNS2-L (70 kDa) fusions were used for expression, as they displayed the highest expression levels in the screen. MBP RdgB LNS2-L was chosen instead of MBP RdgB LNS2-S due to its apparent higher yield under IPTG induction. Additionally, expression of the shorter Trx RdgB LNS2-S fusion (30 kDa), SUMO RdgB LNS2-L (40 kDa), N-terminally $His_6$-tagged Nir1 and LNS2-S (21 kDa and 22 kDa, respectively) was attempted in the hope of obtaining a protein construct directly suitable for structural analysis by NMR spectroscopy, which is challenging with high MW proteins. Large-scale expression of $His_6$-tagged Nir1-2 LNS2-S constructs was carried out from pET-based pNIC28-Bsa4 plasmids.

All three large fusion proteins were successfully expressed in shake-flask cultures and purified by $Ni^{2+}$-affinity chromatography, as evident from the presence of protein bands in the chromatography elution fractions (Fig 3-6). TF Nir2 LNS2-S and TF RdgB LNS2-S were expressed in the same culture volume and purified using identical conditions, whereas MBP RdgB LNS2-L fusion was purified using a different protocol. Thus, the intensity the MBP-fusion band cannot be directly compared with TF-fusions to estimate protein yield. From the gel it can be concluded, however, that the protein yield was similar between the two TF-fusion. The total protein yield for each construct was estimated using UV/Vis spectroscopy and was found to be equal to ~ 14 mg/L of TB or ~ 2 mg/L of LB for both TF Nir2 LNS2-S and TF RdgB LNS2-S and to ~ 6 mg/L of TB for MBP RdgB LNS2-L, which was consistent with the relative expression levels observed in the expression screening.

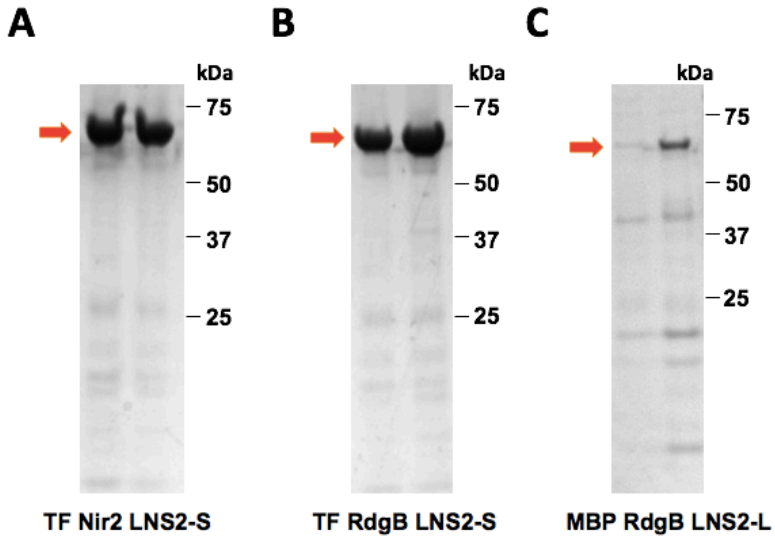



**Figure 3-6. Large-scale expression of trigger factor (TF) and maltose-binding protein (MBP) fusions of Nir2 and RdgB LNS2 domains in *E. coli*.** Coomassie® protein staining. **A.** TF Nir2 LNS2 (70 kDa). **B.** TF RdgB LNS2 (67 kDa). **C.** MBP RdgB LNS2-L (70 kDa). Two elution fractions from $Ni^{2+}$-affinity chromatography are shown for each protein. Recombinant protein bands are highlighted by a red arrow.

As expected, large-scale production of shorter (~ 20-40 kDa) LNS2 constructs was much less successful (see Fig 3-7 for three examples). From SDS-PAGE analysis alone it was not entirely clear whether recombinant protein bands were present in $Ni^{2+}$-affinity elution fractions, as they could not be readily distinguished from the co-purifying *E. coli* proteins. The presence of the $His_6$ Nir2 LNS2-S and Trx RdgB LNS-S recombinant proteins was, however, confirmed by Western blotting using an anti-His-tag antibody (data not shown).

The highest yield of soluble protein was observed with $His_6$-tagged Nir1 LNS2-S (~ 0.1 mg/L of LB), which however, was very low compared to the protein yields obtained with large LNS2 domain fusions. Similar or lower yields of soluble protein were observed for $His_6$ Nir2 LNS2-S and Trx RdgB LNS-S. This is in contrast with the results of the high-throughput expression screen, where a relatively higher amount of soluble protein was obtained with the SUMO- and Trx-fusions of RdgB LNS2 domain fusion compared to the $His_6$-tagged LNS2 domain fragments during *E. coli* expression.

In an attempt to improve the yield of $His_6$ Nir1 LNS2 domain, 10-20% glycerol and/or 50 mM L-arginine and L-glutamine were added to the lysis and affinity chromatography buffers. Although this helped to obtain higher amount of purer soluble protein (Fig 3-7A), it was found that the LNS2 domain fragment was very prone to aggregation and could not be easily purified further by size-exclusion chromatography, exchanged into additive-free buffers by diafiltration or dialysis, or concentrated above ~ 0.5 mg/ml, making it unsuitable

for characterisation by many biochemical and biophysical methods. Nevertheless, the Nir1 fragment was used for characterisation by CD spectroscopy, as described in Chapter 4.

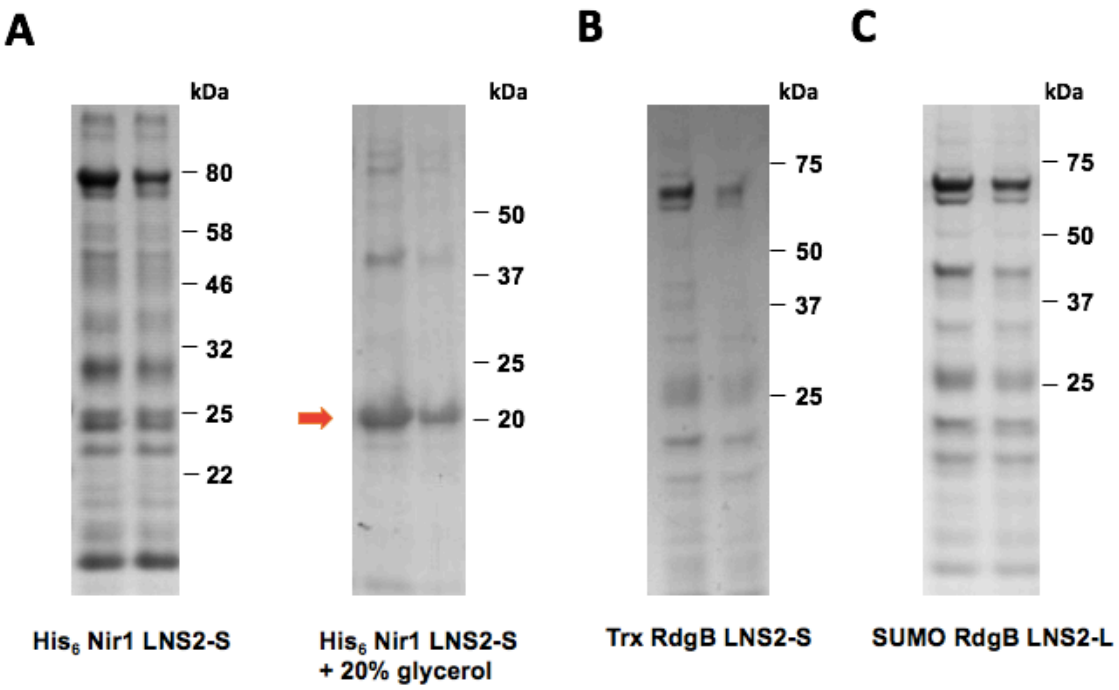



**Figure 3-7. Large-scale expression of $His_6$-tagged Nir1 LNS domain, and thioredoxin (Trx) and SUMO fusions of RdgB LNS2 in *E. coli*.** Coomassie® protein staining. **A.** $His_6$ Nir2 LNS2-S (22 kDa). Left. Protein purified in the absence of 20% glycerol. Right. Protein purified in the presence of 20% glycerol. Position of the recombinant protein band is marked by a red arrow. **B.** Trx RdgB LNS2-S (30 kDa). **C.** SUMO RdgB LNS2-S (40 kDa). Two elution fractions from $Ni^{2+}$-affinity chromatography are shown for each protein.

### 3.1.6.2 Expression in insect cells

In insect cells, only scale-up of the C-terminally $His_6$-tagged Nir2 LNS2-L (C-$His_6$ Nir2 LNS2-L) construct was attempted (Fig 3-8), as this was the fragment with the smallest MW detected in the high-throughput screen. It was found the protein was expressed in insect cells with an adequate yield of ~ 0.5 mg per 100 ml of adherent Sf9 cell culture, which was in contrast with the $His_6$-tag-only LNS2 domain constructs expressed in *E. coli*. Although higher purity of the protein could be achieved already after the affinity chromatography, C-$His_6$ Nir2 LNS2-L also proved to be very prone to aggregation during further manipulations both in the absence and the presence of 10-20% glycerol, and hence could not be successfully used for characterisation of the PA binding by major techniques. Insect cell expression was carried out with support from Dr Jan Petersen (University of Glasgow).

Based on the outcomes of the large-scale expression, characterisation of the LNS2 domain binding properties was carried out mostly with the large fusions of the LNS2 domain,

which could be successfully purified from *E. coli* in markedly higher amounts and did not require additives for stability and solubility. The foldedness and functionality of some $His_6$-tagged LNS2 domain fragments nevertheless investigated, as described in Chapter 4.

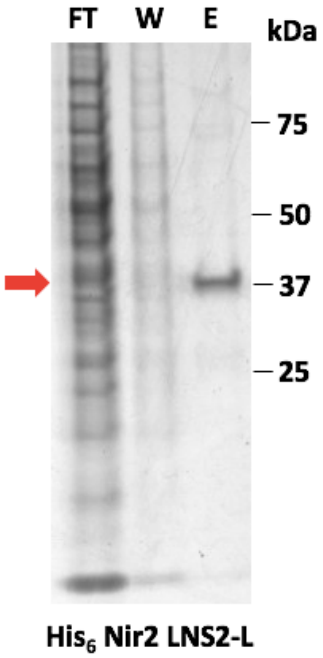



**Figure 3-8. Large-scale expression of $His_6$-tagged Nir2 LNS domain in insect cells.** Coomassie® protein staining. Flow-through (FT), wash (W) and elution (E) fractions from $Ni^{2+}$-affinity chromatography are shown. Position of the recombinant protein band is marked by a red arrow. Theoretical MW of the protein is 37 kDa.

### 3.1.7 Purification and characterisation of the TF and MBP fusions of Nir2 LNS2 domain by size-exclusion chromatography

#### 3.1.7.1 TF fusions of Nir2 and RdgB LNS2 domain

After suitable LNS2 domain constructs were identified, further purification and characterisation of the fragments was a carried out by size-exclusion chromatography (SEC). SEC is a robust technique used for batch purification of proteins and analysis of protein samples to determine the oligomeric state of the protein and presence of aggregates in the sample. In SEC, proteins are separated based on their size or, more accurately, their Stokes radii, which generally correlate with the molecular weight of the proteins. Separation is achieved by filtration of the analytes through a porous matrix, where larger proteins are retained for a longer period of time than smaller proteins.

Analysis of affinity-purified TF Nir2 LNS2-S fusion by SEC has demonstrated that the samples are polydisperse and the protein is found in a number of distinct states (Fig 3-9). A set of peaks corresponding to the fusion protein was observed in the $A_{280}$ trace at the retention volumes of 7.9 ml, 10.2-11.0 ml, 11.8 ml and 13.4 ml. From a calibration curve constructed using a set of known protein standards with molecular weights up to 67 kDa, it

was estimated that the molecular weight of the species eluting at 13.4 ml is approximately 68 kDa, which is close to the theoretical MW of the monomeric protein (70 kDa). The peak that eluted at 7.9 ml was assumed to correspond to a high MW aggregate present in the sample, as it eluted close to the void volume of the column (7.3 ml).

To characterise the TF Nir2 LNS2-S species more accurately, the fusion protein solution was further analysed by multi-angle light scattering (MALS). In MALS, light scattering by a protein is measured at different angles, and the size and molecular weight of the protein are calculated based on the observed light scattering properties. When used in tandem with SEC, MALS allows determination of the molecular weights of individual proteins and distinct protein forms in a complex mixture.

Analysis of TF Nir2 LNS2 by SEC-MALS has revealed four protein peaks eluting at 6.8, 7.1 ml, 7.6 ml and 8.2 ml (Fig 3-10). Unlike in Fig 3-9, only a weak peak corresponding to high MW aggregates was observed at ~ 6.1 ml. The differences in retention volumes of the peaks in Fig 3-10 and Fig 3-9 are due to the use of different SEC columns in the experiments. The differences in the relative intensity of the peaks can be attributed to the variations between protein batches.

MW calculations were carried out for each peak using MALS data. The peaks at 6.8 ml and 7.1 ml were poorly separated and were treated as one peak for the calculation. The average MWs of the eluting protein species were found to be ~ 309 kDa for the 6.8 – 7.1 ml peak, ~ 170 kDa for the 7.6 ml peak ~ 93 kDa for the peak at 8.2 ml. The 93 kDa and 170 kDa peaks were mostly monodisperse [polydispersity index (PDI) = 1.00, 1.01, respectively], whereas the 309 kDa peak was polydisperse (PDI = 1.06) with MWs ranging from ~ 359 kDa to ~ 221 kDa (Fig 3-10, red lines). Based on this data, the 93 kDa peak was assumed to contain the monomeric form of the protein, the 170 kDa peak the dimeric form of the protein, and the 309 kDa peak a mixture of protein microaggregates or trimers/quadromers. The difference between the calculated and the theoretical MWs of the protein monomer (93 kDa and 70 kDa, respectively) and dimer (170 kDa and 140 kDa) can be explained by the elongated shape of Trigger factor (Hoffmann *et al*., 2010) and the resulting problems in the MW calculations which are based on an assumption that the protein is an ideal sphere. From the peak areas, it was estimated that 42% of total protein was found in the putative microaggregate form, with respective proportions of dimeric and monomeric forms equal to 23% and 35%.

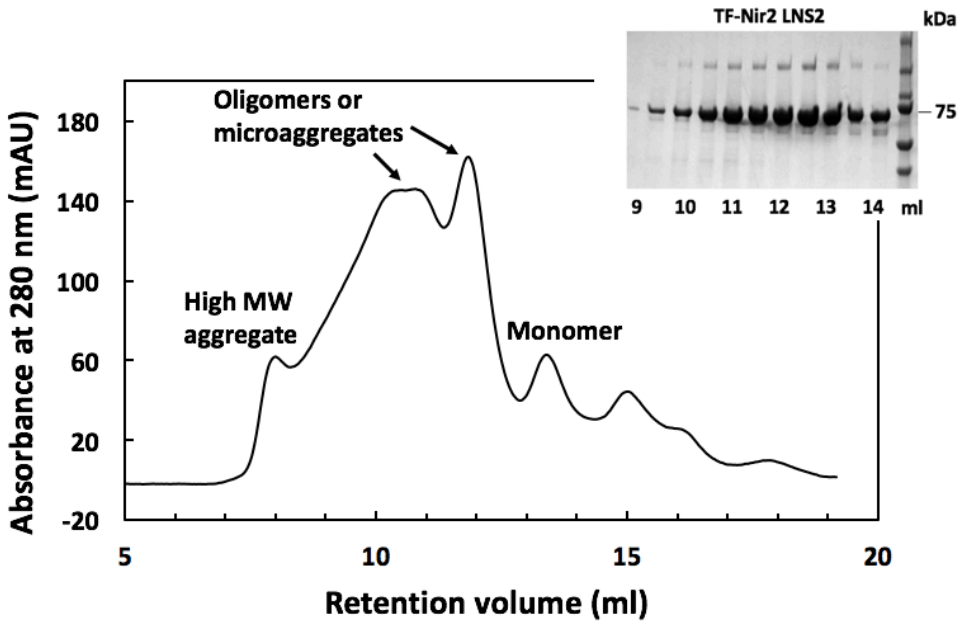



**Figure 3-9. SEC trace of $Ni^{2+}$-affinity purified TF fusion of Nir2 LNS2-S.** Complex elution profile is observed. Assumed contents of the elution peaks are displayed above or near the peaks. Polyacrylamide gel containing 9-14 ml elution fractions is displayed in the upper right corner.

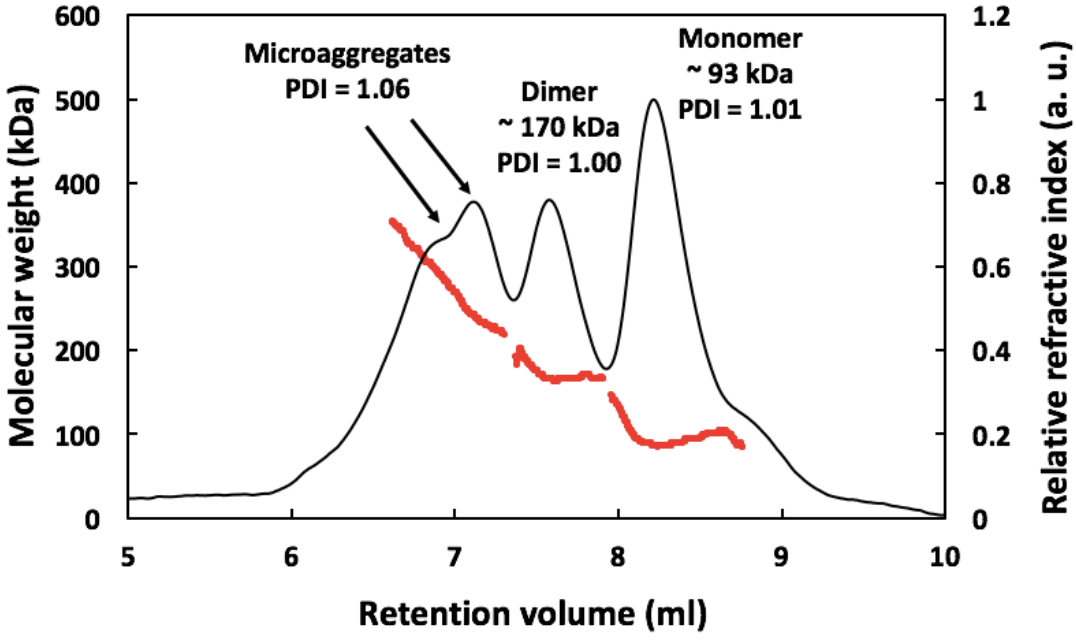



**Figure 3-10. Tandem SEC-multi-angle light scattering (MALS) analysis of $Ni^{2+}$-affinity purified TF Nir2 LNS2-S fusion.** Relative refractive index is plotted in conjunction with calculated molecular weight (MW). The peaks corresponding to the putative monomeric, dimeric and microaggregate forms of the protein are labelled. Average calculated MW and PDI (polydispersity index) of the peak are displayed. SEC-MALS allowed more accurate characterisation of the protein forms than SEC alone (see Fig 3-9). PDI = Mw/Mn.

### 3.1.7.2 MBP fusions of RdgB LNS2 domain

A slightly different picture was observed with MBP RdgB LNS2-L (Fig 3-11). Although a number of distinct peaks were found in the SEC UV trace, from SDS-PAGE analysis it became clear the fusion protein eluted mostly in the peak at 10.2 ml. The peaks 7.3 ml and 8.3 ml were assumed to correspond to protein aggregates due to their likely high MW and their proximity to the void volume (6.8 ml). The peaks eluting after 11 ml corresponded to the co-purifying *E. coli* proteins and fusion protein degradation products. The columns used for SEC purification of MBP RdgB LNS2-L and TF Nir2 LNS2-L were not identical, and hence a direct elution profile comparison cannot be made.

In order to determine the molecular weight of the most abundant MBP RdgB LNS2-L species more accurately, the SEC fractions collected during the elution of the 10.2 ml peak were analysed by MALS. From MALS data (Fig 3-12), the MW of the LNS2 domain MBP fusion species was found to be ~ 402 kDa, which is close to the theoretical MW of the protein hexamer (420 kDa). Therefore, it was assumed that the MBP RdgB LNS2-L is mostly found in a hexameric form after affinity purification. As MBP is a monomeric protein, and does not produce oligomers in solution, it was assumed that the observed hexamer is formed due to the presence of the LNS2 domain in the fusion protein. It was, however, unclear whether the hexamer was a functional form of the fusion protein, a soluble microaggregate or a micelle-like structure formed by the fusion of the soluble MBP and relatively hydrophobic LNS2 domain. The column used for SEC-MALS anaylsis of the MBP fusion of Nir2 LNS2 was identical to that with the TF fusion of Nir2 LNS2.

SEC-MALS was carried out by Mrs June Southall (University of Glasgow).

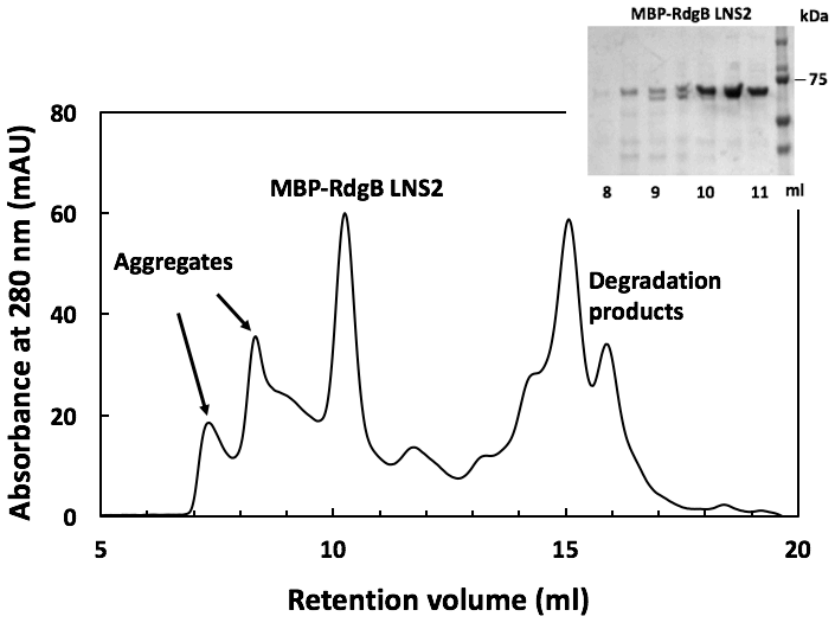



**Figure 3-11. SEC trace of $Ni^{2+}$-affinity purified MBP fusion of RdgB LNS2-L.** Assumed contents of the elution peaks are displayed above or near the peaks. Polyacrylamide gel containing 8-11 ml elution fractions is displayed in the upper right corner.

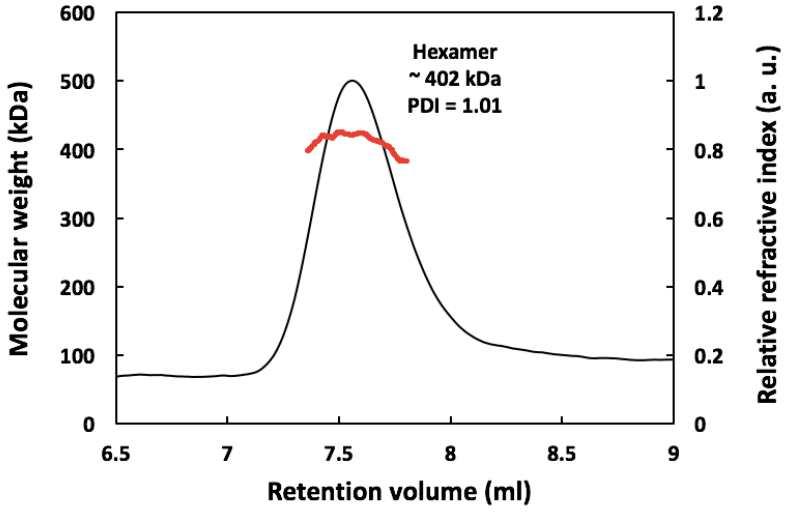



**Figure 3-12. Multi-angle light scattering (MALS) analysis of SEC-purified MBP RdgB LNS2-L fusion.** Relative refractive index is plotted in conjunction with calculated molecular weight (MW). The observed peak likely corresponds to a protein hexamer. Average calculated MW and PDI (polydispersity index) of the peak are displayed. PDI = Mw/Mn.

### 3.1.8 Removal of the solubility tag of TF Nir2 LNS2-S and RdgB LNS2-S fusion

As the TF fusions of the LNS2 domain contained a human rhinovirus (HRV) 3C protease recognition site between the tag and the protein, an attempt was made to proteolytically cleave the tag in order to produce soluble, untagged LNS2 domain fragments. SEC fractions containing the putative TF fusion dimer peak were used for cleavage. About ~50% of the protein was cleaved, as evident from the relative protein bands intensity on

the SDS-PAGE (Fig 3-13). It is not clear why the cleavage has not proceeded to completion. However, it could be an indication of sample heterogeneity and could suggest that the protein is found in at least two states in the sample: a state or multiple states in which HRV 3C is accessible to the protease, and a state or multiple states in which it is inaccessible to the protease. The different states might arise due to the formation of higher order oligomers in the sample or due to potential presence of microaggregates. The incomplete cleavage of the fusion protein is unlikely to be caused by problems with protease's activity, as HRV 3C proteases from two different sources was employed in the reactions. $His_6$-tagged home-produced HRV 3C was used for TF Nir2 LNS2-S cleavage, and commercial GST-tagged HRV 3C PreScission protease (GE Healthcare) was used for TF RdgB LNS2-S cleavage.

Further, although the cleaved LNS2 domain appeared at the correct MW in the SDS-PAGE analysis, it was not possible to separate the cleaved fragment from the fusion protein by SEC. The fusion protein and the cleaved fragments eluted together, which could suggest that the fragment either forms a tight complex with the fusion protein, or, more likely, is aggregated as a result of cleavage, as the LNS2 domain has proved to be unstable in solution in the previous experiments. It should be noted that the cleavage products of the two fusions were purified using different types of SEC column, and therefore the elution profiles are slightly different to each other.

Due to the fact that it was impossible to separate the cleaved LNS2 domains from the fusion proteins, full-length fusions of LNS2 were employed in the experiments in Chapter 4 for characterisation of the ligand binding properties of the LNS2 domain.

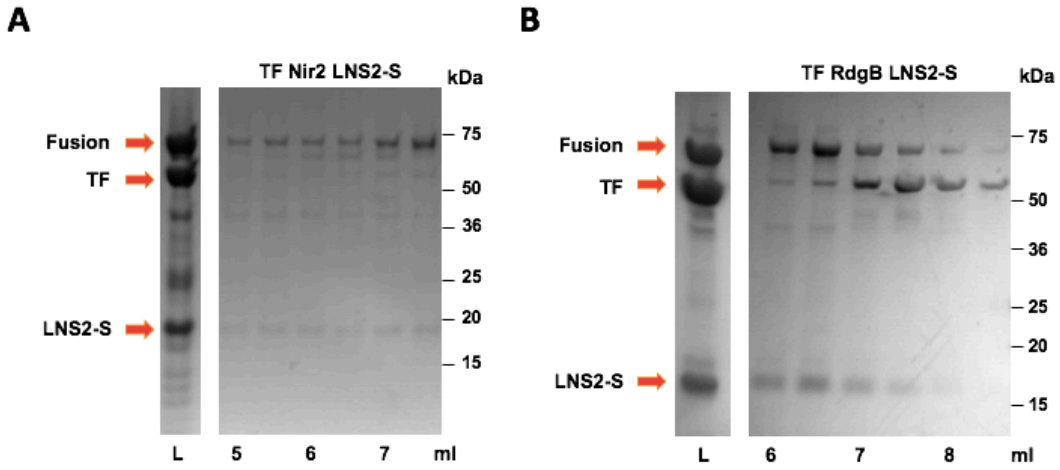



**Figure 3-13. Proteolytic cleavage of TF fusions of Nir2 and RdgB LNS2-S by HRV 3C protease.** Coomassie® protein staining. **A.** SDS-PAGE analysis of TF Nir2 LNS2-S cleavage. SEC fractions (5-7.5 ml) and the protein sample loaded onto the column (L) are shown. LNS2-S fragment elutes together with the fusion protein. **B.** SDS-PAGE analysis of TF RdgB LNS2-S cleavage. SEC fractions (6-8.5 ml) and the protein sample loaded onto the column (L) are shown. The LNS2-S fragment elutes together with the fusion protein. Different columns were used for purification of the two fusions, and hence the elution volumes are not directly comparable.

## 3.2 Na-FAR-1

### 3.2.1 Expression and purification

As mentioned above, previously established for expression and purification of Na-FAR-1 were used. Protein was expressed in *E. coli* from a recombinant pET-11d plasmid produced by Florencia M. Rey-Burusco (Universidad Nacional de La Plata, Argentina). The plasmid encoded a $His_6$-tagged Na-FAR-1 without its N-terminal 14 amino acid secretion signal peptide, as previously described (Rey-Burusco *et al*., 2015). Na-FAR-1 expression was easily repeated here. Na-FAR-1 has demonstrated a high yield of soluble protein (~ 25 mg/L of LB) after $Ni^{2+}$-affinity chromatography. The protein was sufficiently pure after the affinity chromatography and did not require additional purification by SEC, as evident from SDS-PAGE analysis (Fig 3-14A).

### 3.2.2 Removal of co-purifying *E. coli* lipids from recombinant Na-FAR-1

It has been previously demonstrated that recombinant Na-FAR-1 binds native *E. coli* lipid during expression, including a range of fatty acids and phospholipids (Rey-Burusco *et al*., 2015). Hence, further purification step was necessary to remove the co-purifying lipids from the protein. This was achieved by reverse-phase high pressure liquid chromatography (RP-HPLC) using acetonitrile gradient elution. In RP-HPLC, lipids bind the hydrophobic

stationary phase more tightly than the protein, and thus elute at different points in the gradient, allowing the separation of the protein from the lipids.

In order to determine whether removal of the lipids from the protein was successful, lipids were extracted from the 3 mg of HPLC-purified Na-FAR-1 and from 3 mg of Na-FAR-1 that was not purified by HPLC. The lipid extracts were used for thin-layer chromatography (TLC), and the results were compared (Fig 3-14B). From TLC, it was apparent that most of the lipids were successfully removed from the protein by RP-HPLC. Intense lipid spots were observed in the lane containing the extract of Na-FAR-1 not subjected to RP-HPLC, whereas only weak spots were observed in the lane that contained lipids from HPLC-purified Na-FAR-1. From densitometry analysis, it was estimated that ~ 94% of bound lipids were removed from the protein. As the protein was eluted in a mixture of acetonitrile and water from the HPLC, it was lyophilised to remove the solvents. The lipid-free protein was then redissolved in a suitable buffer and used for the experiments described in the chapters that follow.

It should be noted, however, that RP-HPLC purification of the Na-FAR-1 batch used for crystallisation with oleic acid (see Chapter 5) was less successful, and only ~ 82% of lipids were removed from Na-FAR-1, as estimated by TLC and densitometry. This was assumed to be due to an issue with the HPLC instrument experienced during purification.

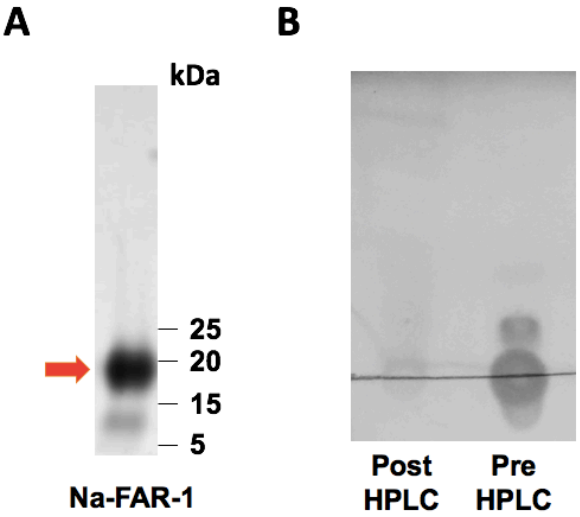



**Figure 3-14. Purification of Na-FAR-1.** **A.** SDS-PAGE analysis of $Ni^{2+}$-affinity purified Na-FAR-1. Recombinant protein band is marked by a red arrow. Coomassie® protein staining. **B.** TLC analysis of lipid extracts from RP-HPLC-purified Na-FAR-1 (Post HPLC lane) and Na-FAR-1 before HPLC purification (Pre HPLC lane). Lipids appear as dark spots on the silica plate. Approximately ~ 94% of lipids were removed by RP-HPLC.

## 3.3 Conclusions

### 3.3.1 The LNS2 domain of Nir1-3 and RdgB

Recombinant protein fragments containing the LNS2 domain of the Nir proteins and RdgB were produced in *E. coli* after identification of the optimal expression constructs in the

high-throughput expression screening. LNS2 domain constructs containing only a $His_6$-tag exhibited very low yields of soluble protein, and were found to be predominantly targeted to inclusion bodies when expressed in *E. coli*. Furthermore, the short $His_6$-tagged constructs purified by affinity chromatography were found to be unstable and prone to aggregation. In contrast, LNS2 domain fusions with > 40 kDa solubility tags have displayed better yields of soluble protein compared to the constructs without a large fusion partner. Of these, the TF fusion of Nir2 LNS2 and the MBP fusion of RdgB LNS2 demonstrated the highest yields of soluble protein in large-scale *E. coli* expression.

Characterisation by SEC has revealed that the TF and MBP fusions of LNS2 have complex elution profiles, indicating heterogeneity of the affinity-purified protein samples. TF Nir2 LNS2-S appeared to be found in several oligomeric states, with ~ 60% of the protein found in putative monomeric and dimeric forms as indicated by SEC-MALS analysis. In contrast, MBP RdgB LNS2-L was mostly found in a putative hexameric state. It is unknown whether the LNS2 domain oligomerises *in vitro* or *in vivo*, and therefore it was not clear whether the observed oligomers are formed by the folded protein or whether they correspond to soluble protein microaggregates containing unfolded or partially unfolded protein. However, as discussed in Chapter 4, oligomers of TF and MBP LNS2 domain fusions demonstrated PA binding, which suggests that they retain at least partial functionality.

### 3.3.2 Na-FAR-1

Na-FAR-1 is a soluble and stable protein, which was expressed and purified from *E. coli* without the need for expression screening or optimisation of expression and purification conditions. The protein was stripped of the ligands that have co-purified from *E. coli* by reverse-phase HPLC and used for the ligand-binding and structural characterisation.

# 4. Characterisation of the recombinant LNS2 domain of Nir2 and the interaction between the LNS2 domain and phosphatidic acid

## 4.1 Introduction

To obtain a deeper knowledge of Nir2 function in cellular homeostasis and disease, it is desirable to elucidate the molecular mechanism of the interaction between the LNS2 domain of Nir2 and phosphatidic acid (PA), which regulates Nir2 localisation in the cell. In this study, several biochemical and biophysical methods were employed for the *in vitro* characterisation of the LNS2-PA interaction.

Most experiments were performed using the Trigger factor (TF) fusion of Nir2 LNS2, as it had demonstrated the highest yield of soluble protein out of the constructs tested, as discussed in Chapter 3. Various forms of PA were used in the investigation, depending on the method employed for characterisation of the LNS2-PA binding. Where possible, lipid vesicles containing long-chain PA species were used as model membranes. In some instances, free short-chain PA species were used instead of the membrane-embedded long-chain PA species due to their high solubility in aqueous buffers.

The study has yielded several important results. First, it was clearly demonstrated that the recombinant TF and MBP fusions of the Nir2 LNS2 produced in *E. coli* retain their PA binding functionality *in vitro*. Secondly, by employing the recombinant protein, it was possible to obtain important insights into the mechanism of PA binding by LNS2. Thirdly, two potential ways of medium-to-high throughput screening for inhibitors of the PA and LNS2 interaction were identified.

Unfortunately, the characterisation of the three-dimensional structure of the LNS2 domain could not be performed, as the TF and MBP fusion of the LNS2 domain failed to crystallise and proved too large to be readily investigated by biomolecular NMR spectroscopy. However, the secondary structure of the LNS2 domain of Nir1 was characterised by circular dichroism spectroscopy.

Below, the results of the characterisation of PA binding will be discussed in detail. A brief introduction to each technique used for analysis of the LNS2-PA binding will be given. At the end of the chapter, conclusions will be drawn from the experiments. A proposed

mechanism for LNS2 and PA interaction at the membrane will be discussed and two strategies for the screening of PA interaction inhibitors will be presented. Future experiments which could be performed to test the proposed hypotheses and investigate the mechanism of LNS2-PA binding in more detail will be also be discussed.

## 4.2 TF and MBP fusions of Nir2 LNS2 domain bind phosphatidic acid

### 4.2.1 Lipid binding assays

Lipids and lipid binding proteins are increasingly coming to prominence as crucial players in many normal and pathological cellular processes, including cell signalling, growth, motility and differentiation. Consequently, a variety of biochemical assays have been developed or adapted for quick analysis of protein-lipid interactions (Zhao and Lappalainen, 2012). Two of the most-widely used assays are the lipid overlay assay and the liposome sedimentation assay. Due to their ease of use, these two assays were employed for analysis of the LNS2 and PA interaction in the first instance.

The lipid overlay assay is a type of immunoblot assay that is, in principle, similar to more widely known enzyme-linked immunosorbent assay and Western blotting. In the lipid overlay assay, lipids of interest are immobilised onto a piece of nitrocellulose membrane, after which the membrane is blocked with a solution of fatty acid free bovine serum albumin or fat free milk and incubated with a lipid binding protein of interest. Next, the membrane is washed and incubated with an antibody specific for the protein. The antibody is typically conjugated to an enzyme, such as alkaline phosphatase or horseradish peroxidase, that allows detection of the signal via a reaction with a specific substrate that produces a coloured product or light. Membranes containing sets of immobilised membrane phospholipids are available commercially. Here, they were produced as required for each experiment. The principle of the assay is visually summarized in Fig 4-1A.

In contrast to the lipid overlay assay, the lipid co-sedimentation assay (LSA) does not use immobilised lipid layers. Instead, LSA is based on detection of lipid binding to lipid vesicles (liposomes). Several types of liposomes can be used for characterisation of protein-lipid binding, which vary by size and the number of lipid bilayers they contain. Most commonly, large unilamellar vesicles (LUVs) or multilamellar vesicles (MLVs) are

used in LSA as their curvature resembles the curvature of natural membranes (Zhao and Lappalainen, 2012).

In LSA, liposomes are mixed with the solution of a lipid binding protein, and the mixture is incubated to allow protein-liposome binding to occur. The mixture is then centrifuged to pellet the liposomes, and the pellet and supernatant fractions are analysed by SDS-PAGE. The presence of the protein in the pellet fraction is typically a sign of interaction between the protein and the liposomes, but the protein can also be pelleted with the liposomes due to a non-specific interaction with the vesicles or aggregation. Thus, an appropriate control should be carried out to determine whether the binding observed in the LSA is specific. In this study, liposomes containing only phosphatidylcholine (PC) were as a control for non-specific protein binding, as PC does not interact with the LNS2 domain of Nir2 (Kim *et al.*, 2013). The experimental procedure of the assay is visually summarized in Fig 4-1B.

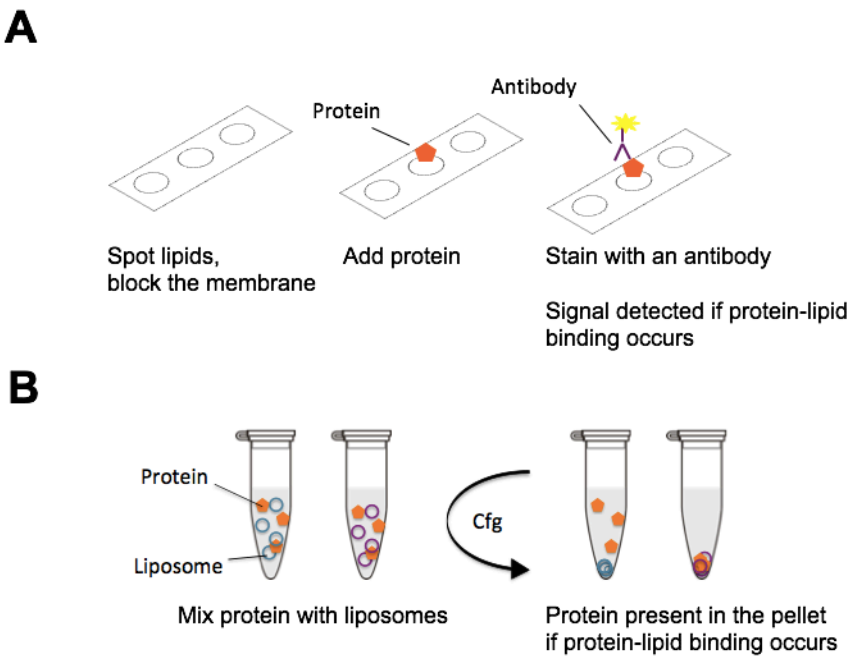



**Figure 4-1. Lipid binding assays. A.** Visual summary of the lipid overlay assay. Protein-lipid interaction is detected by immunostaining. **B.** Visual summary of the liposome co-sedimentation assay. Protein-lipid interaction is detected by analysis of the supernatant and pellet fractions.

### 4.2.2 Recombinant Nir2 and RdgB LNS2 domains interact with PA

Due to the complex elution profile of the TF fusion of Nir2 LNS2 observed during characterisation by SEC-MALS, it was assumed that at least part of the recombinant protein may be in the aggregated or misfolded state. Therefore, in the first step in the characterisation of the PA interaction with the LNS2 domain it had to be confirmed that

the recombinant fusion of the LNS2 domain retains its PA binding function after purification.

Analysis of lipid binding by both lipid overlay and liposome co-sedimentation assays has demonstrated that the recombinant LNS2 domain of Nir2 is able to bind both the natural PA derived from the egg yolk extract, as was reported previously (Chang and Liou, 2015; Kim *et al*., 2013), and synthetic PA species. Furthermore, binding between the MBP fusion of the RdgB LNS2 and natural PA was demonstrated. Synthetic PA species were not used in the RdgB LNS2 binding experiments.

#### 4.2.2.1 Analysis of PA binding by lipid overlay assay

Lipid overlay assay was carried out with the TF fusion of Nir2 LNS2-S and the MBP fusion of Nir2 LNS2-L purified from *E. coli*, and the C-terminally $His_6$-tagged Nir2 LNS2-L purified from insect cells (Fig 4-2). Prior to the assay, the TF and MBP fusions of the LNS2 domain were purified by SEC. Each of the TF Nir2 LNS2 fractions corresponding to putative oligomeric forms of the fusion protein was analysed separately. The MBP fusion of RdgB LNS2-L was used in the predominant putative hexameric form (see Chapter 3). The C-terminally $His_6$-tagged Nir2 LNS2-L was used directly after $Ni^{2+}$-affinity chromatography, without further purification. In addition, a negative control was carried out using recombinant $His_6$-tagged TF purified from *E. coli*. This was done to ensure that any binding observed in the assay was not due to a non-specific interaction of TF with the lipids.

The recombinant LNS2 domain bound to PA but not to PC in all assays. This was evident from the fact that the staining signal was observed only from the PA dot on the lipid blots (Fig 4-2). Interestingly, binding was also seen with the SEC fractions of the recombinant TF Nir2 LNS2 that were believed to contain the protein in the aggregated form (8.5 – 9.0 ml and 9.5 – 10 ml fractions). This indicates that each of the TF Nir2 LNS2 SEC fractions used in the assay contained at least a portion of the recombinant protein that is able to specifically interact with PA. However, it is unlikely that all recombinant protein present in the SEC fractions possesses PA binding functionality. For instance, the observed binding of the putative microaggregates to PA might be explained by the presence of partially folded protein in the microaggregate species. Consistently, the MBP fusion of RdgB LNS2-L and the $His_6$-tagged Nir2 LNS2-L fragment also demonstrated specific binding to PA. Reassuringly, no TF binding to the phospholipids was detected, suggesting

that the binding observed in the assays was occurring specifically due to presence of the LNS2 domain in the TF fusion proteins. MBP is not expected to bind to the phospholipids.

#### 4.2.2.2 Analysis of PA binding by liposome co-sedimentation assay

The results of the LSA were in agreement with the results of the lipid overlay assay. MLVs were employed for the qualitative assay due to the ease of their preparation. MLVs contained either egg yolk PA and PC mixed in 1:2 mass ratio or only PC. Only binding of the TF fusion of Nir2 LNS2 domain to PA was assessed by using LSA.

Protein was clearly found in the pellet in the presence of the PA-containing MLVs (Fig 4-3A). In the representative example shown in Fig 4-3A, no protein band was observed in the PC-only MLV pellet lane. However, the band in the PC-only pellet was occasionally present in the gels after SDS-PAGE. This suggests that either a small degree of protein precipitation was occurring in the sample in the presence of lipid vesicles, which is not uncommon in LSA, or that the protein was interacting with PC in the vesicles, which is unlikely, as it was not observed in the lipid overlay assays. Convincingly, as the intensity of protein band in the PA-containing liposome pellet was always greater than that of the band in the PC-only liposome pellet, it was concluded that the protein was preferentially binding to PA-containing MLVs, which is indicative of a specific protein-PA interaction.

In addition, a semi-quantitative assessment of the apparent dissociation constant ($K_{app}$) of Nir2 LNS2 binding to PA was carried out. As quantification of the molar lipid concentrations was required for $K_{app}$ determination, pure synthetic dioleoyl PA (DOPA or di-18:1 *cis*-9 PA) and dioleoyl PC (DOPC or di-18:1 *cis*-9 PC) species were used in the assay rather than the natural PA and PC extracted from egg yolk, as those comprised a mixture of PA and PC with varying acyl chain lengths and MWs. Also, LUVs were employed in the assay instead of MLVs, as LUVs contain a single membrane layer, allowing estimation of the PA concentration accessible to the protein to be achieved (see below). LUVs were prepared by extrusion through a semi-permeable membrane. The semi-quantitative LSA was carried out using high-speed centrifugation (150,000 x g) to ensure that all potential protein-LUV complexes were pelleted in the experiment. For the assay, PA-containing LUVs were mixed with the TF Nir2 LNS2 fusion to obtain a range of protein-LUV mixtures with varying total lipid concentrations (2 mM, 1 mM, 0.5 mM, 0.3 mM, 0.1 mM and 0.05 mM). Protein concentration was kept constant in the assay at 2 μM.

After the LSA fractions were analysed by SDS-PAGE, it was observed that the intensity of the protein band in the pellet increased with the concentration of liposomes in the sample, as was expected. The intensities of protein bands in the supernatant and pellet fractions were analysed by densitometry using Fiji (Schindelin *et al*., 2012). The percentage of the bound protein was calculated by dividing the integrated intensity of the pellet protein band by the sum of the integrated intensities of the protein bands in the pellet and the supernatant fractions. The maximum percentage of the protein bound to liposomes was observed with 2 mM total lipid and was equal to 33%. In the presence of 1 mM, 0.5 mM, 0.3 mM, 0.1 mM and 0.05 mM total lipid, 22%, 13%, 10% and 0% of protein was found in the pellet, respectively.

In order to estimate the $K_{app}$ for PA binding, the concentration of DOPA accessible to the LNS2 domain in the sample was calculated assuming that DOPA is evenly split between the two layers of the LUV membrane and that the protein is only able to interact with the lipids in the outer membrane layer. The percentage of the protein bound was plotted as function of the concentration of accessible DOPA in the sample, and the binding data were fitted to Equation 4-1 describing one-site binding:

**Equation 4-1.**

$$[PL] = \frac{[PL]_{max} \times [L]_{acc}}{K_{app} + [L]_{acc}}$$

where [PL] is the percentage of the protein bound, $[PL]_{max}$ is the maximum specific binding, $[L]_{acc}$ is the concentration of DOPA accessible to the protein and $K_{app}$ is the apparent dissociation constant. Protein-membrane binding affinities are often reported in terms of $K_{app}$ values instead of the real dissociation constant $K_d$ values. This is because protein-membrane binding is complex and involves multiple types of interactions between the amino acids in the protein and the membrane lipids which are difficult to account for in the calculation of the real $K_d$ (Zhao and Lappalainen, 2012). Based on the data obtained in the experiments, the $K_{app}$ was calculated to be equal to ~ 0.5 mM.

It should be noted that the $K_{app}$ value reported here should only be treated as an estimate. This is because the LSA is in principle prone to underestimation of $K_{app}$ because of the possibility of protein aggregation occurring in the sample that could interfere with the binding analysis. Indeed, as mentioned above, protein precipitation was observed in the presence of PC-only vesicles, which was difficult to account for in the semi-quantitative experiment, and therefore it is likely that the $K_{app}$ value reported here is lower than the real value. Also, due to time constraints, replicates were carried out only with LUVs containing

150 μM accessible PA (1 mM total lipid), which makes the precision of the estimation difficult to assess. The bound protein percentages in the presence of 150 μM PA were, however, relatively consistent, with the mean value equal to 18 ± 4% (n = 4; ± s.e.m.).

In order to obtain a more accurate $K_{app}$ value, a liposome co-flotation assay can be carried out, in which the protein bound to the liposomes is separated from the free protein by centrifugation in a density gradient. Soluble protein bound to the liposomes will also be separated from the aggregated protein, which will sediment to the bottom of the gradient due to its high density. This allows protein aggregation to be accounted for during $K_{app}$ estimation. Nevertheless, the components used for gradient preparation in the co-flotation assay can themselves cause protein aggregation, and hence it might not be suitable for use with Nir2 LNS2.

To summarise, $K_{app}$ estimation by LSA allowed the affinity of DOPA binding by the LNS2 domain to be placed into at least high micromolar range. Since the affinity of LNS2-PA binding was not characterised in the previous studies, this is the first glimpse into the *in vitro* affinity of PA binding by the LNS2 domain.

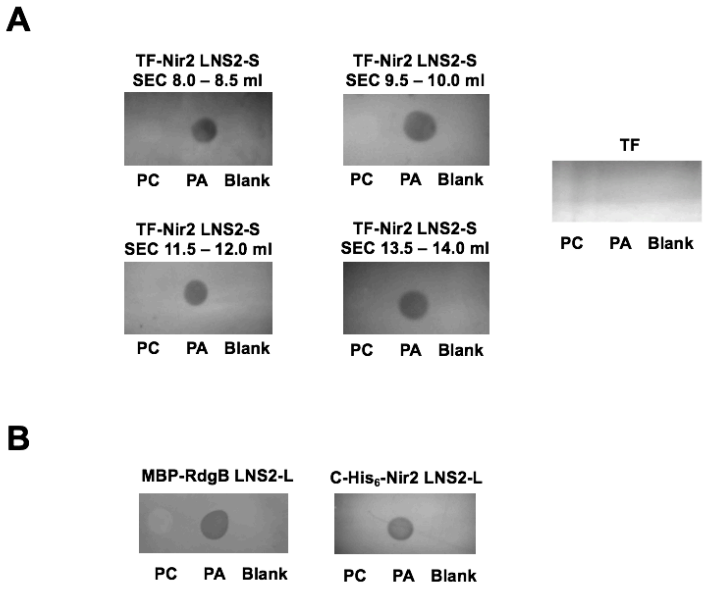



**Figure 4-2. Lipid overlay assay with the recombinant Nir2 and RdgB LNS2 domains.** Representative assay results are shown. Detection was carried out by incubation with an anti-$His_6$-tag antibody. **A**. Lipid overlay assays using four size-exclusion chromatography (SEC) fractions of the TF fusion of Nir2 LNS2 domain. The fraction elution volumes are given above each image. The fractions contained the following putative oligomeric species of the fusion protein: microaggregates/multimers (8.0 – 8.5 ml and 9.5 – 10 ml), dimer (11.5 – 12.0 ml) or monomer (13.5 – 14.0 ml). Binding to PA but not to PC was observed in all assays. The assay with TF was carried out as a negative control. TF did not bind to any phospholipids. **B.** Lipid overlay assays using the MBP fusion of RdgB LNS2-L fragment and C-terminally $His_6$-tagged Nir2 LNS2-L. Both proteins demonstrated binding to PA but not to PC.

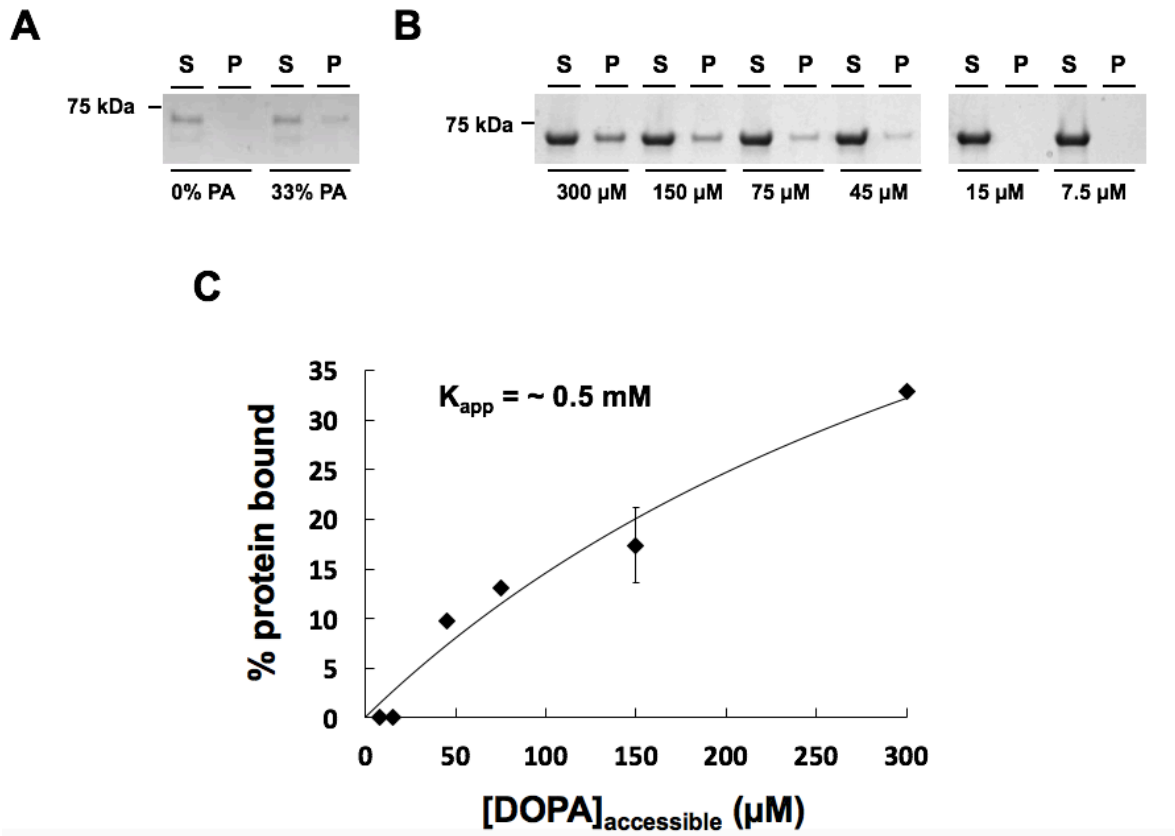



**Figure 4-3. Liposome co-sedimentation assay (LSA) with the recombinant TF fusion of Nir2 LNS2 domain. A.** LSA with MLVs prepared from natural PA and PC from egg yolk. Protein interaction is observed with the MLVs containing 33% PA but not to the MLVs containing 0% PA. S = supernatant, P = pellet. **B.** Semi-quantitative LSA with LUVs consisting of 30% DOPA (di-18:1 *cis*-9 PA) and 70% DOPC (di-18:1 *cis*-9 PC). Protein concentration was kept constant in the assay, while LUV concentration was varied. Concentration of accessible DOPA is displayed below the images. The amount of the protein bound to LUVs increased with increasing concentration of the vesicles. S = supernatant, P = pellet. **C.** Binding isotherm for the semi-quantitative LSA. Data were fitted to a one-site binding equation by least squares method. Apparent dissociation constant ($K_{app}$) was found to be equal to ~ 0.5 mM. n = 1 except for the point at $[DOPA]_{accessible}$ = 150 µM, for which n = 4 independent experiments (the data point represents the mean, error bars represent ± s.e.m.).

## 4.3 Analysis of Nir2 LNS2 domain binding to PA by ligand-observed NMR spectroscopy

### 4.3.1 Ligand-observed NMR spectroscopy

NMR spectroscopy (or simply NMR) is a powerful and versatile technique which can be applied to structural and functional characterisation of proteins and protein-ligand interactions. By using NMR, the signals of both the protein and the ligand can be detected in the protein-ligand complexes, and thus protein-ligand interactions can be investigated

from two angles: by observing the resonance signals of the ligand (ligand-observed NMR) or by observing the resonance signals of the protein (protein-observed NMR).

#### 4.3.1.1 Limitations of protein-observed NMR spectroscopy in protein-ligand binding studies

In protein-observed NMR, chemical shifts of protein atoms are usually monitored. From the chemical shifts, one can obtain a great wealth of information about the protein structure, conformation and folding state, as they report on the immediate chemical environment of the corresponding nuclei. Since proteins are the main targets of biomolecular NMR investigations, a wide variety of NMR experiments have been developed that allow accurate assignment of the resonances in the protein spectra to specific residues and atoms in the protein. One can then use these assignments to determine the structure of the protein in complex with the ligand, map ligand binding sites in the protein or follow conformational changes in the protein structure upon ligand binding.

There are, however, important limitations associated with protein-observed NMR. As mentioned above, in order to obtain an interpretable spectrum, the protein sample needs to be concentrated (normally 0.1-2 mM) and homogenous. This can be challenging to achieve if the protein is poorly expressed in heterologous hosts or is prone to aggregation in solution. Further, due to the complexity of the protein NMR spectra, one needs to observe at least both $^{1}$H and $^{15}$N nuclei, or ideally $^{1}$H, $^{15}$N and $^{13}$C nuclei to carry out accurate assignments of the resonance signals to protein atoms. Since the natural abundance of magnetically active spin-1/2 $^{13}$C and $^{15}$N nuclei is low (1.1% and 0.36%, respectively), the protein typically needs to be uniformly isotopically labelled, which is achieved by expression in a minimal medium supplemented with nutrients synthetically enriched in the relevant isotopes. As isotope-labelled nutrients are relatively expensive and protein yields in minimal media are generally lower that in normal rich media, this can make the production of protein sample prohibitively expensive, especially in the case of difficult-to-express, unstable proteins.

Another limitation of protein-observed NMR spectroscopy lies in protein size. Protein size limitations arise due to effects of signal broadening and crowding in the spectra of large proteins. Although modern methods allow investigation of very large monomeric and multimeric protein systems (100 kDa - 1.1 MDa) (Fiaux *et al*., 2002; Gelis *et al*., 2007; Mainz *et al*., 2013; Sprangers and Kay, 2007), this remains a challenging, non-trivial and

labour-intensive task. Typically, best results are obtained with proteins with molecular weight smaller or equal to 25-35 kDa.

For a more detailed description of protein-observed NMR spectroscopy and the theory of NMR please see Chapter 6.

#### 4.3.1.2 Ligand-observed NMR spectroscopy

In contrast to protein-observed NMR, in ligand-observed NMR protein signals are not the objects of investigation. Thus, ligand-observed NMR does not have any limitations associated with the protein size or spectral crowding, as long as the ligand is a small molecule. Indeed, the use of larger proteins is often beneficial in ligand-observed NMR, as it results in higher sensitivity in certain types of experiments. Importantly, in ligand-observed NMR, proteins do not require isotopic labelling, and experiments can be successfully carried out even with low protein concentrations (~ 0.1-100 μM).

Ligand-observed NMR spectroscopy detects the effects of protein-ligand interaction on ligand signals. Although ligand-observed NMR does not provide any information on protein structure in protein-ligand complexes, it can be used as a tool to qualitatively or quantitatively characterise ligand binding. Due to its low cost, minimal protein sample requirements and the short experimental time required to produce an interpretable spectrum, ligand-observed NMR is widely employed in pharmaceutical industry for drug discovery (Pellecchia *et al*., 2008; Renaud and Delsuc, 2009; Śledź *et al*., 2012; Unione *et al*., 2014; Zhang *et al*., 2015). In the fragment-based drug screening, the use of ligand-observed NMR is especially advantageous, as it can be used to detect a range of weak (~ 1 mM) to moderately strong (~ 0.1 μM) binding events. The advantages of ligand-observed NMR also apply to low-throughput analyses of protein-ligand binding, such as in the study described here. Several types of ligand-observed NMR experiments have been developed, which rely on observation of different NMR parameters of ligands, which include $T_1$ and $T_2$ relaxation rates, NOEs and saturation transfer parameters (Meyer and Peters, 2003).

To record $T_2$ relaxation-edited experiments, the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (Carr and Purcell, 1954; Meiboom and Gill, 1958) is typically used. In a CPMG-based experiment, signals arising from the protein or protein-bound ligand nuclei can be distinguished from the signals of the free ligand due to their short $T_2$ relaxation times compared to the those of the free ligand signals. In the sample containing a ligand transiently bound to a protein, a reduction in the intensity of the ligand resonance signals

will be observed due to the averaging of the signals arising from the ligand molecules bound to the protein and the free ligand molecules. Thus, by comparing small molecule signal intensities in the spectra acquired in the absence and in the presence of the protein, one can determine whether a small molecule interacts with the protein in the experiment.

Saturation transfer difference (STD) (Angulo and Nieto, 2011; Mayer and Meyer, 1999) is a technique which is also widely used in ligand-observed NMR. In STD, protein in the presence of a large excess of a small molecule is irradiated with a radiofrequency pulse selective for the protein methyl protons. If the small molecule interacts with the protein, magnetisation is transferred from the protein methyl protons to the small molecule protons via the Nuclear Overhauser effect (NOE). This results in increased intensity of the corresponding small molecule signals in the recorded $^{1}$H spectrum. By comparing the intensity of the small molecule signals in the presence and in the absence of the protein, protein-ligand binding in the sample can be detected and characterised.

#### 4.3.1.3 WaterLOGSY

Water-Ligand Observed by Gradient SpectroscopY (waterLOGSY) (Dalvit *et al*., 2001) is another common ligand-observed NMR experiment. Like STD, it is also based on saturation transfer. However, in waterLOGSY, the bulk water nuclei are saturated with a selective pulse instead of the protein nuclei. From the bulk water, magnetisation is transferred to the protein and the small molecule in solution. When the magnetisation is transferred to the protein, it can spread through the protein via intramolecular NOE. If a small molecule binds to the protein, the magnetisation from the protein can be further transferred to the small molecule through intermolecular NOE. The direct magnetisation transfer from bulk water to the ligand occurs in the fast tumbling regime, whereas the magnetisation transfer from bulk water to the ligand via the protein occurs in the slow tumbling regime. As a consequence, the sign of protein-ligand NOE will be opposite to that of water-ligand NOE. For a ligand that is transiently binding to the protein, these two transfer processes occur simultaneously, however, the magnetisation transferred from the protein will dominate, as protein-ligand NOE build up is faster than water-ligand NOE build up.

Like in STD, the differences in the waterLOGSY magnetisation transfer pathways can be exploited for characterisation of ligand binding. To determine whether a small molecule interacts with a protein, one can record $^{1}$H waterLOGSY spectra of the molecule in the

absence and the presence of the protein. If small molecule is a ligand, in the presence of a protein its resonance signals will either be of opposite sign or of reduced intensity compared to those in the absence of the protein. If there is no binding between the small molecule and the protein, the small molecule signal will be the same independent of the presence of the protein. WaterLOGSY experiments are usually set up in such a way so that the signals of the nuclei interacting with the protein appear as positive in the spectrum, and the signals of non-interacting nuclei as negative.

In addition to qualitative analysis, waterLOGSY can be employed for quantitative analysis of ligand binding (Asencio-Hernández *et al*., 2016; Dalvit *et al*., 2001; Forget *et al*., 2015; Hopkinson *et al*., 2015). Since the waterLOGSY signal intensity depends on the concentration of protein-ligand complexes in the sample, this can be done by monitoring the changes in the intensity of the ligand signals upon ligand titration into a substoichiometric protein solution.

Signal intensity (I) for a ligand proton *i* is given by the following equation (Dalvit *et al*., 2001):

**Equation 4-2.**

$$I \propto [PL]\left(\sum_j \sigma_{ij}^{bound} + \sum_k \sigma_{ik} + \sum_w \sigma_{iw}^{bound}\right) + [L]\left(\sum_j \sigma_{ij}^{free} + \sum_w \sigma_{iw}^{free}\right)$$

where [PL] is the concentration of the ligand bound to the protein, [L] is the concentration of the free ligand, σ is the cross-relaxation rate constant, j is an index of ligand exchangeable protons, k is an index of protein protons next to the ligand and w is an index of water molecules near ligand. [L] is equal to $[L_{tot}] - [PL]$, where $[L_{tot}]$ is the total concentration of the ligand added to the sample.

Due to the dependence of signal intensity on both [L] and [PL], in cases where the dissociation constant $K_d$ is relatively large and the ligand in high excess, the contribution of the second term of the equation will outweigh the contribution of the first term, which will lead to a reduction in the signal intensity. Thus, for an accurate estimate of binding affinity, a correction must be made for the increasing ligand concentration during the titration. In order to do this, a spectrum of the ligand in the absence of the protein can be recorded, and the observed ligand intensity values can be subtracted from those recorded in the presence of the protein. The corrected signal intensities can then be plotted as a function of [L] to construct a binding isotherm for $K_d$ determination. One caveat is that waterLOGSY has been reported to overestimate the $K_d$ due to re-binding of ligands to proteins after saturation transfer (Fielding *et al*., 2005; Huang *et al*., 2017), but the

estimation was found to be adequate when a low protein concentration (~ < 20 μM) and a short mixing time (~ 0.5 s) was used in the experiments (Huang *et al*., 2017).

### 4.3.2 Nir2 LNS2 interacts with short-chain PA and the fatty acyl chains of short-chain PC

Due to the ease of data interpretation, simplicity and good experimental sensitivity (Antanasijevic *et al*., 2014) of waterLOGSY compared to other ligand-observed NMR techniques, it was chosen as a primary method of probing the interaction between the LNS2 domain of Nir2, PA and other ligands. Additionally, CPMG experiments were carried out to confirm the results of waterLOGSY.

In order to carry out the binding characterisation, a suitable PA species had to be identified for use in the experiments. Since the use of lipid vesicles and micelles is not compatible with ligand-observed NMR, short acyl chain dihexanoyl phosphatidic acid (DHPA or di-6:0 PA) and phosphatidylcholine (DHPC or di-6:0 PC) were employed due to their high solubility in the aqueous buffers and millimolar cmc (Marsh, 1990). DHPC was intended to be used as a negative control in these experiments, as the LNS2 domain of Nir2 has not been previously shown to bind to PC.

The spectra of the waterLOGSY experiments are shown in Fig 4-3. Standard $^{1}H$ spectra with water suppression are shown for reference. Changes in the sign and intensity of DHPA fatty acyl tail signals **1**-**4** and the glycerol backbone signals were observed after addition of TF Nir2 LNS2, which was indicative of DHPA interaction with the LNS2 domain. The signals from the two most downfield shifted backbone proton signals were relatively weak likely because of their close proximity to the $H_2O$ signal (not shown). The hydroxyl resonance signal of the phosphate head group of DHPA was not visible due to fast exchange with the solvent. In the control experiment with TF, no positive ligand signals were observed in the presence of the protein, indicating that there was no interaction between DHPA and TF. The positive broad signals observed in the presence of the protein arise from the TF protons. The lack of DHPA interaction with TF suggests that the DHPA was specifically binding to the LNS2 domain in the experiments using the TF fusion of Nir2 LNS2. Importantly, the data showed that all NMR-observable DHPA protons were interacting with the protein.

Interestingly, in the NMR experiments DHPC also displayed binding to the LNS2 domain, which was unexpected. Signals **1**-**4** corresponding to DHPC fatty acyl protons changed

their sign after addition of the LNS2 fusion protein, although the signal intensities did not increase as strongly as was observed in the DHPA binding experiments. For example, the increase in the intensity of the acyl chain methyl signal **1** after protein addition was approximately 2.5-fold smaller in the DHPC experiments than in the DHPA experiments. As the intensity of proton resonances in waterLOGSY spectra is related to the concentration of protein-ligand complexes in the sample, this observation suggests that the binding of LNS2 domain to DHPC is likely to be of weaker affinity than to DHPA. Significantly, no clear change in the sign and intensity of the DHPC glycerol backbone and head group signals (including the very strong phosphocholine methyl signal **5**) was observed after protein addition, indicating that the DHPC head group is not interacting with the protein or that the interaction is too weak to be detected by waterLOGSY.

Next, CPMG experiments were carried out using DHPA or DHPC with the LNS2 domain of Nir2 to confirm the findings of the waterLOGSY experiments. As described in Section 4.3.1.2, CPMG uses a different principle of detecting protein-ligand interaction than waterLOGSY, and hence it can be used to distinguish any possible artefacts in the waterLOGSY spectra from the real binding events. Reassuringly, the findings in the CPMG experiments were in agreement with the waterLOGSY data. In the experiments using DHPA (Fig 4-5A), the intensity of all the DHPA proton resonance signals observed in the spectrum decreased after addition of the TF fusion of Nir2 LNS2, indicating that DHPA was interacting with the protein. In the CPMG experiments with DHPC (Fig 4-5B), the intensity of the DHPC fatty acyl chain resonances **1**-**4** also decreased, suggesting that an interaction was taking place between the acyl chain of DHPC and the protein, as was observed in waterLOGSY.

The percent decrease in the intensity of DHPA and DHPC resonances after addition of the LNS2 domain in CPMG experiments is summarised in Fig 4-5C. In signal multiplets, the intensity of the most intense component was measured. From the chart, it can be seen that the decrease in the intensity of DHPA resonance signals (~ 40-60%) is greater than that of DHPC resonances (~ 15-35%). This suggests that DHPA spends on average more time bound to the protein than DHPC, and hence that DHPA interacts with the protein with higher affinity than DHPC. In both the DHPA and DHPC experiments, the strongest decrease is observed for acyl chain signals **4a** and **4b**, which arise from the methylene groups bonded to the oxygen atom in the two chains. Interestingly, only a slight decrease in intensity (~ 3%) was observed for the DHPC phosphocholine head group signal **5**, which is a ~ 5- to 7.5-fold smaller decrease in intensity compared to the decrease observed for

other DHPC resonances. This might suggest that the chemical environment and/or mobility of the DHPC head group is different to the rest of the molecule in the presence of the protein, which could indicate that the head group is not participating in the interaction with the protein. The apparent small change in the intensity of signal **5** is also consistent with the waterLOGSY experiments, where the same signal did not appear to change its sign on protein addition in contrast to the fatty acyl resonances of DHPC.

From this data, it can be inferred that the interaction of Nir2 LNS2 with DHPC occurs through the acyl chains of DHPC, whereas the phosphocholine head group of DHPC does not participate in the interaction. Also, the interaction between DHPC and the LNS2 domain is likely of a weaker affinity than the one between DHPA and the LNS2 domain. This might suggest that although the LNS2 domain appears to have weak affinity for the fatty acyl tails of both DHPA and DHPC, the presence of negatively charged phosphate head group of DHPA is responsible for the higher affinity of the LNS2-DHPA interaction compared to the LNS2-DHPC interaction.

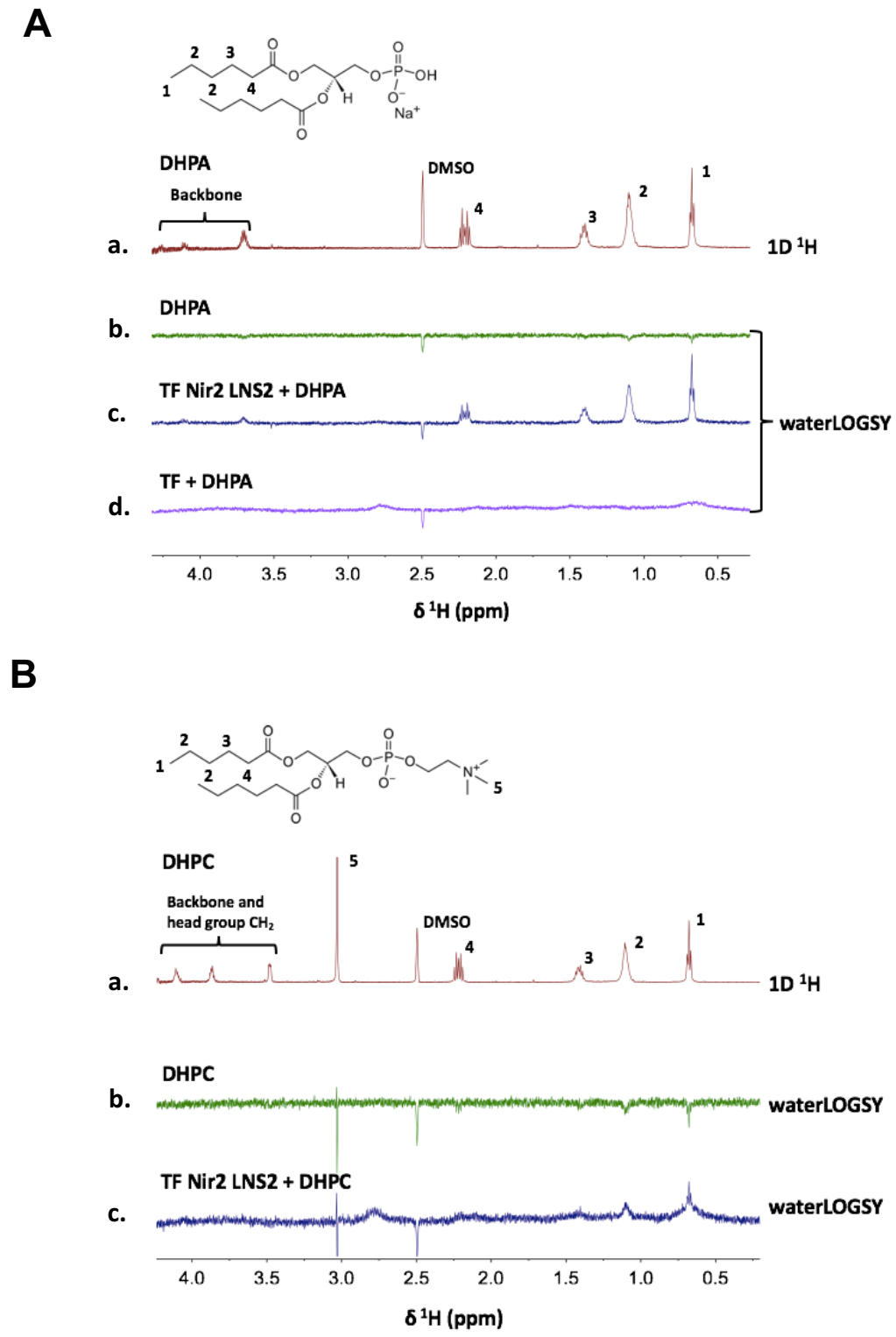



**Figure 4-4. Analysis of DHPA (di-6:0 PA) and DHPC (di-6:0 PC) interaction with the Nir2 LNS2 domain by NMR waterLOGSY.** a. Reference 1D $^1$H spectrum, b. waterLOGSY spectrum in the absence of Nir2 LNS2, c. waterLOGSY spectrum in the presence of TF Nir2 LNS2, d. waterLOGSY spectrum in the presence of trigger factor (TF). Structures of DHPA and DHPC are displayed for reference. DHPA and DHPC protons with corresponding resonance signals are numbered. **A.** DHPA experiments. In waterLOGSY spectra, the sign of all DHPA proton resonances (**1**-**4**) is inverted upon protein addition, indicating that DHPA interacts with the protein. No interaction between TF and DHPA is observed. **B.** DHPC experiments. The sign of acyl proton signals **1**-**4** is inverted upon protein addition, whereas no clear change is observed in the head group methyl signal **5** as well as in the head group and glycerol backbone methylene signals. This indicates an interaction between the fatty acyl tails of the DHPC molecule and the protein, but not between the head group moiety and the protein.

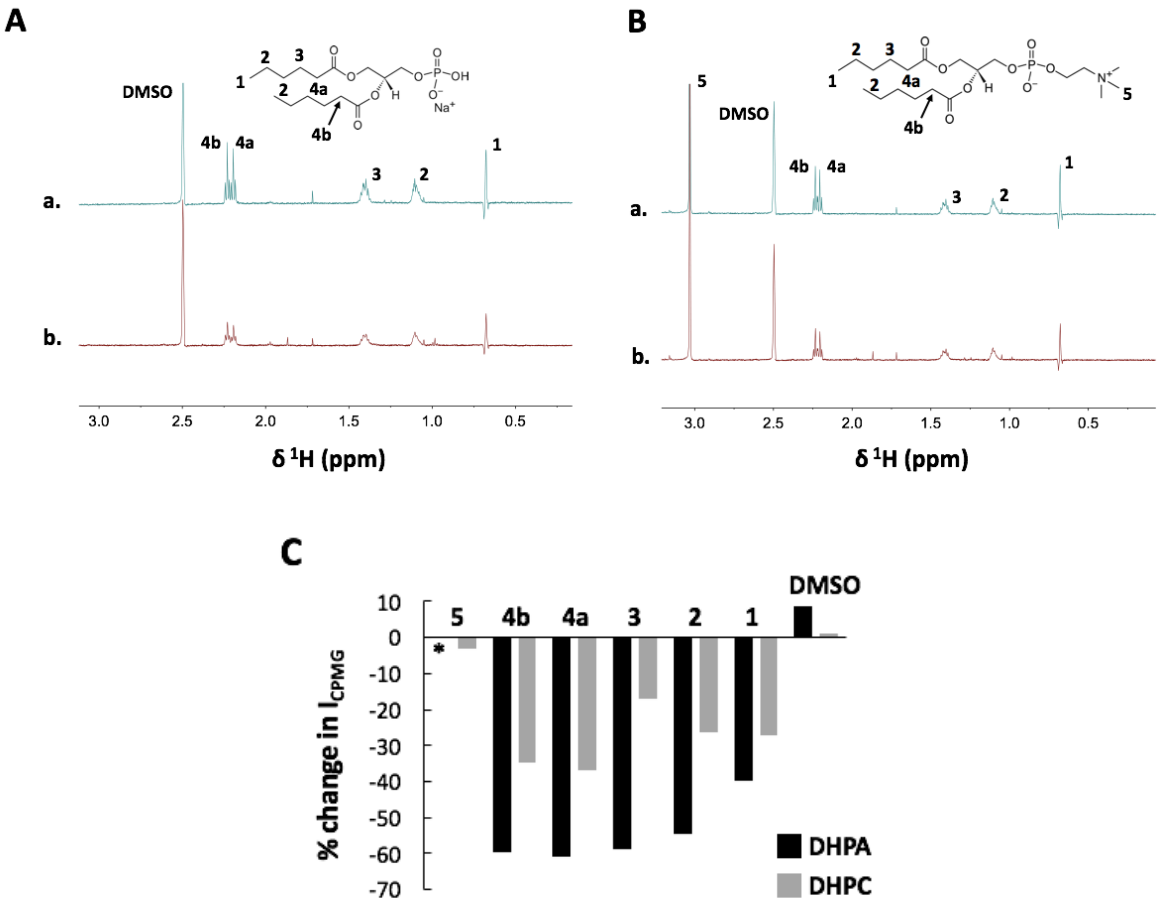



**Figure 4-5. Analysis of DHPA (di-6:0 PA) and DHPC (di-6:0 PC) interaction with the Nir2 LNS2 domain by CPMG.** a. CPMG $^{1}$H spectrum in the absence of Nir2 LNS2, b. CPMG spectrum in the presence of Nir2 LNS2. Structures of DHPA and DHPC are displayed for reference. DHPA and DHPC protons with corresponding resonance signals are numbered. Resonances below ~ 3.2 ppm are not shown due to their poor signal-to-noise ratio. **A.** DHPA experiments. Intensity of the DHPA proton resonances **1-4** is reduced in the presence of the protein, which is indicative of the interaction between the molecule and the protein. **B.** DHPC experiments. Intensity of the DHPC acyl chain resonances **1-4** is reduced in the presence of the protein, but to a lower extent than observed for DHPA. Only slight reduction in intensity is observed for the head group methyl resonance **5** (whole peak is not shown). **C.** Comparison of the observed changes in the intensity ($I_{CPMG}$) of the proton resonances of DHPA (black bars) and DHPC (grey bars) after addition of the Nir2 LNS2. In multiplets, the intensities of the highest peaks were compared. Resonance signals are numbered as in **A** and **B**. Larger intensity changes are observed in the DHPA acyl chain signals (~ 40-60%) than in the DHPC acyl chain signals (~ 15-35%). Very slight decrease in intensity is observed for the DHPC head group methyl signal **5** (~ 3%). Proton resonance equivalent to DHPC resonance **5** is not present in DHPA (absence is marked by *). Changes in the intensity of the DMSO signal reflect slight differences in the ligand concentration between the experiments.

### 4.3.3 Nir2 LNS2 does not interact with free hexanoic acid nor with free glycerol 3-phosphate

As it was observed that the fatty acyl chains of DHPA and DHPC interact with the TF fusion of Nir2 LNS2 domain, next it was decided to determine if the protein is able to bind free short-chain fatty acids. This would help to reveal whether the protein is able to bind lipids non-specifically. In order to investigate short-chain fatty acid binding, a series of NMR waterLOGSY experiments was carried out with hexanoic acid, a C6 fatty acid fragment of DHPA and DHPC. The experiments were carried out under conditions identical to the ones used in the NMR experiments with DHPA and DHPC.

Significantly, no interaction between hexanoic acid and the Nir2 LNS2 domain was observed in the experiments. None of the methyl or methylene proton signals have changed the sign after addition of the LNS2 in the waterLOGSY spectrum (Fig 4-6A), indicating that hexanoic acid is not interacting with the Nir2 LNS2 domain.

In order to determine whether the LNS2 binds the glycerol backbone of phospholipids in the absence of the fatty acyl tails, waterLOGSY experiments were carried out using glycerol 3-phosphate (G3P), which forms the backbone moiety of glycerophospholipids. No evidence of interaction between G3P and the LNS2 domain was observed, as apparent from the absence of positive G3P proton signals after protein addition (Fig 4-6B). In the waterLOGSY spectra, negative G3P peaks are indistinguishable from the background due to the poor signal-to-noise ratio.

Based on this data, it can be concluded that the presence of both the fatty acyl tails and the glycerol backbone is required for LNS2 binding, which indicates that the LNS2 domain of Nir2 interacts with DHPA and the fatty acyl chains of DHPC through a mechanism that allows specific recognition of glycerophospholipids. Furthermore, although Nir2 LNS2 recognises DHPC and interacts with its acyl chains, Nir2 LNS2 does not appear to interact with the backbone and the head group of DHPC, which suggests that the identity of the head group is important for phospholipid binding by the LNS2 domain. Thus, it can also be concluded that the specificity of the interaction between the LNS2 domain and PA is brought about by the presence of unsubstituted, negatively charged phosphate group in the PA molecule. To summarise, it is apparent that both the hydrophobic interactions with the fatty acyl tails as well as polar interactions with the glycerol backbone and the head group of PA play a role in the LNS2-PA binding, and the polar interactions with the head group

of PA are required for the specificity of PA binding.

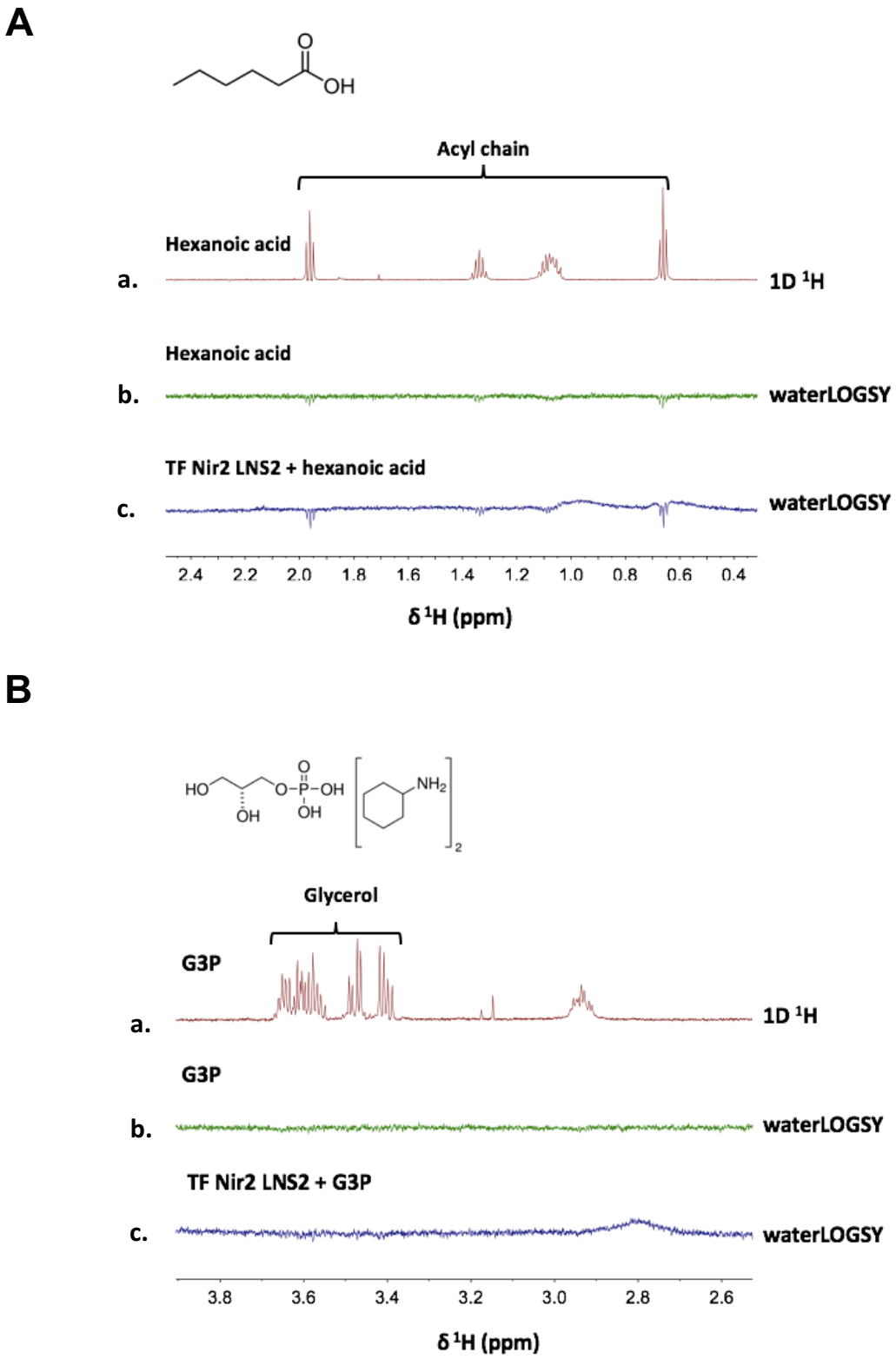



**Figure 4-6. Analysis of hexanoic acid and glycerol 3-phosphate interaction with the Nir2 LNS2 domain.** a) Reference 1D $^{1}$H spectrum, b) waterLOGSY spectrum in the absence of Nir2 LNS2, c) waterLOGSY spectrum in the presence of Nir2 LNS2. Structures of hexanoic acid and glycerol 3-phoshpate bis(cyclohexylammonium) salt used in the experiments are displayed for reference. Acyl and glycerol proton resonance regions are labelled in the relevant spectra **A.** Hexanoic acid experiments. No change in the sign of proton resonances is observed upon protein addition in the waterLOGSY spectra, indicating that no interaction is occurring between hexanoic acid and the LNS2 domain. **B.** Glycerol 3-phoshate (G3P) experiments. No positive ligand signals are observed in the presence of the protein, indicating that no interaction is occurring between G3P and the LNS2 domain.

### 4.3.4 Nir2 LNS2 interacts with DHPA with millimolar affinity

To obtain a more quantitative picture of DHPA binding, the affinity of the interaction between DHPA and Nir2 LNS2 was measured by waterLOGSY. For this, DHPA was added to the protein sample in a series of steps and a waterLOGSY spectrum was recorded after each addition. Mixing time in the experiments was set to 0.5 s and the protein concentration was kept relatively low (10 μM) to increase the likelihood of obtaining a more accurate $K_d$ value, as described in the literature (Huang *et al*., 2017). The intensity of the highest peak in the triplet at 0.68 ppm that corresponds to the methyl protons of the fatty acyl chains of DHPA was measured, as it was the most intense peak in the waterLOGSY spectrum. To be sure that the ligand was not aggregating at higher concentrations, the measurements was first carried out using DHPA in the absence of the protein. The plot of the signal intensity against the ligand concentration was linear (Fig 4-7A), which suggests that no aggregation of the ligand was occurring. After DHPA was titrated into the protein solution, the intensity of the 0.68 ppm peak was plotted as a function of DHPA concentration for $K_d$ determination (Fig 4-7B). Prior to the construction of the plot, the intensity values measured in the presence of the protein were corrected for the signal of the free ligand by subtracting the ligand intensities in the absence of the protein from the ligand intensities in the presence of the protein (see Section 4.3.1.3 for the theory).

After correction, the data were fitted by the least square method to Equation 4-3 describing one-site binding:

$$I = \frac{I_{max} \times [L]_{tot}}{K_d + [L]_{tot}} \qquad \textbf{Equation 4-3.}$$

where $I_{max}$ is the maximum intensity, $[L]_{tot}$ is the total ligand concentration and $K_d$ is the dissociation constant. This binding model assumes that $[L]_{tot}$ is approximately equal to the concentration of free ligand [L], which is true at high excess of ligand to the protein used in the measurements. Curve fitting using the data from three independent experiments has yielded $K_d = 13.0 \pm 0.4$ mM. It should be noted that since the binding did not start to reach saturation in the titration experiments, it is possible that the $K_d$ reported here is underestimated. Nevertheless, LNS2-DHPA binding can be defined as weak and placed into the millimolar range.

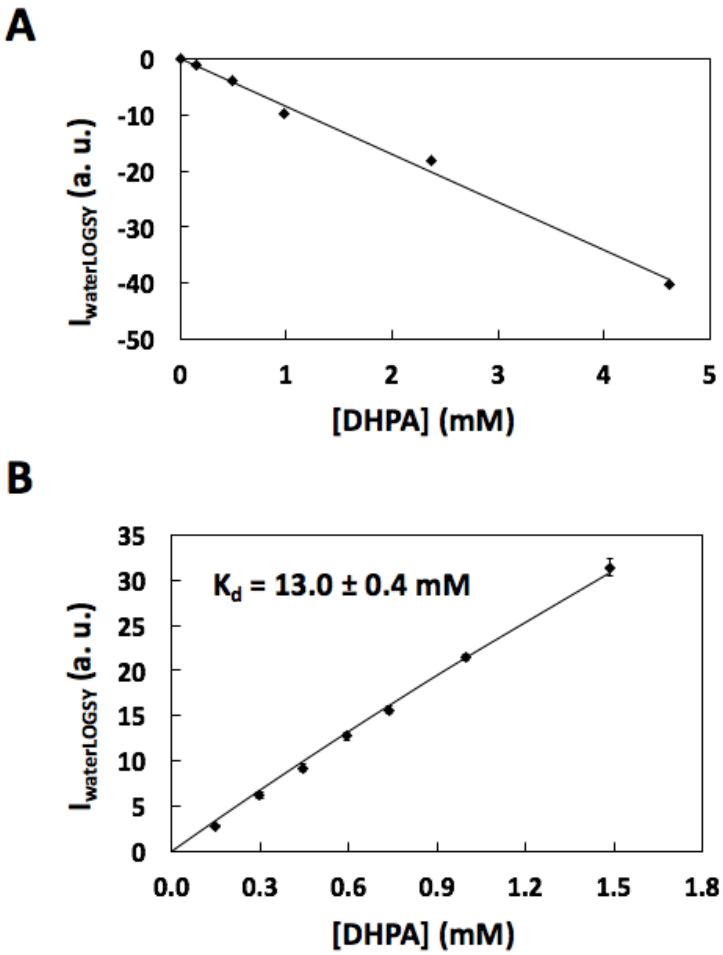



**Figure 4-7. $K_d$ calculation of the Nir2 LNS2-DHPA (di-6:0 PA) interaction using NMR waterLOGSY.** The intensity of the highest peak in the triplet at 0.68 ppm corresponding to the protons of the DHPA methyl group is shown as a function of DHPA concentration. **A.** DHPA titration in the absence of the protein. Intensity changes linearly with the increasing DHPA concentration, indicating that no aggregation of ligand is occurring. The data were fitted to a linear equation. **B.** DHPA titration in the presence of 10 μM TF fusion of Nir2 LNS2. The data were corrected for the free ligand concentration by subtracting the line in **A**. The corrected values were fitted to a hyperbolic one-site binding equation. The mean corrected values from triplicate experiments are shown. Error bars represent ± s.e.m. The calculated $K_d$ value is shown.

## 4.4 Analysis of Nir2 LNS2 domain binding to PA by fluorescence spectroscopy

### 4.4.1 A brief overview of fluorescence spectroscopy

#### 4.4.1.1 Fluorescence

Fluorescence spectroscopy is a mainstay of molecular bioscience that is continuing to find new applications with the development of novel fluorescent probes. Fluorescence spectroscopy is based on the fundamental physical phenomenon termed fluorescence, which is defined as the emission of light by a molecule (called a fluorophore) that has absorbed electromagnetic radiation. Fluorescence is observed when said molecule transitions from a high-energy excited state, which it enters on absorption of electromagnetic radiation, to a low-energy ground state. Since part of the energy of a molecule in an excited state is lost to the environment as a result of non-radiative decay, the energy of the light emitted during transition to the ground state is lower than that of the light absorbed by the fluorophore. Fluorophores remain in an excited state only for a short

period of time ($10^{-7}$-$10^{-9}$ s) before emitting light, which distinguishes fluorescence from phosphorescence, another type of radiative decay.

Importantly, the photophysical properties of a fluorophore, such as the lifetime, wavelength and intensity of fluorescence, often depend on its local environment. For instance, if the fluorophore is in a solution, its properties can be influenced by the physicochemical properties of the solvent (e.g., polarity, pH and ionic strength) and may change if the fluorophore binds to a macromolecule or finds itself in a close proximity to another fluorophore. This environmental sensitivity of fluorophores is harnessed in biomolecular research to analyse the behaviour of biomolecules *in vivo* and *in vitro*.

#### 4.4.1.2 Fluorescence emission, excitation and anisotropy measurements

By using fluorescence spectroscopy, one can measure several properties of the fluorophore. These include fluorescence excitation and emission maxima, fluorescence lifetime and fluorescence anisotropy. Furthermore, fluorescence measurements can be broadly divided into two types: time-resolved measurements and steady-state measurements. As the name implies, steady-state measurements report on the time-and-ensemble average state of the system during the measurement, whereas with time-resolved measurements one is able to follow changes in the system as they occur, and hence obtain more detailed information about its behaviour.

Fluorescence excitation and emission measurements are the most traditional types of fluorescence experiments. They are normally carried out in steady-state mode and require less sophisticated equipment than lifetime or anisotropy measurements. In emission and excitation measurements, light is shone onto the sample and the intensity of the outgoing fluorescence is recorded. During the measurement, monochromators can be used to transmit light with selected wavelength to and from the sample to acquire excitation or emission spectra. Fluorescence emission and excitation spectra can provide information about the environment of the fluorophore and can be used to observe fluorescence quenching (Mátyus *et al*., 2006) or fluorescence resonance energy transfer (Clegg, 2009; Lakowicz, 2006). Emission and excitation measurements are widely employed for a variety of investigations, including analysis of protein-ligand and protein-protein binding, formation of macromolecular complexes, characterisation of enzymatic reaction kinetics, determination of intermolecular distances in the cell, and others.

As measurement of fluorescence emission and excitation spectra relies on the accurate

measurement of fluorescence intensity of the sample, there are certain limitations associated with these types of experiments. Fundamentally, fluorescence intensity is dependent on the fluorophore concentration, intensity of the excitation light and the pathlength of the light in the cell from which the measurement is made. This makes measurements of fluorescence spectra highly dependent on the instrumental settings, sample and cell specifications. Because of this, technical difficulties may arise during intensity measurements when the experimental conditions cannot be tightly controlled. Moreover, the emission and excitation spectra of a fluorophore in two different states or environments are often not sufficiently different from each other to allow discrimination between the different states of the fluorophores that may be present in the sample. Hence, an additional dimension may be required, which can be obtained by anisotropy or lifetime measurements.

Fluorescence anisotropy measurements are based on the photoselection principle in fluorescence. That is, a fluorophore preferentially absorbs light whose electric vector is aligned in a specific way to the molecular axis (Lakowicz, 2006), leading to anisotropy of the emitted light. Anisotropy can be affected by several factors, including rotational diffusion. Crucially, fluorescence anisotropy reports on the mobility and shape of the fluorophore in solution. When a fluorophore associates with a larger molecule, rotational diffusion of the fluorophore will now be influenced by the associated partner, which will lead to changes in the observed fluorescence anisotropy. Fluorescence anisotropy measurements can be carried out in both steady-state and time-resolved modes. Unlike fluorescence intensity, fluorescence anisotropy does normally not depend on the concentration of the fluorophore. Although anisotropy measurements are a great tool in the study of biomolecular interactions, such as protein-protein or protein-ligand binding, their applications are more limited than those of fluorescence lifetime measurements. This is especially true when binding between the fluorophore and its interaction partner does not lead to sufficiently large changes in fluorescence anisotropy that can be detected experimentally.

#### 4.4.1.3 Fluorescence lifetime measurements

In contrast to fluorescence intensity, fluorescence lifetime is an intrinsic property of a fluorophore, and hence is independent of the excitation parameters, instrumentation or fluorophore concentration. Fluorescence lifetime is the average time required for the fluorophore in the excited state to return to the ground state via non-radiative decay and photon emission (Lakowicz, 2006). As mentioned previously, fluorescence lifetimes generally range from picoseconds to nanoseconds, and are sensitive of the fluorophore microenvironment.

Fluorescence lifetimes are measured with time-resolved techniques. The use of lifetime measurements allows to discriminate between several populations of a fluorophore that might be present in the sample and find their respective proportions if they have different lifetimes. As fluorescence decays are very short, sophisticated instrumentation is required for the measurements. One of the most commonly used method of fluorescence lifetime measurement is time-correlated single photon counting (TCSPC) (O'Connor and Phillips, 1984). In TCSPC, single photons emitted by the fluorophore are detected repetitively in a precisely timed manner, where the time of the excitation pulse serves as a reference for photon detection. As only a single photon is detected at a time, light pulses of low energy are used, which helps to prevent sample degradation and non-desirable optical effects. The data acquired from many cycles of single-photon measurements are processed digitally to produce a histogram of photon counts against arrival time of the photons after the pulse. The histogram is then used for lifetime determination.

Fluorescence decay is an exponential process. To calculate the fluorescence lifetime from the histogram obtained in the experiment, the fluorescence decay trace formed by the photon counts is fitted with an exponential function by a non-linear least squares method. In its simplest form, fluorescence decay is expressed by a single exponential equation:

$$I(t) = I_0 \, e^{-t/\tau} \qquad \textbf{Equation 4-3.}$$

where I(t) is time-dependent fluorescence intensity, $I_0$ is intensity at time zero and $\tau$ is fluorescence lifetime. The decay may also be fitted with a multi-exponential function, if several populations of the fluorophore are present in the sample.

Multi-exponential fluorescence decay is expressed by Equation 4-4:

$$I(t) = I_0 \sum_i \alpha_i \, e^{-t/\tau}$$ **Equation 4-4.**

where $\alpha_i$ is the pre-exponential factor or amplitude of a lifetime component i.

As can be seen from Equation 4-3, by fitting the data to a multi-exponential function one can determine the fractional contributions of the fluorophore in different states or environments to the time-dependent fluorescence intensity. This allows quantification of the different fluorophore populations and determination of their lifetimes in the sample. In this way, fluorescence lifetime measurements can provide a window into the intricate details of the molecular environment of the fluorophore in the sample, which are inaccessible by many other methods.

Importantly, fluorescence measurements are becoming increasingly popular in the pharmaceutical industry and academic research for drug and tool compound screening in a high-throughput format. For this purpose, various fluorescence-based assays which involve fluorescence anisotropy and lifetime measurements have been developed (Boettcher *et al*., 2014; Pritz *et al*., 2011; Zhang *et al*., 2015). High-throughput fluorescence experiments are carried out on fluorescence spectrometers equipped with microplate readers, which are available commercially. A common type of assay that can be used for the screening is a competition assay, in which a complex of a fluorescently-labelled ligand with a receptor protein is pre-formed, and competitor molecules from a compound library are added to the mixture to identify strong binders that displace the labelled ligand from the protein. The displacement can be monitored by measuring changes in the fluorescence anisotropy or fluorescence lifetime of the fluorophore.

### 4.4.2 Time-domain fluorescence lifetime measurements confirm DHPA binding to Nir2 LNS2

In this study, fluorescence lifetime measurements were used to further characterise the binding of LNS2 to DHPA and explore the potential for developing an assay for high-to-medium throughput screening of the LNS2-PA interaction inhibitors. In order to determine the feasibility of carrying out such an assay and to further confirm the LNS2-DHPA

interaction, a DHPA molecule labelled with a fluorophore (nitrobenzoxadiazole or NBD) at one of the fatty acyl tails was used. With the fluorescently-labelled ligand, a series of fluorescence lifetime measurements were performed by using picosecond TCSPC. The measurements were carried out in a 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT buffer in a quartz cuvette. The probe was excited at 460 nm, and the fluorescence was measured at 534 nm. The photon counts recorded in the experiments were used to construct the fluorescence decay traces. The decay traces were fitted in FluoFit software (PicoQuant) using an exponential model with reconvolution described by Equation 4-5:

$$I(t) = \int_{-\infty}^{t} IRF(t') \sum_{i}^{n} \alpha_i e^{\frac{t-t'}{\tau_i}} dt' \qquad \textbf{Equation 4-5.}$$

where IRF is the instrument response function, and the other parameters are as in Eq. 4-3 and Eq. 4-4. Accordingly, n = 1 for single exponential decay, and n = 2 for biexponential decay. IRF was deconvoluted from the measured fluorescence response to allow determination of the amplitudes and lifetimes of the fluorophore populations in the sample.

The lifetime measurements of DHPA-NBD fluorescence were performed either in the presence or the absence of TF Nir2 LNS2. When measured in the absence of the protein, the decay trace of the DHPA-NBD fluorescence resembled a straight line when photon counts were plotted on a log scale as a function of the time of photon arrival (Fig 4-8, green). The decay data were fitted to an exponential equation ($\chi^2 = 1.06$), and the average lifetime $\tau_{av}$ was determined to be equal to 1.27 ns. After addition of the protein, a drastic change in the fluorescence decay trace was observed (Fig 4-8, purple). In the presence of the protein, the decay trace was no longer linear, but resembled a curve. Hence, the decay was fitted to a biexponential equation ($\chi^2 = 0.989$), and two distinct amplitude-weighted lifetime values were calculated. The first lifetime $\tau_1$ was equal to 1.21 ns, which was similar to the lifetime of DHPA-NBD fluorescence observed in the absence of the protein. Interestingly, the second lifetime $\tau_2$ was found to be 7.78 ns, which was ~ 6.5-fold longer than $\tau_1$ or the DHPA-NBD lifetime in the absence of the protein. The fractional contributions of the two lifetimes were equal to 97.5% for $\tau_1$ and 2.50% for $\tau_2$, and $\tau_{av}$ was 1.38 ns, suggesting that the fluorophore was mostly found in the state with the shorter lifetime $\tau_1$.

The presence of $\tau_2$ was assumed to be due to the binding of DHPA-NBD to the LNS2 domain and the resulting change in the fluorophore surroundings. This notion is supported

by evidence from literature, as NBD has been previously reported to be highly sensitive to environmental polarity (Chattopadhyay and London, 1988; Chattopadhyay and Mukherjee, 1993), and possess a characteristic lifetime of ~ 7-8 ns when it is incorporated into lipid membranes (Chattopadhyay and Mukherjee, 1993; Mukherjee *et al*., 2004) and a much shorter ~1 ns (Lin and Struve, 1991) lifetime in pure water. Consistently, it can be proposed that the short lifetime $\tau_1$ (1.21 ns) is observed from the free fluorophore population, and the longer lifetime $\tau_2$ (7.78 ns) is observed from the fluorophore population bound to the protein. Therefore, if we consider the fractional contribution of $\tau_2$ to correspond to the fraction of the ligand bound to the protein, it is apparent that only 2.5% of the ligand in the sample was in a bound state at the 1:4 protein:ligand ratio used in the experiment. Such low proportion of bound ligand was likely observed due to the low binding affinity of DHPA-NBD to the LNS2 domain of Nir2, as the binding appears to be away from saturation at this ligand concentration.

In addition, lifetime measurements using TF and DHPA-NBD were carried out to determine if the observed lifetime changes in the presence of the fusion were caused by ligand binding to the LNS2 domain in the fusion and not to TF. TF was added to the sample to reach the 1:4 protein:ligand ratio used in the measurements with the fusion protein. Unexpectedly, the addition of TF has also lead to a change in the decay trace (Fig 4-8, blue), which appeared to be curved in the presence of the protein. The decay trace was fitted to a biexponential equation, and lifetimes $\tau_1$ and $\tau_2$ were identified, which were equal to 1.22 ns and 5.31 ns, respectively. The fact that a second lifetime was observed in the presence of TF indicates that DHPA-NBD also binds to TF in the absence of the LNS2 domain. However, this binding appears to be different from the binding observed between TF and the probe, as the $\tau_2$ values observed in the presence of the fusion protein and TF alone were different (7.78 ns and 5.31 ns, respectively). The fractional contributions of the decay components in the presence of TF were 1.0% for $\tau_2$ and 99% for $\tau_1$, and the $\tau_{av}$ was equal to 1.38 ns.

From the $\tau_2$ fractional contribution we can assume that only 1% of the fluorophore molecules were interacting with TF in the sample, which is a lower proportion of the interacting molecules than the one observed with the fusion protein, suggesting that DHPA-NBD-TF binding is of even lower affinity. Importantly, as in the ligand-observed NMR experiments no binding was observed between DHPA and TF, it is likely that it is the NBD moiety that is interacting with TF rather than the DHPA moiety the DHPA-NBD conjugate. Alternatively, it may be that the DHPA moiety also interacts with TF, but the

interaction is too weak to be detected by the ligand-observed NMR experiments, which points towards its very transient nature. Since $\tau_2$ is shorter in the presence of TF than in the presence of the LNS2 fusion protein, it is also likely that the probe is in a less hydrophobic environment when bound to TF than when it is bound to TF Nir2 LNS2, and is partially exposed to the solvent when interacting with TF. The interaction may therefore be occurring with the exposed hydrophobic patches present of the surface of TF (Hoffmann *et al*., 2010), and hence is likely non-specific.

To summarise, the data show that the recombinant TF fusion of the Nir2 LNS2 domain weakly interacts with DHPA-NBD, which confirms the findings in the ligand-observed NMR experiments (see above). DHPA-NBD also displayed interaction with TF in the absence of the LNS2 domain, but that interaction appears to occur via a different mechanism and is of weaker affinity than the interaction between DHPA-NBD and the LNS2 domain fusion proteins.

The results of fluorescence lifetime measurements are summarised in Table 4-1.

| **Protein/competitor** | **$\tau_1$ (ns)** | **$\tau_2$ (ns)** | **$\tau_{av}$[a] (ns)** | **$\tau_{av}$[b] (ns)** | **$\alpha_1$[c]** | **$\alpha_2$[c]** | **$\chi^2$** |
|---|---|---|---|---|---|---|---|
| None | 1.28 | N/A | 1.28 | 1.28 | 1.00 | N/A | 1.06 |
| TF-LNS2[d] | 1.21 | 7.78 | 1.38 | 2.14 | 0.03 | 0.97 | 0.99 |
| TF-LNS2/5 μM DHPA | 1.22 | 7.48 | 1.32 | 1.83 | 0.02 | 0.98 | 1.01 |
| TF-LNS2/10 μM DHPA | 1.21 | 7.48 | 1.31 | 1.76 | 0.02 | 0.98 | 1.01 |
| TF-LNS2/20 μM DHPA | 1.22 | 7.50 | 1.32 | 1.78 | 0.02 | 0.98 | 1.06 |
| TF | 1.22 | 5.31 | 1.25 | 1.38 | 0.01 | 0.99 | 0.99 |

a. Amplitude-weighted average lifetime.
b. Intensity-weighted average lifetime.
c. Normalised pre-exponential factors (amplitudes).
d. The TF fusion of Nir2 LNS2.

**Table 4-1. Fluorescence decay parameters of DHPA-NBD (di-6:0 PA-NBD).** The decay parameters are shown in the absence or the presence of proteins (TF Nir2 LNS2 or TF), with or without a competitor molecule (DHPA) at 5, 10 or 20 μM concentration. The parameters were obtained from fitting the decay data to single exponential or biexponential equations after correction for instrumental response, as described in the text.

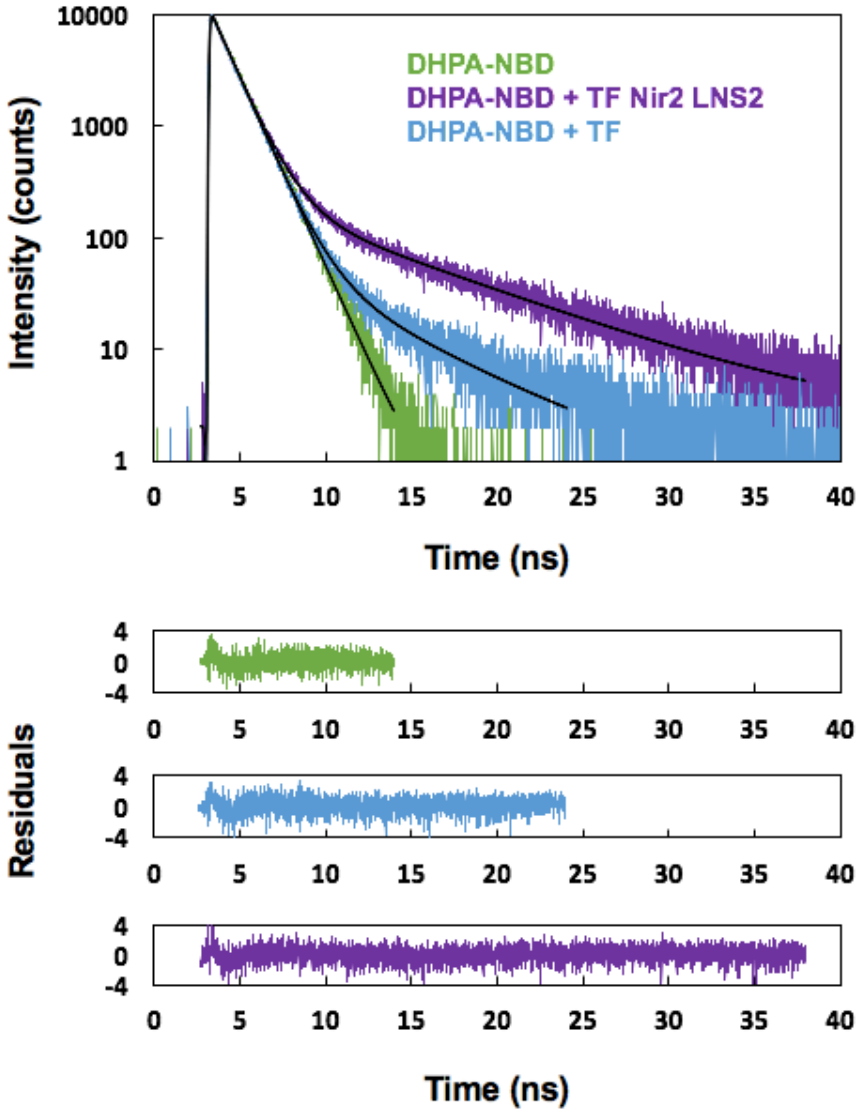



**Figure 4-8. Fluorescence lifetime measurements of DHPA-NBD (di-6:0 PA-NBD).** Fluorescence decay traces of DHPA-NBD alone in buffer (green lines), in the presence of TF Nir2 LNS2 (purple lines) and in the presence of TF (blue lines). In the absence of TF Nir2 LNS2, the decay signal can be fitted to a single exponential equation ($\tau = 1.27$ ns), whereas in the presence of TF Nir2 LNS2 and TF, the decay can be fitted to a biexponential equation ($\tau_1 = 1.21$ ns, $\tau_2 = 7.47$ ns for TF Nir2 LNS2). Fitted decay models are shown as black curves. Residuals of best fits are shown below the decay graph.

Since an interaction between TF and the fluorescently-labelled DHPA-NBD was detected, it was instrumental to confirm experimentally that the binding between DHPA-NBD and the Nir2 LNS2 domain fusion protein is caused by the interaction between the DHPA moiety of DHPA-NBD and the LNS2 domain. For this purpose, displacement of the fluorescently-labelled DHPA-NBD from TF Nir2 LNS2 was attempted with the unlabeled DHPA. Displacement was expected if the interaction between the DHPA part of DHPA-NBD and the LNS2 domain was present.

For the displacement assay, DHPA was titrated into the sample to the final concentrations of approximately 5 µM, 10 µM and 20 µM, corresponding to the ligand:competitor ratios

of 1:2.5, 1:5 and 1:10, respectively. The fluorescence of DHPA-NBD appeared to decay faster after addition of DHPA, as judged from the observed decay traces (Fig 4-9, grey, orange and red lines). Like the other decay traces of DHPA-NBD measured in the presence of TF Nir2 LNS2, they were fitted by a biexponential model. $\chi^2$ was 0.989 for the 5 μM DHPA, 1.01 for 10 μM DHPA and 1.06, representing a good fit. The $\tau_2$ values arising due to the interaction of DHPA-NBD with the protein were calculated to be equal to 7.48 ns in the presence of 5 μM and 10 μM DHPA, and 7.50 ns in the presence of 20 μM DHPA. The $\tau_2$ values were very close to the $\tau_2$ value (7.47 ns) measured in the absence of the competitor, as described above. Similarly, the shorter lifetime $\tau_1$ was equal to 1.21 or 1.22 ns, consistent with the previous measurements. The fractional contribution of $\tau_2$ has decreased from 2.50% in the absence of the competitor to a minimum of 1.54% (38% decrease) after the addition of 10 μM unlabelled DHPA, indicating that DHPA is displacing DHPA-NBD from the protein. An apparent 0.02% increase in the $\tau_2$ fractional contribution was observed after addition of DHPA to 20 μM concentration. As the increase was very low, it was assumed to be an artifact of curve fitting. Consistently, the amplitude-weighted $\tau_{av}$ values have decreased from 1.38 ns in the absence of the competitor to a minimum of 1.31 ns in the presence of DHPA in ~ 5 molar excess to DHPA-NBD.

The fact that complete displacement was not observed in the experiment can be explained by the presence of non-specific binding between DHPA-NBD and TF and/or higher affinity of TF Nir2 LNS2 to DHPA-NBD than to DHPA. Nevertheless, it appears that the binding between DHPA-NBD and TF Nir2 LNS2 is in large part caused by the LNS2 domain. In order to confirm this, fluorescence of NBD not linked to DHPA can be measured in the absence and the presence of the protein to determine whether NBD alone can also interact with the protein.

Importantly, it was shown that DHPA-NBD can be at least partially displaced from the protein by addition of a competitor molecule. Hence, it can be proposed that a competition-based assay using the fluorescence intensity measurements of DHPA-NBD or another fluorescently-labelled PA molecule can be employed for identification of inhibitors of LNS2-PA binding. The considerations for designing such an assay are discussed at the end of this chapter.

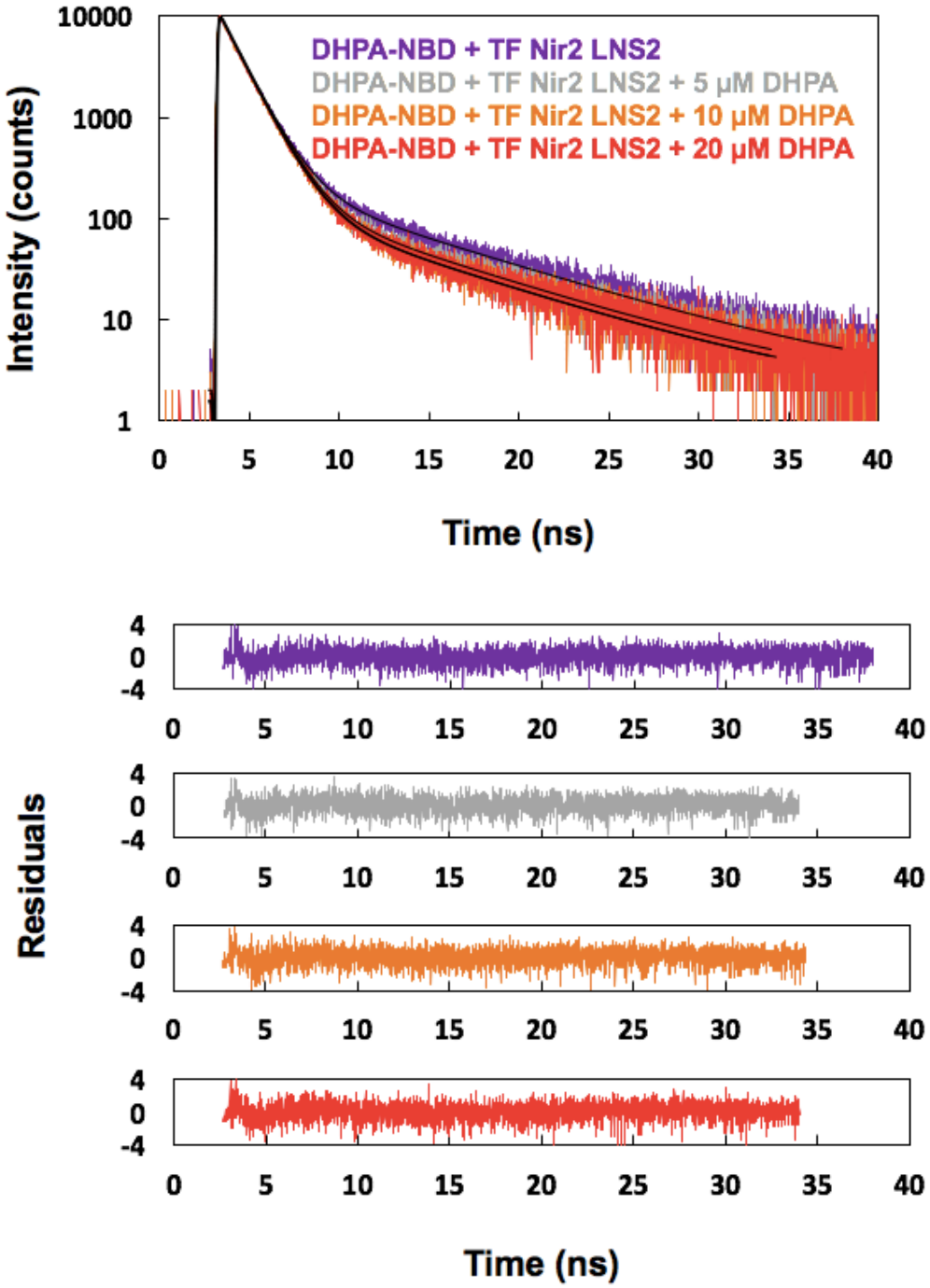



**Figure 4-9. Fluorescence lifetime measurements of DHPA-NBD (di-6:0 PA-NBD) displacement from TF Nir2 LNS2 by DHPA.** Fluorescence decay traces of DHPA-NBD are shown in the presence of TF fusion of Nir2 LNS2 (purple lines) and in the presence of both TF fusion of Nir2 LNS2 and a competitor DHPA molecule at either 5 µM, 10 µM or 20 µM concentrations (grey, orange and red lines, respectively). Fitted decay models are shown as black curves. Residuals of best fits are shown below the decay graph.

## 4.5 Analysis of the interaction between Nir2 LNS2 and PA-containing LUVs by CD spectroscopy

### 4.5.1 CD spectroscopy

Circular dichroism (CD) spectroscopy is a biophysical technique commonly used in protein research for characterisation of the secondary structure of proteins (Greenfield, 2006; Kelly *et al*., 2005). Circular dichroism arises when the two circularly polarised components (left and right) of plane polarised light are absorbed differentially by the sample. In CD spectroscopy, the sample is illuminated with plane polarized light and the difference in absorbance of polarised light components is measured. Differential absorbance is reported in degrees of ellipticity of light transmitted through the sample. Ellipticity refers to the measure of elliptical polarisation of light, which occurs when circularly polarised components of incident light are absorbed unequally by the sample. The degree of ellipticity corresponds to the tangent of the ratio of the two components of plane polarised light.

To obtain a CD spectrum of a protein, one can follow the absorbance of peptide bonds (far UV region, < 240 nm), aromatic amino acid side-chains (near UV region, 260-320 nm) or disulphide bridges (~ 260 nm). Proteins are CD active due to the intrinsic chirality of amino acids and the presence of chiral secondary and tertiary structure components such as α-helices and β-sheets. Conveniently, the common types of protein secondary structures can be distinguished by their CD spectra. A set of representative far UV CD spectra of secondary structure components is shown in Fig 4-10 [reproduced from (Greenfield, 2006)]. Typically, α-helices display positive ellipticity in the region from 190 to 202 nm, with the peak value at 193 nm, and negative ellipticity in the region from 202 to 250 nm, with the lowest value at 222 nm (Greenfield and Fasman, 1969). Typical antiparallel β-sheets demonstrate positive ellipticity in the region from 190 nm to 205 nm, with the maximum at 195 nm, and negative ellipticity between 208 and 238 nm, with a dip in the spectrum at 217 nm. Like α-helices and β-sheets, disordered peptide chain also has a characteristic CD spectrum with negative ellipticity between 190 and 210 nm. Several other protein secondary structures, such as the collagen triple-helix, display other unique spectral shapes (Fig 4-10).

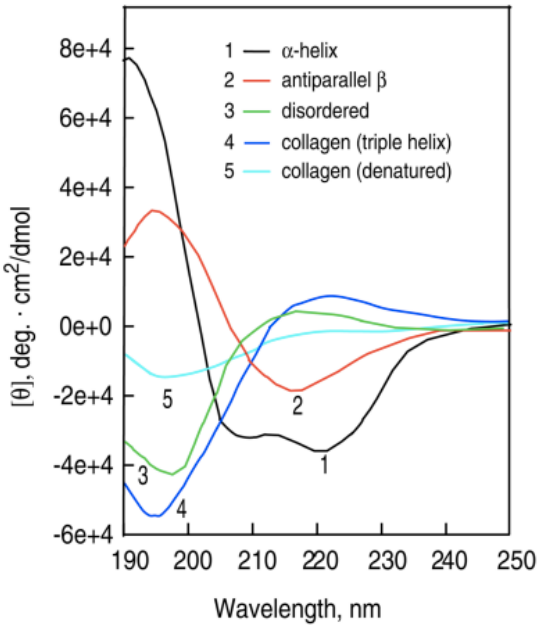



**Figure 4-10. Characteristic CD spectra of peptide and protein secondary structures.** Spectra 1-3 correspond to poly-L-tryptophan in α-helical, anti-parallel β-sheet and disordered forms, respectively. Spectra 4 and 5 correspond to collagen in a triple-helical and denatured forms, respectively.

Reprinted by permission from Springer Customer Service Centre GmbH: Springer Nature, Nature Protocols (Greenfield, N.J. Using circular dichroism spectra to estimate protein secondary structure. 2006. Nature Protocols 1, 2876–2890), copyright (2006).

Information about the secondary structure of protein of interest is typically obtained by comparison of the measured CD spectrum to the reference CD spectra of proteins with known three-dimensional structures. Most commonly, the analysis of CD spectra is carried out online by using DICHROWEB (Whitmore and Wallace, 2004), an online server for protein secondary structure analysis. DICHROWEB provides access to most popular analytical algorithms including SELCON3 (Sreerama and Woody, 1993), CONTIN/LL (Provencher and Gloeckner, 1981), CDSSTR (Johnson, 1999; Sreerama *et al*., 2000) and K2d (Andrade *et al*., 1993), as well as seven reference CD datasets (Sreerama *et al*., 2000).

In protein-ligand binding studies, CD spectroscopy can be used for monitoring the conformational change in the protein upon interaction with the ligand. This is done by recording the CD spectra of the protein in the presence and the absence of the ligand, and comparing the shapes of the spectra and the predicted secondary structure proportions. Changes in the CD spectra can be attributed to a structural change in the protein that occurs upon ligand binding, and could reveal structural details about the ligand binding mechanism. It should be noted, however, that ligand binding does not always result in the conformational change in the protein or, if a conformational change does occur, it might be not reflected in the CD spectrum.

### 4.5.2 No secondary structure change is observed in the TF fusion of Nir2 LNS2 domain upon PA binding

Here, CD spectroscopy was employed to determine whether there is any observable change in the secondary structure of the LNS2 domain of Nir2 upon PA binding. The putative dimeric form of the fusion protein was used in the experiments.

Far UV CD spectra were recorded from samples containing either: 1) TF Nir2 LNS2 only, 2) TF Nir2 LNS2 in the presence of PA-containing LUVs, or 3) TF Nir2 LNS2 in the presence of PC-only LUVs. PC-only LUVs were used as a control, as no specific binding was expected between the LNS2 domain of Nir2 and PC. Spectra recorded in the absence of the protein were used to correct for the background scattering by buffer components and the vesicles. The appearance of the spectra was very similar in the absence and the presence the vesicles (Fig 4-11). The difference between the spectra observed in the region from 190 to 196 nm can be attributed to non-specific effects of vesicle addition or light scattering by the vesicles that was not corrected by subtraction of vesicle-only spectra. Importantly, the CD spectra obtained from the samples containing 30% PA, 70% PC LUVs and PC-only LUVs in the presence of TF Nir2 LNS2 appeared nearly identical, signifying that the presence of PA does not specifically influence the secondary structure of the LNS2 domain. The spectra were analysed by K2d neural network algorithm (Andrade *et al*., 1993) to estimate the fractions of the secondary structure components in the protein (Table 4-2). Consistently, in all three samples, the protein was found to comprise 37% α-helices, 17-18% β-sheets and 45-46% turns and unordered regions, indicating that the secondary structure of the protein did not change on PA binding.

It should be noted, however, that it was the TF fusion of the Nir2 LNS2 domain that was used for CD measurements and not the LNS2 domain alone. Since TF forms the largest portion of the fusion protein (69% by MW), it is a larger contributor to the CD spectrum than the recombinant LNS2 domain, which constitutes only 28% of the fusion. Nevertheless, it would be expected that at least small differences between the spectra obtained in the presence of PA-containing LUVs versus PC-only LUVs would be observed if there was a secondary structure change in the LNS2 domain upon binding to PA.

The analysis of the secondary structure composition of the Nir2 LNS2 domain in the absence of TF could not be carried out due to the low protein expression yield and the tendency of the protein to aggregate in solution, as described in Chapter 3. A way of obtaining the structural contributions of Nir2 LNS2 from the CD spectrum of the fusion

protein would be to record a CD spectrum of TF under the same conditions as used for the fusion protein, subtract the spectrum of TF from that of the fusion and carry out the secondary structure analysis by one of the conventional algorithms. The fractional secondary structure contributions obtained in such way should approximately correspond to those of the Nir2 LNS2 domain, assuming that the secondary structure of TF in the fusion protein and in the free form is not majorly different. This work was not carried out in this study due to time constraints.

| Sample | α-helix | β-sheet | Unordered | Max. error |
|---|---|---|---|---|
| TF-LNS2 alone | 0.37 | 0.17 | 0.45 | 0.08 |
| TF-LNS2 + PA:PC | 0.37 | 0.17 | 0.46 | 0.08 |
| TF-LNS2 + PC | 0.37 | 0.18 | 0.45 | 0.08 |

**Table 4-2. Secondary structure analysis of the TF fusion of Nir2 LNS2 domain.** The CD spectra were recorded in the absence (TF-LNS2) and presence of LUVs containing either 30% PA and 70% PC (TF-LNS2 + PA:PC) or only PC (TF-LNS2 + PC) are shown. Fractional contributions of secondary structure components and maximum errors of the estimates are shown. Analysis was carried out by K2d unsupervised neural network algorithm (Andrade *et al*., 1993). Maximum error is the sum of root mean square deviations of α-helix and β-sheet predictions. The prediction is considered to be reliable if the maximum error is less than 0.227.

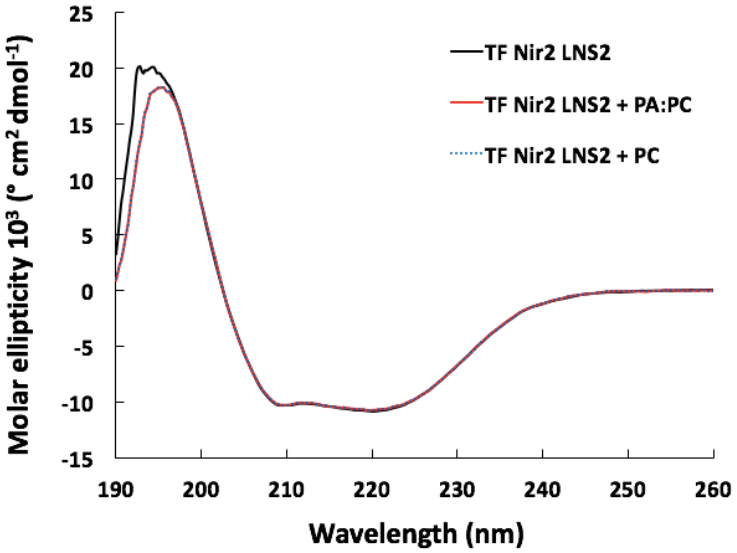



**Figure 4-11. Far UV CD spectra of the TF fusion of Nir2 LNS2.** The spectra were recorded in the absence (black curve) and the presence of LUVs either containing 30% PA and 70% PC (red curve) or only PC (blue dotted curve). All three spectra appear indistinguishable in the region above 200 nm.

### 4.5.3 The LNS2 domain of Nir1 has β-sheet-rich fold in solution

As it was not possible to carry out a direct analysis of the secondary structure of Nir2 LNS2, N-terminally $His_6$-tagged Nir1 LNS2-S fragment was employed for CD spectroscopy to obtain information about the secondary structure of the LNS2 domain from

a human protein. The recombinant Nir1 LNS2 domain was expressed as described in Chapter 3. Purification was carried out by $Ni^{2+}$-affinity chromatography in the presence of 20% glycerol to improve the purity of the recombinant protein. CD spectroscopy in the far UV region revealed that recombinant Nir1 LNS2 has spectral characteristics of a folded protein (Fig 4-12). Structural analysis was successfully carried out by the CDSSTR algorithm (Johnson, 1999; Sreerama *et al*., 2000) despite the noise that was observed in the spectrum due to the low protein concentration in the sample (0.12 mg/ml; higher concentration was difficult to obtain because of protein instability). The 193-260 nm spectral region was used for analysis. Nir1 LNS2 was predicted to comprise 10% α-helices, 48% β-sheets, 12% turns and 29% unordered regions, indicating that the LNS2 domain of Nir1 is rich in β-sheets. The normalised root mean square deviation value was equal to 0.029, indicative of a reliable prediction (Whitmore and Wallace, 2004). Although it is yet to be shown experimentally that the secondary structure of the LNS2 domain is conserved between the human Nirs, this is likely to be the case because of the high sequence similarity between the LNS2 domain of Nir1, Nir2 and Nir3. Ideally, the protein purification procedure should be optimised further to obtain samples with sufficient concentrations of Nir2 and Nir3 LNS2 domains for CD analysis. Buffers containing solubility-enhancing additives (e.g., glycerol) should be considered for CD measurements.

CD measurements and analysis of CD spectra were carried out by Dr Sharon Kelly (University of Glasgow).

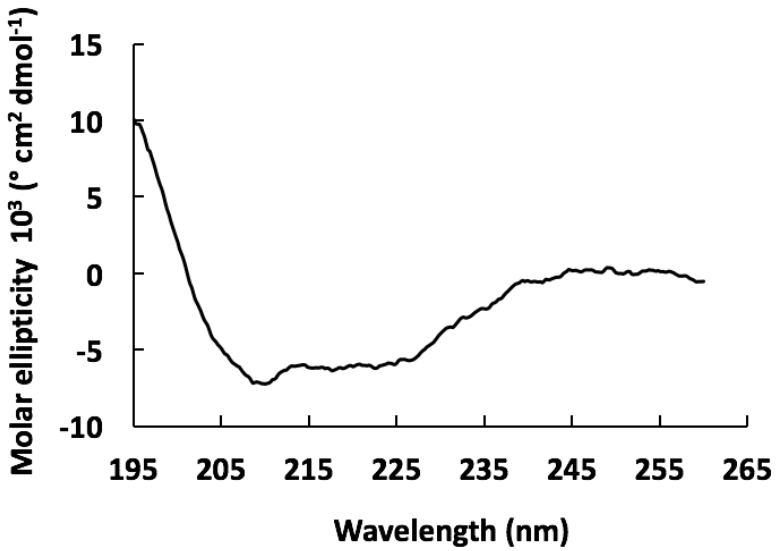



**Figure 4-12. Far UV CD spectrum of the Nir1 LNS2-S fragment.** Noise is observed in the spectrum due to the low protein concentration in the sample (0.12 mg/ml).

## 4.6 Conclusions and discussion

In the experiments described in this chapter several findings were made concerning the interaction of the LNS2 domain of Nir2 and PA. First, the ability of the recombinant fusions of Nir2 LNS2 to bind PA was confirmed by several methods, and the affinity of the Nir2 LNS2-PA interaction was estimated using PA-containing LUVs and short fatty acyl chain PA species. Secondly, it was discovered that the protein interacts with both the hydrophobic and polar regions of PA. Thirdly, it was observed that the protein does not appear to change its secondary structure upon PA binding. Fourthly, the secondary structure composition of Nir1 LNS2 was characterised, and the protein was found to consist mostly of β-sheets. Moreover, two competition-based assays were proposed that could be employed for medium-to-high throughput screening of the LNS2-PA interaction inhibitors.

### 4.6.1 The mechanism of PA binding by the LNS2 domain of Nir2

Taking into account the data obtained from the biophysical and biochemical experiments, several assumptions can be made about the mechanism of the LNS2-PA interaction. As mentioned above, the results of the ligand-observed NMR studies strongly suggest that the LNS2 domain of Nir2 interacts with both the fatty acyl chains and with the glycerophosphate backbone of PA. From these studies, it is also apparent that the protein recognises the presence of the anionic phosphate group in DHPA, as differences were observed between the LNS2-DHPA and LNS2-DHPC interactions where the LNS2 domain of Nir2 bound to the polar region of DHPA but not DHPC. However, the interaction between the phosphate group of PA and the protein could not be demonstrated directly due to the fast chemical exchange of the phosphate hydroxyl proton with the bulk water. A potential way to confirm the interaction would be to carry out $^{31}$P NMR measurements of the chemical shift, $T_2$ relaxation time or the lineshape of the DHPA phosphate signal in the presence and the absence of the protein. It is possible that either a change in the chemical shift, or shortening of the $T_2$ relaxation time and broadening of the linewidth of the phosphate group signal would be observed upon addition of the protein to the phospholipid. Furthermore, as it was observed that Nir2 LNS2 did not interact with free hexanoic acid and glycerol 3-phosphate within the range of concentrations tested, it is clear that the protein specifically recognises the structural features of phospholipids and not of their isolated parts.

From comparing the estimated high micromolar (~ 0.5 mM) apparent dissociation constant for DOPA in LUVs estimated in the liposome co-sedimentation experiments to a ~ 13 mM dissociation constant for DHPA measured in the waterLOGSY experiments, it can be seen that the LNS2 domain appears to interact with PA with ~ 25-fold stronger affinity when PA is embedded in a lipid membrane. Both interactions are, however, weak on the absolute scale. Also, it is unclear whether the affinity of the LNS2-membrane interaction *in vivo* is comparable to the calculated value for the LNS2-LUV binding, or whether the low affinity is observed only *in vitro* with the recombinant TF fusion of Nir2 LNS2. For instance, one reason for the observed low *in vitro* LNS2-PA affinity could be that a significant part of the recombinant LNS2 domain used in the experiments was in a misfolded or partially misfolded state with reduced ability to interact with PA. A definitive conclusion on the folding state of the TF Nir2 LNS2 oligomeric species could not be made in this study.

Interestingly, lower affinity to free lipids compared to membrane-embedded lipids was previously reported for PI(3)P-binding zinc finger FYVE domains and phorbol ester/DAG-binding C1 domains. FYVE domains bind at least 20 times more strongly to PI(3)P in the lipid bilayer than to the short-chain PI(3)P or the free $IP_2$ head group of PI(3)P (Gaullier *et al*., 2000; Kutateladze *et al*., 1999), and C1 domains have a $10^4$-fold higher affinity for phorbol esters embedded in the membrane than for free phorbol esters (Kazanietz *et al*., 1995). Crystal structures of FYVE (Misra and Hurley, 1999) and C1 (Zhang *et al*., 1995) domains have revealed that the domains dock onto the membrane through insertion of specific hydrophobic residues into the lipid bilayer and formation of salt bridges between charged amino acids and the polar functional groups of phospholipids in the membrane. It is therefore possible that the LNS2 domain contains similar structural elements that aid its interaction with phospholipid membranes. This notion is supported by the apparent inability of Nir2 LNS2 to interact with free glycerol 3-phosphate, suggesting that the interaction between the polar region of the membrane and the protein may not be sufficient for protein-membrane binding. Indeed, the ability of the LNS2 domain to interact with the fatty acyl chains of phospholipids indicates that the interaction between the aliphatic region of the membrane and the protein is highly likely. Furthermore, the apparent low affinity of PA binding by the LNS2 domain suggests that it is unlikely that the PA molecule is enclosed in a deep binding pocket or a cavity when bound to the LNS2 domain, as observed for the PITP domain or the nematode FARs (see Sections 1.3 and 1.4). Instead, the PA-binding site of the LNS2 domain is likely located on the surface of the protein or in a shallow pocket. This is further supported by the fact that the protein does not appear to have lipid-solubilising properties, as evident from the fact that it remains

bound to the blot and the lipid vesicles after binding the ligand in the lipid overlay and liposome co-sedimentation assays, respectively.

The potential membrane-interacting hydrophobic residues of Nir2 include W1055 and W1079 which are conserved between the LNS2 domains of Nir1-3 and RdgB (Fig 4-13). Tryptophan residues were singled out as they often play a key role in membrane binding by the membrane-interacting proteins (Das and Rahman, 2014; Feng *et al*., 2002; Glomset, 1999; Lee *et al*., 2006) and possess the highest free energy of partitioning from bilayer interface into water (Wimley and White, 1996). However, other hydrophobic residues, which are abundant in the LNS2 domain, may also participate in the proposed membrane interaction. Of interest here is the hydrophobic linear peptide consisting of Y1060-LIVY-V1065 that is conserved across the Nir1-3 and RdgB LNS2 domains as the Y(L/M)(I/L)XY(I/V) sequence motif where X is a non-aromatic apolar amino acid. It is, however, unknown whether any of these residues are found on the exterior of the protein, where they could contact they membrane.

The LNS2 domain of Nir2 also contains a number of distributed charged residues that are conserved between the Nir1-3 and RdgB LNS2 domains, and which could participate in an electrostatic interaction with the phospholipid head groups in the membrane. The conserved basic residues in Nir2 LNS2 include four lysines (K1043, 1104, 1127 and 1151) and six arginines (R1045, 1053, 1068, 1075, 1147 and 1150). Furthermore, eight acidic residues are also highly conserved (D1028, 1041, 1050, 1057, 1070, 1094, 1099 and 1128). It can be proposed that at least some of the conserved basic residues could form salt bridges to the anionic head group of PA. Correspondingly, the function of the acidic residues might be to orient the protein relative to the membrane in such a way as to allow the PA binding to occur or to coordinate solvent molecules in the binding site that can in turn interact with the polar groups of PA. Since the LNS2 domain was unable to bind the head group of DHPC in the ligand-observed NMR experiments, it is unlikely that the acidic residues interact with cationic head groups of membrane lipids.

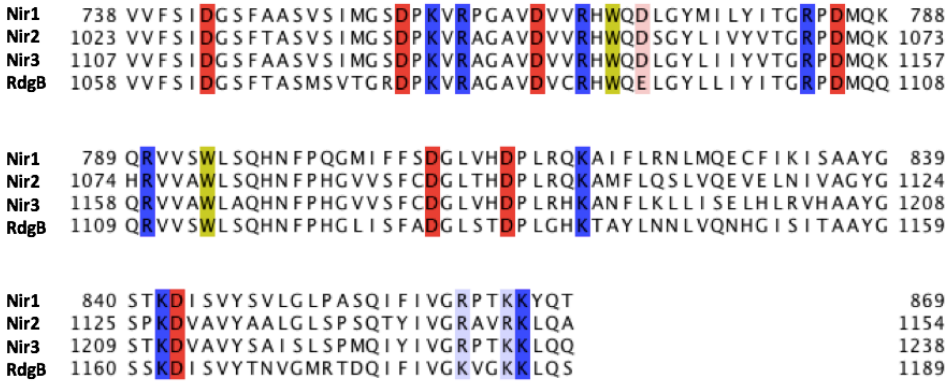



**Figure 4-13. Multiple sequence alignment of the LNS2 domains of Nir1-3 and RdgB.** Amino acid sequences of the LNS2 domains are shown, as defined by SMART (Schultz *et al*., 1998). Conserved tryptophan residues (yellow), as well as basic and acidic residues (blue and red, respectively) are highlighted. Fully conserved residues are highlighted in dark colour, partially conserved residues are in light colour. The alignment was produced in Jalview (Waterhouse *et al*., 2009).

Previously, no common PA binding motif has been identified, and it is widely assumed that PA binding protein rely of the presence of appropriately spaced basic amino acids that bind the phosphate group of PA through electrostatic interactions (Lemmon, 2008; Stace and Ktistakis, 2006). However, at least two PA binding peptides including the PA binding region of Raf-1 (Kraft *et al*., 2008) and synthetic lysine polymers (Takahashi *et al*., 1991) are known to adopt β-sheet-like structure upon PA binding. It is therefore a possibility that the β-sheet-rich structure of the LNS2 domain of Nir1 and, possibly, other Nirs and RdgB somehow contributes to PA binding. Unfortunately, the three-dimensional structure of Nir1 LNS2 could not be determined, as the protein was prone to aggregation in solution and displayed very low protein yield in recombinant expression (see Chapter 3). Disappointingly, crystallisation screening of the more stable TF and MBP fusions of Nir2 LNS2 failed to identify conditions that produced protein crystals.

To summarise, it can be suggested that the hypothetical PA binding site of Nir2 LNS2 is located in a shallow pocket and contains polar amino acids that interact with the head group and the glycerol region of PA via hydrogen bonding and electrostatic interactions, and apolar residues that interact with the aliphatic region of PA and, possibly, of other membrane lipids. It is also possible these or other apolar residues can insert into the hydrophobic interior of the membrane, contributing binding energy to the LNS2-membrane interaction. Further experiments would be required to confirm these hypotheses.

### 4.6.2 Proposed assays for identification of Nir2 LNS2-PA interaction inhibitors

Development of inhibitors of the Nir2 LNS2-PA would provide an efficient way of determining the role of the interaction in phosphoinositide signalling. Based on the experimental findings reported here, at least two simple assays can be proposed for screening of Nir2 LNS2-PA interaction inhibitors, which can be adapted for use in a medium-to-high throughput format. Both assays are competition-based and use short-chain PA molecules due to their high solubility in water and ability to bind the LNS2 domain of Nir2. Short descriptions of the principles are given below.

The first proposed assay employs ligand-observed NMR spectroscopy, which is a very common tool in drug discovery (Klages *et al*., 2007; Mashalidis *et al*., 2013; Pellecchia *et al*., 2008). The assay could be used for the fragment-based screening of the LNS2-PA inhibitors that bind to the LNS2 domain with medium-to-weak affinity. For the assay, short chain PA (e.g., DHPA) should be mixed with a fragment in the presence of a substoichiometric concentration of recombinant LNS2 domain. Additionally, a mixture of the DHPA and the protein should also be prepared in the absence of the fragment. The DHPA concentration should be kept constant in both mixtures. STD, CPMG and waterLOGSY (see Section 4.3) as well as a standard $^{1}$H spectrum of the ligand mixtures should then be recorded. By comparing the DHPA signals in the absence and in the presence of fragments, it can be determined whether displacement of the DHPA from the protein is occurring. The spectra should also be recorded in the absence of the protein to ensure that no aggregation of DHPA is occurring in the presence of the fragments, which could lead to a false negative result in the assay. After identification of displacers, their affinity to the LNS2 domain can be estimated using ligand-observed NMR as described in Section 4.3, or by an orthogonal technique such as surface plasmon resonance (SPR). In order to scale up the screening, automation can be employed for sample preparation and sample insertion into the magnet, as well as data acquisition.

The second proposed assay employs fluorescence lifetime measurements, a robust technique that has recently found its way into drug discovery. The principles of fluorescence lifetime measurements were discussed in Section 4.4. A fluorescently-labeled short-chain PA, such as DHPA-NBD used in this study, can be employed in the assay. To carry out the assay, the ligand should be mixed with the LNS2 domain, a fragment added to the mixture and the lifetime of DHPA-NBD fluorescence measured. If the fragment is a

displacer of DHPA-NBD, the $\tau_2$ lifetime of NBD fluorescence will reduce from ~ 8 ns observed when the ligand is bound to the LNS2 domain closer to ~ 1 ns in the aqueous buffer (see Section 4.4). To avoid interference with the fluorescence of the ligand, fragments should be selected for the lack of intrinsic fluorescence. The assay can be adapted for medium-to-high throughput compound screening by the use of a fluorescence spectrometer fitted with a multi-well microplate reader. Alternatively, instead of the lifetime, anisotropy of a fluorescently-labelled ligand can be measured in the assay. Anisotropy measurements can be performed using polarisation function of, e.g., Analyst™ HT Plate Reader (Molecular Devices, USA).

Other methods of inhibitor identification include structure-guided inhibitor design or *in silico* screening of inhibitors (Zoete *et al*., 2009). However, since structural details of the LNS2 domain are required for such strategies, and no structure of Nir2 LNS2 is available, these methods are not discussed here.

### 4.6.3 Future directions

A number of experiments can be carried out to obtain more details of the mechanism of the LNS2-PA binding.

In order to determine whether the LNS2 domain inserts into the membrane on binding to liposomes, measurements of 1,6-diphenyl-1,3,5-hexatriene (DPH) fluorescence anisotropy can be performed. DPH is a fluorescent dye that incorporates into the lipid membranes and serves as a reporter of membrane fluidity. The reduction of membrane fluidity which can be caused by protein insertion leads to an increase in mobility of DPH, which results in a decrease in anisotropy that can be monitored (Jasniewski *et al*., 2008). Furthermore, steady-state intrinsic fluorescence intensity of the Nir2 LNS2 tryptophans could also be monitored in the absence and presence of lipid vesicles to determine if they participate in the interaction with the membrane.

A quantitative characterisation of the affinity of the LNS2-liposome binding could be achieved by SPR. Sensor chips coated with aliphatic groups could be used for immobilisation of the liposomes, and the protein concentration can be varied to obtain the $K_{app}$. Furthermore, as PA is a known pH sensor in the cell (Shin and Loewen, 2011), dependence of the LNS2-PA binding affinity on the pH of the buffer can be investigated by SPR or the methods used in this study. The effect of salt concentration on the binding affinity can also be assessed to determine the importance of electrostatic interactions in PA

binding. To investigate the role of individual LNS2 residues in the binding, site-directed mutagenesis can be carried out to mutate the residues of interest one-by-one, and assess the PA binding properties of the mutants. This would also require measurements of the CD spectra of the mutants and comparison of the spectra to that of the wild type protein to ensure that the mutations do not lead to significant conformational change in the protein, which could disrupt the binding in a non-specific way.

Lastly, further optimisation of the crystallisation conditions can be carried out to obtain crystals of the LNS2 domain of human Nirs for structure determination by X-ray crystallography. A number of detergents or other additives may be added during purification and/or crystallisation of the protein, which could improve the stability of the protein and potentially promote crystal growth. Screening of detergents/additives can be carried out in the multi-well plate format, as described in the literature for membrane proteins (Parker and Newstead, 2016; Shimizu *et al*., 2008). Detergents with low aggregation number such as CHAPS (Hjelmeland, 1980) may also be used during sample preparation for the structural analysis of the LNS2 domain by protein-observed NMR to prevent or reduce protein aggregation and obtain higher protein concentrations.

# 5. Determination of the crystal structure of Na-FAR-1 complex with oleic acid

## 5.1 Introduction

*Necator americanus* FAR-1 protein (Na-FAR-1) belongs to a family of fatty acid and retinol binding proteins (FARs), which are known to bind a broad spectrum of lipid ligands and have been proposed to play a role in parasitism (see Chapter 1). To understand the mechanisms behind the lipid specificity of FAR proteins that could reveal clues about the biological functions of FARs, structural information is required. To date, the structure of Na-FAR-1 is the most extensively characterised among FARs. Importantly, it has been determined both in the ligand-free form and in the ligand-bound form (Rey-Burusco *et al*., 2015). However, although the structure of ligand-bound Na-FAR-1 was solved, the ligands with which the protein was co-crystallised comprised a mixture of endogenous *E. coli* lipids that were co-purified with the protein after recombinant expression. The heterogeneity of the mixture made identification of the ligands and mapping of the ligand binding sites difficult, and only one binding site was identified with certainty. Hence, it was highly desirable to obtain more structural details about the locations of ligands in the Na-FAR-1-ligand complex and map the protein-ligand interactions with higher accuracy.

In this study, the structure of the Na-FAR-1 complex with one of its natural fatty acid ligands, oleic acid, was determined by X-ray spectroscopy at high resolution (1.29 Å). Oleic acid was chosen as a ligand in the structural investigation as its Na-FAR-1 binding stoichiometry was previously determined by NMR experiments, revealing that the protein is able to bind a maximum of four oleic acid molecules (Rey-Burusco *et al*., 2015). Ligand densities were identified inside the internal cavity of the Na-FAR-1 oleic acid complex, and the interactions between the ligands and Na-FAR-1 amino acids were mapped. As a result, significant clues were obtained about the fatty acid binding mechanism by Na-FAR-1.

In this chapter, a brief overview of protein X-ray crystallography will be given and the structure of the oleic acid complex will be presented. The process of structure determination and identification of the oleic acid binding sites will be described in detail. At the end of the chapter, the fatty acid binding mechanism of Na-FAR-1 will be discussed, and experiments to validate the findings of this study will be suggested.

## 5.2 A brief overview of protein X-ray crystallography

### 5.2.1 The history of protein X-ray crystallography in a flash

It is difficult to find a technique that has played a more significant role in the development of modern biomolecular science than X-ray crystallography. Since the discovery of X-ray diffraction by Max von Laue in 1912 and invention of the X-ray spectrometer by William Henry Bragg in 1913, X-ray crystallography has been an invaluable tool in molecular structure investigations, finding applications in mineralogy, chemistry and other disciplines. Biology and X-ray crystallography first came together when X-ray diffraction was observed from pepsin crystals in 1934 by John Desmond Bernal and his student, Dorothy Hodgkin (Bernal and Crowfoot, 1934), who made many significant contributions to the field of macromolecular X-ray crystallography in the following years (Jaskolski *et al*., 2014). The first protein structure solved by X-ray crystallography was that of myoglobin (1958) (Kendrew *et al*., 1958), shortly followed by the structure of haemoglobin in 1960 (Perutz *et al*., 1960). In last 25 years, technological advances such as emergence of ultra-fast pixel array detectors, tunable synchrotron X-ray radiation sources, cryocrystallography, as well as improvements in the computational power available to researchers allowed protein X-ray crystallography to flourish and become a routine technique in biology. Landmark protein structures determined in this period include high-resolution crystal structures of $F_1$-ATP synthase (Abrahams *et al*., 1994), RNA polymerase II (Cramer *et al*., 2001; Gnatt *et al*., 2001) and the ribosomal 50S and 30S subunits (Ban *et al*., 2000; Harms *et al*., 2001; Wimberly *et al*., 2000). These structures helped to elucidate molecular mechanisms crucial for the life of the cell and lead to several Nobel prizes. In total, more than 130,000 protein and peptide structures determined by X-ray crystallography are deposited on the Protein Data Bank (PDB) today.

### 5.2.2 A brief introduction to the theory of protein X-ray crystallography

#### 5.2.2.1 Diffraction and crystals

X-ray crystallography is based on the fundamental wave-like behaviour of light and the resultant ability of light to diffract when meeting an obstacle, such as a diffraction grating. Crucially, the pattern of the diffracted light and the nature of the obstacle are reciprocally related, and hence the diffraction pattern can be used to obtain information about the obstacle. Like a diffraction grating, molecules are also able to diffract light. Since the

distances between the atoms in a molecule are very short (~ 1-3 Å), X-rays with the wavelength comparable to the interatomic distance are used to observe diffraction.

In molecules, X-rays are scattered primarily by electrons in the atomic electron shells. Scattering occurs when an X-ray beam incident on an electron is absorbed by the electron and then re-emitted as scattered light. Importantly, electrons can scatter X-rays elastically (i.e., the frequency of the incident radiation and the scattered radiation is identical). Due to the small size of molecules, X-ray diffraction from a single molecule is nearly impossible to detect with the current crystallography systems due to the very poor signal-to-noise ratio. Therefore, to obtain information about the structure of a molecule, a crystal is used, where the molecules are arranged in a periodic pattern called the crystal lattice. Since all the molecules in a crystal are identically oriented, the diffraction pattern of a molecule is amplified many times through constructive interference when obtained from a crystal, which makes its detection possible.

The repeated component of the crystal lattice from which the whole crystal can be build is called the unit cell, which is defined by three axial lengths (*a, b, c*) and three interaxial angles (*α, β, γ*). The reciprocals of *a*, *b* and *c* define the dimensions of the reciprocal unit cell, a basic unit of the reciprocal lattice. The reciprocal lattice is a purely imaginary crystallographic concept which, however, is extremely useful for understanding diffraction patterns. The reciprocal lattice can be divided into crystallographic planes specified by the Miller indices (*h, k, l*) that describe the intercepts of the planes with the crystallographic axes. The diffraction spots (reflections) observed in a crystallography experiment correspond to the points on the *(hkl)* planes in the reciprocal lattice, and hence their relative positions and intensities can be related back to the real crystal lattice to obtain the information about the unit cell and its contents, respectively.

#### 5.2.2.2 Bragg's law

There are rules that specify the conditions for observing diffraction from atoms in the crystal lattice, which are to do with the interference effects observed due to the wave-like properties of X-rays. These rules are conveniently formulated by Bragg's law, which can be written in the form of the following equation:

**Equation 5-1.**

$$2dsin\theta = n\lambda$$

where $d$ is the distance between two lattice planes in a crystal, $\theta$ is the angle between the incident X-rays and the planes, $\lambda$ is the wavelength of the incident X-rays and $n$ is an integer. To illustrate Bragg's law, in Fig 5-1, a diagram is shown depicting X-ray scattering from atoms lying on two adjacent lattice planes in a crystal. As can be seen from the diagram, the X-ray beam incident on the bottom plane travels an extra distance before it is scattered, compared to the beam incident on the top plane. The extra distance that the beam travels depends on the separation between the planes in the crystal and the angle of incidence, and is equal to $2dsin\theta$. The X-ray waves scattered from the planes will be out of phase unless $2dsin\theta$ is equal is an integer multiple of the wavelength of the incident X-rays. If the X-rays scatter out of phase, destructive interference will occur, which will lead to cancellation or partial cancellation of the scattered waves. If the $2dsin\theta = n\lambda$ condition is satisfied, the scattered X-rays will be in phase, and constructive interference will lead to magnification of the amplitude of the scattered X-ray waves. As a result, intense reflection spots will be detected. It should be noted that the reflection spots do not only carry information about the atoms lying on the planes diffraction from which satisfies the Bragg's law, but about all the atoms in the crystal that interact with the incident radiation.

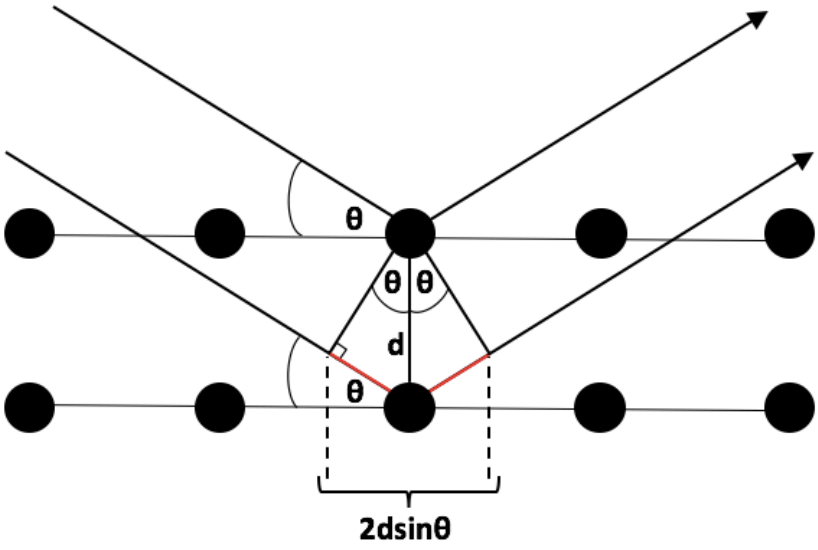



**Figure 5-1. Schematic depiction of X-ray scattering from atoms lying on two lattice planes in a crystal.** The X-ray beam that is incident on the plane shown on the bottom travels more distance than the X-ray beam incident on the plane shown on the top. The difference in the distance travelled by the two waves is highlighted in red and is equal to $2dsin\theta$, where $d$ is the interplanar separation and $\theta$ is the angle of incidence. Intense diffraction spots are observed when $2dsin\theta$ is equal to an integer of the wavelength of the incident X-rays, so that the X-rays scattered from the different lattice planes remain in phase.

### 5.2.2.3 Structure factors and the phase problem

Intensities of the diffraction spots tell us about the distribution of the electron density inside the unit cell, which is crucial for determination of the structure of its contents. In order to extract the information about the electron density from the intensities, structure factors are used. Structures factors are functions that correlate the amplitude and phase of the waves diffracted from crystal lattice planes with the electron density distribution in the unit cell. Structure factor for an (*hkl)* reflection can be expressed by the following equation, where the structure factor is a Fourier transform of the electron density distribution in the unit cell:

**Equation 5-2.**

$$F(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) e^{2\pi i(hx+ ky+ lz)} dxdydz$$

Here, *F(hkl)* is the structure factor, *V* is the volume of the unit cell, *x*, *y*, *z* are the fractional coordinates in the unit cell and *ρ(xyz)* is the electron density at position (*xyz)*. This equation can be transformed to produce the equation describing the electron density distribution in the unit cell as a function of all structure factors *F(hkl)*:

**Equation 5-3.**

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F(hkl)| e^{[-2\pi i(hx+ky+lz)+i\alpha(hkl)]}$$

where |*F(hkl)*| is the amplitude of the structure factor and *α(hkl)* is its phase angle. From Eq. 5-3, it can be seen that in order to calculate the electron density distribution, we need to know both the amplitude of the structure factor, which can be obtained from the intensities of the diffraction spots, and the phase angle. As the information about the phases is lost during detection, the origin of the famous phase problem in crystallography becomes clear.

### 5.2.2.4 Ways to solve the phase problem

Several methods have been developed to solve the phase problem. The most popular ones are molecular replacement (MR), multiple isomorphous replacement (MIR), multiple wavelength anomalous diffraction (MAD) and single wavelength anomalous diffraction (SAD). Of these techniques, MR is by far the easiest and the quickest, but requires access to a structure of a protein with a similar fold. In contrast to MR, each of MIR, MAD and SAD can be used for structure determination *de novo*.

Multiple isomorphous replacement (MIR) is the oldest of the methods, and was famously used to solve the structures of myoglobin and haemoglobin. MIR relies on the attachment of strongly-scattering heavy atoms, such as mercury or lead, to a limited number of sites in the protein. Crucially, the structural arrangement of the crystal and dimensions of the unit cell should not change after attachment of the heavy atoms. The perturbation of the diffraction pattern by the heavy atoms is then used to determine the positions of the heavy atoms in the crystal, which are in turn used to estimate the experimental phases. MAD and SAD also rely on the introduction of heterogeneous atoms into the protein crystal, but instead of strong scatterers, atoms that scatter X-rays anomalously are used, such as selenium or sulphur. Selenium is most commonly introduced into the protein via selenomethionine labelling by recombinant expression in a minimal medium supplemented with selenomethionine using an *E. coli* strain that is a methionine auxotroph. The presence of anomalously scattering atoms in the unit cell leads to the differences in the intensity of observed *(hkl)* reflections versus *(-h-k-l)* reflections, from which the location of the anomalous scatters in the unit cell can be deduced. This information is used to predict the phases for structure determination, in a way similar to MIR. Typically, the SHELX software family (Sheldrick, 2008) is used for obtaining experimental phases from MIR, MAD and SAD data.

Unlike MIR, MAD and SAD, MR does not require any additional experimentation beyond collecting the complete set of reflections from a protein crystal. As mentioned above, MR relies on the use of a previously-solved structure of a protein that is structurally related to the protein of interest. In brief, the known structure is fitted into the unit cell of the crystal via a series of rotations and translations, and the theoretical diffraction pattern predicted for the known structure in the unit cell is compared to the diffraction pattern observed experimentally. If the known structure is positioned correctly, the patterns will match, and the theoretical phases calculated from this model can then be used together with the experimentally determined amplitudes of the structure factors to calculate the electron density distribution in the unit cell and solve the structure of the protein of interest. A number of programs have been developed for performing MR. The most popular ones include MrBUMP (Keegan and Winn, 2008) and Phaser (McCoy *et al*., 2007), which are available in the CCP4 crystallographic software package.

Detailed descriptions of MR, MIR, MAD and SAD are beyond the scope of this thesis, and have been presented elsewhere (Drenth, 2007; Evans and McCoy, 2008; Rossmann, 2001; Rossmann and Blow, 1962; Taylor, 2010).

### 5.2.2.5 Refinement and validation of the model

After the electron density distribution has been calculated, the model of the protein structure can be created by fitting the protein atoms into the determined electron density. The agreement of the model with the experimental data is assessed by comparing the sum of the structure factor amplitudes calculated from the model ($|F_{calc}|$ or $|F_c|$) to the sum of the structure factor amplitudes determined from experimental data ($|F(hkl)|$ or $|F_{obs}|$ or $|F_o|$).

The agreement is typically quantified by calculating the reliability factor (R-factor) using the following equation:

$$R = \frac{\sum \big| |F_o| - |F_c| \big|}{\sum |F_o|} \qquad \textbf{Equation 5-4.}$$

where R is the R-factor. Consequently, if the R-factor is equal to 0, the model and the experimental data are in perfect agreement, and if the R-factor is equal to 1, there is no agreement between the model and the data. R-factor is rarely equal to 0 for protein structures because of their complexity. R-factor also tends to vary with the resolution that can be collected, with high-resolution structures (~ < 1.5 Å) typically demonstrating lower R-factors than medium- or low-resolution structures. Refinement of the model is carried out to increase its agreement with the experimental data. Normally, a small set of reflections is excluded from refinement and used to calculate the so-called free R-factor ($R_{free}$) after the refinement. If no over-fitting of the data has occurred, $R_{free}$ and the R-factor of the refined model ($R_{work}$) should be consistent with each other.

Refinement is performed by changing the coordinates of the protein atoms, temperature factors (B-factors), which describe the degree of displacement of each protein atom due to various molecular effects, and other parameters of the atoms in the model. Refinement can be performed automatically using programs such as REFMAC (Murshudov *et al*., 1999, 1997) and semi-automatically using real-space refinement (Diamond, 1971) in the program Coot (Emsley *et al*., 2010). Typically, both types of refinement strategies are used.

To aid refinement of the model, several programs have been developed to validate the accuracy of the models from the physicochemical perspective. Validation can be carried out by analysing the geometry of the atomic bonds, energetic favourability of the combinations of protein dihedral angles and the presence of interatomic clashes in the model. The most popular tools for structure validation are MolProbity (Chen *et al*., 2010)

and PROCHECK (Laskowski *et al*., 1993); several validation tools are also available in Coot.

Electron density maps are used to visually represent the electron density distribution during refinement. Two most common types of the electron density maps are the $F_o$-$F_c$ map and the $2F_o$-$F_c$ map. The $F_o$-$F_c$ map is often called the difference map, as it simply shows the differences in the electron density distribution calculated from the model and the experimental data. The $2F_o$-$F_c$ map is a composite map which contains the information about the density distribution calculated from the experimental data in addition to the information about the differences between the model and the data. As it represents the protein electron density more accurately, it is commonly used for fitting of the protein atoms during model refinement in Coot.

## 5.3 High-resolution crystal structure of Na-FAR-1 in complex with oleic acid

### 5.3.1 Data collection and structure determination

For structure determination by X-ray crystallography, Na-FAR-1 was co-crystallised with four molar equivalents of oleic acid sodium salt. The details of crystallisation conditions and data collection are given in Chapter 2. High quality data was collected from the largest crystal (approximate dimension ~ 150 μm x 150 μm x 150 μm), with the resolution reaching 1.29 Å. The unit cell was found to have cubic symmetry, and was in the same space group (P 432) as in the previously solved structure of Na-FAR-1 complex with *E. coli* lipids (PDB ID 4XCP) (Rey-Burusco *et al*., 2015). The diffraction pattern recorded from the crystal is shown in Fig 5-2.

The three-dimensional crystal structure of the Na-FAR-1 oleic acid complex was determined by MR using 4XCP as a search model in Phaser. Ligands were excluded from the search model by setting the ligand atom occupancies to zero. The model was refined using iterative cycles of anisotropic refinement in REFMAC, as well as real-space refinement and model building in Coot. The data collection and refinement statistics are summarised in Table 5-1. The final $R_{work}$ and $R_{free}$ values were equal to 0.137 and 0.162, respectively, suggesting good agreement between the model and the experimental data.

Remote data collection at the synchrotron was carried out by Dr Aleksander Roszak (University of Glasgow).

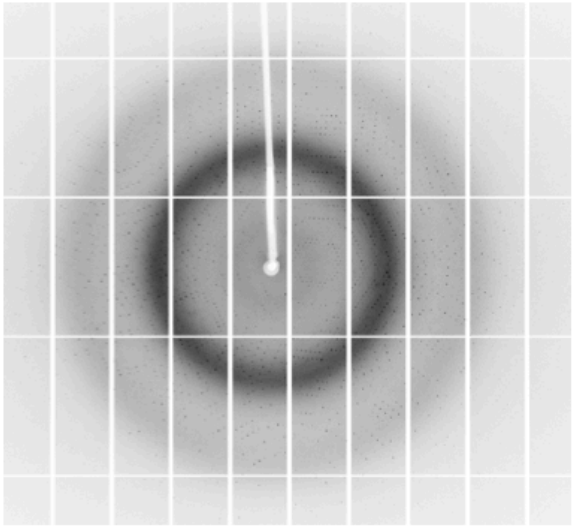


**Figure 5-2. Diffraction pattern recorded from the crystal of Na-FAR-1-oleic acid complex.** Well-resolved reflections can be observed.

| | **Na-FAR-1 oleic acid complex** |
|---|---|
| **Data collection** | |
| Space group | P 432 |
| Cell dimensions | |
| *a, b, c* (Å) | 121.38, 121.38, 121.38 |
| *α, β, γ* (°) | 90.0, 90.0, 90.0 |
| Resolution (Å) | 42.91-1.29 (1.32-1.29) * |
| I/σI | 25.9 (1.5) |
| $R_{pim}$ | 0.016 (0.697) |
| Completeness | 100.0 (100.0) |
| Multiplicity | 21.3 (8.3) |
| | |
| **Refinement** | |
| Resolution (Å) | 42.91-1.29 |
| No. reflections | 73075 |
| $R_{work}/R_{free}$ | 0.137/0.162 |
| Number of atoms | |
| Protein | 1284 |
| Ligand | 80 |
| Water | 313 |
| B-factors | |
| Protein | 23.2 |
| Ligand | 46.0 |
| Water | 53.9 |
| RMS deviations | |
| Bond lengths (Å) | 0.015 |
| Bond angles (°) | 1.800 |

* Highest resolution shell is shown in parentheses.

**Table 5-1. Data collection and refinement statistics for the Na-FAR-1 complex with oleic acid.**

### 5.3.2 Structural features of the Na-FAR-1-oleic acid complex

It was observed that the overall conformation of Na-FAR-1 in complex with oleic acid was very similar to that of 4XCP (Fig 5-3, RMSD for all heavy atoms = 0.52 Å). Like in the 4XCP structure, the protein contained eleven α-helices (α1-11, numbered from the N-terminal to the C-terminal). The helices formed the internal cavity of Na-FAR-1, which serves as the binding site for ligands. The positions of α-helices were nearly identical in the two complexes. The most noticeable difference was in a flexible loop formed by T101, G102 and R103, located between α7 and α8, which had a different conformation in the oleic acid complex, compared to 4XCP. When viewed in the surface representation, the structure of the oleic acid complex appeared slightly more expanded than that of 4XCP. Several entrances to the cavity present in 4XCP either had reduced area or were absent in the oleic acid complex (an example is given in Fig 5-3C).

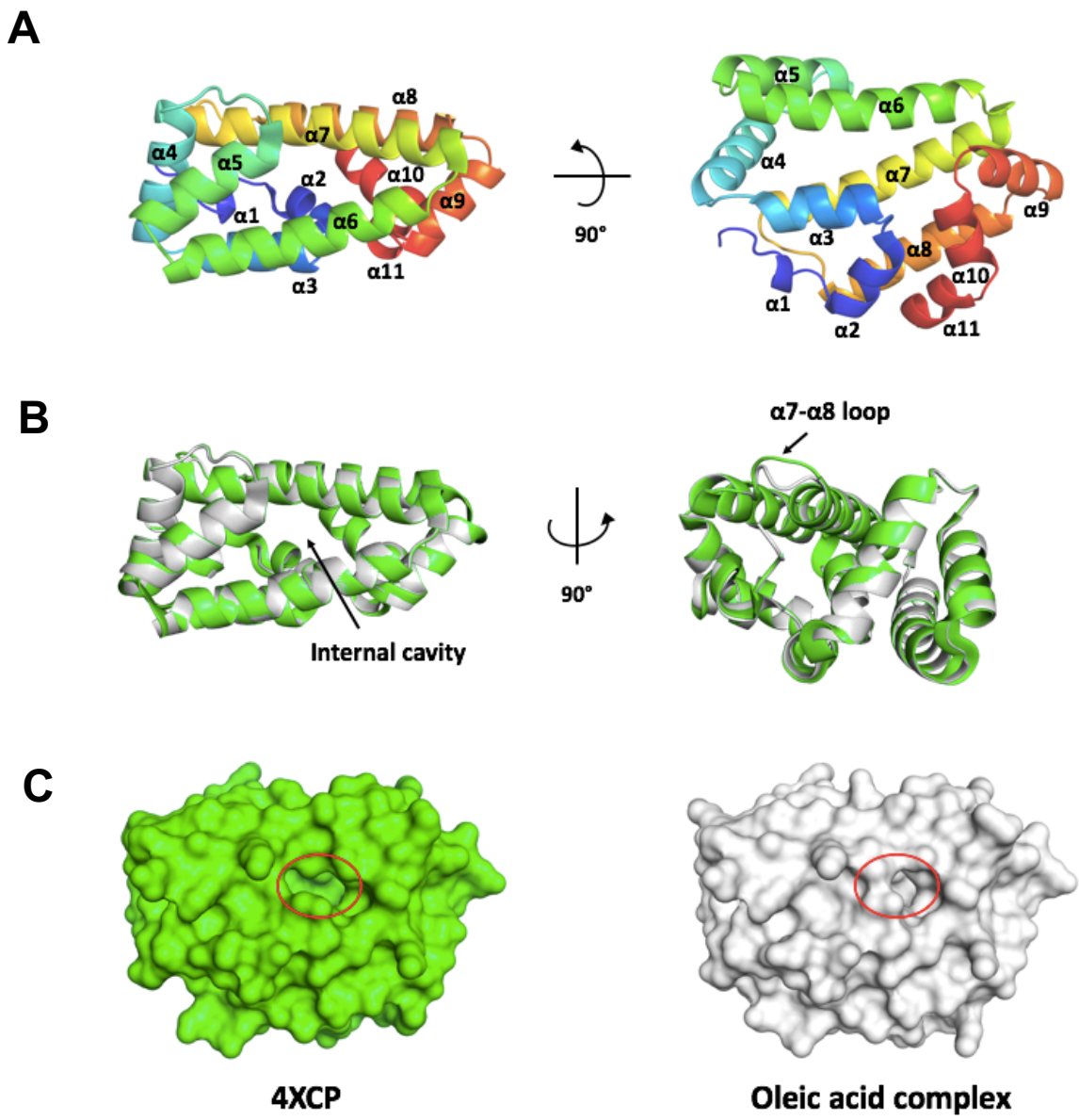



**Figure 5-3. The structure of the Na-FAR-1 oleic acid complex and its comparison with the structure of the Na-FAR-1 complex with *E. coli* lipids (PDB ID 4XCP). A**. The structure of the oleic acid complex is shown in cartoon representation, coloured from blue to red (N-terminus to C-terminus) with the α-helices numbered. **B.** Superimposed cartoons of Na-FAR-1 in complex with oleic acid (grey) and with *E. coli* lipids (green) [PDB ID 4XCP, (Rey-Burusco *et al*., 2015)]. The location of the internal cavity of Na-FAR-1 is marked in the structures of the left. In the structures on the right, α7-α8 loop that has different conformations in the two complexes is marked. **C.** A representative small conformational difference on the surface of the two complexes. One of the entrances to the internal cavity is partially closed in the oleic acid complex. The location of the entrance is highlighted in red. The colour-coding is identical to **B**.

Consistent with these observations, analysis by the protein surface topology tool CASTp (Binkowski *et al*., 2003) revealed that the internal cavity volume of Na-FAR-1 oleic acid complex was slightly larger than that of 4XCP when measured both with a 1.4 Å probe equivalent to a water molecule (2983 $Å^3$ vs 2570 $Å^3$, respectively) and with a 1.925 Å probe equivalent to a $CH_2$ group (2437 $Å^3$ vs 2170 $Å^3$).

Previously, it was demonstrated that Na-FAR-1 expands on ligand binding (Rey-Burusco *et al*., 2015). From the data presented here, it can be suggested that the degree of Na-FAR-1 expansion depends on the nature of the ligands bound to the protein or the stoichiometry of the protein-ligand complex. In 4XCP, Na-FAR-1 was believed to contain a 16-carbon palmitic acid molecule inside its cavity, along with other unidentified lipids (Rey-Burusco *et al*., 2015). In contrast to 4XCP, the structure reported here contained 18-carbon oleic acid molecules, as the majority of the *E. coli* lipids were removed by the reverse-phase HPLC (see Chapter 3), and any remaining lipids were expected to have been displaced by the excess of oleic acid. It appears that in the oleic acid complex, the internal cavity of Na-FAR-1 has expanded further to accommodate either the larger ligands, the higher loading, or both. The apparent ability of Na-FAR-1 to adjust the volume of its internal cavity to accommodate different types of molecules might be one of the reasons behind the observed diversity ligands that Na-FAR-1 can bind.

### 5.3.3 Location of the oleic acid binding sites in the Na-FAR-1 internal cavity

#### 5.3.3.1 Ligand fitting

Volumes of electron density were observed inside the internal cavity of Na-FAR-1-oleic acid complex in the $2F_o$-$F_c$ map, into which the ligands were fitted. Prior to fitting of the ligands, water molecules were introduced into the model in Coot. As the electron density of the ligands appeared less well defined than the electron density of the protein atoms, fitting of some of the ligands proved to be a challenging task. Ligand modelling was carried out by first locating the electron density patches that could be fitted with the carboxylic head groups of oleic acid. The preference was given to the patches located near the polar groups of the protein residues, as it was postulated that it would be energetically favourable for the oleic acid head groups to make polar contacts to the protein. Subsequently, the oleic acid aliphatic chains were extended into the electron density protruding from the identified head group densities. If negative or positive density peaks surrounding the ligand atoms were observed in the in the $F_o$-$F_c$ map after refinement, the positions of the ligands or the occupancies of the ligand atoms were adjusted manually, and the automatic refinement was repeated until the difference peaks either disappeared or at least reduced were in size in the $F_o$-$F_c$ map. The change in the R-factor was also monitored after each refinement step, and the models that displayed the lowest $R_{work}$ and $R_{free}$ values that were in agreement with each other were selected. In the end, four oleic acid molecules were fitted into the cavity (OLA1-4, Fig 5-4A).

It should be noted that inconsistencies between the model and the experimental data were still expected to be present in the final model. For instance, after the final model was refined, many fragments of positive electron density were still found inside the cavity in the $F_o$-$F_c$ map, especially near OLA2, 3 and 4 (Fig 5-4B). This density could originate from different conformations of the oleic acid aliphatic tails, which are expected to be highly flexible. However, since it was difficult to find unambiguous solution to model the alternative conformations, these remaining positive $F_o$-$F_c$ density regions near the modelled ligands were left unfilled.

In the cavity, all four oleic acid molecules were oriented in a similar way. Their carboxylic acid head groups were contacting either the solvent or the polar groups at the surface of the cavity, whereas their aliphatic tails were hidden in the interior of the cavity. The aliphatic tails of the ligands formed an apparent hydrophobic core inside the cavity, the presence of which may prevent collapse of the cavity structure observed in the ligand-free form of Na-FAR-1 (Rey-Burusco *et al*., 2015). The positions of each of the ligand molecules are presented below.

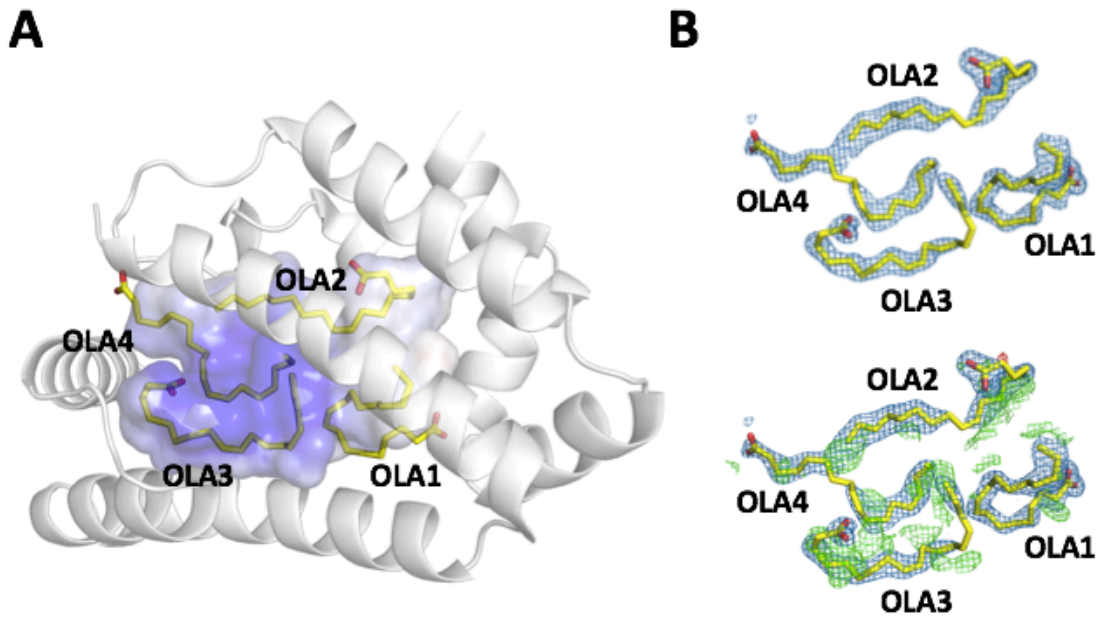



**Figure 5-4. Ligand molecules inside the Na-FAR-1 complex with oleic acid. A**. Cartoon representation of the Na-FAR-1-oleic acid complex structure. The surface of the Na-FAR-1 internal cavity is shown within the backbone cartoon, coloured according to its electrostatic potential (blue and red represent positive and negative potential, respectively). The oleic acid molecules inside the cavity (OLA1-4) are depicted as sticks. **B**. Electron density of OLA1-4. Top. The $2F_o$-$F_c$ map (blue mesh). Bottom. The $2F_o$-$F_c$ map (blue mesh) superimposed with the $F_o$-$F_c$ difference map (green and red mesh for positive and negative peaks, respectively). Ligands are depicted as sticks. $2F_o$-$F_c$ map is contoured at 1σ and carved at 1.6 Å around the ligands. The $F_o$-$F_c$ map is contoured at 3σ and carved at 2.0 Å around the ligands.

### 5.3.3.2 Ligand binding sites

OLA1 was fitted into the most well-defined electron density inside the internal cavity of Na-FAR-1 in the $2F_o$-$F_c$ map (Fig 5-4B). The ligand was located close to an opening in the protein surface between helices α3, α6 and α10. The OLA1 head group was in a position to make hydrogen bonds to the backbone amide groups of F138 (2.9 Å) and L139 (3.0 Å), and a polar contact with a solvent molecule (Fig 5-5A). The fatty acyl tail of OLA1 contacted the apolar groups in the side-chains of P15, M69, V70, S88, I89, F136, F138 and L139. It should be noted that the protonation state of the carboxyl acid group of oleic acid molecules in the protein-ligand complex was unclear, as the pH of the mother liquor (~ 4.5) was too close to the theoretical pKa of oleic acid in water (5.0) predicted by ChemAxon model (www.chemaxon.com). However, the ligands were predicted to be deprotonated by the Protoss software (Bietz *et al*., 2014) based on the structural features of the complex, and hence were drawn accordingly in Fig 5-5.

In contrast to OLA1, the electron density of OLA2 was less defined. For instance, in the final model, no electron density was observed around ligand carbons 2, 3 and 10 in the $2F_o$-$F_c$ map contoured at 1σ, which may be explained by a higher conformational flexibility of the ligand regions containing these atoms compared to the rest of the molecule. A small negative density peak was observed close to carbon 2 in the $F_o$-$F_c$ map contoured at 3.0σ, which was not completely removed by refinement. This peak might have arisen due to the problems in the model caused by the incomplete modelling of alternative conformations of oleic acid tails, as mentioned above. Unlike OLA1, the head group of OLA2 was not located near a cavity entrance, but was hidden inside the cavity, contacting a cavity wall formed by α7 and α8. The carboxylic acid group of OLA2 made hydrogen bonds to the hydroxyl groups of Y10 (2.7 Å) and S113 (2.6 Å) (Fig 5-5B). The aliphatic tail of the molecule was curved at carbon 4, after which point it extended away from the wall into the central cavity. The ligand had hydrophobic contacts to the side-chains of I6, L13, M14, A92, I95, Y99, V114 and V117, as well as to the aliphatic tail of OLA4.

The carboxylic acid group of OLA3 was fitted into a region of electron density near the side-chain of K96 in conformation A (K96a). OLA3 was located near a large region of positive electrostatic potential in the cavity next to the largest cavity opening between α4, α5 and α7. In this position, the carboxylic acid group of OLA3 appeared to make hydrogen bonds to the side-chain amine group of K96a (2.8 Å) and the hydroxyl group of Y100 (2.7 Å) (Fig 5-5C), as well as contacted a solvent molecule. In addition, the ligand made hydrophobic contacts to the side-chains of L33, V36, F37, T45, S48, I49, V63, H67 and

the fatty acyl tail of OLA4. The aliphatic tail of OLA3 was curved in way similar to the OLA2 tail, making a turn after carbon 3 and extending into the cavity.

Finally, OLA4 was placed into the electron density close to OLA2 and OLA3. The head group of OLA4 was located close to the cavity opening found between α4, α7 and the N-terminus of the protein. Its carboxylic oxygen atoms made hydrogen bonds with the backbone amide of F1 (3.1 and 3.2 Å) (Fig 5-5D) and a polar contact with a solvent molecule. The ligand also contacted the hydrophobic side-chains of F1, F21, L22, K30, L33, K34 and Y100, as well as the fatty acyl tails of OLA2 and OLA3. It should be noted that the electron density of OLA4's head group was rather poorly defined in the $2F_o$-$F_c$ map compared to the other ligands. This might suggest that at least the head group region of OLA4 is relatively mobile and might occupy an alternative, unknown binding site in the cavity. It should also be noted that F1 is the N-terminal residue in the secreted form of Na-FAR-1, as the residues preceding F1 in the structure reported here belong to the recombinant tag. However, as in the native form of the protein the amino group of F1 is expected to carry a positive charge and thus would be able to form a salt bridge with the negatively charged group of oleic acid, it is likely that the interaction between F1 and OLA4 will also be observed in the native protein.

To summarise, four oleic acid binding sites in the cavity were mapped. The head groups of all the ligand molecules appeared to be hydrogen bonded to either backbone amide groups or the polar side-chain groups of the protein residues. The head groups of three oleic acid molecules (OLA1, OLA3 and OLA4) were located close to the cavity openings, whereas the head group of one ligand (OLA2) was hidden inside the cavity. The head groups of OLA1, OLA2 and OLA4 contacted structured waters in their binding sites. The aliphatic tails of all ligands made a large number of contacts with the hydrophobic amino acid side-chains in the internal cavity of the protein. Hence, fatty acid binding by Na-FAR-1 is expected to be predominantly hydrophobic in nature, although polar interactions also appear to play a role in positioning the ligands.

The proposed mechanism of Na-FAR-1-fatty acid binding is discussed below.

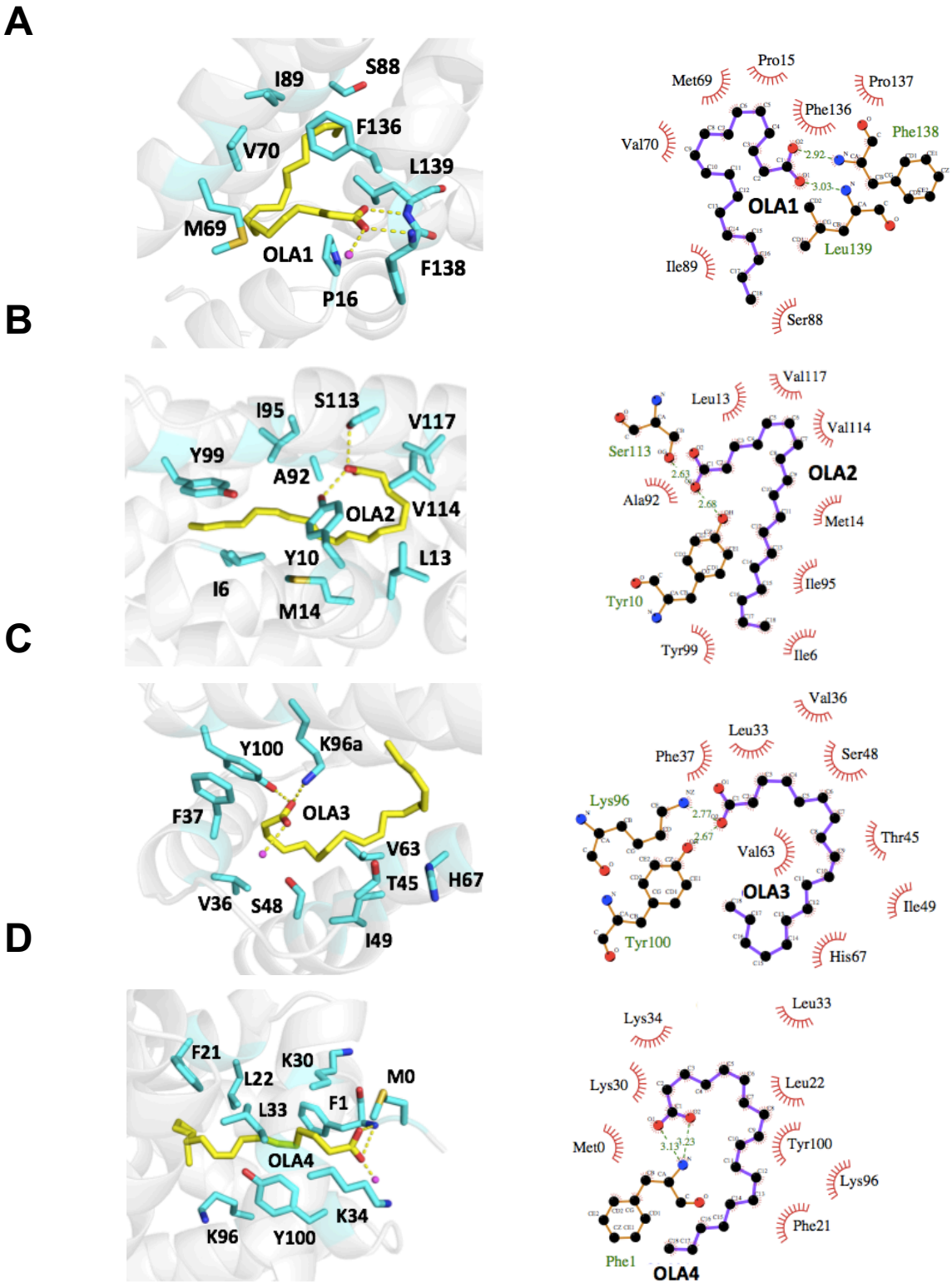



**Figure 5-5. Oleic acid binding sites in the Na-FAR-1-oleic acid complex. A**. OLA1 binding site. **B**. OLA2 binding site. **C**. OLA3 binding site. **D**. OLA4 binding site. Left column: OLA1-4 molecules and Na-FAR-1 residues contacting OLA1-4 are shown as sticks (yellow and teal, respectively). Predominant conformations of Na-FAR-1 residue side-chains are shown. In cases when the side-chain conformations were modelled in two conformations, the conformation closest to the ligand is shown. Water molecules that make polar contacts to OLA1-4 are shown as spheres (magenta). Polar contacts between the carboxylic groups of OLA1-4 and either protein residues or water molecules are depicted as yellow dashed lines. Right column: LigPlot depictions (Wallace *et al*., 1995) of the interactions between OLA1-4 and the protein. Residues making polar contacts to the ligands are shown as balls-and-sticks. Hydrogen bonds between the protein residues and OLA1-4 are shown as green dashed lines (bond length is in Å). Hydrophobic contacts are represented as red arcs with spikes oriented towards the contacting atom.

## 5.4 Conclusions and discussion

### 5.4.1 Comparison of the Na-FAR-1 ligand complexes and implications for the mechanism of fatty acid binding by Na-FAR-1

The structure of the Na-FAR-1 oleic acid complex was determined at a high resolution (1.29 Å), which allowed fitting of protein residues with high accuracy. As reported above, the conformation of Na-FAR-1 in the oleic acid complex was found to be largely similar to that of Na-FAR-1 in complex with *E. coli* lipids (PDB ID 4XCP). An important difference, however, was in the volume of the internal cavity, which was larger in the oleic acid complex than in 4XCP. As noted above, it is likely that the ability to change the volume of its cavity might help Na-FAR-1 to accommodate lipids of varying sizes and shapes or form lipid complexes with different protein-ligand stoichiometries. It is, however, unclear what the maximum volume that the internal cavity of Na-FAR-1 can expand to is.

The increase in the cavity volume was accompanied by reduction in the sizes of the cavity openings in the oleic acid complex, which was a consequence of the conformational changes in the side-chains of Na-FAR-1 residues. However, cavity openings were still present in the structure, and the head groups of three oleic acid molecules were located close to the openings, likely due to the local environmental polarity. Furthermore, three oleic acid molecules appeared to contact structured water molecules, one of which was located in the interior of the cavity. Hence, it can be suggested that water plays a specific role in the oleic acid binding by Na-FAR-1 by making polar contacts to the oleic acid head groups, and thereby aiding the positioning of the ligands inside the cavity in a thermodynamically favourable manner.

The aliphatic tails of different oleic acid molecules contacted one another in the internal cavity of Na-FAR-1. If we assume that oleic acid molecules bind Na-FAR-1 in a sequential manner, which is apparent from the ability of Na-FAR-1 to form a number of distinct complexes with different protein-ligand stoichiometries (Rey-Burusco *et al*., 2015), it is possible that binding of at least some of the ligands might occur via a cooperative mechanism. Here, binding of one ligand might lead to an increase in affinity for binding of the next ligand due to the increase in the hydrophobicity of the cavity interior as a result of ligand binding. The same is likely to hold true for other fatty acids with chain lengths similar to that of oleic acid.

As mentioned above, differences in the conformations of many amino acid side-chains in the cavity were observed between 4XCP and the oleic acid complex. The most interesting example here is K96, which was located in the region of the cavity with a large positive electrostatic potential and was involved in ligand binding. The side-chain of K96 was found to have two conformations in the oleic acid complex (K96a and K96b) and only one in 4XCP (Fig 5-6). In the oleic acid complex, K96a was the predominant conformer (70% occupancy), which made a hydrogen bond with the carboxylic head group of OLA3. In 4XCP, the side-chain of K96 also appeared to contact the ligand, but the amino acid was in a conformation that was equivalent to K96b in the oleic acid complex and not to K96a. As the ligand in 4XCP was believed to be palmitic acid rather than oleic acid, the change in the position of the side-chain could have been caused by the structural differences between the ligands in the two structures. Hence, it can be proposed that the side-chain of K96 may help the protein to accommodate fatty acids with different aliphatic tail lengths by flipping between at least two conformations, and changing the spatial arrangement and the local charge distribution of the cavity in this way. Additionally, K96 appears to regulate the size of the largest cavity entrance, which is partially obstructed by K96's side-chain in the oleic acid complex.

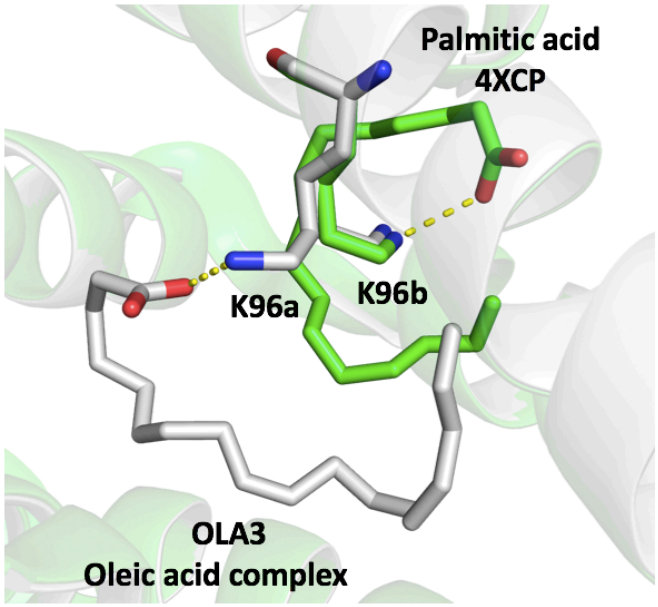



**Figure 5-6. Conformation of Na-FAR-1 K96 side-chain in the oleic acid complex and in the *E. coli* lipid complex containing palmitic acid (PDB ID 4XCP).** In the oleic acid complex (grey cartoon), K96 is found in two conformations (K96a and K96b). K96a is the predominant conformation (70% occupancy), in which the residue makes a hydrogen bond with an oleic acid molecule (OLA3, grey sticks). In 4XCP (green cartoon) (Rey-Burusco *et al*., 2015), K96 is only found in a conformation equivalent to K96b, in which the residue makes a hydrogen bond with a palmitic acid molecule (green sticks). The change in the K96 side-chain conformation might be required to accommodate different types of ligands. Hydrogen bonds between the ligands and K96 are represented as yellow dashes.

To summarise, from the structures of Na-FAR-1 complexes with oleic acid and *E. coli* lipids, a number of propositions can be made regarding the Na-FAR-1 mechanism of fatty acid binding. First, the internal cavity of Na-FAR-1 can expand to varying degrees to accommodate different types of ligands or allow formation of complexes with varying protein-ligand stoichiometries. Second, fatty acid binding by Na-FAR-1 involves both polar and apolar interactions. Apolar interactions are, likely, the major contributors to the binding energy of Na-FAR-1, whereas polar interactions might be responsible for fixing the ligands in an energetically favourable conformation inside the cavity. Third, the conformation of certain protein residues, such as K96, may regulate the ligand binding properties of Na-FAR-1, and may change the local arrangement of the cavity to accommodate fatty acids of different chain length. Fourth, the binding of fatty acids by Na-FAR-1 is likely to be cooperative due to the hydrophobic effect associated with the aliphatic tails of the fatty acids.

Apparent differences between the mechanisms of fatty acid and phospholipid binding by Na-FAR-1 will be discussed in Chapter 6.

### 5.4.2 Future directions

The results presented here suggest several experimental avenues that could be explored to test the hypotheses put forward.

The role of specific Na-FAR-1 residues in oleic acid binding could be tested by site-directed mutagenesis. For that purpose, the residues whose side-chains appear to make polar to the ligands head groups in the crystal structure can be mutated to, e.g., alanine residues that lack polar side-chains. The influence of the mutations on binding affinity could then be assessed, for example, by a fluorescence-based assay which was used previously for analysis of lipid binding by FARs and other lipid binding proteins (Basavaraju *et al*., 2003; Mei *et al*., 1997), and the principle of which is described in Chapter 6. To complement the functional analysis, structures of Na-FAR-1 mutants in complex with oleic acid can be determined by X-ray crystallography to characterise the effects of mutations on the positioning of ligands in the complex. It should be noted that three Na-FAR-1 residues (F1, F138 and L139) appeared to make hydrogen bonds to oleic acid via their backbone amides. Therefore, it would be difficult to use site-directed mutagenesis to confirm the role of these residue in oleic acid binding, as mutation of the residues to prolines would be required to disrupt the hydrogen bonds between the

backbone amides of the residues and the oleic acid head groups, which could have a significant impact on the protein backbone conformation.

As the fitting of some of the ligands proved to be challenging, oleic acid labelled with strongly-scattering bromine atoms at specific aliphatic tail carbons could be used to confirm the location of the ligands in the internal cavity of Na-FAR-1 and potentially simplify the fitting procedure. Previously, brominated lipids were successfully used to identify lipid binding sites in the photosynthetic reaction center of *Rhodobacter sphaeroides* (Roszak *et al*., 2007). However, it should be noted that the presence of the large bromine atoms in the aliphatic tail of the ligands might affect the positioning of the lipids inside the Na-FAR-1 cavity, and hence the structure of the Na-FAR-1 complex with brominated oleic acid might not appear in agreement with the structure reported in this study.

Furthermore, the mechanism of the interaction between Na-FAR-1 and fatty acids of different chain lengths could be investigated by X-ray crystallography. However, due to the structural similarities between the long-chain fatty acids that Na-FAR-1 shows a preference for, their mechanism of binding is not expected to be radically different from the one observed with oleic acid, and, hence, is unlikely to provide new information. However, determination of the crystal structure with, for example, 14-carbon myristic acid would help to confirm or refute the proposed role of K96 side-chain conformation in ligand binding by Na-FAR-1.

# 6. Analysis of phospholipid binding by Na-FAR-1

## 6.1 Introduction

The structure of *Necator americanus* FAR-1 protein (Na-FAR-1) in complex with one of its natural fatty acid ligand, oleic acid, was solved in this study. As a result, clues about the fatty acid binding mechanism of Na-FAR-1 were obtained, which were discussed in Chapter 5.

Recently, it has also been demonstrated that Na-FAR-1 is also able bind phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) (Rey-Burusco *et al*., 2015). PG and PE belong to a class of molecules termed phospholipids, which are structurally distinct from fatty acids and other known ligands of Na-FAR-1, such as retinol and retinoic acid. Although the biological importance of phospholipid binding by Na-FAR-1 is unknown, lipid binding by secreted nematode lipid binding proteins has been previously proposed to play a role in parasitism by facilitating nutrient acquisition from the host and/or by host immunomodulation through sequestration of signalling lipids (Basavaraju *et al*., 2003; Kennedy, 2000). Therefore, it can be proposed that phospholipid binding by Na-FAR-1 might play a similar role in the parasite infections, especially since soluble phospholipids, such as lysophosphatidic acid (LPA) (Yun and Kumar, 2015; Zhao *et al*., 2015; Zhao and Natarajan, 2013), are known mediators of inflammatory responses.

In order to obtain insights into the mechanism that Na-FAR-1 employs for phospholipid binding, structural and functional characterisation of a Na-FAR-1-phospholipid complex was carried out. Protein nuclear magnetic resonance (NMR) spectroscopy and a fluorescence-based lipid binding assay were employed to characterise Na-FAR-1-phospholipid binding. Phosphatidic acid (PA) was chosen as the model phospholipid for the binding analysis. In addition, Na-FAR-1 binding to LPA was demonstrated.

In the beginning of the chapter, the results of the fluorescence-based assay used to demonstrate binding of PA and LPA to Na-FAR-1 will be presented, and the principle of the assay will be described. Next, an overview of protein NMR spectroscopy will be given, and the results of the NMR experiments will be presented. The chapter will end with the discussion of the results and proposals for future study of phospholipid binding by Na-FAR-1.

## 6.2 Na-FAR-1 binds phosphatidic acid and lysophosphatidic acid

In order to characterise phospholipid binding by Na-FAR-1, a fluorescence-based competition assay was carried out that was previously used for characterisation of the Na-FAR-1-oleic acid interaction. The assay employs 11-(Dansylamino)undecanoic acid (DAUDA), an environment-sensitive fluorescent dye that is weakly fluorescent in polar solvents and strongly fluorescent in a hydrophobic environment. When transferred from apolar into a polar environment, a red shift of DAUDA emission maximum is observed.

The principle of the assay is based on the fluorescence properties of DAUDA. For the assay, protein-DAUDA complexes are premade in an aqueous buffer, and the fluorescence emission spectrum of DAUDA is measured. When bound to a protein, DAUDA is in a hydrophobic environment, and hence will fluoresce with high intensity. Subsequently, a ligand of interest is added to the premade protein-DAUDA complexes, and the fluorescence emission spectrum is recorded again. If the binding site of the ligand and DAUDA overlap, the ligand will at least partially displace DAUDA from the protein. As a certain population of DAUDA molecules will now be found in a polar environment, the total measured fluorescence intensity will reduce, and the fluorescence emission maximum of DAUDA will undergo a red shift. In this way, protein-ligand binding can be characterised by following the intensity and wavelength of DAUDA fluorescence emission during ligand titration.

Dioleoyl PA (DOPA or di-18:1 *cis*-9 PA) was chosen as a model phospholipid for analysis of Na-FAR-1 phospholipid binding, as PA forms the basic unit of all glycerophospholipids in the cell. DOPA was titrated into the mixture of Na-FAR-1 and DAUDA, and displacement of DAUDA by DOPA was monitored by measuring the emission spectrum of DAUDA. It was observed that DOPA displaced DAUDA from Na-FAR-1, as was evident from a decrease in DAUDA fluorescence emission intensity and the shift in the fluorescence emission maximum from 470 nm to 530 nm after sequential additions of DOPA (Fig 6-1). After the concentration of DOPA was brought to 8 μM, the intensity of the DAUDA emission maximum was found to be equal to 11% of its intensity in the absence of DOPA, and did not decrease in the further titration steps.

Next, it was decided to investigate whether Na-FAR-1 binds oleoyl lysophosphatidic acid (OLPA or 18:1 *cis*-9 LPA), a soluble phospholipid involved in inflammatory signalling that shares structural similarity with DOPA and that Na-FAR-1 is likely to encounter in its natural environment after secretion by the worm. It was observed that OLPA was also able

to displace DAUDA from Na-FAR-1, as was evident from the decrease in DAUDA fluorescence observed on OLPA addition. The ability of OLPA to displace DAUDA suggests that OLPA also binds to Na-FAR-1 and shares a binding site with DAUDA.

Once it was determined that Na-FAR-1 is able to bind DOPA and OLPA, characterisation of DOPA binding by Na-FAR-1 was carried out by using protein NMR spectroscopy.

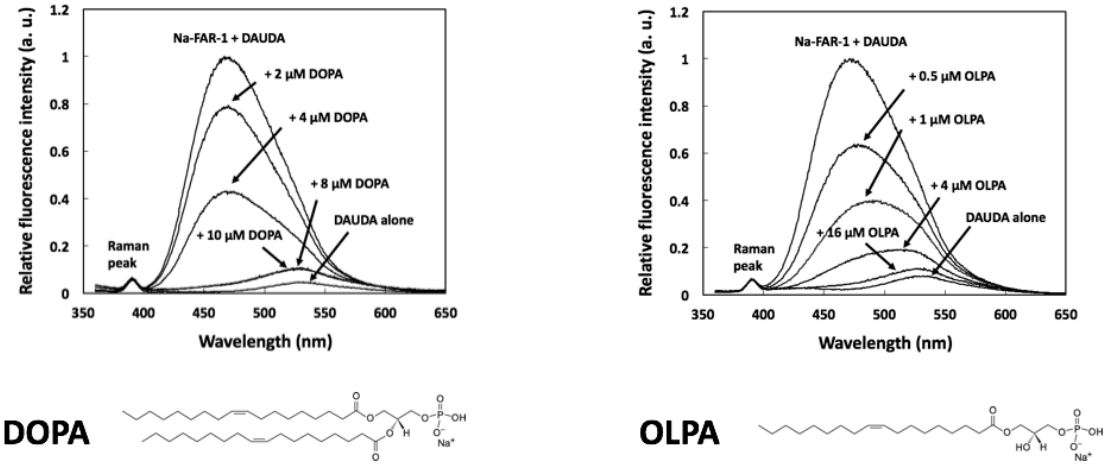



**Figure 6-1. DAUDA displacement assay using di-oleoyl phosphatidic acid (DOPA, di-18:1 *cis*-9 PA) and oleoyl lysophosphatidic acid (OLPA, 18:1 *cis*-9 LPA).** Fluorescence emission spectrum of DAUDA is shown ($\lambda_{ex}$ = 345 nm). **A**. DAUDA is displaced from Na-FAR-1 by sequential additions of DOPA to 2, 4, 8 and 10 µM total DOPA concentration. **B**. DAUDA is displaced from Na-FAR-1 by sequential additions of OLPA to 0.5, 1, 4 and 16 µM total OLPA concentration. A Raman scattering peak is observed at 390 nm in both spectra. Structures of DOPA and OLPA are shown for reference.

## 6.3 Analysis of phosphatidic acid binding by Na-FAR-1 using protein-observed NMR spectroscopy

### 6.3.1 A brief overview of protein NMR spectroscopy

#### 6.3.1.1 Introduction

Compared to X-ray crystallography described in Chapter 5, NMR spectroscopy is a relative newcomer to biology. Although the phenomenon of NMR in bulk materials was first observed separately by Bloch (Bloch *et al*., 1946) and Purcell (Purcell *et al*., 1946) in 1946, it took decades of technological and scientific advances for NMR spectroscopy to become an established technique in biomolecular science and structural biology. Today, however, it is widely used for various type of biochemical and biophysical investigations, including analyses of protein structure, dynamics and protein-ligand interactions. Although

structure determination by NMR spectroscopy is fundamentally difficult for biomolecules with MW larger than ~ 25-35 kDa (see Section 4.3.1.1 for a brief discussion of protein NMR spectroscopy limitations), numerous structures of protein and protein domains have been solved by NMR spectroscopy. To date, Protein Data Bank (PDB) includes more than 10,000 solution structures of protein and peptides determined using NMR spectroscopy.

#### 6.3.1.2 From nuclear spin to NMR spectrum

NMR spectroscopy uses magnetic properties of atomic nuclei to obtain information about molecular structure and dynamics. To understand how this is achieved, the concept of spin has to be introduced. Since there is no macroscopic equivalent to spin, it is difficult to understand the concept of spin intuitively. In basic terms, spin can be defined as quantum mechanical property of elementary particles that can be understood as a form of angular momentum. It should be noted, however, that in contrast to classical angular momentum, spin angular momentum does not arise due to the rotation of a particle, but is its intrinsic property. As nuclei consist of elementary particles, they also carry spin angular momentum. Spin angular momentum of particles and nuclei is characterised by spin quantum number. The nuclear spin quantum number depends on the spin quantum numbers of the elementary particles that form the nucleus. If the nuclear spin quantum number is non-zero, then the nucleus possesses a spin magnetic moment and can be investigated by NMR spectroscopy. Typically, nuclei with spin quantum number equal to 1/2 (further, spin 1/2 nuclei) are used in NMR spectroscopy, as they have convenient magnetic properties for investigation by NMR spectroscopy. The ratio of the magnetic moment of a nucleus to its spin angular momentum is called the gyromagnetic (or magnetogyric) ratio, the value of which is unique to each isotope.

When a nucleus with non-zero spin is placed into a magnetic field, its spin magnetic moment will precess around the magnetic field. The precession occurs at a frequency termed the Larmor frequency ($\omega$), which is defined by the following equation:

$$\omega = -\gamma B \qquad \textbf{Equation 6-1.}$$

where $\gamma$ is the gyromagnetic ratio and $B$ is the magnetic field strength at the position of the nucleus. The energy of the interaction between spin magnetic moment of a nucleus and the magnetic field depends on the orientation of the magnetic moment in relation to the direction of the magnetic field. The energy of the interaction will be lowest when the spin

magnetic moment is oriented in parallel to the magnetic field, and the highest when the orientations of the magnetic moment and the magnetic field are opposite.

Next, let us consider a bulk sample containing numerous identical nuclei. In the absence of the external magnetic field, spin magnetic moments of individual nuclei will be oriented randomly due to the thermal motion of the nuclei. If we place the sample into an external magnetic field, slight anisotropy of the spin magnetic moment orientations will be observed, as the magnetic moments of the nuclei will tend to align in parallel to the magnetic field. This will lead to build up of a net nuclear magnetic moment, which will be oriented in the direction of the magnetic field. In NMR spectroscopy, this net magnetic moment is commonly referred to as longitudinal magnetisation. However, due to the small energy difference between the highest energy and the lowest energy arrangements of the nuclear magnetic moments in the sample, the net magnetic moment is relatively small compared to the diamagnetism of the sample caused by the interaction of atomic electrons with the magnetic field. Hence, longitudinal magnetisation is essentially immeasurable directly.

In order to detect nuclear magnetisation, the sample can be irradiated with a radiofrequency pulse of a specific length and energy that will rotate the net nuclear magnetisation vector into the plane perpendicular to the external magnetic field. As a result, the sample will acquire transverse magnetisation. As the nuclei will continue to experience the external magnetic field, the net transverse magnetic moment will precess at the Larmor frequency of the nuclei, generating a small magnetic field. If a coil is wound around the sample, this precession will induce an oscillating electric current in the coil, which can be amplified and measured. The oscillation will decay over time as the transverse magnetisation of the sample decays away due to transverse (or $T_2$) relaxation. The oscillating current is called the free-induction decay (FID) and corresponds to the NMR signal in the time domain. In order to convert the signal into the frequency domain to produce an NMR spectrum, a Fourier transform is performed on the FID. Phasing of the signal is then carried out to obtain a traditional Lorentzian absorption peak. The position of the peak in the spectrum will depend on the Larmor frequency of the corresponding nuclei, whereas the intensity of the peak will depend on the number of the nuclei with the identical Larmor frequency in the sample.

### 6.3.1.3 Chemical shift, J-couplings and nuclear Overhauser effect (NOE)

When NMR measurements are made of a solution of molecules, nuclei of the same isotope in a molecule do not necessarily experience identical magnetic field and therefore may not have identical Larmor frequencies. This is because the strength of the external magnetic field that the nuclei experience is affected by the local magnetic effects in the molecule. These local magnetic effects are caused by the magnetism of the atomic electrons, or of the neighbouring nuclei, which interferes with the external magnetic field. The information provided by the local magnetic field inhomogeneities in a molecule is thus of immense importance, as it reports on the chemical structure of the molecule. Reduction in the Larmor frequency of a nucleus as a result of the local magnetic effects is commonly referred to as shielding (from the external magnetic field), whereas an increase in the Larmor frequency of a nucleus is called deshielding.

From Eq. 6-1 it can be seen that nuclear Larmor frequency depends on the strength of the magnetic field that a nucleus experiences. Therefore, the same nucleus will display NMR signals at different frequencies when placed into NMR spectrometers that produce magnetic fields of different strengths. To standardise NMR measurements, the positions of NMR signals in the spectrum are usually not presented in Larmor frequencies, but in field-independent chemical shifts. To determine a chemical shift of a nucleus, its Larmor frequency is measured in a magnetic field of a particular strength and is then compared to a Larmor frequency of an isotopic nucleus in a reference compound that was measured in a magnetic field of the same strength. The chemical shift ($\delta$) of a nucleus is expressed by the following equation:

**Equation 6-2.**

$$\delta = \frac{\omega - \omega_R}{\omega_R}$$

where $\omega$ is the Larmor frequency of the nucleus of interest and $\omega_R$ is the Larmor frequency of the nucleus in a reference compound. Most commonly, trimethylsilane is used as a reference compound in $^{1}H$ and $^{13}C$ NMR spectroscopy.

Nuclear spins in a molecule can interact primarily by two mechanisms: scalar coupling (J-coupling) and dipole-dipole coupling (DD-coupling). As J-coupling involves interactions between nuclei in atoms connected by covalent bonds and the bonding electrons, it is sometimes also called indirect DD-coupling. Importantly, J-coupling provides information about through-bond connectivities between nuclei in a molecule, which is extremely useful for determination of molecular structure. J-coupling also leads to

spin-spin splitting effects, which are used by organic chemists to determine what functional groups are present in a molecule. Due to the complexity of the spectra of biomolecules, spin-spin splitting is typically undesirable in biomolecular NMR spectroscopy, and decoupling radiofrequency pulses are used during signal acquisition to eliminate the splitting effect.

In contrast to J-coupling, DD-coupling occurs directly between the nuclei and acts through space. Although DD-coupling is approximately three orders of magnitude larger than J-coupling, it is not observable in isotropic liquids due to its orientation dependence and the averaging effects of molecular tumbling. Nevertheless, DD-coupling is very important in NMR spectroscopy, as it constitutes a major source of relaxation. DD-coupling between two nuclei can lead to observation of the nuclear Overhauser effect (NOE). The NOE is defined as a change in the intensity of a resonance when spin transitions of another resonance are perturbed. As the NOE occurs due to through-space interaction between nuclear spins, it can be employed for the determination of inter- and intramolecular distances. The NOE plays a crucial role in protein structure determination by NMR spectroscopy, where NMR experiments based on the NOE are used to obtain distance restraints between nuclei to guide protein structure calculation. The NOE reduces quickly with distance, and is observable only between nuclei separated by a distance shorter than approximately 5-6 Å.

More detailed description of the theory of NMR spectroscopy is beyond the scope of this thesis, and has been given in a number of highly respectable textbooks (Cavanagh *et al*., 2007; Keeler, 2010; Levitt, 2008). The NMR experiments used in this study to obtain functional and structural information about PA binding by Na-FAR-1 will be described in brief below.

#### 6.3.1.4 Multidimensional experiments in protein NMR spectroscopy

Proteins are large molecules containing thousands of atomic nuclei, each of which is theoretically able to contribute to an NMR spectrum. Even a small 20 kDa protein that can be conveniently investigated by NMR spectroscopy (further, simply NMR) will consist of approximately 200 amino acid residues and 4000 atoms. Furthermore, as the chemical structure of natural amino acids is largely similar, the chemical shift distributions of the amino acid nuclei are rather narrow. If we also consider that a single amino acid type is often found multiple times in the polypeptide chain of a protein, it will be clear that a significant overlap between the NMR signals of distinct atomic nuclei will be observed in

protein NMR spectra. Hence, one-dimensional $^{1}H$ and $^{13}C$ NMR experiments popular in organic chemistry will reveal little to no information about protein structure.

In order to resolve the signals of protein nuclei, multidimensional NMR experiments are employed. Multidimensional NMR experiments can either be homonuclear or heteronuclear. In homonuclear experiments, chemical shifts of the nuclei of the same isotope are recorded, whereas in heteronuclear experiments, chemical shifts of the nuclei of, typically, two or three isotopes are measured. Multidimensional experiments report on the interactions between nuclei via either through-bond (J-coupling) or through-space (NOE) mechanisms. The chemical shifts of the interacting nuclei are plotted on several frequency axes (e.g., $F_1$ and $F_2$ in a two-dimensional experiment), giving rise to a cross-peak. Homonuclear experiments, such as 2D $^{1}H$, $^{1}H$ COrrelation SpectroscopY (COSY), TOtal Correlation SpectroscopY (TOCSY) or NOE SpectroscopY (NOESY), are of limited use in NMR of proteins over ~ 10 kDa, as they do not provide sufficient signal resolution. Therefore, heteronuclear experiments are typically used to observe protein NMR signals. Since all amino acids contain hydrogen, carbon and nitrogen atoms, the typical isotopes used in heteronuclear protein NMR experiments are $^{1}H$, $^{13}C$ and $^{15}N$. As the natural abundances of $^{13}C$ and $^{15}N$ are very low (1.1% and 0.37%, respectively), uniform isotope labelling of proteins is normally required for $^{13}C$ and $^{15}N$ experiments (see Section 4.3.1.1).

Experiments that report on through-bond correlations of protein nuclei are used for the assignment of protein NMR signals to particular atomic nuclei in the protein residues. Usually, a combination of 2D, 3D or even 4D heteronuclear double- and triple-resonance experiments (i.e., experiments involving magnetisation transfer between two and three different isotopes, respectively) are used for this purpose.

A typical 2D double-resonance experiment used for determination of the chemical shifts of $^{15}N$ and $^{1}H^{N}$ nuclei of the protein amide and amine groups is the $^{1}H$, $^{15}N$ heteronuclear single quantum coherence (HSQC) experiment. Typically, $^{1}H$, $^{15}N$ HSQC provides sufficient resolution to distinguish between the most directly bonded $^{15}N$-$^{1}H$ pairs in the protein. The chemical shifts of the backbone amide $^{15}N$-$^{1}H^{N}$ pairs are then used as a starting point for assigning the nuclei to particular residues in the protein with the help of more complex 3D experiments. Examples of such 3D triple-resonance experiments include $^{1}H$, $^{13}C$, $^{15}N$ HNCO (Grzesiek and Bax, 1992a; Kay *et al*., 1990; Muhandiram and Kay, 1994), HNCA (Farmer *et al*., 1992; Grzesiek and Bax, 1992a; Kay *et al*., 1990), HNcaCO (Clubb *et al*., 1992), HNcoCA (Bax and Ikura, 1991), CBCANH (Grzesiek and Bax, 1992b), CBCAcoNH (Grzesiek and Bax, 1992c), HBHANH (Wang *et al*., 1994) and

HBHAcoNH (Grzesiek and Bax, 1993). These experiments can be used for establishing the sequential links between groups of nuclei assigned to particular types of residues in order to deduce the positions of the residues in the protein primary sequence. A 3D double-resonance experiment very useful for determination of through-bond connectivities of protein nuclei is HCCH-TOCSY (Bax *et al*., 1990; Kay *et al*., 1993), which can be used to establish correlations between all aliphatic $^{13}C$ and $^{1}H$ nuclei in a residue, and hence is useful for assigning the NMR signals of the residue side-chain nuclei.

Although experiments based on through-space nuclear interactions can also be used during assignment of the protein resonance signals, their major role in protein NMR spectroscopy is to reveal the distances between the protein atoms for protein structure determination. Typical 3D experiments that reveal through-space connectivities in the protein are the $^{1}H$, $^{1}H$, $^{15}N$ NOESY-HSQC and $^{1}H$, $^{1}H$, $^{13}C$ NOESY-HSQC (Marion *et al*., 1989a, 1989b; Zuiderweg and Fesik, 1989). In these experiments, the NOE is first used to obtain through-space correlations between neighbouring $^{1}H$ nuclei. Subsequently, one-bond correlations of $^{1}H$ nuclei to either $^{15}N$ or $^{13}C$ are obtained via J-couplings. As a result, the distances between the neighbouring $^{1}H$ nuclei in the protein can be estimated and used to guide structure determination.

The heteronuclear 2D and 3D experiments employed in this study as well as the principle of protein structure determination from NMR data are described in more detail in the following sections of this chapter.

#### 6.3.1.5 2D HSQC and HMQC

The HSQC is one of the most commonly used 2D experiments in protein NMR spectroscopy. It is employed to correlate chemical shifts of a $^{1}H$ nucleus and a heteronucleus ($^{13}C$ or $^{15}N$) that are linked by a covalent bond. HSQC uses an Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) pulse sequence element to create transverse magnetisation on $^{1}H$ and transfer the magnetisation from $^{1}H$ to the heteronucleus through J-coupling. The magnetisation is then evolved on the heteronucleus, and the reverse-INEPT sequence is used to transfer the magnetisation back to $^{1}H$ for detection. Signal detection at $^{1}H$ rather than at the heteronucleus enhances the sensitivity of the experiment, as $^{1}H$ is more magnetically susceptible than $^{13}C$ and $^{15}N$ due to its higher gyromagnetic ratio. Heteronuclear multiple quantum coherence (HMQC) experiment is very similar to HSQC, and is used to obtain the same information as HSQC. The difference between the two is in the coherence pathways involved during the magnetisation evolution.

In protein NMR spectroscopy, $^{1}H$, $^{15}N$ HSQC is used to obtain a chemical shift correlation map of the protein amide and amine groups (of amine groups, typically, only tryptophan and arginine side-chain amines can be observed). The $^{1}H$, $^{15}N$ HSQC spectrum of a protein can be considered one of its "fingerprints", as the spectra obtained for different proteins are usually fairly distinguishable. The chemical shifts of the protein NH groups measured in HSQC are used during protein resonance assignment. $^{1}H$, $^{15}N$ HSQC/HMQC spectra are also commonly used for investigation of protein-ligand binding. As chemical shifts are influenced by their chemical environment, conformational change or ligand binding by a protein typically leads to the change in the position of at least some of the cross-peaks in the $^{1}H$, $^{15}N$ HSQC spectrum. By following the cross-peak movements in the spectrum during ligand titration, information about the binding affinity and stoichiometry can be obtained.

#### 6.3.1.6 3D triple-resonance experiments for backbone assignment

In protein NMR spectroscopy, 3D $^{1}H$, $^{13}C$, $^{15}N$ triple-resonance experiments, such as HNCO, HNCA, HNcaCO, CBCANH, CBCAcoNH, HBHANH and HBHAcoNH, are used to obtain chemical shifts correlations of the backbone $^{15}N$ and $^{1}H^{N}$ nuclei to the backbone $^{13}C\alpha$, $^{13}CO$, $^{1}H\alpha$ as well as the side-chain $^{13}C\beta$ and $^{1}H\beta$ nuclei. In all these experiments one-bond or two-bond J-couplings are used to transfer the magnetisation between the nuclei. The magnetisation starts on $^{1}H^{N}$ or on $^{1}H\alpha$ and $^{1}H\beta$ and the signal is typically detected on $^{1}H^{N}$. As in the HSQC, INEPT sequences are employed during initial and the final steps of the magnetisation transfer. The names of the triple-resonance experiments typically reference the nuclei involved in the magnetisation transfer pathway, where N refers to the backbone $^{15}N$, H to the backbone $^{1}H^{N}$, CA to the backbone $^{13}C\alpha$, CB to the side-chain $^{13}C\beta$, HA to $^{1}H\alpha$ and HB to $^{1}H\beta$. The nuclei whose chemical shifts are measured in the experiment (i.e., on which the chemical shift is evolved) are written in upper-case, whereas the nuclei that are involved in the magnetisation transfer but on which the chemical shift is not evolved are written in lower-case or in parentheses [e.g, $^{13}C\alpha$ in HNcaCO, which can also be written as HN(CA)CO].

Importantly, by the use of different combinations of the triple-resonance experiments, protein intra- and interresidue correlations can be identified. For instance, the HNCO/HNcaCO pair of experiments allows the intra- and interresidue $^{1}H^{N}$, $^{15}N$, $^{13}CO$ correlations to be distinguished. This is achieved because magnetisation transfer in an HNcaCO occurs through one-bond and two-bond $^{15}N$-$^{13}C\alpha$ J-couplings ($J_{NC\alpha}$), which allows correlations between the amide group of a residue *i* and both the $^{13}CO$ of the same

residue and the preceding residue *i*-1 to be obtained. Since the HNCO does not use $J_{NC\alpha}$, in HNCO the correlations are obtained only for $^{1}H^{N}$, $^{15}N$ and $^{13}CO$ of the residue *i*-1. By comparing the HNCO and HNcaCO spectra, the cross-peaks can be assigned to either residue *i* or *i*-1. In a similar way, other pairs of complimentary triple-resonance experiments can be used to obtain intra- and interresidue correlations of other backbone nuclei. These pairs include HNCA/HNcoCA, HNCACB/CBCAcoNH and HBHANH/HBHAcoNH, which are used to obtain $^{1}H^{N}$, $^{15}N$ correlations to $^{13}C\alpha$, $^{13}C\alpha$ and $^{13}C\beta$, and $^{1}H\alpha$ and $^{1}H\beta$, respectively. After links between short stretches of residues are established, they can be connected into a longer polypeptide sequence until assignment of the whole backbone chain is obtained. In many cases, however, complete assignment is not possible, as some of the backbone nuclei signals may not be observable due to relaxation as a result of chemical exchange or conformation flexibility of certain regions of the protein. Once assignment of the majority of the protein backbone nuclei is established, assignment of the side-chain nuclei can be carried out. The magnetisation transfer pathways of several common 3D triple-resonance NMR experiments are summarised in Figure 6-2.

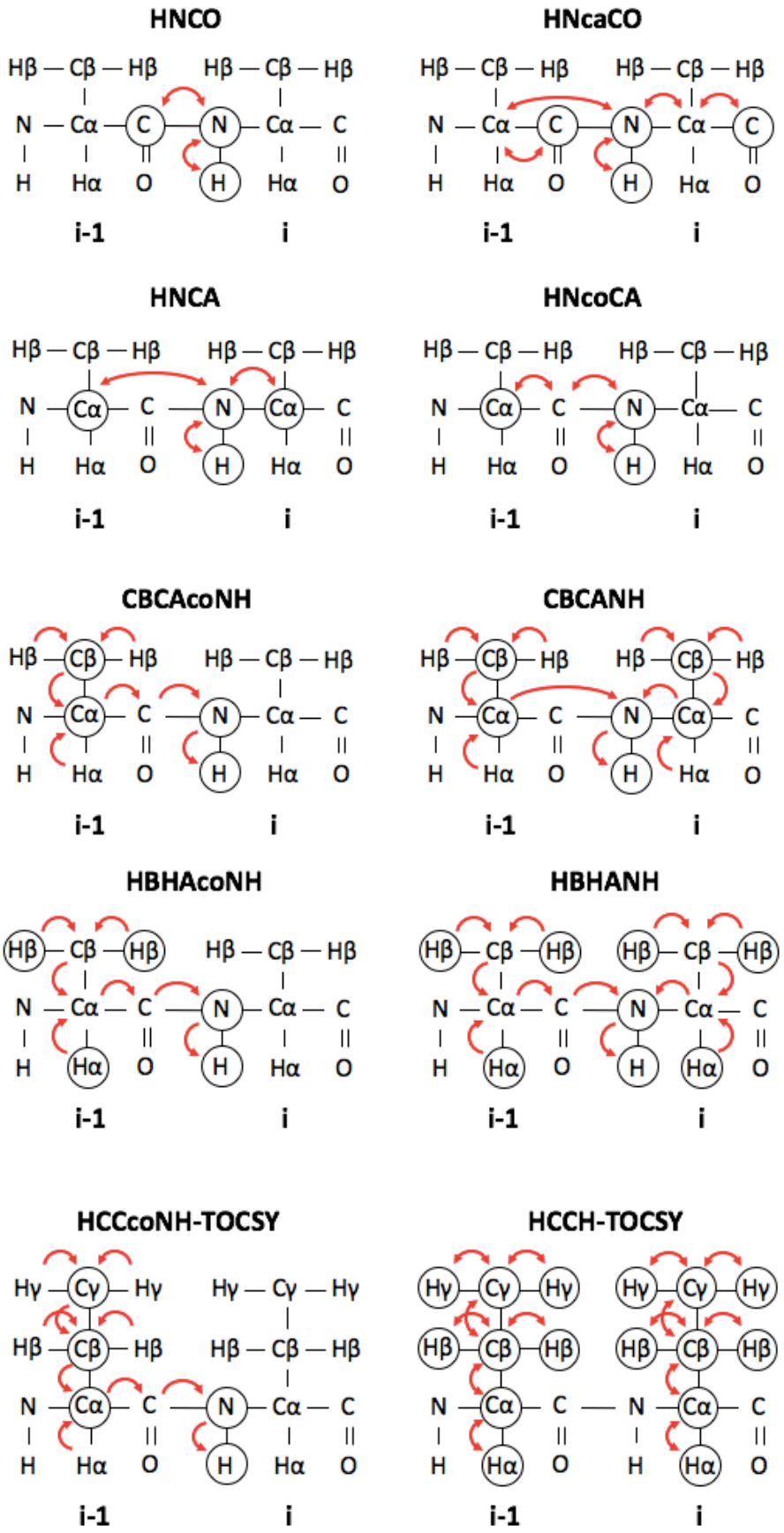



**Figure 6-2. Magnetisation transfer pathways in common 3D protein NMR spectroscopy experiments.** Nuclei whose chemical shifts are recorded are shown in circles. The magnetisation transfer direction is shown with red arrows.

#### 6.3.1.7 3D TOCSY

Typically, TOCSY experiments are used for assignment of the side-chain nuclei in the protein. All TOCSY experiments employ isotropic mixing to transfer the magnetisation between nuclei via strong J-coupling Hamiltonian. In protein NMR this is typically achieved by using broad band mixing pulses such as the Decoupling In the Presence of Scalar Interactions (DIPSI) pulse sequence. The initial and last steps of magnetisation transfer are typically performed by using INEPT and reverse-INEPT sequences, respectively. Two of the most popular TOCSY experiments for the residue side-chain assignment are the double-resonance HCCH-TOCSY (Bax *et al*., 1990; Kay *et al*., 1993) and the triple-resonance HCCcoNH-TOCSY (Montelione *et al*., 1992).

The conventional HCCH-TOCSY developed by Bax *et al*. allows correlations between all aliphatic $^{13}C$ nuclei in a residue to all aliphatic $^{1}H$ nuclei in the same residue to be observed. As a result, complex spectra are obtained that in the absence of relaxation will contain cross-peaks between any aliphatic $^{1}H$-$^{13}C$ pair in a residue. Three 3D variants of the HCCH-TOCSY exist (hCCH-, HcCH- and HCcH-TOCSY), which are distinguished by the positions of the delays during which the chemical shift evolution takes place in the experiment. The use of both the hCCH and HcCH/HCcH variants of HCCH-TOCSY during side-chain resonance assignment can help to distinguish between the nuclei with overlapping chemical shifts in the protein. Recently, H($C^{ali}C^{aro}$)H-TOCSY has been developed (Kovacs and Gossert, 2014), which allows correlations between the aromatic and the aliphatic $^{1}H$-$^{13}C$ pairs in the residue to be obtained.

In contrast to the double-resonance HCCH-TOCSY, the HCCcoNH-TOCSY is a triple-resonance experiment that correlates the $^{1}H^{N}$ and $^{15}N$ amide pair of a residue *i* to aliphatic backbone and side-chain $^{1}H$ and $^{13}C$ nuclei of the residue *i*-1. Typically, 3D forms of HCCcoNH-TOCSY (hCCcoNH- and HcccoNH-TOCSY) are employed to reveal correlations of the backbone amide nuclei to the aliphatic $^{13}C_{i-1}$ and $^{1}H_{i-1}$ nuclei, respectively. HCCcoNH-TOCSY is typically of lower sensitivity than HCCH-TOCSY. The magnetisation transfer pathways of HCCH-TOCSY and HCCcoNH-TOCSY are schematically presented in Fig 6-2.

#### 6.3.1.8 2D experiments for assignment of aromatic side-chains

Assignment of aromatic side-chains nuclei is typically more complicated than that of aliphatic side-chain nuclei, as conventional HCCH-TOCSY does not reveal the correlations between the aliphatic and the aromatic nuclei in the residue. As an alternative

or a complement to H($C^{ali}C^{aro}$)H-TOCSY, 2D double-resonance hbCBcgcdHD and hbCBcgcdceHE (Yamazaki *et al*., 1993) are often used, which reveal the aliphatic-aromatic $^{13}$Cβ-$^{1}$Hδ and $^{13}$Cβ-$^{1}$Hε correlations, respectively. After the aliphatic $^{13}$C-aromatic $^{1}$H correlations have been obtained in the 2D experiments, one can use the chemical shifts of the aromatic $^{1}$H nuclei to identify the chemical shifts of the correlated aromatic $^{13}$C nuclei from $^{1}$H, $^{13}$C-HSQC, $^{1}$H, $^{13}$C aromatic Transervse Relaxation Optimised SpectroscopY (aroTROSY) (Pervushin *et al*., 1997) and 3D NOESY-HSQC spectra (see below) to complete the assignment. As aromatic groups have relatively high chemical shift anisotropy compared to aliphatic groups, aroTROSY is a more sensitive experiment for detection of the aromatic $^{1}$H-$^{13}$C correlations and produces better resolved peaks than HSQC.

#### 6.3.1.9 3D NOESY-HSQC

After the protein backbone and side-chain assignments have been carried out with the help of double- and triple-resonance experiments, distances between the protein nuclei need to be determined to obtain distance restraints for protein structure calculation. Distances between the protein $^{1}$H nuclei are typically obtained from 3D NOESY-HSQC experiments. 3D NOESY-HSQC pulse sequence combines the homonuclear 2D NOESY sequence with an HSQC sequence. The heteronucleus employed in the HSQC part of the NOESY-HSQC experiment can either be $^{13}$C or $^{15}$N. By using $^{13}$C NOESY-HSQC and $^{15}$N NOESY-HSQC experiments, one can obtain through-space correlations between $^{1}$H nuclei in $^{1}$H-$^{13}$C pairs and their neighbouring $^{1}$H nuclei, and between $^{1}$H nuclei of $^{1}$H and $^{15}$N pairs and their neighbouring $^{1}$H nuclei, respectively. The 3D NOESY-HSQC experiments typically provide a good resolution of the protein NOE signals and allow a sufficient number of $^{1}$H-$^{1}$H distance restraints to be obtained for the calculation of a protein's 3D structure.

As NOESY-HSQC spectra reveal through-space correlation of the neighbouring $^{1}$H nuclei, they can also be used to guide the assignment of protein resonances or to confirm the validity of the assignments. For instance, $^{1}H_N$-$^{1}H_N$ and $^{1}H_N$-$^{1}$Hα correlations that can be obtained from $^{15}$N NOESY-HSQC can provide information about the sequential links between the residues that help assignment of the protein backbone. Similarly, $^{13}$C NOESY-HSQC can be used to obtain intraresidue correlations between the aromatic $^{1}$H and $^{13}$C nuclei for the aromatic side-chain assignment, as mentioned above.

#### 6.3.1.10 Structure calculation

Information about three-dimensional protein structure cannot be directly obtained from the chemical shifts of the protein nuclei. Instead, sets of restraints, the main ones being the $^{1}$H-$^{1}$H distance restraints obtained from NOESY experiments, are used to produce a model of the protein structure. The process of model generation from the sets of restraints is called structure calculation. As well as $^{1}$H-$^{1}$H distance restraints, peptide dihedral angles, residual dipolar couplings, hydrogen bond restraints and other types of restrains can be used to constraint the structure calculation, together with empirical "forcefield" terms about bond length and angles. Usually, the restraints do not carry enough information to unambiguously define a single structural model. Therefore, as a result of structure calculation from NMR data, an ensemble of models is generated, where each model satisfies the restraints used during the calculation. Typically, a small set of final models with the lowest energies is presented after structure determination by protein NMR.

Several algorithms have been developed for automatic assignment of NOESY spectra, generation of $^{1}$H-$^{1}$H distance restraints and structure calculation. One of the most popular programs is ARIA (Ambiguous Restraints for Iterative Assignment) (Linge *et al*., 2003a, 2001), which employs an algorithm based on the use of ambiguous distance restraints for assignment of NOE signals (Nilges, 1995; Nilges *et al*., 1997) and CNS software for performing the structure calculation (Brünger *et al*., 1998). In ARIA, calibration of NOESY spectra, generation of distance restraints, structure calculation and analysis of violated restraints are carried out in an iterative manner. In the first round of structure calculation by ARIA, unambiguous restrains obtained from the manually assigned NOEs and ambiguous restraints generated by the program are used to calculate a set of preliminary structures. The structures are then utilised to guide assignment and selection of restraints by analysis of restraint violations. This process is repeated until the structures obtained in the calculations converge. After the final iteration, the structures are refined in explicit solvent (water or DMSO) (Linge *et al*., 2003b) to yield a set of final models. The program outputs an ensemble of the refined structures (typically, 5 or 10), the list of violated restraints in the final structures, and the results of the stereochemical structure analysis by PROCHECK (Laskowski *et al*., 1993) and similar programs, if required. The list of violations can then be examined manually to correct any obvious misassignments of the NOE signals in the NOESY spectra. Typically, at least several rounds of structure calculations by ARIA and manual analysis of NOESY spectra are required to obtain a satisfactory model.

CYANA is another popular program used for structure calculation from NMR data (Güntert *et al*., 1997; Herrmann *et al*., 2002), which uses a torsion angle dynamics algorithm for structure calculation and an algorithm similar to ARIA for assignment of NOEs. Since CYANA was not employed in this study, it will not be discussed further.

### 6.3.2 Chemical shift perturbation analysis of Na-FAR-1-PA binding

#### 6.3.2.1 Na-FAR-1 forms at least three distinct complexes with PA

In order to investigate the mechanism of PA binding by Na-FAR-1, either $^{1}H$, $^{15}N$ HSQC or HMQC spectra, which provide identical information, were recorded during DOPA titration into the solution of apo Na-FAR-1. Gradual changes in the chemical shift of the majority of backbone amide cross-peaks were observed during DOPA addition (Fig 6-3), which was assumed to be due to the conformational change in the protein occurring on ligand binding. After addition of two molar equivalents of DOPA to the protein, four distinct cross-peaks were observed for many of the backbone amides, indicating that the protein was in slow exchange between four different states. One of the cross-peaks was present at the chemical shift identical to the one observed in the apo protein spectrum, which was assumed to correspond to the ligand-free state of the protein. The other three cross-peaks were shifted compared to the cross-peak coming from the ligand-free protein. These were assumed to arise from three different ligand-bound forms of the protein, likely corresponding to the protein bound to one, two and three or more molecules of DOPA. The cross-peaks originating from the Na-FAR-1 complex with two and three or more DOPA molecule were overlapping with each other and were more intense than the rest of the peaks, suggesting that the Na-FAR-1-DOPA complex mostly contained two molecules of DOPA when DOPA was present in a two-fold molar excess to the protein.

In the subsequent titration step, peaks corresponding to the ligand-free protein and to the Na-FAR-1 complex with a single DOPA were no longer visible, suggesting that the protein was now in the fast exchange regime between the states containing two and three or more DOPA molecules. Once a four-fold excess of ligand was added to Na-FAR-1, the protein appeared to enter a fast exchange regime where separate ligand-bound forms of Na-FAR-1 could no longer be distinguished and only one cross-peak was observed for most residues. The chemical shifts of the amide cross-peaks observed at this ligand concentration were identical to those of the cross-peaks originating from Na-FAR-1 complex with three or more DOPA recorded when the ligand was in two-fold excess. The binding appeared to be saturated at this point, and little change in chemical shifts was observed after further

additions of DOPA. From this data, it can be concluded that Na-FAR-1 is able to form at least three distinct complexes with PA that have different protein:ligand stoichiometries. Although it could not be determined what is the maximum number of PAs that Na-FAR-1 can bind, the data indicate that Na-FAR-1 can accommodate at least three DOPA molecules.

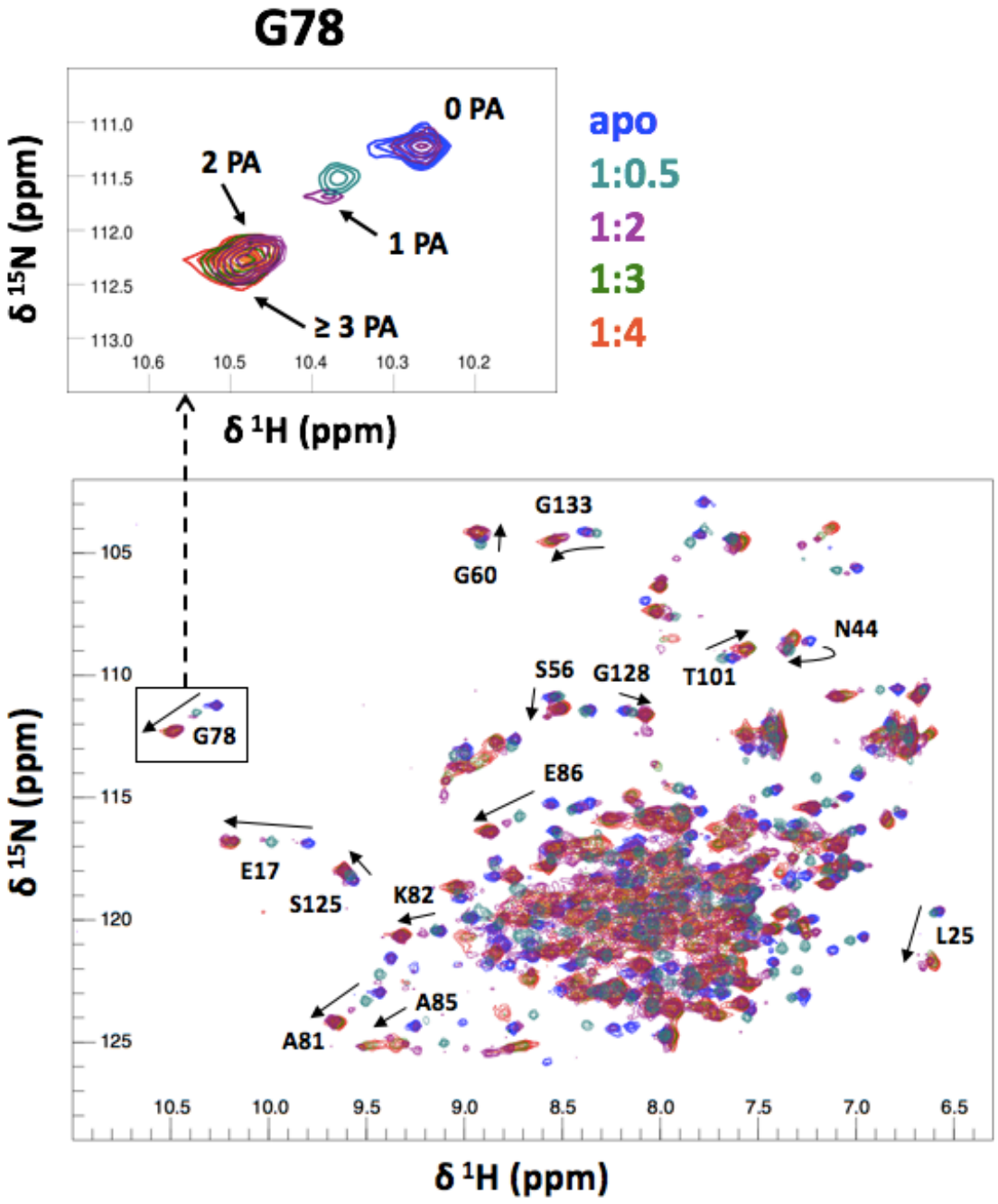



**Figure 6-3. Chemical shift perturbation analysis of the Na-FAR-1 interaction with PA.** Superimposed $^{1}$H, $^{15}$N-HSQC spectra of the ligand-free Na-FAR-1 (blue cross-peaks), Na-FAR-1 in the presence of 2, 3 and 4 molar equivalents of PA (purple, orange and green cross-peaks, respectively), and an $^{1}$H, $^{15}$N-HMQC spectrum of Na-FAR-1 in the presence of 0.5 molar equivalent of PA (teal cross-peaks) are shown. Changes in the chemical shifts of amide cross-peaks are observed with increasing PA concentration. Several examples of cross-peak movements are highlighted. An expansion of the region containing the G78 cross-peaks is shown above. The cross-peaks are proposed to originate from the ligand-free Na-FAR-1 (0 PA) and distinct Na-FAR-1 complexes containing 1, 2 and 3 or more PA molecules.

### 6.3.2.2 Mapping of the chemical shift changes in Na-FAR-1 induced by DOPA binding

For the next step in the analysis of Na-FAR-1-DOPA binding, Na-FAR-1 complex with DOPA was prepared by mixing Na-FAR-1 with seven-molar excess of dry DOPA. $^{1}$H, $^{15}$N HSQC spectrum of the DOPA complex was recorded and compared to the $^{1}$H, $^{15}$N HSQC spectrum of apo Na-FAR-1 to map the differences between the chemical shifts of the backbone amide nuclei in the presence and the absence of the ligand (Fig 6-4). Since the chemical shift of the nuclei depends on their chemical environment, changes in the chemical shift of the backbone amide nuclei report on the conformation change in the regions where the nuclei are located and/or on binding of the ligand in proximity of the nuclei. The resonances of the Na-FAR-1 complex with PA were assigned with the help of 3D triple- and double-resonance experiments, as described in the section below (see Fig 6-8 for the spectrum). The spectrum of the apo Na-FAR-1 protein produced in this study was assigned with the help of the apo Na-FAR-1 spectrum previously assigned by Dr Florencia M. Rey-Burusco (Instituto Nacional de Tecnologia Agropecuaria, Buenos Aires, Argentina) (Rey-Burusco *et al*., 2014).

The geometric distances between the cross-peaks in the spectra of apo Na-FAR-1 and Na-FAR-1-PA complex were calculated by using the following equation:

$$d\delta = \sqrt{[\Delta\delta_{1H}^2 + (0.15 \times \Delta\delta_{15N})^2]} \qquad \textbf{Equation 6-3.}$$

where $d\delta$ is the chemical shift distance between two cross-peaks in the $^{1}$H, $^{15}$N HSQC spectra, $\Delta\delta_{1H}$ is the chemical shift difference in $^{1}$H dimension, $\Delta\delta_{15N}$ is the chemical shift difference in $^{15}$N dimension, and 0.15 is the scaling factor for $^{15}$N. Chemical shift distance could be calculated for ~ 75% of residues in the protein (116 out of 155, not counting the residues in the recombinant tag). For the remaining ~ 25% of the protein residues, chemical shift distances could not be calculated due to the absence of the corresponding cross-peaks either in the apo Na-FAR-1 spectrum, the DOPA complex spectrum or both. The majority of such residues were located in the loop regions between the helices and in the helix α10, which could indicate the high conformational flexibility of these regions or the rapid hydrogen exchange rates between the relevant backbone amide groups and the solvent. The residues whose backbone chemical shifts changed in the presence of PA are highlighted on the previously determined structure of apo Na-FAR-1 (PDB ID 4UET) (Fig 6-4B). Large changes in the cross-peak positions were observed for most residues,

with 20 residues demonstrating *dδ* larger than 0.5 ppm, 51 residues demonstrating *dδ* larger than 0.3 ppm and 76 residues demonstrating *dδ* larger than 0.2 ppm. Since from this and the previous studies (Rey-Burusco *et al*., 2014) it was known that the protein expands on ligand binding and that many protein residues contact ligand atoms in the central cavity of Na-FAR-1, extensive changes in the positions of the cross-peaks in the HSQC spectrum of Na-FAR-1 were expected after ligand addition.

The largest *dδ* (0.93 ppm) was observed for T101, which is located in a flexible loop between α7 and α8, and the smallest *dδ* (0.01 ppm) was observed for the C-terminal residue N155. From Fig 6-4B it can be observed that most of the residues that displayed the largest *dδ* were located in the regions of the helices that form the largest portions of the cavity walls in the ligand-bound forms of Na-FAR-1 (see Chapter 5), such as the central regions of α6 and α7, and the region of α4 proximal to the C-terminus. In contrast, small *dδ* were observed for the backbone amides of the residues located in α9, which forms only a small portion of the central cavity in the ligand-bound Na-FAR-1.

The residues with the largest *dδ* that were located in the cavity-forming regions of the helices in the ligand-bound forms of Na-FAR-1 included V70 (0.70 ppm, α6), F37 (0.62 ppm, α4), E35 (0.62 ppm, α4), A85 (0.60 ppm, α7), R98 (0.58 ppm, α7), A68 (0.58 ppm, α6), R93 (0.52 ppm, α7), K71 (0.51 ppm, α6) and I95 (0.48 ppm, α7). It possible that the large chemical shift perturbations observed for these residues in the presence of PA are caused by the binding of PA molecules close to the residues. As it was observed that the structure of Na-FAR-1 is very similar in both the oleic acid complex and the *E. coli* lipid complex, it is expected that the structure of Na-FAR-1-PA complex will not be radically different from the previously determined ligand-bound structures. Interestingly, from these residues, the side-chains of I95, R93, A85, V70 and F37 line the cavity of the ligand-bound forms of Na-FAR-1, and the side-chains of I95, V70, F37 made apolar contacts to the ligands in the Na-FAR-1-oleic acid complex. If the same structural arrangement of the cavity is present in the Na-FAR-1-PA complex, it can be proposed that the side-chains of I95, R93, A85, V70 and F37 could be making contacts to the PA molecules in the PA complex. It should be noted that the residues with smaller *dδ* than the residues listed above could also interacting with PA in the complex. This is because the involvement of residues in ligand binding cannot be predicted directly from the magnitude of *dδ,* as the chemical shift of the protein nuclei can be affected by many factors apart from direct interaction with the ligand.

To summarise, the chemical shift perturbation analysis of Na-FAR-1-PA binding has revealed that Na-FAR-1 is able to form at least three DOPA complexes with distinct stoichiometries. Furthermore, from comparison of the backbone amide cross-peak positions in the $^{1}H$, $^{15}N$ HSQC spectra of Na-FAR-1 in the presence and the absence of PA, clues about the location of PA binding sites were obtained. Many of the largest backbone amide chemical shift changes were observed for the residues located in the central regions of α6 and α7 and the region of α4 proximal to the C-terminus, indicating that these regions might be involved in PA binding.

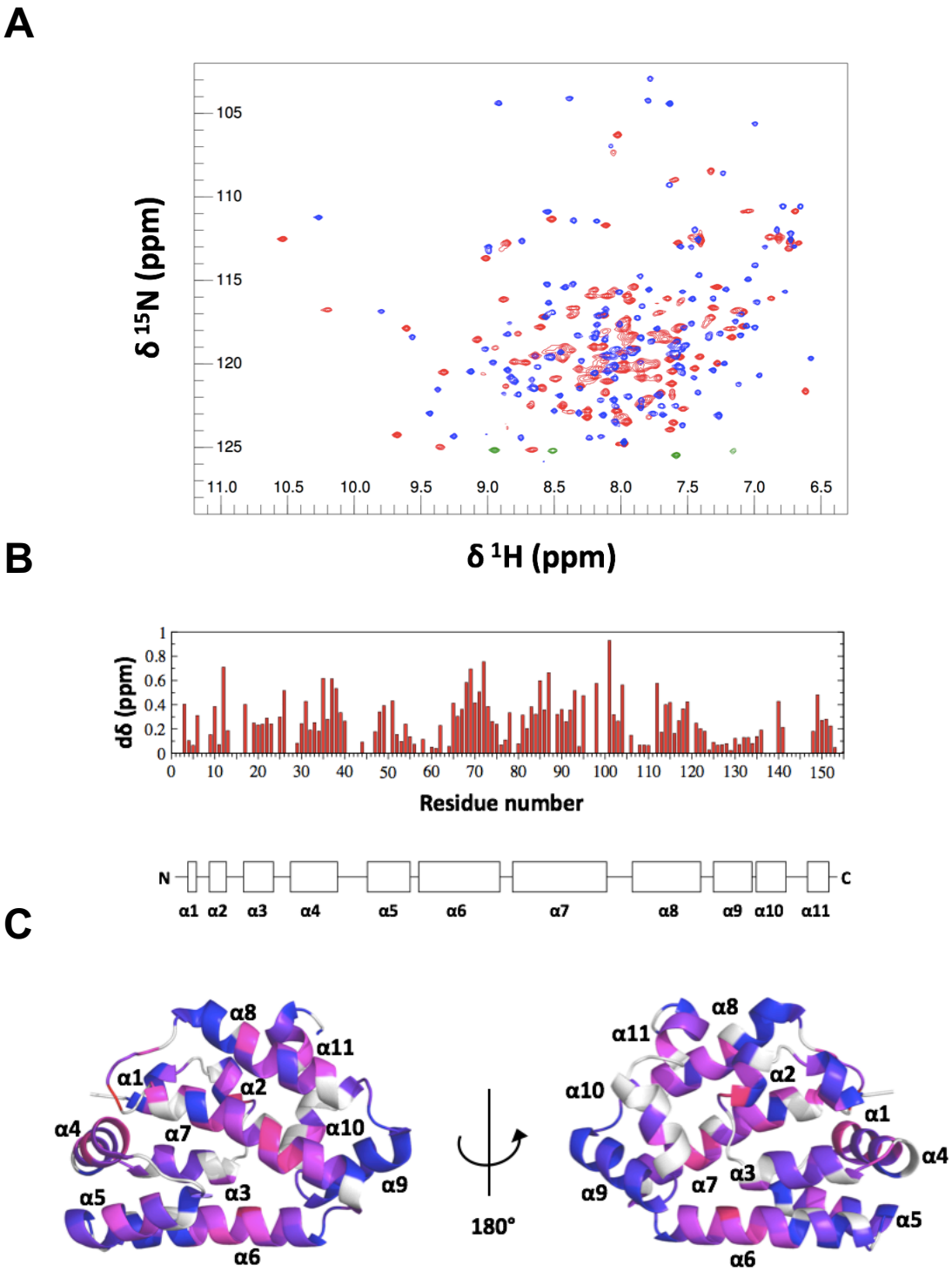



**Figure 6-4. Backbone amide chemical shift perturbations between the $^{1}$H, $^{15}$N HSQC spectra of apo Na-FAR-1 and Na-FAR-1 in complex with PA. A.** The $^{1}$H, $^{15}$N HSQC spectra of apo Na-FAR-1 (blue) and Na-FAR-1 in the presence of 7 molar equivalent of PA (red; aliased cross-peaks are shown in green). Assignments are not shown. **B**. A histogram displaying the $^{1}$H, $^{15}$N chemical shift distances ($d\delta = \sqrt{[\Delta\delta_{1H}^2 + (0.15 \times \Delta\delta_{15N})^2]}$) between the cross-peaks in the spectra in A. In cases when $d\delta$ is equal to zero, either no chemical shift change was observed or the chemical shift change was not calculated due to the absence of the assigned cross-peaks in one or both spectra. A schematic chart of the secondary structure elements in apo Na-FAR-1 is shown below the graph for reference. **C.** A cartoon representation of apo Na-FAR-1 [PDB ID 4UET (Rey-Burusco *et al*., 2015)] highlighting the residues which displayed the chemical shift distances presented in B. The residues are coloured according to their $d\delta$ from blue (lowest $d\delta$) to red (highest $d\delta$) through magenta (medium $d\delta$). The residues for which $d\delta$ could not be determined are shown in light grey. The residues which showed the largest $d\delta$ are predominantly located in the cavity-forming helices.

### 6.3.3 Resonance assignment of the Na-FAR-1 complex with PA

Clues about the mechanism of PA binding by Na-FAR-1 were obtained from the chemical shift perturbation analysis. To obtain molecular details about the PA binding mechanism and the residues involved in PA binding, it was decided to determine the structure of the Na-FAR-1 complex. An attempt to co-crystallise Na-FAR-1 with PA for X-ray crystallography was made. However, crystallisation screening failed to identify conditions suitable for crystallisation of the Na-FAR-1-PA complex, and hence, crystals could not be obtained for structural analysis. Since the complex produced good quality HSQC spectra, protein NMR spectroscopy was chosen as a means of structure determination. The Na-FAR-1-PA complex was prepared by mixing Na-FAR-1 with seven molar equivalent of dry DOPA. As a first step in the structure determination process, assignment of the Na-FAR-1 backbone and side-chain resonances in the PA complex was carried out.

#### 6.3.3.1 Backbone resonance assignment

Backbone resonance assignment was carried out with the help of $^{1}H$, $^{15}N$ HSQC, HNCA, CBCANH/CBCAcoNH, HNCO/HNcaCO, HBHANH/HBHAcoNH and $^{15}N$ NOESY-HSQC (see Section 6.3.1 for the overview of protein NMR experiments) in the CcpNmr Analysis software (Vranken *et al*., 2005). Acquisition parameters of the experiments used in this study for structure determination of the Na-FAR-1-PA complex are summarised in Table 2-1.

In the first step of the assignment, $^{1}H$, $^{15}N$ HSQC spectrum of the Na-FAR-1-PA complex was peak picked and each cross-peak was assigned to a separate spin system. The concept of spin system is central to CcpNmr Analyis, where it is understood as a collection of nuclei that are J-coupled with one another. In the assignment of protein resonances, one typically uses a spin system to represent a particular residue in the protein. $^{1}H$, $^{15}N$ HSQC cross-peaks arising from the side-chain $NH_2$ groups of glutamines and asparagines were identified based on their characteristic chemical shifts and their appearance as pairs of cross-peaks with the same $^{15}N$ chemical shift, and grouped accordingly. The reference chemical shifts were obtained from RefDB (Zhang *et al*., 2003), a database of uniformly referenced chemical shifts derived from Biological Magnetic Resonance Bank (BMRB) (Ulrich *et al*., 2008). After exclusion of the side-chain cross-peaks, the rest of the peaks were assumed to arise from the backbone amide groups of the protein residues, and were assigned accordingly.

In the next step of backbone assignment, the chemical shifts of the amide resonance pairs were used to navigate into the corresponding regions of the HNCA, CBCANH, CBCAcoNH, HNCO, HNcaCO, HBHANH and HBHAcoNH spectra. The cross-peaks present at the frequencies of the amide resonance pairs were peak picked and assigned to $^{13}C\alpha/^{13}C\beta$ or $^{13}CO$ in the HNCA/CBCANH/CBCAcoNH and HNCO/HNcaCO spectra, respectively, and to $^{1}H\alpha/^{1}H\beta$ in the HBHANH/HBHAcoNH spectra, depending on the type of the spectrum in which the cross-peak was present and its chemical shifts. Furthermore, the cross-peaks in the triple-resonance spectra were assigned either to the same spin system as the correlated amide resonances (corresponding to residue *i*) or to a new spin system (corresponding to residue *i*-1). This was done by comparing the complimentary pairs of the triple-resonance spectra, such as the CBCANH and CBCAcoNH. For example, since it was known that the CBCANH reveals both the inter- and intra-residue correlations of $^{1}H_N$-$^{15}N$ with $^{13}C\alpha$ and $^{13}C\beta$, whereas the CBCAcoNH reveals only the inter-residue correlations of the same nuclei, the cross-peaks that appeared only in the CBCAcoNH spectrum were assigned to $^{13}C\alpha_{i-1}$ and $^{13}C\beta_{i-1}$, and the cross-peaks that appeared only in the CBCANH spectrum were assigned to $^{13}C\alpha_i$ and $^{13}C\beta_i$. The HNCA experiment, which is more sensitive than the CBCANH, was used to confirm the assignments of $^{13}C\alpha_i$. In a similar way, the chemical shifts of $^{13}CO_i$ and $^{13}CO_{i-1}$ were distinguished by using HNCO and HNcaCO spectra, and the chemical shifts of $^{1}H\alpha_i$, $^{1}H\alpha_{i-1}$, $^{1}H\beta_i$ and $^{1}H\beta_{i-1}$ were distinguished by using the HBHANH and HBHAcoNH.

After the atom type assignments were established, the spin systems were semi-automatically linked into short sequences by using the Protein Sequence Assignment tool in CcpNmr Analysis. Protein Sequence Assignment matches all $^{13}C_i$ and $^{13}C_{i-1}$ resonances in the complimentary triple-resonance spectra, i.e., the $^{13}C\alpha_i$ and $^{13}C\beta_i$ resonances of one spin system in the CBCANH spectrum with the $^{13}C\alpha_{i-1}$ and $^{13}C\beta_{i-1}$ resonances of another spin system in the CBCAcoNH spectrum. If the resonances match, then it is likely that the spin systems are linked sequentially. Since unambiguous resonance matching is complicated by the fact that the chemical shifts of the same type of nuclei can often be similar in different residues, $^{13}CO_i$ and $^{13}CO_{i-1}$ resonances were also used in the matching process to increase the likelihood of finding the correct link between the spin systems. Optimal matches were chosen manually in each case. An example of a short sequence of residues (F82, G83 and K84) connected by using the $^{13}C\alpha$, $^{13}C\beta$ and $^{13}CO$ resonances in the triple-resonance spectra is shown in Fig 6-5.

After the short linked sequences were established, assignment of several spin systems to particular types of residues was carried out. This was possible because some types of amino acids have characteristic $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts distinct from other amino acids. Typical examples include serine and threonine, which have unusually large $^{13}C\beta$ shifts, glycine which only has $^{13}C\alpha$ and alanine which has a relatively small $^{13}C\beta$ shift. The patterns of residue types appearing in the spin system sequences were used to link the short sequences into the longer spin system stretches by using the primary Na-FAR-1 sequence to guide the process. Based on the match between the residue pattern in the spin system stretch and the primary protein sequence, the spin systems were assigned to particular residues in the protein. If the spin system sequences could not be linked together in an unambiguous way, the sequential links between the spin systems were re-analysed by using the Protein Sequence Assignment tool and corrected, if needed. This process was repeated many times until assignments of as many spin systems as possible were obtained. Finally, $^{15}N$ NOESY-HSQC was used to confirm the links between the residues and identify several missing links primarily by analysis of the $^{1}H^{N}$-$^{1}H^{N}$ through-space correlations.

Since prolines lack a backbone NH group when they are a part of a polypeptide, no correlations to the proline backbone amide group can be obtained. Hence, proline $^{13}C\alpha$, $^{13}C\beta$ and $^{13}CO$ chemical shifts were primarily obtained from the CBCAcoNH and HNCO spectra of the next residue in the primary sequence. Proline assignment was further complicated by the fact that 13% of the Na-FAR-1 residues, including several that follow prolines in the primary sequence, did not produce observable backbone amide resonance signals. Hence, although several proline resonance assignments were obtained from the $^{13}C$ NOESY-HSQC spectrum, only 4 out of 8 proline residues were assigned in the Na-FAR-1 sequence.

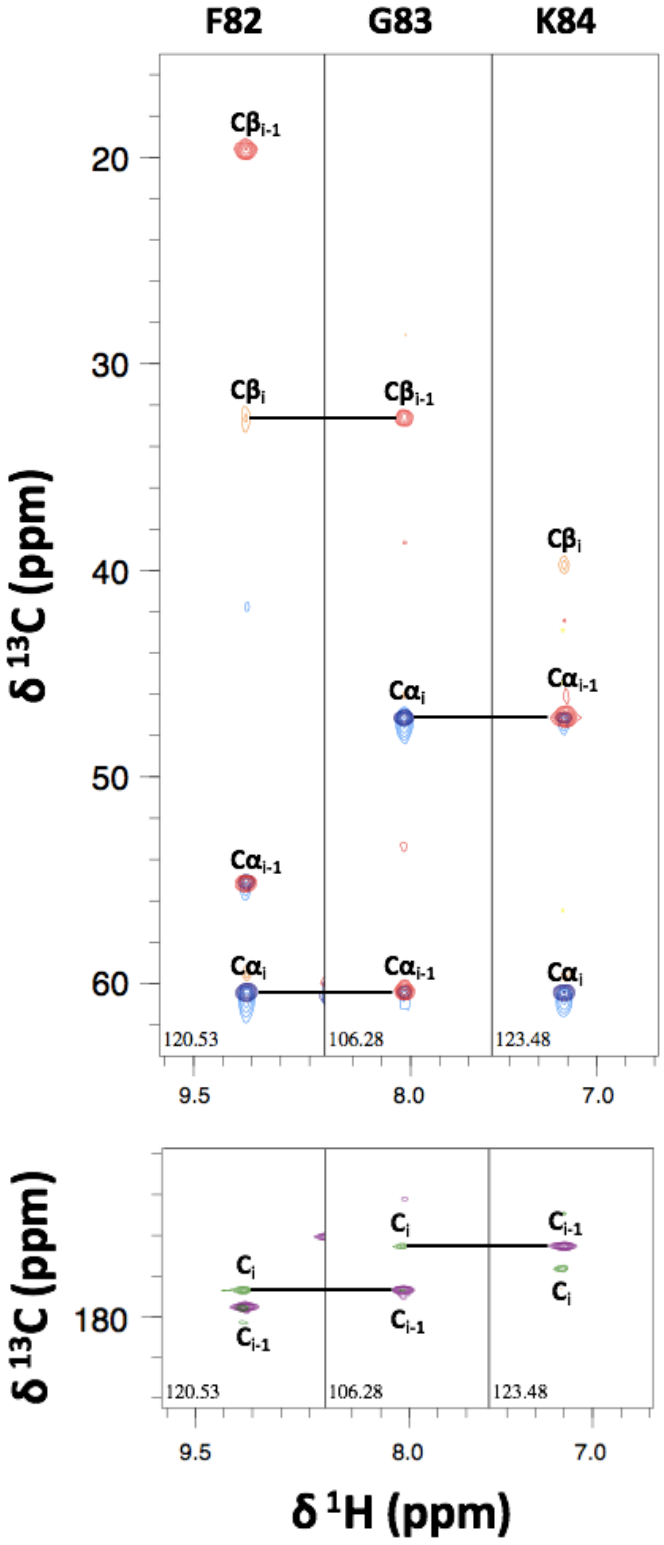



**Figure 6-5. Backbone resonance assignment using triple-resonance NMR experiments.** Strips showing F82, G83 and K84 cross-peaks in the overlaid HNCA (dark blue), CBCANH (light blue for positive peaks, orange for negative peaks), CBCAcoNH (red), HNCO (purple) and HNcaCO (green) spectra. The chemical shifts of the cross-peaks in the $^{15}$N dimension are shown in the bottom left corner of each strip. Cross-peaks marked *i* arise due to intraresidue correlations, cross-peaks marked *i*-1 arise due to interresdidue correlations. By identifying the matching *i* and *i*-1 cross-peaks across the spectrum, sequential connections between the residues can be made (marked by solid horizontal lines between the strips).

### 6.3.3.2 Side-chain resonance assignment

Once the majority of the backbone resonance assignments were established, assignment of the residue side-chains was carried out. In this study, primarily HcCH-TOCSY, hCCH-TOCSY and hCCcoNH were used for the assignment of the aliphatic side-chains, and hbCBcgcdHD, hbCBcgcdceHE, $^{1}$H, $^{13}$C TROSY and $^{13}$C NOESY-HSQC for assignment of the aromatic side-chains.

Aliphatic side-chain assignment from the HCcH-TOCSY and hCCH-TOCSY spectra was carried out as follows. The chemical shifts of the $^{1}$Hα/$^{13}$Cα and $^{1}$Hβ/$^{13}$Cβ pairs determined from the triple-resonance experiments were used to navigate into relevant the parts of the TOCSY spectra, and strips of the spectrum were created that were centered at the chemical shifts of each $^{1}$Hα/$^{13}$Cα and $^{1}$Hβ/$^{13}$Cβ pair. As HCCH-TOCSY-type experiments reveal correlations between all aliphatic $^{1}$H-$^{13}$C pairs in the residue, in each $^{1}$Hα/$^{13}$Cα and $^{1}$Hβ/$^{13}$Cβ strip, correlations to all intraresidue aliphatic $^{1}$H-$^{13}$C pairs were observed. In HCcH-TOCSY, the cross-peaks that appeared at each $^{1}$H-$^{13}$C strip revealed the chemical shifts of the $^{1}$H nuclei in the correlated aliphatic pairs, whereas in hCCH-TOCSY, the cross-peaks revealed the $^{13}$C chemical shifts of the correlated pairs. In order to identify the $^{13}$C chemical shifts of all correlated aliphatic pairs in HCcH-TOCSY, the spectrum was inspected for repetition of the cross-peak pattern observed for the $^{1}$Hα/$^{13}$Cα or $^{1}$Hβ/$^{13}$Cβ pair used for initial analysis of the spectrum. The $^{13}$C chemical shift at which the identical cross-peak pattern appeared corresponded to the $^{13}$C chemical shift of the correlated pair. Similarly, $^{1}$H correlations were obtained from the hCCH-TOCSY spectra. The correlated $^{1}$H and $^{13}$C resonances were assigned to particular types of nuclei by using the reference chemical shifts in RefDB and by inspection of the correlations between the nuclei. The use of both HCcH-TOCSY and hCCH-TOCSY allowed better resolution of many overlapped signals in the spectra, especially in the case of the lysine and glutamate side-chains. See Figure 6-6 for an example of the HCcH spectrum of V70, and the intraresidue correlations observed.

As the hCCcoNH-TOCSY provides correlations between the amide nuclei of residue *i* and the aliphatic $^{13}$C nuclei of residue *i*-1, the spectrum was analysed in the manner similar to the triple-resonance experiments described in the backbone assignments section of the chapter. The hCCcoNH-TOCSY was used in parallel with HCCH-TOCSY to obtain the $^{13}$C aliphatic shifts for the analysis of the HCCH-TOCSY spectra and to confirm the side-

chain assignments obtained from HCCH-TOCSY. In addition, the use of hCCcoNH-TOCSY allowed the accuracy of the sequential links between the residues to be verified.

As neither the HCCH-TOCSY nor hCCcoNH-TOCSY reveal correlations between the aliphatic and aromatic CH groups in the protein, a different strategy was adopted for assignment of the aromatic side-chain resonances, and hbCBcgcdHD, hbCBcgcdceHE and $^{1}$H, $^{13}$C TROSY aromatic spectra were recorded. Since $^{13}$Cβ chemical shifts of aromatic residues were obtained from the HCCH-TOCSY and hCCcoNH-TOCSY spectra, it was possible to find the chemical shifts of the aromatic $^{1}$Hδ and $^{1}$Hε nuclei by identified the intraresidue cross-peaks at the $^{13}$Cβ shifts in hbCBcgcdHD and hbCBcgcdceHE, respectively. After the aromatic $^{1}$Hδ and $^{1}$Hε shifts were identified, the correlated $^{13}$Cδ and $^{13}$Cε shifts could be identified from the cross-peaks in the $^{1}$H, $^{13}$C TROSY spectrum. However, unambiguous assignment was difficult to achieve due to the overlap between the chemical shifts of the aromatic side-chain nuclei in the protein. Hence, the aromatic side-chain assignments were also supported by analysis of the intra- and interresidue through-space correlations in the $^{13}$C NOESY-HSQC spectrum. Similarly, the $^{13}$C NOESY-HSQC was also used to confirm the aliphatic side-chain resonance assignments.

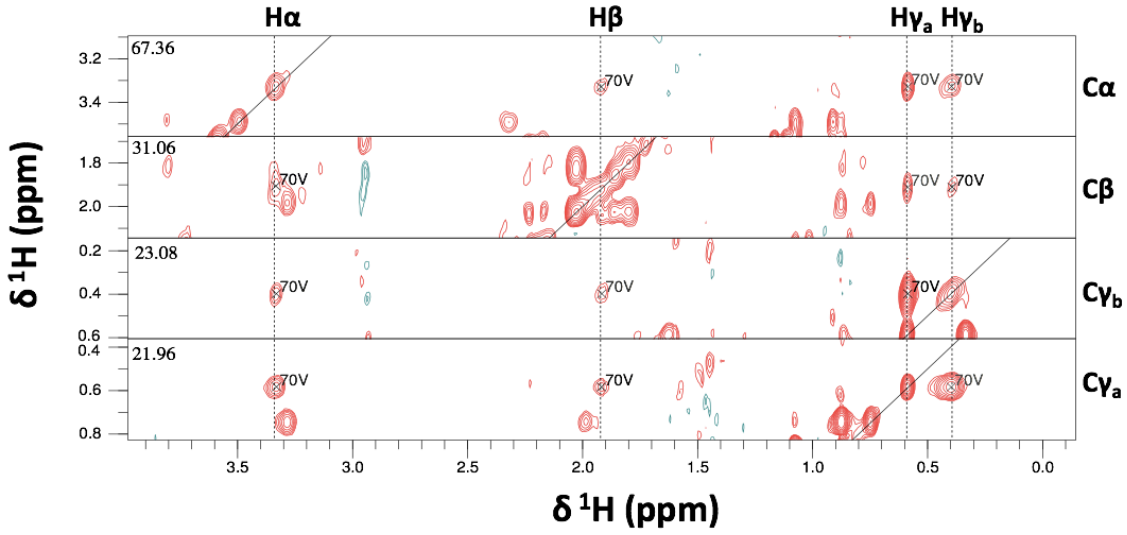



**Figure 6-6. Side-chain resonance assignment from HCcH-TOCSY.** Strips showing intraresidue V70 cross-peaks in the HCcH-TOCSY spectrum. The chemical shifts of the cross-peaks in the $^{13}$C dimension are shown in the top left corner of each strip. Cross-peaks arise due to correlations between the $^{1}$H-$^{13}$C pairs in the residue. In a strip centred at the $^{1}$H and $^{13}$C chemical shits of a single $^{1}$H-$^{13}$C pair, cross-peaks to all side-chain $^{1}$H nuclei in the residue are visible. The chemical shifts of the correlated $^{13}$C nuclei can be identified by locating the matching patterns of cross-peaks across the $^{13}$C dimension. The identical chemical shift information can be obtained from the hCCH-TOCSY spectrum.

### 6.3.3.3 Assignment of DOPA resonances

For assignment of the ligand signals, a model DOPA molecule was introduced into the CcpNmr Analysis project of the Na-FAR-1 PA complex. The model DOPA molecule was created by Dr Brian Smith (University of Glasgow) in CCPN ChemBuild software. For the assignment of ligand resonances, a 2D NOESY experiment $^{13}$C-filtered in $F_2$ was used, which is a NOESY variant that only reveals through-space correlations between the ligand $^{1}$H nuclei or the ligand and the protein $^{1}$H nuclei, but not between the protein $^{1}$H nuclei. Several ligand-ligand NOEs were observed in the spectrum (Fig 6-7). The cross-peaks were assigned to several types of aliphatic $^{1}$H nuclei in the ligand based on their chemical shifts. For instance, the most upfield cross-peaks in the aliphatic region of the spectrum were assigned to the methyl protons of the ligand, the most downfield aliphatic cross-peaks were assigned to the $COCH_2$ protons, and the cross-peaks at ~ 5.2 ppm were assigned to the methine protons. Since only four distinct DOPA $CH_2$ proton chemical shifts could be determined, the chemical shifts of the majority of $CH_2$ protons were assumed to be degenerate. The ligand chemical shifts were used to identify potential contacts between the ligand and the protein in the NOESY-HSQC spectra, as described in Section 6.3.4.

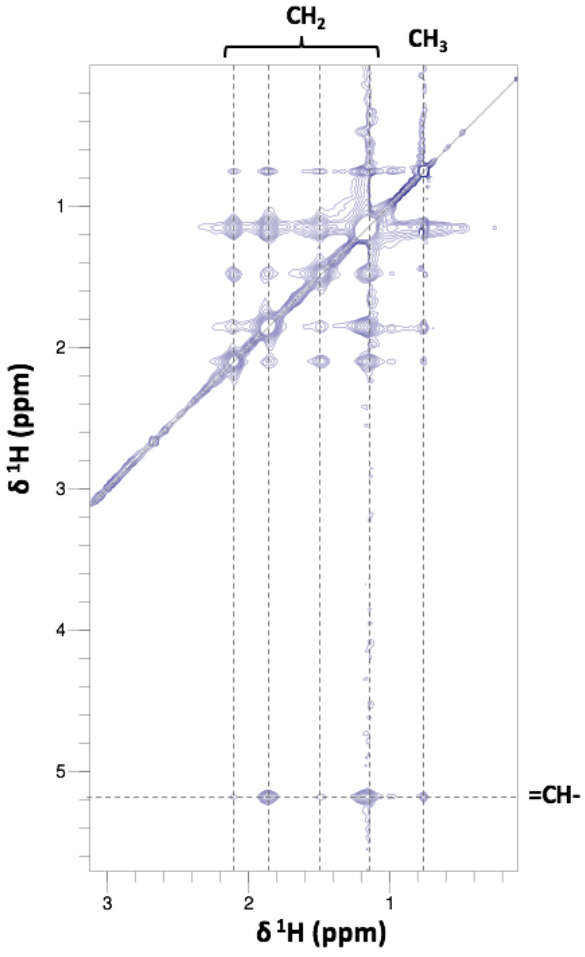



**Figure 6-7. A fragment of the $^{13}$C-filtered 2D NOESY spectrum showing ligand-ligand NOEs.** NOE signals (NOEs) between the ligand protons were used for determination of the ligand proton chemical shifts. NOEs were observed between the ligand fatty acyl tail methyl ($CH_3$), methylene ($CH_2$) and methine (=CH-) group protons. Dashed lines are shown at the chemical shift of the distinct ligand protons.

### 6.3.3.4 Protein resonance assignment summary

Final resonance assignment statistics of the Na-FAR-1-PA complex are presented in Table 6-2 and the assigned $^{1}H$, $^{15}N$ HSQC spectrum is shown in Figure 6-8. In total, 155 Na-FAR-1 residues were available for assignment, excluding the residues in the recombinant tag. No assignments were obtained for P7, P15, D27, P47, P105, F138, E144 and G146. Furthermore, as mentioned above, backbone amide resonances of 13% of residues were not assigned as the residues lacked identifiable cross-peaks in the spectra of the experiments that reveal NH correlations. Nevertheless, assignment of 86% hydrogen atoms was achieved in this study, providing a good basis for generation of $^{1}H$-$^{1}H$ distance restraints for structure determination.

| Category | Available | Assigned | % Assigned |
|---|---|---|---|
| Residues | 155 | 147 | 95 |
| C atoms | 744 | 613 | 82 |
| H atoms | 937 | 802 | 86 |
| N atoms | 200 | 133 | 67 |
| Backbone amides | 465 | 404 | 87 |
| Side-chain H atoms | 624 | 522 | 84 |
| Side-chain non-H atoms | 479 | 341 | 71 |

**Table 6-1. Assignment statistics of the Na-FAR-1 residues in the Na-FAR-1-PA complex.** The residues in the recombinant tag are not included.

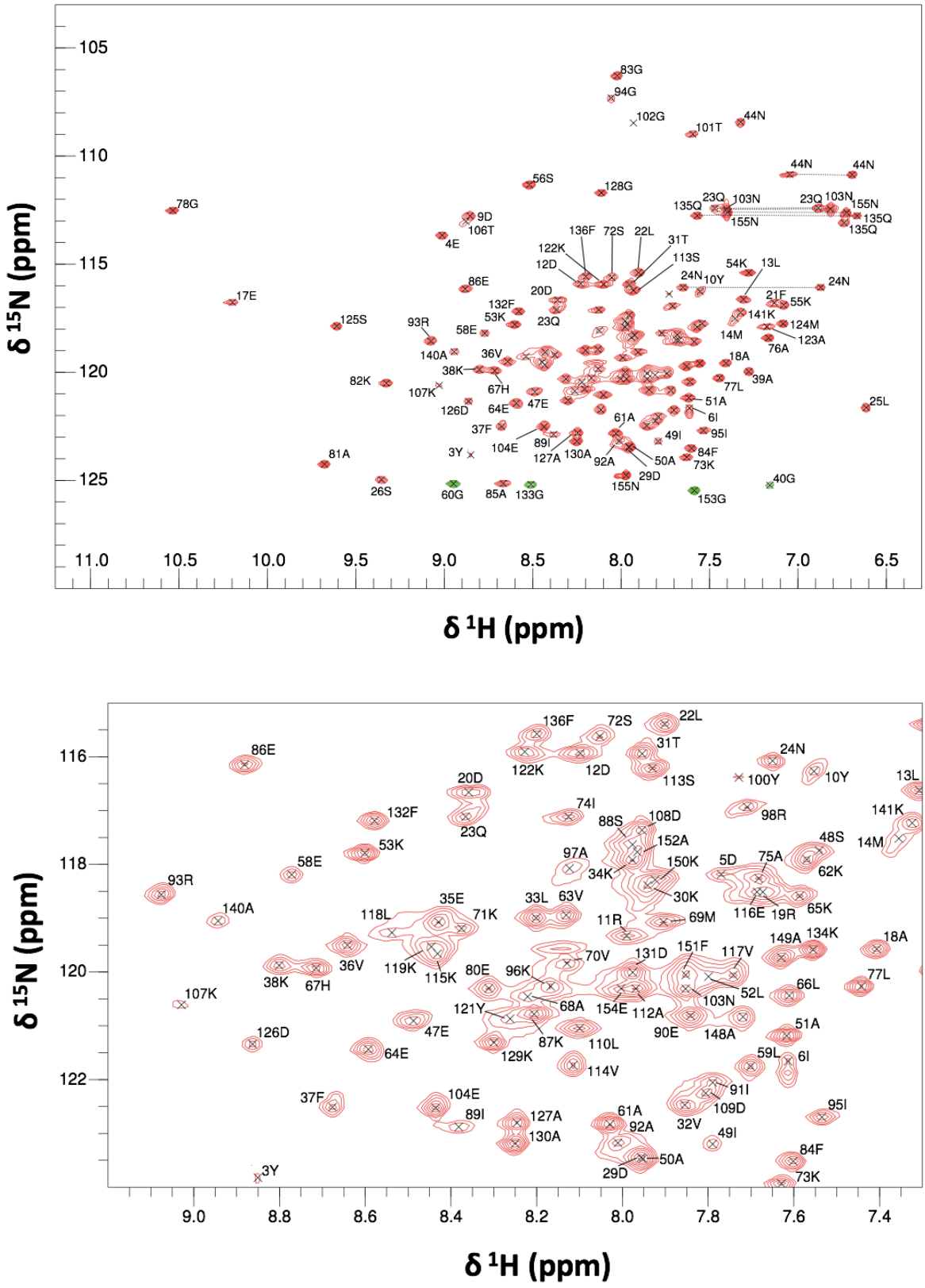



**Figure 6-8. Assigned $^{1}$H, $^{15}$N HSQC spectrum of Na-FAR-1 in complex with PA.** Top: Full view of the spectrum. Side-chain amide cross-peaks are connected by dashed lines. Assignment of the central region is not shown due to crowding. Aliased cross-peaks are shown in green. Bottom: Close-up view of the central region of the spectrum. Full assignment is shown.

### 6.3.4 Structural analysis of the Na-FAR-1-PA complex

#### 6.3.4.1 Structure calculation in ARIA

After as many protein resonances as possible were assigned, structure determination of the Na-FAR-1 complex with PA was attempted. Structure calculation was carried out in ARIA software (see Section 6.3.1.10 for description of the software).

In order to obtain $^{1}$H-$^{1}$H distance restraints for the structure calculation, the cross-peaks in the $^{15}$N and $^{13}$C NOESY-HSQC spectra were picked in CcpNmr Analysis. Care was taken to avoid picking artefact peaks, especially near the water frequency. Typically, the cross-peaks were left unassigned in the indirect dimension, to avoid introducing incorrect assignments into the structure calculation process. In cases when neither of the dimensions could be assigned unambiguously, peaks were left completely unassigned. After the peak-picking was completed, lists of ambiguous and unambiguous restraints were generated in CcpNmr Analysis and merged together according to the type of the spectrum from which they were obtained. The NOESY cross-peaks whose chemical shifts in the indirect $^{1}$H dimension were within 0.3 ppm of the ligand $^{1}$H chemical shifts were not used for restraint generation. This was done to avoid misguiding the structure calculation, as the Na-FAR-1 structure calculation was carried out without the ligand.

In addition to distance restraints, a set of dihedral angle restraints was generated by DANGLE software in CcpNmr Analysis. DANGLE uses Bayesian inference to predict $\varphi$ and $\psi$ angles and the secondary structure of the protein of interest by comparing the chemical shifts and the primary sequence of the protein to a database of proteins with known dihedral angles and chemical shifts (Cheung *et al*., 2010). Consistent with the previously determined structure of Na-FAR-1, DANGLE predicted the structure of Na-FAR-1 in the PA complex to be predominantly α-helical.

Next, the sets of restraints were imported into ARIA and the structure calculation was carried out. Nine iterations of structure calculation were carried out in each separate run. The structure calculations were performed in the torsion angle mode. For the most part, the standard ARIA protocol was used for structure calculation. Changes were made to the molecular dynamics parameters, where the number of steps at 1000 K and 50 K were increased to 8000 and 10000, respectively, and to the iteration protocols, which are summarised in Table 6-3. Since dihedral restraints were obtained from parameters that are not directly correlated with the real values of $\varphi$ and $\psi$, but from the values predicted by

DANGLE, the restraints were not used in the final stage of each iteration of the structure calculation to avoid introducing bias into the final set of models.

Although a set of model structures was obtained from ARIA, visual inspection of the models has revealed that they did not fully resemble the previously determined structures of Na-FAR-1. Furthermore, it was observed that a significant number of restraints were violated during the structure calculation. This suggested that the accurate models of the Na-FAR-1 could not be obtained due to the presence of a significant number of missassigned or artefact cross-peaks in the NOESY spectra used for restraint generation, so that too many restraints were rejected during the structure calculation and the number of the remaining restraints was not sufficient to guide the calculation. Also, it is a possibility a significant number of long-range NOEs were simply absent from the spectrum, and therefore could not be used for restraint generation. The second notion is supported by the fact that although the number of restraint violations decreased after three rounds of manual violation analysis and a degree of convergence between the models was observed, a satisfactory final model of globular Na-FAR-1 could not be obtained. Due to the time constraints, the structure determination of Na-FAR-1-PA complex was not taken further in this study. The strategies for obtaining the structure of the complex will be discussed in Section 6.3.6.

| **Iteration** | **No. structures** | **Viol. tolerance (Å)** | ***p*[a]** |
|---|---|---|---|
| 0 | 20 | 1000 | 1.0 |
| 1 | 20 | 5.0 | 0.99 |
| 2 | 20 | 3.0 | 0.99 |
| 3 | 20 | 1.0 | 0.99 |
| 4 | 20 | 1.0 | 0.98 |
| 5 | 20 | 1.0 | 0.96 |
| 6 | 20 | 0.3 | 0.95 |
| 7 | 40 | 0.3 | 0.95 |
| 8 | 40 | 0.3 | 0.95 |

a. Partial assignment cut-off.

**Table 6-2. Iteration parameters of the structure calculation in ARIA.**

### 6.3.4.2 Analysis of protein-ligand contacts from NOEs

Although the structure of the Na-FAR-1-PA complex was not determined in this study, further clues were obtained about the positions of the ligand in the complex by analysis of through-space correlations between the ligand and the protein nuclei.

$^{13}$C-filtered 2D NOESY, $^{13}$C NOESY-HSQC and $^{15}$N NOESY-HSQC were employed to determine the protein-ligand correlations. Since the $^{13}$C-filtered 2D NOESY was carried out in $D_2O$, ligand $^{1}$H correlations were primarily observed to the aromatic side-chain $^{1}$H nuclei in the protein (Fig 6-9) due to the chemical exchange of the protein amides and $D_2O$. However, three cross-peaks could also be distinguished in the chemical shift range corresponding to the backbone amide protons. Two sets of cross-peaks were assigned to A92 and V117, as these residues had observable amide cross-peaks in the $^{1}$H, $^{15}$N HSQC spectrum in $D_2O$ whose chemical shifts closely matched the chemical shifts of the $^{13}$C-filtered NOESY cross-peaks. The third set of cross-peaks could not be unambiguously assigned to any residue, as several backbone amides had similar $^{1}$H chemical shifts in the $D_2O$ $^{15}$N HSQC spectrum. Similarly, the cross-peaks arising from the correlations of the aliphatic ligand protons and the aromatic side-chain protons could not be assigned due to their broadness and the high degree of overlap between the chemical shifts of the aromatic side-chain protons.

It was expected that the through-space correlations between the ligand and the protein would also be observed in the $^{13}$C NOESY-HSQC and $^{15}$N NOESY-HSQC spectra. In the NOESY spectra, the cross-peaks potentially arising due to the protein-ligand correlations were selected based on the match between the chemical shift of the cross-peaks in the indirect $^{1}$H dimension and the ligand $^{1}$H chemical shifts. Before assigning the cross-peaks to the ligand nuclei, assignments to protein resonances were considered. If no neighbouring protein atom with a closely matching chemical shift ($\Delta\delta < 0.03$ ppm) was available for assignment, the cross-peak was assigned to the ligand atoms. The judgement about the potential distance between the protein atoms was made based on their relative position in the primary protein sequence and their separation in the previously reported structures of Na-FAR-1. Based on the NOESY data, ligand correlations to L22, L33, T45, V70, S88, I89, A92, R93, I95, L110 can be proposed. These are partially consistent with the $^{13}$C-filtered NOESY data, where correlations between the ligand and A92 were observed. More putative protein-ligand NOEs were observed in the $^{15}$N HSQC-NOESY than in the $^{13}$C-filtered NOESY, as the former experiment was carried out in $H_2O$ instead of $D_2O$. However, it should be noted that since full assignment of the Na-FAR-1-PA complex was

not obtained, at least some of the correlations observed might be to the unassigned protein $^{1}$H nuclei, and not the ligand nuclei.

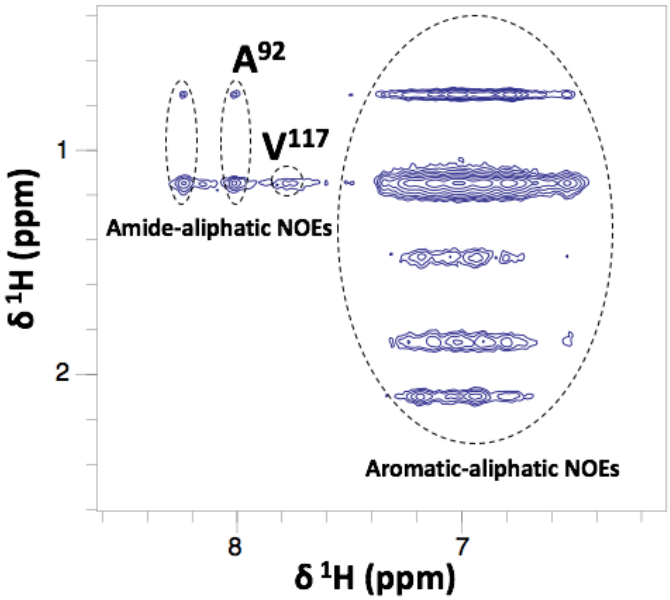



**Figure 6-9. A region of the $^{13}$C-filtered 2D NOESY spectrum showing Na-FAR-1-PA NOEs.** NOEs between the Na-FAR-1 aromatic protons and the ligand (aromatic-aliphatic NOEs), as well as between the Na-FAR-1 amide protons and the ligand (amide-aliphatic NOEs) are observed. Two of the amide-aliphatic cross-peak sets were assigned to the backbone amides of A92 and V117 in the indirect dimension. Unambiguous assignment of other cross-peaks could not be carried out.

## 6.4 Conclusions and discussion

In this study, phospholipid binding by Na-FAR-1 was characterised using a fluorescence-based assay and protein NMR spectroscopy. It was determined that Na-FAR-1 binds PA, which is in agreement with the previous experiments that demonstrated PE and PC binding by Na-FAR-1 (Rey-Burusco *et al.*, 2015). Importantly, this study is the first to demonstrate lysophosphatidic acid (LPA) binding by Na-FAR-1. The interaction between LPA and Na-FAR-1 is especially interesting due to the known role of LPA in cell signalling and the fact that LPA is present in the extracellular environment where Na-FAR-1 can encounter it after being secreted by the parasite. The biological implications of the Na-FAR-1 phospholipid binding will be discussed in more detail in Chapter 7.

From the chemical shift perturbation analysis of the Na-FAR-1-PA interaction, it was apparent that PA binding induces a conformational change in the protein. By monitoring the exchange regimes during the DOPA titration, it was also apparent that the protein is able to form three distinct complexes with the ligand, which likely contain one, two and three or more DOPA molecules. Backbone amide chemical shift changes were observed with the majority of the protein residues, but were mostly concentrated in the central

regions of the cavity-forming helices α4, α6 and α7. Although the structure of the Na-FAR-1-PA could not be determined, 86% hydrogen atoms and 84% carbon atoms in Na-FAR-1 were assigned from the NMR spectra of the Na-FAR-1-PA complex, and several chemical shifts of the ligand $^{1}$H nuclei were obtained from the $^{13}$C-filtered 2D NOESY spectrum. Based on the assignments, putative contacts between the ligands and L22, L33, T45, V70, S88, I89, A92, R93, I95, L110 and V117 were established from the $^{13}$C and $^{15}$N NOESY-HSQC experiments. Of these residues, L22 is located in α3, L33 in α4, V70 in α6, S88, I89, A92, R93 and I95 in α7, L100 and V117 in α8. The fact that most putative ligand correlations were observed to the residues in α7 is consistent with the chemical shift changes observed in α7 in the chemical shift perturbation analysis. Interestingly, some of the largest chemical shift changes in Na-FAR-1 induced by DOPA addition were observed for V70, R93 and I95, which provides more support for their hypothetical role in DOPA binding.

Since strict criteria were used for identification of the protein-ligand correlations from NOEs, it is a possibility that at least a number of cross-peaks arising from to the Na-FAR-1-PA contacts were selected against during the analysis of the NOESY-HSQC spectra. In order to map the protein-ligand contacts with greater accuracy, determination of the structure of Na-FAR-1 complex with PA will be required. From the Na-FAR-1 structure in complex with PA, the distances between the protein atoms can be estimated more accurately and the conformation of the residue side-chains will be apparent, which will help to analyse the through-space correlations in the NOESY spectra and identify the real protein-ligand correlations. Furthermore, structure determination of the complex can be attempted by including the ligand model into the structure calculation.

### 6.4.1 Comparison of the Na-FAR-1 oleic acid and PA binding mechanisms

Based on the Na-FAR-1 lipid binding data obtained in this study, comparison between the Na-FAR-1 interaction mechanisms with oleic acid and PA can be made. From the previous chemical shift perturbation studies of Na-FAR-1-oleic acid binding, it was clear that Na-FAR-1 can make four distinct complexes with oleic acid that likely contain one, two, three and four or more oleic acid molecules (Rey-Burusco *et al*., 2015). A similar picture was observed in Na-FAR-1-PA binding, where Na-FAR-1 appeared to make distinct complexes with one, two and three or more DOPA molecules. The differences in the number of distinct ligand complexes that Na-FAR-1 can form with oleic acid and DOPA likely stems from the structural differences between the ligands. It should be noted that since the volume of DOPA is more than two times larger than the volume of oleic acid, it is likely

that the volume of the Na-FAR-1 central cavity saturated with PA is larger than the cavity volume in the structure of Na-FAR-1 in complex with oleic acid.

Further conclusions about the mechanism of ligand binding by Na-FAR-1 can be made from the exchange regimes observed during oleic acid and PA titration. During the oleic acid titration, it was observed that Na-FAR-1 is in a slow exchange regime in the presence of 1, 2 and 3 molar equivalents of oleic acid, and in fast exchange regime in the presence of 4 oleic acid equivalents, suggesting that the protein binds three oleic acid molecules with high affinity, and the fourth oleic acid molecule with lower affinity (Rey-Burusco *et al*., 2015). In contrast, during DOPA titration, slow exchange was only observed when the ligand was added in up to two-molar equivalents to the protein, indicating that the protein binds two DOPA molecules with high affinity, and a third DOPA molecule with lower affinity. The proposed differences in the affinities of DOPA and oleic acid binding by Na-FAR-1 can also be explained by the structural differences between the ligands. Since DOPA is bulkier than oleic acid, it is likely that the high affinity lipid binding sites in Na-FAR-1 are saturated by a smaller number of DOPA molecules than oleic acid molecules.

As many of the backbone amide chemical shifts of Na-FAR-1 are different in the oleic acid and PA complexes, it is likely that the conformation of Na-FAR-1 is dissimilar in the two complexes. A hint as to the conformational dissimilarity can also be obtained from the different crystallisation properties of the Na-FAR-1 complexes with oleic acid and DOPA. Na-FAR-1 in complex with oleic acid crystallised readily in this study (see Chapter 2 for conditions), whereas no crystals of the Na-FAR-1 DOPA complex were obtained in the identical or any other conditions tested in this study. This could suggest that the Na-FAR-1 complex with DOPA has a different conformation and/or higher conformational heterogeneity compared to the oleic acid complex, which can affect crystal packing.

Since the structure of the Na-FAR-1-DOPA complex was not solved, direct comparison of the Na-FAR-1 structure in the two complexes cannot be made. However, from the analysis of the protein-ligand NOEs in the Na-FAR-1-DOPA complex it can be concluded that at least several Na-FAR-1 residues are involved in the binding of both DOPA and oleic acid. These likely include L22, L33, S88, I89, A92, I95 and V117, apparent ligand contacts to which were observed in both complexes. Furthermore, S88, I89, A92, R93, I95, which displayed putative contacts to the ligand in the DOPA complex, are found in proximity to K96, which made hydrogen bonds to oleic acid head group in the Na-FAR-1-oleic acid complex. Since K96 is found in a region of the cavity surface with a positive electrostatic potential and carries a positive charge itself, it is likely that K96 can also interact with the

negatively charged head group of DOPA. If the DOPA head group is indeed bound to K96, DOPA fatty acyl tails might extend along the cavity wall where they could contact S88, I89, A92, R93 and I95, giving rise to the correlations observed in the NOESY spectra.

### 6.4.2 Future directions

In order to complete the structure determination of Na-FAR-1-DOPA complex, further manual analysis of the restraint violation lists produced by ARIA can be carried out to eliminate any remaining misassignments in the NOESY spectra. This will result in a higher number of restraints available for structure calculation which might then lead to generation of a more accurate model of the Na-FAR-1 structure. Furthermore, paramagnetic relaxation enhancement (PRE) (Nadaud *et al*., 2007) can be used to obtain additional long-range distance restraints, and residual dipolar coupling (RDC) (Chen and Tjandra, 2012) can be used to obtain directional information about the dipole-dipole interactions between the protein nuclei. These two types of restraints are often very useful in constraining the structure calculation when the restraints obtained from NOEs are not sufficient to guide the calculation. With the use of PRE and/or RDC restraints, and an increased number of NOE restraints, it is highly likely that the solution structure of the Na-FAR-1-PA complex can be solved in the future.

To determine whether Na-FAR-1 shows preference for certain classes of phospholipids, the affinity of the interaction between various phospholipids and Na-FAR-1 can be assessed by, e.g., the DAUDA displacement assay used in this study. It would be especially interesting to compare the affinity of Na-FAR-1 to lysophospholipids and diacylphospholipids, as well as to phospholipids with different net charges. The comparison of Na-FAR-1 affinity to different types of phospholipids could reveal which phospholipids Na-FAR-1 is more likely to interact with and hence what the biological function of the phospholipid binding of Na-FAR-1 may be. Further, Na-FAR-1 binding to other classes of bioactive lipids could be characterised, including prostaglandins, thromboxanes and specialised pro-resolving lipid mediators. Of special interest here are the pro-inflammatory lipid messengers such as 2-arachidonoglycerol, arachidonic acid, prostaglandin $E_2$ and prostaglandin $D_2$, by binding to which Na-FAR-1 could modulate immune response in the host upon parasite infection.

# 7. Final summary

## 7.1 The LNS2 domain of Nir2

This study has provided first insights into the mechanism of phosphatidic acid (PA) binding by the LNS2 domain of Nir2, which has been previously proposed to regulate the plasma membrane association and lipid transport activity of Nir2 (Chang and Liou, 2015; Kim *et al*., 2013, 2015). For the biochemical and structural characterisation, the LNS2 domain was expressed in *E. coli* as a recombinant Trigger factor fusion protein. This allowed the yield, stability and solubility of the recombinant LNS2 domain to be enhanced, as the LNS2 domain was found to be highly prone to aggregation when expressed without a large fusion partner.

By employing a number of biochemical and biophysical techniques, the PA binding activity of the recombinant LNS2 domain was demonstrated *in vitro*. Further ligand-observed NMR spectroscopy experiments using short-chain phospholipids and phospholipid fragments showed that the LNS2 domain of Nir2 interacts with both the fatty acyl and the glycerol backbone regions of the PA molecule, and that the presence of both these regions in a molecule was required for the interaction to occur. Moreover, the LNS2 domain was found to interact with the polar region of PA and not with the polar region of PC, which contains a positively charged head group. Although the interaction between the fatty acyl tails of PC and the LNS2 domain was also observed, it appeared to be weaker than that with the fatty acyl tails of PA, indicating that the negatively charged head group of PA is involved in binding. Furthermore, as it was previously demonstrated that the Nir2 LNS2 domain binds only to lipid vesicles containing PA but not to lipid vesicles containing PS (Kim *et al*., 2013) which, like PA, is negatively charged, it is highly likely that the LNS2 domain is a specific PA binder. The structural mechanism behind the PA specificity of LNS2 remains unclear, but the binding could occur via the hydrogen bond/electrostatic switch mechanism previously proposed for other PA binding proteins (Kooijman *et al*., 2007).

The fact that Nir2 LNS2 binds to both the fatty acyl chains and the polar region of PA and does not bind to free glycerol-3-phosphate, which constitutes the polar region of PA, indicates that the plasma membrane association mechanism of Nir2 LNS2 might involve both an interaction with the hydrophobic region of the lipid bilayer and hydrogen bonding or electrostatic interactions with polar region of PA. Interestingly, the mechanism of the

LNS2 domain phospholipid binding appears to be different to that of the PH domain, which has been shown to interact with the isolated PI(4,5)$P_2$ head group with higher affinity than with the intact phospholipid. In contrast, the ligand binding properties of the LNS2 domain appear to be similar to those of the PI(3)P-binding FYVE domain and DAG and phorbol ester-binding C1 domain, which bind the membrane-embedded ligands with higher affinity than the free ligands (Gaullier *et al*., 2000; Kazanietz *et al*., 1995; Kutateladze *et al*., 1999). A distinguishing feature of the membrane association by the FYVE and C1 domains is the insertion of specific hydrophobic residues into the membrane upon ligand binding, which does not normally occur during membrane association by the PH domain. Hence, due to the inability of the LNS2 domain to bind the isolated polar region of PA, it can be proposed that membrane insertion plays a role in membrane binding by the LNS2 domain. Consistent with this notion, the LNS2 domain was estimated to have a higher affinity for the membrane-embedded PA than for the free PA. However, since the measurement of the LNS2 affinity to membrane-embedded PA was carried out only in a semi-quantitative manner, further experiments would be required to confirm the findings. A strategy to test the LNS2 membrane insertion hypothesis was proposed in Section 4.6.3, along with several strategies for further analysis of the LNS2-PA interaction.

Due to the large size of the Trigger factor fusion of the LNS2 domain that impeded its analysis by protein NMR spectroscopy, the failure of the fusion protein to crystallise, and the difficulty of obtaining high amounts of pure and homogenous Nir2 LNS2 domain without the fusion partner, the three-dimensional structure of the LNS2 domain could not be determined in this study. However, analysis of the Nir1 LNS2 domain by CD spectroscopy was carried out, which revealed that the LNS2 domain has a β-sheet-rich structure.

Since atomic details of the LNS2-PA binding could not be obtained, a tool compound targeting the interaction could not be designed rationally. However, based on the results of the fluorescence lifetime measurements and ligand-observed NMR experiments, two assays were proposed for the screening of a small library of compounds or fragments for identification of the LNS2-PA interaction inhibitors (see Section 4.5.2 for the description of the assays).

## 7.2 Na-FAR-1

The high-resolution structure of Na-FAR-1 in complex with oleic acid was determined in this study. Four ligand molecules were identified in the complex, and the contacts between

the ligands and the protein residues were mapped. It was observed that the central cavity of Na-FAR-1 was more expanded when the protein was complexed with oleic acid than when it was complexed with a mixture of *E. coli* lipids (Rey-Burusco *et al*., 2015), suggesting that the volume of the Na-FAR-1 cavity might change to allow accommodation of different ligand types or formation of ligand complexes with varying stoichiometries.

Furthermore, it was demonstrated that Na-FAR-1 interacts with PA and lysophosphatidic acid (LPA), which provides the first evidence of Na-FAR-1 interaction with lysophospholipids. If Na-FAR-1 also binds LPA *in vivo*, this interaction might have wide-reaching biological implications. This is because LPA is a well-known mediator of inflammatory signalling (Yung *et al*., 2014) that also facilitates wound healing in the gut (Khurana *et al*., 2008). Hence, by binding to and, possibly, sequestering LPA, Na-FAR-1 could affect the inflammatory responses in the host, thereby allowing the hookworm to modulate the host's immune system to its benefit. Potential LPA sequestration by Na-FAR-1 might also inhibit the healing of the feeding wounds produced by the worm in the gut epithelium, which could enhance hookworm survival in the host. However, it is clear that LPA is only one molecule from the wide repertoire of lipids that Na-FAR-1 is known to interact with. Therefore, if the Na-FAR-1-LPA interaction does have a biological role, it is expected that the probable Na-FAR-1 function in parasitism is not restricted to LPA binding.

Analysis of the Na-FAR-1-PA binding mechanism was carried out by NMR spectroscopy. From the chemical shift perturbation experiments it was observed that Na-FAR-1 formed three distinct complexes with PA, which were assumed to contain one, two and at least three PA molecules. The appearance of the slow chemical exchange regime upon addition of two molar equivalents of PA to Na-FAR-1 and the later transition into the fast exchange regime upon further additions of PA indicated that Na-FAR-1 binds two molecules of PA with higher affinity than the third and any subsequent ones. Similar findings were previously reported for oleic acid binding by Na-FAR-1 (Rey-Burusco *et al*., 2015), suggesting that Na-FAR-1 employs similar mechanisms for binding of different lipids classes. To map the contacts between the ligand and the protein, structure determination of the Na-FAR-1-PA complex was attempted by NMR spectroscopy. Although assignment of 82% of carbon atoms and 85% of hydrogen atoms of Na-FAR-1 was achieved, structure determination could not be completed due to the time constraints. Nevertheless, it is very likely that the structure of the complex can be obtained in the future, providing further insights into the lipid binding functionality of Na-FAR-1.

# References

Abrahams, J.P., Leslie, A.G.W., Lutter, R., Walker, J.E., 1994. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Nature 370, 621–628.

Ahearn, I.M., Haigis, K., Bar-Sagi, D., Philips, M.R., 2012. Regulating the regulator: post-translational modification of RAS. Nat. Rev. Mol. Cell. Biol. 13, 39–51.

Altomare, D.A., Testa, J.R., 2005. Perturbations of the AKT signaling pathway in human cancer. Oncogene 24, 7455–7464.

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

Altucci, L., Leibowitz, M.D., Ogilvie, K.M., de Lera, A.R., Gronemeyer, H., 2007. RAR and RXR modulation in cancer and metabolic disease. Nat. Rev. Drug. Discov. 6, 793–810.

Amarilio, R., Ramachandran, S., Sabanay, H., Lev, S., 2005. Differential Regulation of Endoplasmic Reticulum Structure through VAP-Nir Protein Interaction. J. Biol. Chem. 280, 5934–5944.

Andrade, M., Chacón, P., Merelo, J., Morán, F., 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng. 6, 383–390.

Angulo, J., Nieto, P.M., 2011. STD-NMR: application to transient interactions between biomolecules—a quantitative approach. Eur. Biophys. J. 40, 1357–1369.

Antanasijevic, A., Ramirez, B., Caffrey, M., 2014. Comparison of the Sensitivities of WaterLOGSY and Saturation Transfer Difference NMR Experiments. J. Biomol. NMR 60, 37–44.

Asencio-Hernández, J., Kieffer, B., Delsuc, M.-A., 2016. NMR WaterLOGSY Reveals Weak Binding of Bisphenol A with Amyloid Fibers of a Conserved 11 Residue Peptide from Androgen Receptor. PLoS One 11, e0161948.

Baier, L.J., Sacchettini, J.C., Knowler, W.C., Eads, J., Paolisso, G., Tataranni, P.A., Mochizuki, H., Bennett, P.H., Bogardus, C., Prochazka, M., 1995. An amino acid substitution in the human intestinal fatty acid binding protein is associated with

increased fatty acid binding, increased fat oxidation, and insulin resistance. J. Clin. Invest. 95, 1281–1287.

Balla, T., 2013. Phosphoinositides: Tiny Lipids With Giant Impact on Cell Regulation. Physiol. Rev. 93, 1019–1137.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A., 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. Science 289, 905–920.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol. 10, 411–421.

Basavaraju, S., Zhan, B., Kennedy, M.W., Liu, Y., Hawdon, J., Hotez, P.J., 2003. Ac-FAR-1, a 20 kDa fatty acid- and retinol-binding protein secreted by adult Ancylostoma caninum hookworms: gene transcription pattern, ligand binding properties and structural characterisation. Mol. Biochem. Parasitol. 126, 63–71.

Bax, A., Clore, G.M., Gronenborn, A.M., 1990. 1H 1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. J. Magn. Reson. 88, 425–431.

Bax, A., Ikura, M., 1991. An efficient 3D NMR technique for correlating the proton and15N backbone amide resonances with the α-carbon of the preceding residue in uniformly15N/13C enriched proteins. J. Biomol. NMR 1, 99–104.

Bernal, J., Crowfoot, D., 1934. X-Ray Photographs of Crystalline Pepsin. Nature 133, 794–795.

Berridge, M.J., 1983. Rapid accumulation of inositol trisphosphate reveals that agonists hydrolyse polyphosphoinositides instead of phosphatidylinositol. Biochem. J. 212, 849–858.

Bian, D., Su, S., Mahanivong, C., Cheng, R.K., Han, Q., Pan, Z.K., Sun, P., Huang, S., 2004. Lysophosphatidic Acid Stimulates Ovarian Cancer Cell Migration via a Ras-MEK Kinase 1 Pathway. Cancer Res. 64, 4209–4217.

Bietz, S., Urbaczek, S., Schulz, B., Rarey, M., 2014. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. J. Cheminform. 6, 12.

Binkowski, T.A., Naghibzadeh, S., Liang, J., 2003. CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res. 31, 3352–3355.

Bird, L.E., 2016. OPPF-UK Standard Protocols: Cloning and Expression Screening [WWW Document].

Bird, L.E., 2011. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. Methods 55, 29–37.

Bloch, F., Hansen, W.W., Packard, M., 1946. Nuclear Induction. Phys. Rev. 69, 127.

Boettcher, A., Gradoux, N., Lorthiois, E., Brandl, T., Orain, D., Schiering, N., Cumin, F., Woelcke, J., Hassiepen, U., 2014. Fluorescence Lifetime–Based Competitive Binding Assays for Measuring the Binding Potency of Protease Inhibitors In Vitro. J. Biomol. Screen. 19, 870–877.

Bravo, J., Karathanassis, D., Pacold, C.M., Pacold, M.E., Ellson, C.D., Anderson, K.E., Butler, P.J.G., Lavenir, I., Perisic, O., Hawkins, P.T., Stephens, L., Williams, R.L., 2001. The Crystal Structure of the PX Domain from p40phox Bound to Phosphatidylinositol 3-Phosphate. Mol. Cell 8, 829–839.

Brünger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., Warren, G., 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–21.

Bullen, H.E., Jia, Y., Yamaryo-Botté, Y., Bisio, H., Zhang, O., Jemelin, N.K., Marq, J.-B., Carruthers, V., Botté, C.Y., Soldati-Favre, D., 2016. Phosphatidic Acid-Mediated Signaling Regulates Microneme Secretion in Toxoplasma. Cell Host Microbe 19, 349–360.

Bunney, T.D., Katan, M., 2010. Phosphoinositide signalling in cancer: beyond PI3K and PTEN. Nat. Rev. Cancer 10, 342–352.

Burger, K.N., Demel, R.A., Schmid, S.L., de Kruijff, B., 2000. Dynamin is membrane-active: lipid insertion is induced by phosphoinositides and phosphatidic acid. Biochemistry 39, 12485–12493.

Cantorna, M.T., Nashold, F.E., Hayes, C.E., 1994. In vitamin A deficiency multiple mechanisms establish a regulatory T helper cell imbalance with excess Th1 and insufficient Th2 function. J. Immunol. 152, 1515–1522.

Carlson, E.D., Gan, R., Hodgman, C.E., Jewett, M.C., 2012. Cell-Free Protein Synthesis: Applications Come of Age. Biotechnol. Adv. 30, 1185–1194.

Carman, J., Pond, L., Nashold, F., Wassom, D., Hayes, C., 1992. Immunity to *Trichinella spiralis* infection in vitamin A-deficient mice. J. Exp. Med. 175, 111–120.

Carr, H.Y., Purcell, E.M., 1954. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. Phys. Rev. 94, 630–638.

Cavanagh, J., Skelton, N., Fairbrother, W., Rance, M., Palmer, A.I., 2007. Protein NMR Spectroscopy, 2nd Edition. Principles and Practice. Academic Press.

Chang, C.-L., Hsieh, T.-S., Yang, T.T., Rothberg, K.G., Azizoglu, D.B., Volk, E., Liao, J.-C., Liou, J., 2013. Feedback Regulation of Receptor-Induced Ca2+ Signaling Mediated by E-Syt1 and Nir2 at Endoplasmic Reticulum-Plasma Membrane Junctions. Cell Rep. 5, 813–825.

Chang, C.-L., Liou, J., 2015. Phosphatidylinositol 4,5-Bisphosphate Homeostasis Regulated by Nir2 and Nir3 Proteins at Endoplasmic Reticulum-Plasma Membrane Junctions. J. Biol. Chem. 290, 14289–14301.

Chattopadhyay, A., London, E., 1988. Spectroscopic and ionization properties of N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-labeled lipids in model membranes. Biochim. Biophys. Acta - Biomembr. 938, 24–34.

Chattopadhyay, A., Mukherjee, S., 1993. Fluorophore environments in membrane-bound probes: A red edge excitation shift study. Biochemistry 32, 3804–3811.

Cheever, M.L., Sato, T.K., de Beer, T., Kutateladze, T.G., Emr, S.D., Overduin, M., 2001. Phox domain interaction with PtdIns(3)P targets the Vam7 t-SNARE to vacuole membranes. Nat. Cell. Biol. 3, 613-618.

Chen, J., Yao, Y., Gong, C., Yu, F., Su, S., Chen, J., Liu, B., Deng, H., Wang, F., Lin, L., Yao, H., Su, F., Anderson, K.S., Liu, Q., Ewen, M.E., Yao, X., Song, E., 2011. CCL18 from Tumor-Associated Macrophages Promotes Breast Cancer Metastasis via PITPNM3. Cancer Cell 19, 541–555.

Chen, K., Tjandra, N., 2012. The Use of Residual Dipolar Coupling in Studying Proteins by NMR. Top. Curr. Chem. 326, 47–67.

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

Cheung, M.-S., Maguire, M.L., Stevens, T.J., Broadhurst, R.W., 2010. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J. Magn. Reson. 202, 223–233.

Chmurzyńska, A., 2006. The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. J. Appl. Genet. 47, 39–48.

Choi, J.W., Herr, D.R., Noguchi, K., Yung, Y.C., Lee, C.-W., Mutoh, T., Lin, M.-E., Teo, S.T., Park, K.E., Mosley, A.N., Chun, J., 2010. LPA Receptors: Subtypes and Biological Actions. Annu. Rev. Pharmacol. Toxicol. 50, 157–186.

Clapham, D.E., 2007. Calcium Signaling. Cell 131, 1047–1058.

Clegg, R.M., 2009. Chapter 1 Förster resonance energy transfer—FRET what is it, why do it, and how it's done. Lab. Tech. Biochem. Mol. Biol. 33, 1–57.

Clubb, R.T., Thanabal, V., Wagner, G., 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N, 13C-labelled proteins. J. Magn. Reson. 97, 213–217.

Cockcroft, S., 1999. Mammalian phosphatidylinositol transfer proteins: emerging roles in signal transduction and vesicular traffic. Chem. Phys. Lipids 98, 23–33.

Cockcroft, S., Carvou, N., 2007. Biochemical and biological functions of class I phosphatidylinositol transfer proteins. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1771, 677–691.

Cockcroft, S., Garner, K., 2013. Potential role for phosphatidylinositol transfer protein (PITP) family in lipid transfer during phospholipase C signalling. Adv. Biol. Regul. 53, 280–291.

Cockcroft, S., Garner, K., 2011. Function of the phosphatidylinositol transfer protein gene family: is phosphatidylinositol transfer the mechanism of action? Crit. Rev. Biochem. Mol. Biol. 46, 89–117.

Cockcroft, S., Garner, K., Yadav, S., Gomez-Espinoza, E., Raghu, P., 2016. RdgBα reciprocally transfers PA and PI at ER–PM contact sites to maintain PI(4,5) P2 homoeostasis during phospholipase C signalling in *Drosophila* photoreceptors. Biochem. Soc. Trans. 44, 286–292.

Cockcroft, S., Raghu, P., 2016. Topological organisation of the phosphatidylinositol 4,5-

bisphosphate–phospholipase C resynthesis cycle: PITPs bridge the ER–PM gap. Biochem. J. 473, 4289–4310.

Compton, L.A., Johnson, W.C., 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal. Biochem. 155, 155–167.

Cosker, K.E., Shadan, S., van Diepen, M., Morgan, C., Li, M., Allen-Baume, V., Hobbs, C., Doherty, P., Cockcroft, S., Eickholt, B.J., 2008. Regulation of PI3K signalling by the phosphatidylinositol transfer protein PITPα during axonal extension in hippocampal neurons. J. Cell Sci. 121, 796–803.

Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science 292, 1863–1876.

Creba, J.A., Downes, C.P., Hawkins, P.T., Brewster, G., Michell, R.H., Kirk, C.J., 1983. Rapid breakdown of phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate in rat hepatocytes stimulated by vasopressin and other Ca2+-mobilizing hormones. Biochem. J. 212, 733–747.

Cullen, P.J., Cozier, G.E., Banting, G., Mellor, H., 2001. Modular phosphoinositide-binding domains; their role in signalling and membrane trafficking. Curr. Biol. 11, R882–R893.

Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., Stockman, B., 2001. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. J. Biomol. NMR 21, 349–359.

Das, J., Rahman, G.M., 2014. C1 Domains: Structure and Ligand-Binding Properties. Chem. Rev. 114, 12108–12131.

Davis, G.D., Elisee, C., Newham, D.M. & Harrison, R.G., 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol. Bioeng. 65, 382–388.

Dawson, H., Solano-Aguilar, G., Beal, M., Beshah, E., Vangimalla, V., Jones, E., Botero, S., Urban, J.F., 2009. Localized Th1-, Th2-, T Regulatory Cell-, and Inflammation-Associated Hepatic and Pulmonary Immune Responses in *Ascaris suum*-Infected Swine Are Increased by Retinoic Acid. Infect. Immun. 77, 2576–2587.

de Silva, N.R., Brooker, S., Hotez, P.J., Montresor, A., Engels, D., Savioli, L., 2003. Soil-

transmitted helminth infections: updating the global picture. Trends Parasitol. 19, 547–551.

Dennis, E.A., Norris, P.C., 2015. Eicosanoid Storm in Infection and Inflammation. Nat. Rev. Immunol. 15, 511–523.

Diamond, R., 1971. A real-space refinement procedure for proteins. Acta Crystallogr. Sect. A 27, 436–452.

Di Paolo, G., De Camilli, P., 2006. Phosphoinositides in cell regulation and membrane dynamics. Nature 443, 651–657.

Donkor, J., Sariahmetoglu, M., Dewald, J., Brindley, D.N., Reue, K., 2007. Three Mammalian Lipins Act as Phosphatidate Phosphatases with Distinct Tissue Expression Patterns. J. Biol. Chem. 282, 3450–3457.

Drenth, J., 2007. Principles of Protein X-Ray Crystallography, 3rd ed., Springer-Verlag New York.

Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J., 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43, W389–W394.

Edavettal, S.C., Hunter, M.J., Swanson, R. V, 2012. Genetic Construct Design and Recombinant Protein Expression for Structural Biology. In: Structure-Based Drug Discovery. Methods in Molecular Biology (Methods and Protocols). Humana Press, pp. 29-47.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D 66, 486–501.

Evans, P., McCoy, A., 2008. An introduction to molecular replacement. Acta Crystallogr. Sect. D Biol. Crystallogr. 64, 1–10.

Fahy, E., Cotter, D., Sud, M., Subramaniam, S., 2011. Lipid classification, structures and tools. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1811, 637–647.

Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L., Dennis, E.A., 2005. A comprehensive classification system for lipids. J. Lipid Res. 46, 839–862.

Fahy, E., Subramaniam, S., Murphy, R.C., Nishijima, M., Raetz, C.R.H., Shimizu, T.,

Spener, F., van Meer, G., Wakelam, M.J.O., Dennis, E.A., 2009. Update of the LIPID MAPS comprehensive classification system for lipids. J. Lipid Res. 50, S9–S14.

Fang, Y., Vilella-Bach, M., Bachmann, R., Flanigan, A., Chen, J., 2001. Phosphatidic acid-mediated mitogenic activation of mTOR signaling. Science 294, 1942-1945.

Farmer, B.T., Venters, R.A., Spicer, L.D., Wittekind, M.G., Müller, L., 1992. A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. J. Biomol. NMR 2, 195–202.

Feng, J., Wehbi, H., Roberts, M.F., 2002. Role of Tryptophan Residues in Interfacial Binding of Phosphatidylinositol-specific Phospholipase C. J. Biol. Chem. 277, 19867–19875.

Ferguson, K.M., Lemmon, M.A., Schlessinger, J., Sigler, P.B., 1995. Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. Cell 83, 1037–1046.

Fernandis, A.Z., Wenk, M.R., 2007. Membrane lipids as signaling molecules. Curr. Opin. Lipidol. 18, 121-128.

Fessler, M.B., Rudel, L.L., Brown, M., 2009. Toll-like receptor signaling links dietary fatty acids to the metabolic syndrome. Curr. Opin. Lipidol. 20, 379–385.

Fiaux, J., Bertelsen, E.B., Horwich, A.L., Wuthrich, K., 2002. NMR analysis of a 900K GroEL-GroES complex. Nature 418, 207–211.

Fielding, L., Rutherford, S., Fletcher, D., 2005. Determination of protein–ligand binding affinity by NMR: observations from serum albumin model systems. Magn. Reson. Chem. 43, 463–470.

Fishman, D.A., Liu, Y., Ellerbroek, S.M., Stack, M.S., 2001. Lysophosphatidic Acid Promotes Matrix Metalloproteinase (MMP) Activation and MMP-dependent Invasion in Ovarian Cancer Cells. Cancer Res. 61, 3194–3199.

Flesch, F.M., Yu, J.W., Lemmon, M.A., Burger, K.N.J., 2005. Membrane activity of the phospholipase C-δ(1) pleckstrin homology (PH) domain. Biochem. J. 389, 435–441.

Forget, S.M., Jee, A., Smithen, D.A., Jagdhane, R., Anjum, S., Beaton, S.A., Palmer, D.R.J., Syvitski, R.T., Jakeman, D.L., 2015. Kinetic evaluation of glucose 1-phosphate analogues with a thymidylyltransferase using a continuous coupled enzyme

assay. Org. Biomol. Chem. 13, 866–875.

Franchini, G.R., Pórfido, J.L., Ibáñez Shimabukuro, M., Rey Burusco, M.F., Bélgamo, J.A., Smith, B.O., Kennedy, M.W., Córsico, B., 2015. The unusual lipid binding proteins of parasitic helminths and their potential roles in parasitism and as therapeutic targets. Prostaglandins, Leukot. Essent. Fat. Acids 93, 31–36.

Gabrielsen, M., Rey-Burusco, M.F., Griffiths, K., Roe, A.J., Cooper, A., Smith, B.O., Kennedy, M.W., Corsico, B., 2012. Two crystal forms of a helix-rich fatty acid- and retinol-binding protein, Na-FAR-1, from the parasitic nematode Necator americanus . Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 68, 835–838.

Garofalo, A., Kläger, S.L., Rowlinson, M.-C., Nirmalan, N., Klion, A., Allen, J.E., Kennedy, M.W., Bradley, J.E., 2002. The FAR proteins of filarial nematodes: secretion, glycosylation and lipid binding characteristics. Mol. Biochem. Parasitol. 122, 161–170.

Garofalo, A., Rowlinson, M.-C., Amambua, N.A., Hughes, J.M., Kelly, S.M., Price, N.C., Cooper, A., Watson, D.G., Kennedy, M.W., Bradley, J.E., 2003. The FAR Protein Family of the Nematode *Caenorhabditis elegans*: Differential Lipid Binding Properties, Structural Characteristics, and Developmental Regulation. J. Biol. Chem. 278, 8065–8074.

Gaullier, J.-M., Rønning, E., Gillooly, D.J., Stenmark, H., 2000. Interaction of the EEA1 FYVE Finger with Phosphatidylinositol 3-Phosphate and Early Endosomes: Role of Conserved Residues. J. Biol. Chem. 275, 24595–24600.

Gaullier, J.-M., Simonsen, A., D'Arrigo, A., Bremnes, B., Stenmark, H., Aasland, R., 1998. FYVE fingers bind PtdIns(3)P. Nature 394, 432–433.

Gelis, I., Bonvin, A.M.J.J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., Kalodimos, C.G., 2007. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. Cell 131, 756–769.

Ghosh, S., Strum, J.C., Sciorra, V.A., Daniel, L., Bell, R.M., 1996. Raf-1 kinase possesses distinct binding domains for phosphatidylserine and phosphatidic acid. Phosphatidic acid regulates the translocation of Raf-1 in 12-O-tetradecanoylphorbol-13-acetate-stimulated Madin-Darby canine kidney cells. J. Biol. Chem. 271, 8472-8480.

Giansanti, M.G., Bonaccorsi, S., Kurek, R., Farkas, R.M., Dimitri, P., Fuller, M.T., Gatti, M., 2006. The Class I PITP Giotto Is Required for Drosophila Cytokinesis. Curr. Biol. 16, 195–201.

Glomset, J.A., 1999. Lipids protein-lipid interactions on the surfaces of cell membranes. Curr. Opin. Struct. Biol. 9, 425–427.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D., 2001. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. Science 292, 1876–1882.

Gobeil, F., Bernier, S.G., Vazquez-Tello, A., Brault, S., Beauchamp, M.H., Quiniou, C., Marrache, A.M., Checchin, D., Sennlaub, F., Hou, X., Nader, M., Bkaily, G., Ribeiro-da-Silva, A., Goetzl, E.J., Chemtob, S., 2003. Modulation of Pro-inflammatory Gene Expression by Nuclear Lysophosphatidic Acid Receptor Type-1. J. Biol. Chem. 278, 38875–38883.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1, 2876–2890.

Greenfield, N.J., Fasman, G.D., 1969. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 8, 4108–4116.

Grzesiek, S., Bax, A., 1993. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J. Biomol. NMR 3, 185–204.

Grzesiek, S., Bax, A., 1992a. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. J. Magn. Reson. 96, 432–440.

Grzesiek, S., Bax, A., 1992b. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. J. Magn. Reson. 99, 201–207.

Grzesiek, S., Bax, A., 1992c. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J. Am. Chem. Soc. 114, 6291–6293.

Gulli, M.-P., Peter, M., 2001. Temporal and spatial regulation of Rho-type guanine-nucleotide exchange factors: the yeast perspective. Genes Dev. 15, 365–379.

Güntert, P., Mumenthaler, C., Wüthrich, K., 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. E. J. Mol. Biol. 273, 283–298.

Guo, J., Friedman, S.L., 2010. Toll-like receptor 4 signaling in liver injury and hepatic fibrogenesis. Fibrogenesis Tissue Repair 3, 21.

Han, G.-S., Wu, W.-I., Carman, G.M., 2006. The Saccharomyces cerevisiae Lipin Homolog Is a $Mg^{2+}$-dependent Phosphatidate Phosphatase Enzyme. J. Biol. Chem. 281, 9210–9218.

Hancock, J.F., 2006. Lipid rafts: contentious only from simplistic standpoints. Nat. Rev. Mol. Cell Biol. 7, 456–462.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., Yonath, A., 2001. High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. Cell 107, 679–688.

Harper, S., Speicher, D.W., 2001. Expression and Purification of GST Fusion Proteins. In: Current Protocols in Protein Science. John Wiley & Sons.

Harris, W.A., Stark, W.S., 1977. Hereditary retinal degeneration in Drosophila melanogaster. A mutant defect associated with the phototransduction process. J. Gen. Physiol. 69, 261–291.

Haugland, R.P., Spence, M.T.Z., 1996. Handbook of fluorescent probes and research chemicals. Molecular Probes.

Herrmann, T., Güntert, P., Wüthrich, K., 2002. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. J. Mol. Biol. 319, 209–227.

Hjelmeland, L.M., 1980. A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis. Proc. Natl. Acad. Sci. 77, 6368–6370.

Hoffmann, A., Bukau, B., Kramer, G., 2010. Structure and function of the molecular chaperone Trigger Factor. Biochim. Biophys. Acta - Mol. Cell Res. 1803, 650–661.

Hopkinson, R.J., Leung, I.K.H., Smart, T.J., Rose, N.R., Henry, L., Claridge, T.D.W., Schofield, C.J., 2015. Studies on the Glutathione-Dependent Formaldehyde-Activating Enzyme from Paracoccus denitrificans. PLoS One 10, e0145085.

Huang, R., Bonnichon, A., Claridge, T.D.W., Leung, I.K.H., 2017. Protein-ligand binding affinity determination by the waterLOGSY method: An optimised approach considering ligand rebinding. Sci. Rep. 7, 43727.

Huang, S., Rutkowsky, J.M., Snodgrass, R.G., Ono-Moore, K.D., Schneider, D.A., Newman, J.W., Adams, S.H., Hwang, D.H., 2012. Saturated fatty acids activate TLR-mediated proinflammatory signaling pathways. J. Lipid Res. 53, 2002–2013.

Huang, W.-C., Nagahashi, M., Terracina, K.P., Takabe, K., 2013. Emerging Role of Sphingosine-1-phosphate in Inflammation, Cancer, and Lymphangiogenesis. Biomolecules 3, 408–434.

Hurley, J.H., Tsujishita, Y., Pearson, M.A., 2000. Floundering about at cell membranes: a structural view of phospholipid signaling. Curr. Opin. Struct. Biol. 10, 737–743.

Hurst, R.J.M., Else, K.J., 2012. Retinoic acid signalling in gastrointestinal parasite infections: lessons from mouse models. Parasite Immunol. 34, 351–359.

Iberkleid, I., Sela, N., Brown Miyara, S., 2015. Meloidogyne javanica fatty acid- and retinol-binding protein (Mj-FAR-1) regulates expression of lipid-, cell wall-, stress- and phenylpropanoid-related genes during nematode infection of tomato. BMC Genomics 16, 272.

Iberkleid, I., Vieira, P., de Almeida Engler, J., Firester, K., Spiegel, Y., Horowitz, S.B., 2013. Fatty Acid-and Retinol-Binding Protein, Mj-FAR-1 Induces Tomato Host Susceptibility to Root-Knot Nematodes. PLoS One 8, e64586.

International Union of Biochemistry and Molecular Biology, 1992. Biochemical Nomenclature and Related Documents. Portland Press.

Irvine, R.F., 2003. 20 years of Ins(1,4,5)P3, and 40 years before. Nat. Rev. Mol. Cell Biol. 4, 580–585.

Itoh, F., Divecha, N., Brocks, L., Oomen, L., Janssen, H., Calafat, J., Itoh, S., Dijke, P. ten, 2002. The FYVE domain in Smad anchor for receptor activation (SARA) is sufficient for localization of SARA in early endosomes and regulates TGF-β/Smad signalling. Genes to Cells 7, 321–331.

Jäger, V., Büssow, K., Schirrmann, T., 2015. Transient Recombinant Protein Expression in Mammalian Cells. In: Animal Cell Culture. Springer International Publishing, pp. 27-64

Jarvis, D.L., 2009. Chapter 14 Baculovirus–Insect Cell Expression Systems. Methods Enzymol. 463, 191–222.

Jaskolski, M., Dauter, Z., Wlodawer, A., 2014. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. FEBS J. 281, 3985–4009.

Jasniewski, J., Cailliez-Grimal, C., Younsi, M., Millière, J.-B., Revol-Junelles, A.-M., 2008. Fluorescence anisotropy analysis of the mechanism of action of mesenterocin 52A: speculations on antimicrobial mechanism. Appl. Microbiol. Biotechnol. 81, 339–347.

Johnson, W.C., 1999. Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins Struct. Funct. Bioinforma. 35, 307–312.

Jones, J.T., Kumar, A., Pylypenko, L.A., Thirugnanasambandan, A., Castelli, L., Chapman, S., Cock, P.J.A., Greiner, E., Lilley, C.J., Phillips, M.S., Blok, V.C., 2009. Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. Mol. Plant Pathol. 10, 815–828.

Jones, S.M., Alb, J.G., Phillips, S.E., Bankaitis, V.A., Howell, K.E., 1998. A Phosphatidylinositol 3-Kinase and Phosphatidylinositol Transfer Protein Act Synergistically in Formation of Constitutive Transport Vesicles from the Trans-Golgi Network. J. Biol. Chem. 273, 10349–10354.

Jordanova, R., Groves, M.R., Kostova, E., Woltersdorf, C., Liebau, E., Tucker, P.A., 2009. Fatty Acid- and Retinoid-binding Proteins Have Distinct Binding Pockets for the Two Types of Cargo. J. Biol. Chem. 284, 35818–35826.

Jose Lopez-Andreo, M., Gomez-Fernandez, J.C., Corbalan-Garcia, S., 2003. The Simultaneous Production of Phosphatidic Acid and Diacylglycerol Is Essential for the Translocation of Protein Kinase Cε to the Plasma Membrane in RBL-2H3 Cells. Mol. Biol. Cell 14, 4885–4895.

Kabsch, W., 2010. XDS. Acta Crystallogr. Sect. D 66, 125–132.

Kapust, R.B., Waugh, D.S., 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8, 1668–1674.

Karathanassis, D., Stahelin, R. V, Bravo, J., Perisic, O., Pacold, C.M., Cho, W., Williams, R.L., 2002. Binding of the PX domain of p47(phox) to phosphatidylinositol 3,4-

bisphosphate and phosphatidic acid is masked by an intramolecular interaction. EMBO J. 21, 5057–5068.

Katan, M., 2005. New insights into the families of PLC enzymes: looking back and going forward. Biochem. J. 391, e7.

Kay, L.E., Ikura, M., Tschudin, R., Bax, A., 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J. Magn. Reson. 89, 496–514.

Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., Formankay, J.D., 1993. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations in $H_2O$ Samples of Proteins. J. Magn. Reson. Ser. B 101, 333–337.

Kazanietz, M.G., Barchi, J.J., Omichinski, J.G., Blumberg, P.M., 1995. Low Affinity Binding of Phorbol Esters to Protein Kinase C and Its Recombinant Cysteine-rich Region in the Absence of Phospholipids. J. Biol. Chem. 270, 14679–14684.

Keegan, R.M., Winn, M.D., 2008. MrBUMP: an automated pipeline for molecular replacement. Acta Crystallogr. Sect. D Biol. Crystallogr. 64, 119–124.

Keeler, J., 2010. Understanding NMR Spectroscopy, 2nd Edition. Wiley-Blackwell.

Keinan, O., Kedan, A., Gavert, N., Selitrennik, M., Kim, S., Karn, T., Becker, S., Lev, S., 2014. The lipid-transfer protein Nir2 enhances epithelial-mesenchymal transition and facilitates breast cancer metastasis. J. Cell Sci. 127, 4740–4749.

Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. Biochim. Biophys. Acta - Proteins Proteomics 1751, 119–139.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C., 1958. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature 181, 662–666.

Kennedy, M.W., 2011. The polyprotein allergens of nematodes (NPAs) – Structure at last, but still mysterious. Exp. Parasitol. 129, 81–84.

Kennedy, M.W., 2000. The polyprotein lipid binding proteins of nematodes. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1476, 149–164.

Kennedy, M.W., Brass, A., McCruden, A.B., Price, N.C., Kelly, S.M., Cooper, A., 1995. The ABA-1 Allergen of the Parasitic Nematode *Ascaris suum*: Fatty Acid and Retinoid Binding Function and Structural Characterization. Biochemistry 34, 6700–6710.

Kersten, S., Desvergne, B., Wahli, W., 2000. Roles of PPARs in health and disease. Nature 405, 421–424.

Khurana, S., Tomar, A., George, S.P., Wang, Y., Siddiqui, M.R., Guo, H., Tigyi, G., Mathew, S., 2008. Autotaxin and lysophosphatidic acid stimulate intestinal cell motility by redistribution of the actin modifying protein villin to the developing lamellipodia. Exp. Cell Res. 314, 530–542.

Kim, S., Kedan, A., Marom, M., Gavert, N., Keinan, O., Selitrennik, M., Laufman, O., Lev, S., 2013. The phosphatidylinositol-transfer protein Nir2 binds phosphatidic acid and positively regulates phosphoinositide signalling. EMBO Rep. 14, 891–899.

Kim, Y.J., Guzman-Hernandez, M.-L., Wisniewski, E., Balla, T., 2015. Phosphatidylinositol-Phosphatidic Acid Exchange by Nir2 at ER-PM Contact Sites Maintains Phosphoinositide Signaling Competence. Dev. Cell 33, 549–561.

Klages, J., Coles, M., Kessler, H., 2007. NMR-based screening: a powerful tool in fragment-based drug discovery. Analyst 132, 692–705.

Klinkenberg, D., Long, K.R., Shome, K., Watkins, S.C., Aridor, M., 2014. A cascade of ER exit site assembly that is regulated by p125A and lipid signals. J. Cell Sci. 127, 1765–1778.

Kohn, L., Kadzhaev, K., Burstedt, M.S.I., Haraldsson, S., Hallberg, B., Sandgren, O., Golovleva, I., 2007. Mutation in the PYK2-binding domain of PITPNM3 causes autosomal dominant cone dystrophy (CORD5) in two Swedish families. Eur. J. Hum. Genet. 15, 664–671.

Kooijman, E.E., Carter, K.M., van Laar, E.G., Chupin, V., Burger, K.N., de Kruijff, B., 2005a. What makes the bioactive lipids phosphatidic acid and lysophosphatidic acid so special? Biochemistry 44, 17007–17015.

Kooijman, E.E., Chupin, V., de Kruijff, B., Burger, K.N., 2003. Modulation of membrane curvature by phosphatidic acid and lysophosphatidic acid. Traffic 4, 162–174.

Kooijman, E.E., Chupin, V., Fuller, N.L., Kozlov, M.M., de Kruijff, B., Burger, K.N.J.,

Rand, P.R., 2005b. Spontaneous Curvature of Phosphatidic Acid and Lysophosphatidic Acid. Biochemistry 44, 2097–2102.

Kooijman, E.E., Tieleman, D.P., Testerink, C., Munnik, T., Rijkers, D.T., Burger, K.N., de Kruijff, B., 2007. An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. J. Biol. Chem. 282, 11356–11364.

Kovacs, H., Gossert, A., 2014. Improved NMR experiments with 13C-isotropic mixing for assignment of aromatic and aliphatic side chains in labeled proteins. J. Biomol. NMR 58, 101–112.

Kraft, C.A., Garrido, J.L., Fluharty, E., Leiva-Vega, L., Romero, G., 2008. Role of Phosphatidic Acid in the Coupling of the ERK Cascade. J. Biol. Chem. 283, 36636–36645.

Krohn, S., Garin, A., Gabay, C., Proudfoot, A., 2013. The Activity of CCL18 is Principally Mediated through Interaction with Glycosaminoglycans. Front. Immunol. 4, 193.

Kruijff, B. de, 1997. Lipid polymorphism and biomembrane function. Curr. Opin. Chem. Biol. 1, 564–569.

Kurooka, T., Yamamoto, Y., Takai, Y., Sakisaka, T., 2011. Dual Regulation of RA-RhoGAP Activity by Phosphatidic Acid and Rap1 during Neurite Outgrowth. J. Biol. Chem. 286, 6832–6843.

Kutateladze, T., Overduin, M., 2001. Structural Mechanism of Endosome Docking by the FYVE Domain. Science 291, 1793–1796.

Kutateladze, T.G., 2007. Mechanistic similarities in docking of the FYVE and PX domains to phosphatidylinositol 3-phosphate containing membranes. Prog. Lipid Res. 46, 315–327.

Kutateladze, T.G., Ogburn, K.D., Watson, W.T., de Beer, T., Emr, S.D., Burd, C.G., Overduin, M., 1999. Phosphatidylinositol 3-Phosphate Recognition by the FYVE Domain. Mol. Cell 3, 805–811.

Ladokhin, A.S., Isas, J.M., Haigler, H.T., White, S.H., 2002. Determining the Membrane Topology of Proteins: Insertion Pathway of a Transmembrane Helix of Annexin 12. Biochemistry 41, 13617–13626.

Lakowicz, J.R., 2006. Principles of Fluorescence Spectroscopy, 3rd ed. Springer US.

Lal, P.G., James, E.R., 1996. *Onchocerca* retinol- and ivermectin-binding protein activity. Parasitology 112, 221–225.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

LaVallie, E.R., Lu, Z., Diblasio-Smith, E.A., Collins-Racie, L.A., McCoy, J.M., 2000. Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. Methods Enzymol. 326, 322–340.

Lebendiker, M., Danieli, T., 2014. Production of prone-to-aggregate proteins. FEBS Lett. 588, 236–246.

Lee, S.A., Kovacs, J., Stahelin, R. V, Cheever, M.L., Overduin, M., Setty, T.G., Burd, C.G., Cho, W., Kutateladze, T.G., 2006. Molecular Mechanism of Membrane Docking by the Vam7p PX Domain. J. Biol. Chem. 281, 37091–37101.

Lemmon, M.A., 2008. Membrane recognition by phospholipid-binding domains. Nat. Rev. Mol. Cell Biol. 9, 99–111.

Lemmon, M.A., 2004. Pleckstrin homology domains: not just for phosphoinositides. Biochem. Soc. Trans. 32, 707–711.

Lemmon, M.A., 2003. Phosphoinositide Recognition Domains. Traffic 4, 201–213.

Lemmon, M.A., Ferguson, K.M., 2000. Signal-dependent membrane targeting by pleckstrin homology (PH) domains. Biochem. J. 350, 1–18.

Lemmon, M.A., Ferguson, K.M., O'Brien, R., Sigler, P.B., Schlessinger, J., 1995. Specific and high-affinity binding of inositol phosphates to an isolated pleckstrin homology domain. Proc. Natl. Acad. Sci. U. S. A. 92, 10472–10476.

Letunic, I., Doerks, T., Bork, P., 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 43, D257–D260.

Lev, S., 2004. The role of the Nir/rdgB protein family in membrane trafficking and cytoskeleton remodeling. Exp. Cell Res. 297, 1–10.

Lev, S., Hernandez, J., Martinez, R., Chen, A., Plowman, G., Schlessinger, J., 1999. Identification of a Novel Family of Targets of PYK2 Related to *Drosophila* Retinal Degeneration B (rdgB) Protein. Mol. Cell. Biol. 19, 2278–2288.

Levitt, M., 2008. Spin Dynamics: Basics of Nuclear Magnetic Resonance, 2nd Edition. Wiley-Blackwell.

Li, X., Romero, P., Rani, M., Dunker, A., Obradzovic, Z., 1999. Predicting protein disorder for N-, C-, and internal regions. Genome Informatics 10, 30–40.

Lin, S., Struve, W.S., 1991. Time-resolved fluorescence of nitrobenzoxadiazole-aminohexanoic acid: effect of intermolecular hydrogen-bonding on non-radiative decay. Photochem. Photobiol. 54, 361–365.

Linge, J.P., Habeck, M., Rieping, W., Nilges, M., 2003a. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19, 315–316.

Linge, J.P., O'Donoghue, S.I., Nilges, M., 2001. Automated Assignment of Ambiguous Nuclear Overhauser Effects with ARIA. Methods Enzymol. 339, 71–90.

Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J., Nilges, M., 2003b. Refinement of protein structures in explicit solvent. Proteins Struct. Funct. Bioinforma. 50, 496–506.

Lingwood, D., Simons, K., 2009. Lipid Rafts As a Membrane-Organizing Principle. Science 327, 46–50.

Litvak, V., Argov, R., Dahan, N., Ramachandran, S., Amarilio, R., Shainskaya, A., Lev, S., 2004. Mitotic Phosphorylation of the Peripheral Golgi Protein Nir2 by Cdk1 Provides a Docking Mechanism for Plk1 and Affects Cytokinesis Completion. Mol. Cell 14, 319–330.

Litvak, V., Dahan, N., Ramachandran, S., Sabanay, H., Lev, S., 2005. Maintenance of the diacylglycerol level in the Golgi apparatus by the Nir2 protein is critical for Golgi secretory function. Nat. Cell Biol. 7, 225–234.

Litvak, V., Tian, D., Carmon, S., Lev, S., 2002. Nir2, a Human Homolog of *Drosophila melanogaster* Retinal Degeneration B Protein, Is Essential for Cytokinesis. Mol. Cell. Biol. 22, 5064–5075.

Liu, L.X., Serhan, C.N., Weller, P.F., 1990. Intravascular filarial parasites elaborate cyclooxygenase-derived eicosanoids. J. Exp. Med. 172, 993–996.

Loewen, C.J., Roy, A., Levine, T.P., 2003. A conserved ER targeting motif in three families of lipid binding proteins and in Opi1p binds VAP. EMBO J. 22, 2025-2035.

Loewen, C.J.R., Levine, T.P., 2005. A Highly Conserved Binding Site in Vesicle-associated Membrane Protein-associated Protein (VAP) for the FFAT Motif of Lipid-binding Proteins. J. Biol. Chem. 280, 14097–14104.

Mainz, A., Religa, T.L., Sprangers, R., Linser, R., Kay, L.E., Reif, B., 2013. NMR Spectroscopy of Soluble Protein Complexes at One Mega-Dalton and Beyond. Angew. Chemie Int. Ed. 52, 8746–8751.

Majerus, P.W., York, J.D., 2009. Phosphoinositide phosphatases and disease. J. Lipid Res. 50, S249–S254.

Makrides, S.C., 1996. Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol. Rev. 60, 512–538.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., Butt, T.R., 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J. Struct. Funct. Genom. 5, 75-86.

Manavalan, P., Johnson, W.C., 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal. Biochem. 167, 76–85.

Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X., Butt, T.R., 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. Protein Sci. 15, 182–189.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., Clore, G.M., 1989a. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Biochemistry 28, 6150–6156.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., Bax, A., 1989b. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. J. Am. Chem. Soc. 111, 1515–1517.

Marsh, D., 1990. CRC handbook of lipid bilayers. CRC Press.

Martin, G.S., 2003. Cell signaling and cancer. Cancer Cell 4, 167–174.

Mashalidis, E.H., Śledź, P., Lang, S., Abell, C., 2013. A three-stage biophysical screening

cascade for fragment-based drug discovery. Nat. Protoc. 8, 2309–2324.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D., 2012. Recombinant Protein Production in Yeast. In: Recombinant Gene Expression. Humana Press, pp. 329-358.

Mátyus, L., Szöllősi, J., Jenei, A., 2006. Steady-state fluorescence quenching applications for studying protein structure and dynamics. J. Photochem. Photobiol. B Biol. 83, 223–236.

Mayer, M., Meyer, B., 1999. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angew. Chemie Int. Ed. 38, 1784–1788.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

McLaughlin, S., Murray, D., 2005. Plasma membrane phosphoinositide organization by protein electrostatics. Nature 438, 605–611.

Meenan, N.A.G., Ball, G., Bromek, K., Uhrín, D., Cooper, A., Kennedy, M.W., Smith, B.O., 2011. Solution Structure of a Repeated Unit of the ABA-1 Nematode Polyprotein Allergen of *Ascaris* Reveals a Novel Fold and Two Discrete Lipid-Binding Sites. PLoS Negl. Trop. Dis. 5, e1040.

Mei, B., Kennedy, M.W., Beauchamp, J., Komuniecki, P.R., Komuniecki, R., 1997. Secretion of a Novel, Developmentally Regulated Fatty Acid-binding Protein into the Perivitelline Fluid of the Parasitic Nematode, *Ascaris suum*. J. Biol. Chem. 272, 9933–9941.

Meiboom, S., Gill, D., 1958. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. Rev. Sci. Instrum. 29, 688–691.

Meyer, B., Peters, T., 2003. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors . Angew. Chemie Int. Ed. 42, 864–890.

Michalski, M.L., Monsey, J.D., Cistola, D.P., Weil, G.J., 2002. An embryo-associated fatty acid-binding protein in the filarial nematode *Brugia malayi*. Mol. Biochem. Parasitol. 124, 1–10.

Mietkiewska, E., Siloto, R.M.P., Dewald, J., Shah, S., Brindley, D.N., Weselake, R.J., 2011. Lipins from plants are phosphatidate phosphatases that restore lipid synthesis in

a pah1Δ mutant strain of *Saccharomyces cerevisiae*. FEBS J. 278, 764–775.

Misra, S., Hurley, J.H., 1999. Crystal Structure of a Phosphatidylinositol 3-Phosphate-Specific Membrane-Targeting Motif, the FYVE Domain of Vps27p. Cell 97, 657–666.

Mizukami, M., Miyauchi, H.H. and A., 2010. *Brevibacillus* Expression System: Host-Vector System for Efficient Production of Secretory Proteins. Curr. Pharm. Biotechnol. 11, 251-258.

Montelione, G.T., Lyons, B.A., Emerson, S.D., Tashiro, M., 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J. Am. Chem. Soc. 114, 10974–10975.

Mora, A., Komander, D., van Aalten, D.M.F., Alessi, D.R., 2004. PDK1, the master regulator of AGC kinase signal transduction. Semin. Cell Dev. Biol. 15, 161–170.

Mora, J.R., Iwata, M., von Andrian, U.H., 2008. Vitamin effects on the immune system: vitamins A and D take centre stage. Nat. Rev. Immunol. 8, 685–698. doi:10.1038/nri2378

Muhandiram, D.R., Kay, L.E., 1994. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. J. Magn. Reson. Ser. B 103, 203–216.

Mukherjee, S., Raghuraman, H., Dasgupta, S., Chattopadhyay, A., 2004. Organization and dynamics of N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-labeled lipids: a fluorescence approach. Chem. Phys. Lipids 127, 91–101.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr. Sect. D 53, 240–255.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. Sect. D 55, 247–255.

Nadaud, P.S., Helmus, J.J., Höfer, N., Jaroniec, C.P., 2007. Long-Range Structural Restraints in Spin-Labeled Proteins Probed by Solid-State Nuclear Magnetic Resonance Spectroscopy. J. Am. Chem. Soc. 129, 7502–7503.

Nakanishi, H., de los Santos, P., Neiman, A.M., 2004. Positive and negative regulation of a SNARE protein by control of intracellular localization. Mol. Biol. Cell 15, 1802-1815

Nettleship, J.E., 2016. OPPF-UK Standard Protocols: Insect Expression [WWW Document].

Nilges, M., 1995. Calculation of Protein Structures with Ambiguous Distance Restraints. Automated Assignment of Ambiguous NOE Crosspeaks and Disulphide Connectivities. J. Mol. Biol. 245, 645–660.

Nilges, M., Macias, M.J., O'Donoghue, S.I., Oschkinat, H., 1997. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. J. Mol. Biol. 269, 408–422.

Nishihara, K., Kanemori, M., Yanagi, H., Yura, T., 2000. Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia coli*. Appl. Environ. Microbiol. 66, 884–889

O'Connor, D. V, Phillips, D., 1984. Time-Correlated Single Photon Counting. Academic Press.

Obal, G., Ramos, A.L., Silva, V., Lima, A., Batthyany, C., Bessio, M.I., Ferreira, F., Salinas, G., Ferreira, A.M., 2012. Characterisation of the Native Lipid Moiety of Echinococcus granulosus Antigen B. PLoS Negl. Trop. Dis. 6, e1642.

Ooms, L.M., Horan, K.A., Rahman, P., Seaton, G., Gurung, R., Kethesparan, D.S., Mitchell, C.A., 2009. The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. Biochem. J. 419, 29–49.

Osanai, M., Nishikiori, N., Murata, M., Chiba, H., Kojima, T., Sawada, N., 2006. Cellular Retinoic Acid Bioavailability Determines Epithelial Integrity: Role of Retinoic Acid Receptor α Agonists in Colitis. Mol. Pharmacol. 71, 250–258.

Palmetshofer, A., Robson, S.C., Nehls, V., 1999. Lysophosphatidic Acid Activates Nuclear Factor Kappa B and Induces Proinflammatory Gene Expression in Endothelial Cells. Thromb. Haemost. 82, 1532–1537.

Papackova, Z., Cahova, M., 2015. Fatty Acid Signaling: The New Function of Intracellular Lipases. Int. J. Mol. Sci. 16, 3831–3855.

Park, J.B., Lee, C.S., Jang, J.-H., Ghim, J., Kim, Y.-J., You, S., Hwang, D., Suh, P.-G.,

Ryu, S.H., 2012. Phospholipase signalling networks in cancer. Nat. Rev. Cancer 12, 782–792.

Parker, J.L., Newstead, S., 2016. Membrane protein crystallization: Current trends and future perspectives. Adv. Exp. Med. Biol. 922, 61–72.

Patki, V., Lawe, D.C., Corvera, S., Virbasius, J. V, Chawla, A., 1998. A functional PtdIns(3)P-binding motif. Nature 394, 433–434.

Paulick, M.G., Bertozzi, C.R., 2008. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. Biochemistry 47, 6991–7000.

Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T.L., Homans, S.W., Kessler, H., Luchinat, C., Meyer, B., Oschkinat, H., Peng, J., Schwalbe, H., Siegal, G., 2008. Perspectives on NMR in drug discovery: a technique comes of age. Nat. Rev. Drug Discov. 7, 738–745.

Pendaries, C., Tronchère, H., Plantavid, M., Payrastre, B., 2003. Phosphoinositide signaling disorders in human diseases. FEBS Lett. 546, 25–31.

Perutz, M.F., Rossmann, M.G., Cullis, A.N.N.F., Muirhead, H., Will, G., North, A.C.T., 1960. Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. Nature 185, 416–422.

Pervushin, K., Riek, R., Wider, G., Wüthrich, K., 1997. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc. Natl. Acad. Sci. 94, 12366–12371.

Ponting, C.P., 1996. Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3-kinases: binding partners of SH3 domains? Protein Sci. 5, 2353–2357.

Popeijus, M., Blok, V.C., Cardle, L., Bakker, E., Phillips, M.S., Helder, J., Smant, G., Jones, J.T., 2000. Analysis of genes expressed in second stage juveniles of the potato cyst nematodes *Globodera rostochiensis* and *G. pallida* using the expressed sequence tag approach. Nematology 2, 567–574.

Pritz, S., Doering, K., Woelcke, J., Hassiepen, U., 2011. Fluorescence lifetime assays: current advances and applications in drug discovery. Expert Opin. Drug Discov. 6, 663–670.

Provencher, S.W., Gloeckner, J., 1981. Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20, 33–37.

Pullan, R.L., Smith, J.L., Jasrasaria, R., Brooker, S.J., 2014. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit. Vectors 7, 37.

Purcell, E.M., Torrey, H.C., Pound, R. V, 1946. Resonance Absorption by Nuclear Magnetic Moments in a Solid. Phys. Rev. 69, 37–38.

Pyne, N.J., McNaughton, M., Boomkamp, S., MacRitchie, N., Evangelisti, C., Martelli, A.M., Jiang, H.-R., Ubhi, S., Pyne, S., 2016. Role of sphingosine 1-phosphate receptors, sphingosine kinases and sphingosine in cancer and inflammation. Adv. Biol. Regul. 60, 151–159.

Raiborg, C., Schink, K.O., Stenmark, H., 2013. Class III phosphatidylinositol 3–kinase and its catalytic product PtdIns3P in regulation of endocytic membrane traffic. FEBS J. 280, 2730–2742.

Raman, M., Martin, K., 2014. One solution for cloning and mutagenesis: In-Fusion[reg] HD Cloning Plus. Nat. Meth. 11.

Ranjit, N., Jones, M.K., Stenzel, D.J., Gasser, R.B., Loukas, A., 2006. A survey of the intestinal transcriptomes of the hookworms, *Necator americanus* and *Ancylostoma caninum*, using tissues isolated by laser microdissection microscopy. Int. J. Parasitol. 36, 701–710.

Rebecchi, M.J., Pentyala, S.N., 2000. Structure, Function, and Control of Phosphoinositide-Specific Phospholipase C. Physiol. Rev. 80, 1291–1335.

Renaud, J.-P., Delsuc, M.-A., 2009. Biophysical techniques for ligand screening and drug design. Curr. Opin. Pharmacol. 9, 622–628.

Resh, M.D., 2013. Covalent lipid modifications of proteins. Curr. Biol. 23, R431–R435.

Reue, K., Dwyer, J.R., 2009. Lipin proteins and metabolic homeostasis. J. Lipid Res. 50, S109–S114.

Rey-Burusco, M.F., Ibañez-Shimabukuro, M., Cooper, A., Kennedy, M.W., Córsico, B., Smith, B.O., 2014. (1)H, (13)C and (15)N chemical shift assignments of Na-FAR-1, a helix-rich fatty acid and retinol binding protein of the parasitic nematode *Necator*

*americanus*. Biomol. NMR Assign. 8, 19–21.

Rey-Burusco, M.F., Ibáñez-Shimabukuro, M., Gabrielsen, M., Franchini, G.R., Roe, A.J., Griffiths, K., Zhan, B., Cooper, A., Kennedy, M.W., Córsico, B., Smith, B.O., 2015. Diversity in the structures and ligand-binding sites of nematode fatty acid and retinol-binding proteins revealed by Na-FAR-1 from *Necator americanus*. Biochem. J. 471, 403–414.

Ricciotti, E., FitzGerald, G.A., 2011. Prostaglandins and Inflammation. Arterioscler. Thromb. Vasc. Biol. 31, 986–1000.

Rivera, R., Chun, J., 2008. Biological effects of lysophospholipids. In: Reviews of Physiology Biochemistry and Pharmacology. Springer Berlin Heidelberg, pp. 25-46.

Rizzo, M.A., Shome, K., Watkins, S.C., Romero, G., 2000. The recruitment of Raf-1 to membranes is mediated by direct interaction with phosphatidic acid and is independent of association with Ras. J. Biol. Chem. 275, 23911–23918.

Rojas, J.M., Oliva, J.L., Santos, E., 2011. Mammalian Son of Sevenless Guanine Nucleotide Exchange Factors: Old Concepts and New Perspectives. Genes Cancer 2, 298–305.

Rossmann, M.G., 2001. Molecular replacement historical background. Acta Crystallogr. Sect. D 57, 1360–1366.

Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr. 15, 24–31.

Roszak, A.W., Gardiner, A.T., Isaacs, N.W., Cogdell, R.J., 2007. Brominated Lipids Identify Lipid Binding Sites on the Surface of the Reaction Center from *Rhodobacter sphaeroides*. Biochemistry 46, 2909–2916.

Rother, E., Brandl, R., Baker, D.L., Goyal, P., Gebhard, H., Tigyi, G., Siess, W., 2003. Subtype-Selective Antagonists of Lysophosphatidic Acid Receptors Inhibit Platelet Activation Triggered by the Lipid Core of Atherosclerotic Plaques. Circulation 108, 741–747.

Rubenfeld, J., Guo, J., Sookrung, N., Chen, R., Chaicumpa, W., Casolaro, V., Zhao, Y., Natarajan, V., Georas, S., 2005. Lysophosphatidic acid enhances interleukin-13 gene expression and promoter activity in T cells. Am. J. Physiol. - Lung Cell. Mol. Physiol. 290, L66–L74.

Sacchettini, J.C., Gordon, J.I., Banaszak, L.J., 1989. Crystal structure of rat intestinal fatty-acid-binding protein. J. Mol. Biol. 208, 327–339.

Saio, T., Guan, X., Rossi, P., Economou, A., Kalodimos, C.G., 2014. Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. Science 344, 1250494.

Sani, B.P., Vaid, A., 1988. Specific interaction of ivermectin with retinol-binding protein from filarial parasites. Biochem. J. 249, 929–932.

Sasaki, T., Takasuga, S., Sasaki, J., Kofuji, S., Eguchi, S., Yamazaki, M., Suzuki, A., 2009. Mammalian phosphoinositide kinases and phosphatases. Prog. Lipid Res. 48, 307–343.

Schanda, P., Brutscher, B., 2005. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. J. Am. Chem. Soc. 127, 8014–8015.

Scheffzek, K., Welti, S., 2012. Pleckstrin homology (PH) like domains – versatile modules in protein–protein interaction platforms. FEBS Lett. 586, 2662–2673.

Schenning, M., Goedhart, J., Gadella Jr., T.W.J., Avram, D., Wirtz, K.W.A., Snoek, G.T., 2008. The anti-apoptotic activity associated with phosphatidylinositol transfer protein α activates the MAPK and Akt/PKB pathway. Biochim. Biophys. Acta - Mol. Cell Res. 1783, 1700–1706.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Meth. 9, 676–682.

Schouten, A., Agianian, B., Westerman, J., Kroon, J., Wirtz, K.W.A., Gros, P., 2002. Structure of apo-phosphatidylinositol transfer protein α provides insight into membrane association. EMBO J. 21, 2117–2121.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. 95, 5857–5864.

Sever, R., Brugge, J.S., 2015. Signal Transduction in Cancer. Cold Spring Harb. Perspect. Med. 5, a006098.

Sheldrick, G.M., 2008. A short history of SHELX. Acta Crystallogr. Sect. A 64, 112–122.

Shimizu, H., Nihei, C., Inaoka, D.K., Mogi, T., Kita, K., Harada, S., 2008. Screening of detergents for solubilization, purification and crystallization of membrane proteins: a case study on succinate:ubiquinone oxidoreductase from *Escherichia coli*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 64, 858–862.

Shin, J.J.H., Loewen, C.J.R., 2011. Putting the pH into phosphatidic acid signaling. BMC Biol. 9, 85.

Siess, W., 2002. Athero- and thrombogenic actions of lysophosphatidic acid and sphingosine-1-phosphate. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1582, 204–215.

Simon, J.-P., Morimoto, T., Bankaitis, V.A., Gottlieb, T.A., Ivanov, I.E., Adesnik, M., Sabatini, D.D., 1998. An essential role for the phosphatidylinositol transfer protein in the scission of coatomer-coated vesicles from the trans-Golgi network. Proc. Natl. Acad. Sci. U. S. A. 95, 11181–11186.

Simons, K., Sampaio, J.L., 2011. Membrane Organization and Lipid Rafts. Cold Spring Harb. Perspect. Biol. 3, a004697.

Śledź, P., Abell, C., Ciulli, A., 2012. Ligand-Observed NMR in Fragment-Based Approaches. In: NMR of Biomolecules. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 264-280.

Smathers, R.L., Petersen, D.R., 2011. The human fatty acid-binding protein family: Evolutionary divergences and functions. Hum. Genomics 5, 170–191.

Smyth, J.D., Wakelin, D., 1994. Introduction to animal parasitology. Cambridge University Press.

Snoek, G., 2004. Phosphatidylinositol Transfer Proteins: Emerging Roles in Cell Proliferation, Cell Death and Survival. IUBMB Life 56, 467–475.

Sorof, S., 1994. Modulation of mitogenesis by liver fatty acid binding protein. Cancer Metastasis Rev. 13, 317–336.

Spiegel, S., Milstien, S., 2011. The outs and the ins of sphingosine-1-phosphate in immunity. Nat. Rev. Immunol. 11, 403–415.

Spiegel, S., Milstien, S., 2003. Sphingosine-1-phosphate: an enigmatic signalling lipid.

Nat. Rev. Mol. Cell Biol. 4, 397–407.

Sprangers, R., Kay, L.E., 2007. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445, 618–622.

Sreerama, N., Venyaminov, S.Y., Woody, R.W., 2000. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Inclusion of Denatured Proteins with Native Proteins in the Analysis. Anal. Biochem. 287, 243–251.

Sreerama, N., Woody, R.W., 2000. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. Anal. Biochem. 287, 252–260.

Sreerama, N., Woody, R.W., 1993. A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. Anal. Biochem. 209, 32–44.

Stace, C.L., Ktistakis, N.T., 2006. Phosphatidic acid- and phosphatidylserine-binding proteins. Biochim. Biophys. Acta 1761, 913–826.

Stephenson, L.S., Latham, M.C., Ottensen, E.A., 2000. Malnutrition and parasitic helminth infections. Parasitology 121, S23–S38.

Stracke, M.L., Krutzsch, H.C., Unsworth, E.J., Arestad, A., Cioce, V., Schiffmann, E., Liotta, L.A., 1992. Identification, purification, and partial sequence analysis of autotaxin, a novel motility-stimulating protein. J. Biol. Chem. 267, 2524–2529.

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Subramaniam, S., 2007. LMSD: LIPID MAPS structure database. Nucleic Acids Res. 35, D527–D532.

Sugiki, T., Takeuchi, K., Yamaji, T., Takano, T., Tokunaga, Y., Kumagai, K., Hanada, K., Takahashi, H., Shimada, I., 2012. Structural Basis for the Golgi Association by the Pleckstrin Homology Domain of the Ceramide Trafficking Protein (CERT). J. Biol. Chem. 287, 33706–33718.

Takahashi, H., Matuoka, S., Kato, S., Ohki, K., Hatta, I., 1991. Electrostatic interaction of poly(l-lysine) with dipalmitoylphosphatidic acid studied by X-ray diffraction. Biochim. Biophys. Acta - Biomembr. 1069, 229–234.

Tanford, C., 1980. The Hydrophobic effect : formation of micelles and biological membranes. John Wiley & Sons.

Tang, Y.T., Gao, X., Rosa, B.A., Abubucker, S., Hallsworth-Pepin, K., Martin, J., Tyagi, R., Heizer, E., Zhang, X., Bhonagiri-Palsikar, V., Minx, P., Warren, W.C., Wang, Q., Zhan, B., Hotez, P.J., Sternberg, P.W., Dougall, A., Gaze, S.T., Mulvenna, J., Sotillo, J., Ranganathan, S., Rabelo, E.M., Wilson, R.K., Felgner, P.L., Bethony, J., Hawdon, J.M., Gasser, R.B., Loukas, A., Mitreva, M., 2014. Genome of the human hookworm Necator americanus. Nat. Genet. 46, 261–269.

Taylor, G.L., 2010. Introduction to phasing. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 325–338.

Thompson, J., Winter, N., Terwey, D., Bratt, J., Banaszak, L., 1997. The Crystal Structure of the Liver Fatty Acid-binding Protein: A complex with two bound oleates. J. Biol. Chem. 272, 7140–7150.

Tian, D., Litvak, V., Toledo-Rodriguez, M., Carmon, S., Lev, S., 2002. Nir2, a Novel Regulator of Cell Morphogenesis. Mol. Cell. Biol. 22, 2650–2662.

Tigyi, G., Hong, L., Yakubu, M., Parfenova, H., Shibata, M., Leffler, C.W., 1995. Lysophosphatidic acid alters cerebrovascular reactivity in piglets. Am. J. Physiol. - Hear. Circ. Physiol. 268, H2048-H2055.

Tilley, S.J., Skippen, A., Murray-Rust, J., Swigart, P.M., Stewart, A., Morgan, C.P., Cockcroft, S., McDonald, N.Q., 2007. Structure-Function Analysis of Phosphatidylinositol Transfer Protein Alpha Bound to Human Phosphatidylinositol. Structure 12, 317–326.

Tomlinson, L.A., Christie, J.F., Fraser, E.M., McLaughlin, D., McIntosh, A.E., Kennedy, M.W., 1989. MHC restriction of the antibody repertoire to secretory antigens, and a major allergen, of the nematode parasite Ascaris. J. Immunol. 143, 2349–2356.

Trivedi, D., Padinjat, R., 2007. RdgB proteins: Functions in lipid homeostasis and signal transduction. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1771, 692–699.

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Kent Wenger, R., Yao, H., Markley, J.L., 2008. BioMagResBank. Nucleic Acids Res. 36, D402–D408.

Unione, L., Galante, S., Diaz, D., Canada, F.J., Jimenez-Barbero, J., 2014. NMR and molecular recognition. The application of ligand-based NMR methods to monitor

molecular interactions. Medchemcomm 5, 1280–1289.

Vance, J.E., 2015. Phospholipid Synthesis and Transport in Mammalian Cells. Traffic 16, 1–18.

Vanhaesebroeck, B., Ali, K., Bilancio, A., Geering, B., Foukas, L.C., 2005. Signalling by PI3K isoforms: insights from gene-targeted mice. Trends Biochem. Sci. 30, 194–204.

Várnai, P., Rother, K.I., Balla, T., 1999. Phosphatidylinositol 3-Kinase-dependent Membrane Association of the Bruton's Tyrosine Kinase Pleckstrin Homology Domain Visualized in Single Living Cells. J. Biol. Chem. 274, 10983–10989.

Veerkamp, J.H., Zimmerman, A.W., 2001. Fatty acid-binding proteins of nervous tissue. J. Mol. Neurosci. 16, 133–142.

Vihtelic, T.S., Goebl, M., Milligan, S., O, Tousa, J.E., Hyde, D.R., 1993. Localization of Drosophila retinal degeneration B, a membrane-associated phosphatidylinositol transfer protein. J. Cell Biol. 122, 1013–1022.

Vordtriede, P.B., Doan, C.N., Tremblay, J.M., Helmkamp George M., Yoder, M.D., 2005. Structure of PITPβ in Complex with Phosphatidylcholine: Comparison of Structure and Lipid Transfer to Other PITP Isoforms. Biochemistry 44, 14760–14771.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct. Funct. Bioinforma. 59, 687–696.

Wallace, A.C., Laskowski, R.A., Thornton, J.M., 1995. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 8, 127–134.

Wang, A.C., Lodi, P.J., Qin, J., Vuister, G.W., Gronenborn, A.M., Clore, G.M., 1994. An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins. J. Magn. Reson. Ser. B 105, 196–198.

Wang, X., Devaiah, S.P., Zhang, W., Welti, R., 2006. Signaling functions of phosphatidic acid. Prog. Lipid Res. 45, 250–278.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., 2009. Jalview Version 2 - A multiple sequence alignment editor and analysis workbench. Bioinformatics 25.

Watton, S.J., Downward, J., 1999. Akt/PKB localisation and phosphoinositide generation

at sites of epithelial cell-matrix and cell-cell interaction. Curr. Biol. 9, 433–436.

Whitmore, L., Wallace, B.A., 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 32, W668–W673.

Williams, R.L., 1999. Mammalian phosphoinositide-specific phospholipase C. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1441, 255–267.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V., 2000. Structure of the 30S ribosomal subunit. Nature 407, 327–339.

Wimley, W.C., White, S.H., 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Mol. Biol. 3, 842–848.

Winter, G., 2010. xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. 43, 186–190.

Wolfrum, C., Borrmann, C.M., Börchers, T., Spener, F., 2001. Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors α- and γ-mediated gene expression via liver fatty acid binding protein: A signaling path to the nucleus. Proc. Natl. Acad. Sci. 98, 2323–2328.

Wymann, M.P., Pirola, L., 1998. Structure and function of phosphoinositide 3-kinases. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1436, 127–150.

Wymann, M.P., Schneiter, R., 2008. Lipid signalling in disease. Nat. Rev. Mol. Cell Biol. 9, 162–176.

Xie, Y., Ding, Y.-Q., Hong, Y., Feng, Z., Navarre, S., Xi, C.-X., Zhu, X.-J., Wang, C.-L., Ackerman, S.L., Kozlowski, D., Mei, L., Xiong, W.-C., 2005. Phosphatidylinositol transfer protein-α in netrin-1-induced PLC signalling and neurite outgrowth. Nat. Cell Biol. 7, 1124–1132.

Yadav, S., Garner, K., Georgiev, P., Li, M., Gomez-Espinosa, E., Panda, A., Mathre, S., Okkenhaug, H., Cockcroft, S., Raghu, P., 2015. RDGBα, a PtdIns-PtdOH transfer protein, regulates G-protein-coupled PtdIns(4,5)P(2) signalling during Drosophila phototransduction. J. Cell Sci. 128, 3330–3344.

Yamazaki, T., Forman-Kay, J.D., Kay, L.E., 1993. Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. J. Am. Chem. Soc. 115, 11054–11055.

Yoder, M.D., Thomas, L.M., Tremblay, J.M., Oliver, R.L., Yarbrough, L.R., Helmkamp, G.M., 2001. Structure of a Multifunctional Protein: Mammalian Phosphatidylinositol Transfer Protein Complexed With Phosphatidylcholine. J. Biol. Chem. 276, 9246–9252.

Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P., Fesik, S.W., 1994. Solution structure of a pleckstrin-homology domain. Nature 369, 672–675.

Young, B.P., Shin, J.J., Orij, R., Chao, J.T., Li, S.C., Guan, X.L., Khong, A., Jan, E., Wenk, M.R., Prinz, W.A., Smits, G.J., Loewen, C.J., 2010. Phosphatidic acid is a pH biosensor that links membrane biogenesis to metabolism. Science 329, 1085–1088.

Yun, C.C., Kumar, A., 2015. Diverse roles of LPA signaling in the intestinal epithelium. Exp. Cell Res. 333, 201–207.

Yung, Y.C., Stoddard, N.C., Chun, J., 2014. LPA receptor signaling: pharmacology, physiology, and pathophysiology. J. Lipid Res. 55, 1192–1214.

Zhang, B., Yin, C., Li, H., Shi, L., Liu, N., Sun, Y., Lu, S., Liu, Y., Sun, L., Li, X., Chen, W., Qi, Y., 2013. Nir1 promotes invasion of breast cancer cells by binding to chemokine (C–C motif) ligand 18 through the PI3K/Akt/GSK3β/Snail signalling pathway. Eur. J. Cancer 49, 3900–3913.

Zhang, G., Kazanietz, M.G., Blumberg, P.M., Hurley, J.H., 1995. Crystal structure of the Cys2 activator-binding domain of protein kinase Cδ in complex with phorbol ester. Cell 81, 917–924.

Zhang, H., Neal, S., Wishart, D.S., 2003. RefDB: A database of uniformly referenced protein chemical shifts. J. Biomol. NMR 25, 173–195.

Zhang, H., Wu, Q., Berezin, M.Y., 2015. Fluorescence anisotropy (polarization): from drug screening to precision medicine. Expert Opin. Drug Discov. 10, 1145–1161.

Zhao, C., Du, G., Skowronek, K., Frohman, M.A., Bar-Sagi, D., 2007. Phospholipase D2-generated phosphatidic acid couples EGFR stimulation to Ras activation by Sos. Nat. Cell. Biol. 9, 706–712.

Zhao, H., Lappalainen, P., 2012. A simple guide to biochemical approaches for analyzing protein–lipid interactions. Mol. Biol. Cell 23, 2823–2830.

Zhao, J., Wei, J., Weathington, N., Jacko, A.M., Huang, H., Tsung, A., Zhao, Y., 2015. Lysophosphatidic acid receptor 1 antagonist ki16425 blunts abdominal and systemic inflammation in a mouse model of peritoneal sepsis. Transl. Res. 166, 80–88.

Zhao, Y., Natarajan, V., 2013. Lysophosphatidic acid (LPA) and its receptors: Role in airway inflammation and remodeling. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1831, 86–92.

Zhao, Y., Tong, J., He, D., Pendyala, S., Evgeny, B., Chun, J., Sperling, A.I., Natarajan, V., 2009. Role of lysophosphatidic acid receptor LPA2 in the development of allergic airway inflammation in a murine model of asthma. Respir. Res. 10, 114.

Zheng, J., Chen, R.-H., Corblan-Garcia, S., Cahill, S.M., Bar-Sagi, D., Cowburn, D., 1997. The Solution Structure of the Pleckstrin Homology Domain of Human Sos1: A Possible Structural Role for the Sequential Association of Diffuse B Cell Lymphoma and Pleckstrin Homology Domains. J. Biol. Chem. 272, 30340–30344.

Zheng, Y., Voice, J.K., Kong, Y., Goetzl, E.J., 2000. Altered expression and functional profile of lysophosphatidic acid receptors in mitogen-activated human blood T lymphocytes. FASEB J. 14, 2387–2389.

Zimmerman, A.W., Veerkamp, J.H., 2002. New insights into the structure and function of fatty acid-binding proteins. Cell. Mol. Life Sci. C. 59, 1096–1116.

Zoete, V., Grosdidier, A., Michielin, O., 2009. Docking, virtual high throughput screening and in silico fragment-based drug design. J. Cell. Mol. Med. 13, 238–248.

Zuiderweg, E.R.P., Fesik, S.W., 1989. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. Biochemistry 28, 2387–2391.

# Appendix A: Primers used in the expression screening of the LNS2 domains of Nir2, Nir3 and RdgB

## Nir2

| Name | Plasmid | Sequence (5'-3') |
|---|---|---|
| LNS2-S F | pOPINE-3C-HALO7 | AGGAGATATACCATGGTCTATCCGGTG CGTATGGTCGTGC |
| LNS2-S R | pOPINE-3C-HALO7 | CAGAACTTCCAGTTTTTCCAGCTGACCC AGGTGTGCCAC |
| LNS2-L F | pOPINE-3C-HALO7 | AGGAGATATACCATGGTTAAAATCCGT AACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINE-3C-HALO7 | CAGAACTTCCAGTTTTTCTTCGCTGTCC AGTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINEneo | AGGAGATATACCATGGTCTATCCGGTG CGTATGGTCGTGC |
| LNS2-S R | pOPINEneo | GTGATGGTGATGTTTTTCCAGCTGACCC AGGTGTGCCAC |
| LNS2-L F | pOPINEneo | AGGAGATATACCATGGTTAAAATCCGTA ACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINEneo | GTGATGGTGATGTTTTTCTTCGCTGTCCA GTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINF | AAGTTCTGTTTCAGGGCCCGGTCTATCCG GTGCGTATGGTCGTGC |
| LNS2-S R | pOPINF | ATGGTCTAGAAAGCTTTATTCCAGCTGAC CCAGGTGTGCCAC |
| LNS2-L F | pOPINF | AAGTTCTGTTTCAGGGCCCGGTTAAAATC CGTAACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINF | ATGGTCTAGAAAGCTTTATTCTTCGCTGTC CAGTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINHALO7 | AAGTTCTGTTTCAGGGCCCGGTCTATCCGG TGCGTATGGTCGTGC |
| LNS2-S R | pOPINHALO7 | ATGGTCTAGAAAGCTTTATTCCAGCTGACC CAGGTGTGCCAC |
| LNS2-L F | pOPINHALO7 | AAGTTCTGTTTCAGGGCCCGGTCTATCCGG TGCGTATGGTCGTGC |
| LNS2-L R | pOPINHALO7 | ATGGTCTAGAAAGCTTTATTCCAGCTGACC CAGGTGTGCCAC |
| LNS2-S F | pOPINJ | AAGTTCTGTTTCAGGGCCCGGTCTATCCGG TGCGTATGGTCGTGC |
| LNS2-S R | pOPINJ | ATGGTCTAGAAAGCTTTATTCCAGCTGACC CAGGTGTGCCAC |
| LNS2-L F | pOPINJ | AAGTTCTGTTTCAGGGCCCGGTTAAAATCC GTAACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINJ | ATGGTCTAGAAAGCTTTATTCTTCGCTGTC CAGTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINM | AAGTTCTGTTTCAGGGCCCGGTCTATCCGG TGCGTATGGTCGTGC |
| LNS2-S R | pOPINM | ATGGTCTAGAAAGCTTTATTCCAGCTGACC CAGGTGTGCCAC |
| LNS2-L F | pOPINM | AAGTTCTGTTTCAGGGCCCGGTTAAAATCC GTAACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINM | ATGGTCTAGAAAGCTTTATTCTTCGCTGTCC AGTTTCAGAGAAATGCTGC |

| | | |
|---|---|---|
| LNS2-S F | pOPINNUSA | AAGTTCTGTTTCAGGGCCCGGTCTATCCGG TGCGTATGGTCGTGC |
| LNS2-S R | pOPINNUSA | ATGGTCTAGAAAGCTTTATTCCAGCTGACC CAGGTGTGCCAC |
| LNS2-L F | pOPINNUSA | AAGTTCTGTTTCAGGGCCCGGTTAAAATCC GTAACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINNUSA | ATGGTCTAGAAAGCTTTATTCTTCGCTGTCC AGTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINS3C | AAGTTCTGTTTCAGGGCCCGGTCTATCCGGT GCGTATGGTCGTGC |
| LNS2-S R | pOPINS3C | ATGGTCTAGAAAGCTTTATTCCAGCTGACCC AGGTGTGCCAC |
| LNS2-L F | pOPINS3C | AAGTTCTGTTTCAGGGCCCGGTTAAAATCCG TAACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINS3C | ATGGTCTAGAAAGCTTTATTCTTCGCTGTCCA GTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINTF | AAGTTCTGTTTCAGGGCCCGGTCTATCCGGTG CGTATGGTCGTGC |
| LNS2-S R | pOPINTF | ATGGTCTAGAAAGCTTTATTCCAGCTGACCCA GGTGTGCCAC |
| LNS2-L F | pOPINTF | AAGTTCTGTTTCAGGGCCCGGTTAAAATCCGT AACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINTF | ATGGTCTAGAAAGCTTTATTCTTCGCTGTCCA GTTTCAGAGAAATGCTGC |
| LNS2-S F | pOPINTRX | AAGTTCTGTTTCAGGGCCCGGTCTATCCGGTG CGTATGGTCGTGC |
| LNS2-S R | pOPINTRX | ATGGTCTAGAAAGCTTTATTCCAGCTGACCCA GGTGTGCCAC |
| LNS2-L F | pOPINTRX | AAGTTCTGTTTCAGGGCCCGGTTAAAATCCGT AACGTCACCAGCAATCATCGCG |
| LNS2-L R | pOPINTRX | ATGGTCTAGAAAGCTTTATTCTTCGCTGTCCA GTTTCAGAGAAATGCTGC |

## Nir3

| Name | Plasmid | Sequence |
|---|---|---|
| LNS2-S F | pOPINEneo | AGGAGATATACCATGACCCTGGTCACGAAC AATAGCGGC |
| LNS2-S R | pOPINEneo | GTGATGGTGATGTTTCGCACGGTGACTGTAT TTCAGCTGC |
| LNS2-S F | pOPINF | AAGTTCTGTTTCAGGGCCCGACCCTGGTCAC GAACAATAGCGGC |
| LNS2-S R | pOPINF | ATGGTCTAGAAAGCTTTACGCACGGTGACTG TATTTCAGCTGC |
| LNS2-S F | pOPINHALO7 | AGGAGATATACCATGGTCTATCCGGTG CGTATGGTCGTGC |
| LNS2-S R | pOPINHALO7 | GTGATGGTGATGTTTTTCCAGCTGACCC AGGTGTGCCAC |
| LNS2-S F | pOPINJ | AAGTTCTGTTTCAGGGCCCGACCCTGGT CACGAACAATAGCGGC |
| LNS2-F R | pOPINJ | ATGGTCTAGAAAGCTTTACGCACGGTGA CTGTATTTCAGCTGC |
| LNS2-S F | pOPINM | AAGTTCTGTTTCAGGGCCCGACCCTGGTCAC GAACAATAGCGGC |
| LNS2-S R | pOPINM | ATGGTCTAGAAAGCTTTACGCACGGTGACTG TATTTCAGCTGC |
| LNS2-L F | pOPINS3C | AAGTTCTGTTTCAGGGCCCGACCCTGGTCAC GAACAATAGCGGC |
| LNS2-L R | pOPINS3C | ATGGTCTAGAAAGCTTTACGCACGGTGACTG TATTTCAGCTGC |
| LNS2-S F | pOPINTRX | AAGTTCTGTTTCAGGGCCCGACCCTGGTCAC GAACAATAGCGGC |
| LNS2-S R | pOPINTRX | ATGGTCTAGAAAGCTTTACGCACGGTGACTG TATTTCAGCTGC |

## RdgB

| Name | Plasmid | Sequence |
|---|---|---|
| LNS2-S F | pOPINE-3C-HALO7 | AGGAGATATACCATGGACTGCTACATG GCCGTTGTCCCGC |
| LNS2-L F | pOPINE-3C-HALO7 | AGGAGATATACCATGGGTCAATGGACG TTCCTGAGTACCG |
| LNS2-L R | pOPINE-3C-HALO7 | CAGAACTTCCAGTTTATTTTCGTTCGTG GCCTGTTCGTGC |
| LNS2-S F | pOPINEneo | AGGAGATATACCATGGACTGCTACATG GCCGTTGTCCCGC |
| LNS2-S R | pOPINEneo | GTGATGGTGATGTTTACCATCGCTCAG CACCGTCGC |
| LNS2-L F | pOPINEneo | AGGAGATATACCATGGGTCAATGGAC GTTCCTGAGTACCG |
| LNS2-L R | pOPINEneo | GTGATGGTGATGTTTATTTTCGTTCGT GGCCTGTTCGTGC |
| LNS2-S F | pOPINF | AAGTTCTGTTTCAGGGCCCGGACTGC TACATGGCCGTTGTCCCGC |
| LNS2-S R | pOPINF | ATGGTCTAGAAAGCTTTAACCATCGC TCAGCACCGTCGC |
| LNS2-L F | pOPINF | AAGTTCTGTTTCAGGGCCCGGGTCAA TGGACGTTCCTGAGTACCG |
| LNS2-L R | pOPINF | ATGGTCTAGAAAGCTTTAATTTTCGTT CGTGGCCTGTTCGTGC |
| LNS2-S F | pOPINHALO7 | AAGTTCTGTTTCAGGGCCCGGACTGCT ACATGGCCGTTGTCCCGC |
| LNS2-S R | pOPINHALO7 | ATGGTCTAGAAAGCTTTAACCATCGCT CAGCACCGTCGC |
| LNS2-L F | pOPINHALO7 | AAGTTCTGTTTCAGGGCCCGGGTCAAT GGACGTTCCTGAGTACCG |
| LNS2-L R | pOPINHALO7 | ATGGTCTAGAAAGCTTTAATTTTCGTTC GTGGCCTGTTCGTGC |
| LNS2-S F | pOPINJ | AAGTTCTGTTTCAGGGCCCGGGTCAATG GACGTTCCTGAGTACCG |
| LNS2-S R | pOPINJ | ATGGTCTAGAAAGCTTTAATTTTCGTTCG TGGCCTGTTCGTGC |
| LNS2-L F | pOPINJ | AAGTTCTGTTTCAGGGCCCGGGTCAATGG ACGTTCCTGAGTACCG |
| LNS2-L R | pOPINJ | ATGGTCTAGAAAGCTTTAATTTTCGTTCGT GGCCTGTTCGTGC |
| LNS2-S F | pOPINM | AAGTTCTGTTTCAGGGCCCGGACTGCTACA TGGCCGTTGTCCCGC |
| LNS2-S R | pOPINM | ATGGTCTAGAAAGCTTTAACCATCGCTCAG CACCGTCGC |
| LNS2-L F | pOPINM | AAGTTCTGTTTCAGGGCCCGGGTCAATGGA CGTTCCTGAGTACCG |
| LNS2-L R | pOPINM | ATGGTCTAGAAAGCTTTAATTTTCGTTCGTG GCCTGTTCGTGC |
| LNS2-S F | pOPINNUSA | AAGTTCTGTTTCAGGGCCCGGACTGCTACA TGGCCGTTGTCCCGC |
| LNS2-S R | pOPINNUSA | ATGGTCTAGAAAGCTTTAACCATCGCTCAG CACCGTCGC |
| LNS2-L F | pOPINNUSA | AAGTTCTGTTTCAGGGCCCGGGTCAATGGA CGTTCCTGAGTACCG |
| LNS2-L R | pOPINNUSA | ATGGTCTAGAAAGCTTTAATTTTCGTTCGTG |

| | | |
|---|---|---|
| | | GCCTGTTCGTGC |
| LNS2-S F | pOPINS3C | AAGTTCTGTTTCAGGGCCCGGACTGCTACAT GGCCGTTGTCCCGC |
| LNS2-S R | pOPINS3C | ATGGTCTAGAAAGCTTTAACCATCGCTCAGC ACCGTCGC |
| LNS2-L F | pOPINS3C | AAGTTCTGTTTCAGGGCCCGGGTCAATGGAC GTTCCTGAGTACCG |
| LNS2-L R | pOPINS3C | ATGGTCTAGAAAGCTTTAATTTTCGTTCGTGG CCTGTTCGTGC |
| LNS2-S F | pOPINTF | AAGTTCTGTTTCAGGGCCCGGACTGCTACATG GCCGTTGTCCCGC |
| LNS2-S R | pOPINTF | ATGGTCTAGAAAGCTTTAACCATCGCTCAGCA CCGTCGC |
| LNS2-L F | pOPINTF | AAGTTCTGTTTCAGGGCCCGGGTCAATGGAC GTTCCTGAGTACCG |
| LNS2-L R | pOPINTF | ATGGTCTAGAAAGCTTTAATTTTCGTTCGTGG CCTGTTCGTGC |
| LNS2-S F | pOPINTRX | AAGTTCTGTTTCAGGGCCCGGACTGCTACATG GCCGTTGTCCCGC |
| LNS2-S R | pOPINTRX | ATGGTCTAGAAAGCTTTAACCATCGCTCAGCA CCGTCGC |
| LNS2-L F | pOPINTRX | AAGTTCTGTTTCAGGGCCCGGGTCAATGGACG TTCCTGAGTACCG |
| LNS2-L R | pOPINTRX | ATGGTCTAGAAAGCTTTAATTTTCGTTCGTGGC CTGTTCGTGC |

# Appendix B: Na-FAR-1-PA complex chemical shift table

| | H | N | $H_a$ | $H_b$ | $H_g$ | $H_d$ | C | $C_a$ | $C_b$ | $C_g$ | $C_d$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -13 G | | | | | | | | | | | | |
| -12 S | | | | | | | | | | | | |
| -11 S | | | | | | | | | | | | |
| -10 H | | | | | | | | | | | | |
| -9 H | | | | | | | | | | | | |
| -8 H | | | | | | | | | | | | |
| -7 H | | | | | | | | | | | | |
| -6 H | | | | | | | | | | | | |
| -5 H | | | | | | | | | | | | |
| -4 S | | | | | | | | | | | | |
| -2 G | | | | | | | | | | | | |
| -1 H | | | | | | | | | | | | |
| 0 M | | | | | | | | | | | | |
| 1 F | | | 4.54 | 2.93<br>2.94 | | 7.11 | | 57.81 | 40.79 | | 131.98 | $H^{e*}$ 7.10,<br>$C^{e*}$ 131.38 |
| 2 K | | | 4.46 | 1.64<br>1.84 | 1.30<br>1.33 | 1.63<br>1.63 | 176.19 | 54.46 | 34.90 | 24.32 | 28.92 | $H^{ea}$ 2.90,<br>$H^{eb}$ 2.91,<br>$C^{e}$ 41.78 |
| 3 Y | 8.86 | 123.81 | 3.86 | 2.90<br>2.82 | | 6.85 | 176.31 | 61.72 | 38.74 | | 133.52 | $H^{e*}$ 6.60,<br>$C^{e*}$ 118.99 |
| 4 E | 9.01 | 113.67 | 3.67 | 1.86<br>2.09 | 2.36<br>2.20 | | 175.98 | 58.82 | 28.57 | 37.55 | | |
| 5 D | 7.76 | 118.19 | 4.39 | 2.64<br>2.73 | | | 177.38 | 55.58 | 41.24 | | | |
| 6 I | 7.61 | 121.73 | 3.97 | 1.66 | 0.91<br>1.86<br>0.95 | 0.82 | 177.35 | 59.85 | 38.25 | 28.21<br>17.64 | 13.76 | |
| 7 P | | | | | | | | | | | | |
| 8 A | | | 3.74 | 1.44 | | | 178.14 | 56.21 | 18.73 | | | |
| 9 D | 8.86 | 112.76 | 4.29 | 2.55<br>2.38 | | | 177.84 | 56.05 | 39.12 | | | |
| 10 Y | 7.55 | 116.27 | 4.13 | | | 6.85 | 177.97 | 60.20 | 38.63 | | 132.88 | $H^{e*}$ 6.60,<br>$C^{e*}$ 118.29 |
| 11 R | 7.99 | 119.30 | 3.80 | 1.81<br>2.02 | 1.59<br>1.60 | 3.34<br>3.15 | 178.65 | 60.55 | 30.62 | 28.07 | 43.76 | |
| 12 D | 8.10 | 115.93 | 4.44 | 2.63<br>2.56 | | | 176.74 | 56.37 | 40.65 | | | |
| 13 L | 7.31 | 116.62 | 4.31 | 2.03<br>1.55 | 1.80 | 0.82<br>0.81 | 177.69 | 54.92 | 43.70 | 26.50 | 22.83 | |
| 14 M | 7.36 | 117.57 | 3.88 | 2.01<br>1.73 | | | | | | | | |
| 15 P | | | | | | | | | | | | |
| 16 P | | | 4.01 | 2.33<br>1.94 | 2.02<br>2.18 | 3.84<br>3.79 | 178.17 | 65.82 | 31.88 | 27.72 | 50.77 | |
| 17 E | 10.20 | 116.77 | 4.05 | 1.85 | | | 180.12 | 60.39 | 29.41 | 37.25 | | |
| 18 A | 7.41 | 119.58 | 4.16 | 1.36 | | | 177.75 | 54.27 | 19.24 | | | |
| 19 R | 7.66 | 118.51 | 3.80 | 1.82<br>1.82 | 1.47<br>1.22 | 3.18<br>3.23 | 177.61 | 60.56 | 30.46 | 28.50 | 43.63 | |
| 20 D | 8.35 | 116.66 | 4.09 | 2.63<br>2.49 | | | 177.94 | 57.29 | 41.60 | | | |
| 21 F | 7.13 | 116.78 | 4.14 | 3.33<br>3.04 | | 6.99 | 176.68 | 60.19 | 39.78 | | 132.29 | $H^{e*}$ 7.01,<br>$C^{e*}$ 131.11 |
| 22 L | 7.90 | 115.39 | 3.69 | 1.25<br>1.71 | | 0.85<br>0.86 | 178.67 | 57.11 | 43.11 | 25.83 | 25.80 | |
| 23 Q | 8.36 | 117.09 | 3.41 | | 1.93<br>1.93 | | 176.60 | 58.50 | 28.76 | 34.07 | | $N^{e2}$ 112.40,<br>$H^{e2a}$ 6.88,<br>$H^{e2b}$ 7.47 |
| 24 N | | | 4.68 | 2.58<br>2.81 | | 7.65<br>6.87 | 174.32 | 53.15 | 39.81 | | | $N^{d2}$ 116.09 |
| 25 L | 6.61 | 121.65 | 4.23 | 1.43<br>1.52 | 1.58 | 0.49<br>0.49 | 176.59 | 55.56 | 42.39 | 25.22 | 25.29<br>25.29 | |
| 26 S | 9.35 | 124.98 | 4.70 | 3.99<br>4.34 | | | 175.42 | 56.70 | 66.66 | | | |
| 27 D | | | | | | | | | | | | |
| 28 G | | | 3.84<br>3.84 | | | | 176.43 | 47.09 | | | | |
| 29 D | 7.95 | 123.45 | 4.33 | 2.44<br>3.03 | | | 178.30 | 57.22 | 41.84 | | | |
| 30 K | 7.94 | 118.39 | 3.78 | 2.03<br>1.85 | | 1.66 | 178.30 | 60.92 | 32.35 | 25.60 | 29.76 | $H^{ea}$ 2.68,<br>$H^{eb}$ 2.80,<br>$C^{e}$ 41.35 |
| 31 T | 7.95 | 115.93 | 3.87 | 4.39 | 1.25 | | 176.14 | 67.20 | 68.72 | 21.98 | | |
| 32 V | 7.85 | 122.49 | 3.71 | 2.21 | 1.08<br>0.87 | | 177.81 | 66.89 | 31.43 | 22.32<br>23.27 | | |
| 33 L | 8.20 | 118.97 | 3.82 | 1.29<br>2.07 | 2.05 | 0.89<br>0.89 | 178.51 | 58.86 | 41.76 | 26.34 | 24.13 | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 34 K | 7.98 | 117.90 | 3.91 | 1.97<br>1.97 | 1.30<br>1.30 | 1.75<br>1.75 | 178.30 | 60.64 | 32.68 | 25.96 | 29.70 | $H^{ea}$ 2.88,<br>$H^{eb}$ 2.95,<br>$C^{e}$ 41.96 |
| 35 E | 8.43 | 119.05 | 3.94 | 2.04<br>2.25 | 2.53<br>2.53 | | 180.49 | 59.74 | 29.69 | 36.72 | | |
| 36 V | 8.64 | 119.52 | 3.72 | 2.16 | 0.84<br>1.02 | | 180.16 | 66.72 | 31.58 | 23.98 | | |
| 37 F | 8.68 | 122.48 | 4.31 | 3.22<br>3.29 | | 7.06 | 179.20 | 62.85 | 39.29 | | 132.02 | $H^{e*}$ 7.15,<br>$C^{e*}$ 132.51 |
| 38 K | 8.80 | 119.87 | 4.05 | 1.81<br>2.00 | 1.70<br>1.61 | 1.44<br>1.42 | 176.26 | 59.03 | 32.35 | 25.67 | 29.03 | $H^{ea}$ 2.77,<br>$H^{eb}$ 2.99,<br>$C^{e}$ 41.91 |
| 39 A | 7.27 | 119.94 | 4.08 | 1.50 | | | 176.87 | 51.94 | 18.73 | | | |
| 40 G | 7.15 | 104.20 | 3.83<br>3.83 | | | | | 43.80 | | | | |
| 41 P | | | | | | | | | | | | |
| 42 Y | | | | | | | | | | | | |
| 42 Y | | | 4.61 | 3.16<br>2.59 | | 7.09 | | 56.80 | 40.70 | | 133.23 | $H^{e*}$ 6.65,<br>$C^{e*}$ 117.41 |
| 43 K | | | 3.89 | 1.77<br>1.65 | 1.41 | 1.52 | 176.37 | 59.04 | 33.66 | 25.19 | 29.33 | $H^{ea}$ 2.94,<br>$H^{eb}$ 3.01,<br>$C^{e}$ 41.94 |
| 44 N | 7.32 | 108.45 | 4.69 | 3.21<br>3.29 | | 7.05<br>6.69 | 175.94 | 51.61 | 39.56 | | | $N^{d2}$ 110.84 |
| 45 T | | | 3.89 | 3.81 | 0.83 | | | 65.94 | 68.23 | 22.69 | | |
| 45 T | | | | | | | | | | | | |
| 46 E | | | 3.65 | 2.06<br>1.88 | 2.21<br>2.38 | | 179.29 | 61.53 | 28.28 | 37.71 | | |
| 47 E | 8.48 | 120.90 | 4.06 | 2.23<br>2.09 | 2.35<br>2.50 | | 179.05 | 59.53 | 30.17 | 37.47 | | |
| 48 S | 7.54 | 117.76 | 3.94 | 3.50<br>3.41 | | | | 61.31 | 62.73 | | | |
| 49 I | 7.78 | 123.25 | 3.38 | 1.72 | 0.43<br>0.59<br>0.76 | 0.43 | 177.57 | 65.43 | 37.44 | 17.48<br>30.11 | 13.78 | |
| 50 A | 7.95 | 123.50 | 3.91 | 1.41 | | | 180.56 | 55.31 | 17.68 | | | |
| 51 A | 7.62 | 121.19 | 3.99 | 1.30 | | | 180.70 | 55.03 | 18.04 | | | |
| 52 L | 7.81 | 120.13 | 3.96 | 1.49<br>2.05 | 1.44 | 0.92<br>0.92 | 177.69 | 57.95 | 41.90 | 27.01 | 24.36<br>24.35 | |
| 53 K | 8.60 | 117.79 | 3.76 | 1.75<br>1.85 | 1.25<br>1.56 | 1.54<br>1.54 | 177.62 | 59.74 | 32.82 | 27.26 | 30.04 | $H^{ea}$ 2.62,<br>$H^{eb}$ 2.69,<br>$C^{e}$ 41.87 |
| 54 K | 7.28 | 115.39 | 3.96 | 1.85 | 1.56<br>1.42 | 1.66<br>1.75<br>1.66<br>1.68 | 178.02 | 58.62 | 32.66 | 25.38 | 29.43 | $H^{e2}$ 2.95,<br>$H^{e3}$ 2.95,<br>$C^{e}$ 42.18 |
| 55 K | 7.08 | 116.88 | 4.24 | 1.75<br>1.74 | 1.45<br>1.45 | 1.64<br>1.64 | 177.40 | 57.28 | 33.94 | 25.35 | 28.91 | $H^{e2}$ 2.94,<br>$H^{e3}$ 2.94,<br>$C^{e}$ 42.06 |
| 56 S | 8.52 | 111.31 | 4.83 | 3.68<br>3.76 | | | 177.41 | 54.68 | 62.86 | | | |
| 57 P | | | 4.34 | 2.40<br>1.97 | 2.15<br>2.15 | 3.96<br>3.49 | 179.82 | 66.19 | 32.29 | 27.24 | 50.25 | |
| 58 E | 8.77 | 118.17 | 4.08 | 1.95<br>2.03 | 2.28<br>2.28 | | 179.12 | 59.96 | 29.37 | 36.84 | | |
| 59 L | 7.70 | 121.75 | 4.19 | 1.71<br>1.71 | 1.62 | 0.93<br>0.97 | 179.78 | 57.74 | 41.89 | 27.26 | 25.88<br>23.67 | |
| 60 G | 8.94 | 104.15 | 3.36<br>3.55 | | | | 174.30 | 47.65 | | | | |
| 61 A | 8.02 | 122.86 | 4.22 | 1.53 | | | 180.58 | 55.14 | 18.32 | | | |
| 62 K | 7.57 | 117.89 | 4.06 | 2.20<br>1.89 | 1.38<br>1.65 | | 179.76 | 60.02 | 33.51 | 26.22 | 30.63 | $H^{e2}$ 2.93,<br>$H^{e3}$ 2.93,<br>$C^{e}$ 42.06 |
| 63 V | 8.14 | 118.97 | 3.28 | 1.99 | 0.75<br>0.87 | | 177.11 | 66.87 | 31.23 | 24.16<br>22.06 | | |
| 64 E | 8.59 | 121.45 | 3.94 | 2.23<br>2.09 | 2.24<br>2.24 | | 178.63 | 60.32 | 30.16 | 36.59 | | |
| 65 K | 7.58 | 118.59 | 4.13 | 1.92<br>1.93 | 1.47<br>1.61 | 1.70<br>1.70 | 179.58 | 59.40 | 32.34 | 25.35 | 29.19 | $H^{e2}$ 2.95,<br>$H^{e3}$ 2.95,<br>$C^{e}$ 42.19 |
| 66 L | 7.61 | 120.45 | 4.01 | 1.75<br>1.28 | 1.63 | 0.33<br>0.59 | 178.11 | 58.09 | 41.97 | 26.74 | 23.25<br>25.01 | |
| 67 H | 8.73 | 119.93 | 4.05 | 2.92<br>3.26 | | | 176.58 | 60.63 | 31.38 | | | |
| 68 A | 8.23 | 120.45 | 3.81 | 1.50 | | | 180.70 | 55.25 | 17.85 | | | |
| 69 M | 7.89 | 119.07 | 4.06 | 2.00<br>2.00 | 2.77<br>2.25 | | 178.73 | 58.88 | 32.36 | 31.24 | | $H^{e*}$ 1.86 |
| 70 V | 8.13 | 119.84 | 3.33 | 1.91 | 0.40<br>0.58 | | | 67.36 | 31.06 | 21.96<br>23.08 | | |
| 71 K | 8.37 | 119.18 | | | | | | | | | | |
| 72 S | 8.06 | 115.61 | 4.16 | 3.93<br>3.93 | | | 176.63 | 61.74 | 62.86 | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 73 K | 7.63 | 123.95 | 3.89 | 1.06<br>1.71 | | 1.66<br>1.74 | 178.39 | 60.01 | 34.94 | 25.18 | 29.42 | $H^{ea}$ 2.90,<br>$H^{eb}$ 3.08,<br>$C^{e}$ 41.56 |
| 74 I | 8.13 | 117.12 | 3.80 | 1.91 | 2.24<br>1.06<br>0.98 | 0.87 | 178.36 | 65.36 | 38.14 | 17.38<br>30.10 | 14.20 | |
| 75 A | 7.69 | 118.27 | 4.06 | 1.45 | | | 178.20 | 54.21 | 18.31 | | | |
| 76 A | 7.16 | 118.41 | 4.30 | 1.39 | | | 178.15 | 52.10 | 19.18 | | | |
| 77 L | 7.44 | 120.27 | 4.37 | 2.15<br>1.74 | 1.22 | 1.06 | 177.96 | 55.32 | 43.25 | 26.85 | 23.52<br>25.49 | |
| 78 G | 10.53 | 112.53 | 3.86<br>4.32 | | | | 178.02 | 44.35 | | | | |
| 79 P | | | 4.15 | 2.36<br>1.96 | 2.17<br>2.04 | 3.71<br>3.71 | 179.84 | 66.35 | 32.14 | 27.64 | 48.94 | |
| 80 E | 8.31 | 120.30 | 4.06 | 2.02 | 2.13<br>2.23 | | 179.33 | 59.75 | 29.68 | 35.98 | | |
| 81 A | 9.68 | 124.26 | 4.61 | 1.94 | | | 179.55 | 55.19 | 19.68 | | | |
| 82 K | 9.33 | 120.51 | 3.97 | 1.93<br>1.73 | 1.54<br>1.43 | | 178.70 | 60.33 | 32.52 | 25.56 | 29.74 | $H^{ea}$ 2.99,<br>$H^{eb}$ 3.06,<br>$C^{e}$ 42.09 |
| 83 G | 8.02 | 106.28 | 3.95<br>3.64 | | | | 176.52 | 47.12 | | | | |
| 84 F | 7.60 | 123.52 | 4.37 | 3.28<br>3.43 | | 6.83 | 177.62 | 60.36 | 39.78 | | 131.89 | $H^{e*}$ 6.83 |
| 85 A | 8.66 | 125.17 | 3.34 | 1.44 | | | 179.10 | 55.48 | 17.78 | | | |
| 86 E | 8.87 | 116.14 | 3.87 | 2.19<br>2.01 | 2.56<br>2.56 | | 180.30 | 60.07 | 29.48 | 36.89 | | |
| 87 K | 8.21 | 120.77 | 4.05 | 1.84<br>1.91 | 1.50<br>1.42 | 1.70<br>1.70 | 179.73 | 59.22 | 32.52 | 25.38 | 28.99 | $H^{e2}$ 2.92,<br>$H^{e3}$ 2.92,<br>$C^{e}$ 41.89 |
| 88 S | 7.98 | 117.63 | 3.93 | | | | 175.89 | 63.65 | 62.53 | | | |
| 89 I | 8.39 | 122.88 | 3.51 | 1.92 | 0.89<br>0.89<br>1.79 | 0.78 | 177.65 | 65.64 | 37.76 | 17.47<br>30.52 | 13.52 | |
| 90 E | 7.84 | 120.79 | 4.04 | 2.02<br>2.18 | 2.35<br>2.35 | | 179.99 | 59.70 | 29.25 | 35.74 | | |
| 91 I | 7.79 | 122.06 | 3.73 | 2.04 | 0.90<br>1.66<br>1.66 | 0.80 | 179.49 | 65.45 | 37.77 | 17.55<br>30.22 | 13.98 | |
| 92 A | 8.00 | 123.16 | 3.95 | 1.48 | | | 179.02 | 55.59 | 18.45 | | | |
| 93 R | 9.07 | 118.52 | 3.80 | | | | 179.50 | 60.62 | 30.40 | 25.64 | 41.29 | |
| 94 G | 8.05 | 107.30 | 3.97<br>3.97 | | | | 176.41 | 47.08 | | | | |
| 95 I | 7.53 | 122.70 | 3.83 | 2.20 | 1.39<br>1.65<br>0.95 | 0.80 | 177.76 | 63.57 | 36.84 | 27.86<br>18.16 | 11.56 | |
| 96 K | 8.17 | 120.28 | 3.76 | 1.86 | 1.47<br>1.47 | 1.67<br>1.73 | 177.41 | 60.56 | 32.58 | 25.87 | 29.74 | $H^{e2}$ 2.68,<br>$H^{e3}$ 2.68,<br>$C^{e}$ 41.81 |
| 97 A | 8.11 | 118.05 | 4.27 | | | | 181.28 | 55.20 | 18.14 | | | |
| 98 R | 7.71 | 116.95 | 4.07 | | 1.70<br>1.70 | 3.23<br>3.05 | 178.97 | 59.11 | 29.94 | 27.84 | 43.78 | |
| 99 Y | 7.94 | 118.52 | 3.98 | 3.02<br>2.89 | | 6.79 | 177.73 | 62.75 | 38.03 | | 132.27 | $H^{e*}$ 6.81,<br>$C^{e*}$ 118.66 |
| 100 Y | 7.73 | 116.41 | | 3.24<br>2.86 | | 7.14 | 176.77 | 60.02 | 38.14 | | 133.28 | $H^{e*}$ 6.54,<br>$C^{e*}$ 117.59 |
| 101 T | 7.59 | 108.97 | 4.52 | 4.48 | 1.31 | | 174.94 | 61.95 | 70.66 | 21.65 | | |
| 102 G | 7.93 | 108.47 | | | | | | | | | | |
| 102 G | | | 3.84<br>4.13 | | | | 173.53 | 45.62 | | | | |
| 103 N | 7.85 | 120.33 | 4.91 | 2.47<br>2.64 | | 6.82<br>7.40 | 173.88 | 51.71 | 38.92 | 38.69 | | $N^{d2}$ 112.47 |
| 104 E | 8.43 | 122.49 | 3.83 | 1.87<br>1.87 | 2.22<br>2.22 | | 174.04 | 55.74 | 28.74 | 36.48 | | |
| 105 P | | | | | | | | | | | | |
| 106 T | 8.88 | 113.00 | | | | | | 60.11 | | | | |
| 106 T | | | 4.54 | 4.66 | 1.30 | | 176.06 | 59.99 | 72.18 | 21.82 | | |
| 107 K | 9.03 | 120.60 | 3.93 | 1.90 | 1.40<br>1.33 | 1.72<br>1.68 | 178.49 | 60.50 | 31.99 | 25.48 | 29.56 | $H^{e2}$ 2.95,<br>$H^{e3}$ 2.95,<br>$C^{e}$ 41.81 |
| 108 D | 7.95 | 117.36 | 4.35 | 2.46<br>2.60 | | | 179.06 | 57.76 | 40.41 | | | |
| 109 D | 7.80 | 122.23 | 4.42 | 2.56<br>3.05 | | | 179.54 | 57.58 | 40.93 | | | |
| 110 L | 8.10 | 121.05 | 3.79 | 1.53<br>1.76 | 1.75 | 0.87<br>0.72 | 179.16 | 58.04 | 41.09 | 26.94 | 25.60<br>23.59 | |
| 111 K | | | | | | | | | | 26.00 | | |
| 112 A | 7.97 | 120.28 | | | | | 181.10 | 55.41 | 18.05 | | | |
| 113 S | 7.93 | 116.21 | 4.33 | 3.73<br>3.98 | | | 181.12 | 63.19 | 63.18 | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 114 V | 8.11 | 121.74 | 3.49 | 2.31 | 0.91 | | 177.50 | 67.26 | 31.32 | 23.88 | | |
| | | | | | 1.07 | | | | | | | |
| 115 K | 8.43 | 119.64 | 3.77 | 2.00 | 1.52 | 1.59 | 178.58 | 60.83 | 32.29 | 25.91 | 29.74 | $H^{e2}$ 2.79, |
| | | | | | 1.54 | 1.62 | | | | | | $H^{e3}$ 2.79, |
| | | | | | | | | | | | | $H^{ea}$ 2.73, |
| | | | | | | | | | | | | $H^{eb}$ 2.80, |
| | | | | | | | | | | | | $C^{e}$ 41.78 |
| 116 E | 7.69 | 118.47 | 4.04 | 2.20 | 2.37 | | 178.99 | 59.60 | 29.41 | 35.74 | | |
| | | | | 2.24 | 2.22 | | | | | | | |
| 117 V | 7.74 | 120.05 | 3.58 | 2.18 | 1.08 | | 177.88 | 67.27 | 31.85 | 23.33 | | |
| | | | | | 0.88 | | | | | | | |
| 118 L | 8.53 | 119.26 | 4.06 | 1.51 | 1.83 | 0.73 | 179.15 | 58.17 | 41.55 | 26.95 | 22.67 | |
| | | | | | | 0.73 | | | | | 22.67 | |
| 119 K | 8.44 | 119.53 | 3.77 | 1.86 | 1.20 | | 178.27 | 60.93 | 32.48 | 26.96 | 30.29 | $H^{ea}$ 2.72, |
| | | | | 2.00 | 1.20 | | | | | | | $H^{eb}$ 2.80, |
| | | | | | | | | | | | | $C^{e}$ 42.01 |
| 120 L | | | 3.97 | 1.62 | 1.95 | 0.87 | 179.07 | 57.61 | 42.07 | 26.51 | 22.76 | |
| | | | | 2.33 | | 0.87 | | | | | 22.76 | |
| 121 Y | 8.26 | 120.88 | 2.77 | 2.58 | | 6.06 | 177.35 | 61.48 | 37.32 | | 133.36 | $H^{e*}$ 6.47, |
| | | | | 2.30 | | | | | | | | $C^{e*}$ 116.70 |
| 122 K | 8.22 | 115.92 | 3.42 | 1.82 | 1.65 | 1.66 | 177.66 | 58.96 | 32.36 | 25.74 | 29.44 | $H^{e2}$ 2.94, |
| | | | | 1.82 | 1.51 | 1.66 | | | | | | $H^{e3}$ 2.94, |
| | | | | 1.82 | 1.51 | | | | | | | $C^{e}$ 41.95 |
| | | | | 1.99 | 1.52 | | | | | | | |
| 123 A | 7.18 | 117.92 | 4.08 | 1.38 | | | 178.12 | 52.41 | 19.07 | | | |
| 124 M | 7.08 | 117.75 | 4.15 | 2.03 | 2.45 | | 176.91 | 57.71 | 37.48 | 32.06 | | |
| | | | | 1.71 | 2.45 | | | | | | | |
| 125 S | 9.61 | 117.87 | 4.26 | | | | 174.93 | 58.53 | 64.74 | | | |
| 126 D | 8.86 | 121.35 | 4.19 | 2.63 | | | 179.16 | 58.04 | 39.60 | | | |
| | | | | 2.63 | | | | | | | | |
| 127 A | 8.25 | 122.82 | 4.17 | 1.44 | | | 180.99 | 55.24 | 18.27 | | | |
| 128 G | 8.11 | 111.69 | 4.03 | | | | 176.08 | 47.29 | | | | |
| | | | 3.64 | | | | | | | | | |
| 129 K | 8.30 | 121.31 | 3.76 | 1.77 | 1.68 | 1.62 | 180.26 | 60.95 | 33.06 | 27.02 | 29.59 | $H^{e2}$ 2.69, |
| | | | | | 1.17 | 1.71 | | | | | | $H^{e3}$ 2.69, |
| | | | | | | | | | | | | $C^{e}$ 42.44 |
| 130 A | 8.25 | 123.17 | 4.14 | 1.48 | | | 180.09 | 54.92 | 17.94 | | | |
| 131 D | 7.98 | 120.01 | 4.41 | 2.91 | | | 177.86 | 57.48 | 42.70 | | | |
| | | | | 3.04 | | | | | | | | |
| 132 F | 8.58 | 117.17 | 4.23 | 3.16 | | 7.37 | 177.72 | 62.76 | 39.97 | | 131.74 | $H^{e*}$ 6.93, |
| | | | | 3.28 | | | | | | | | $C^{e*}$ 131.02 |
| 133 G | 8.51 | 104.18 | 3.60 | | | | 174.41 | 46.51 | | | | |
| | | | 3.78 | | | | | | | | | |
| 134 K | 7.55 | 119.59 | 4.06 | 2.01 | 1.43 | 1.73 | 178.54 | 58.67 | 32.51 | 25.30 | 29.08 | $H^{e2}$ 2.96, |
| | | | | | 1.43 | 1.67 | | | | | | $H^{e3}$ 2.96, |
| | | | | | | | | | | | | $C^{e}$ 41.87 |
| 135 Q | 6.74 | 113.07 | 4.17 | 1.57 | 2.42 | | 174.67 | 55.25 | 29.83 | 32.45 | | $N^{e2}$ 112.75, |
| | | | | 1.65 | 2.43 | | | | | | | $H^{e2a}$ 6.66, |
| | | | | | | | | | | | | $H^{e2b}$ 7.57 |
| 136 F | 8.20 | 115.59 | 4.61 | 2.68 | | | 171.41 | 55.14 | 39.26 | | | $H^{e*}$ 6.93, |
| | | | | 2.57 | | | | | | | | $C^{e*}$ 132.50 |
| 137 P | | | 4.34 | 2.33 | 1.95 | 3.62 | | 65.63 | 31.68 | 27.43 | 50.15 | |
| | | | | 1.86 | 1.95 | 2.86 | | | | | | |
| 138 F | | | | | | | | | | | | |
| 139 L | | | 3.38 | 1.97 | 1.97 | 0.73 | 178.60 | 57.70 | 41.54 | 26.97 | 22.74 | |
| | | | | 0.97 | | 0.73 | | | | | 22.74 | |
| 140 A | 8.93 | 119.02 | 3.86 | 1.60 | | | 178.94 | 56.15 | 19.09 | | | |
| 141 K | 7.32 | 117.23 | 3.89 | 1.73 | 1.38 | 1.66 | 180.17 | 58.96 | 33.57 | 24.80 | 28.95 | $H^{e2}$ 2.95, |
| | | | | 1.68 | 1.38 | 1.66 | | | | | | $H^{e3}$ 2.95, |
| | | | | 1.73 | | | | | | | | $C^{e}$ 41.82 |
| | | | | 1.79 | | | | | | | | |
| 142 V | | | 3.60 | 2.23 | 1.16 | | | 67.44 | 31.66 | 22.38 | | |
| | | | | | 1.11 | | | | | 23.74 | | |
| 143 F | | | 4.21 | 2.80 | | 7.22 | | 61.74 | 37.27 | | 133.20 | $H^{e*}$ 7.21, |
| | | | | 3.44 | | | | | | | | $C^{e*}$ 131.39 |
| 144 E | | | | | | | | | | | | |
| 145 S | | | 4.36 | 4.09 | | | | 59.86 | 64.27 | | | |
| | | | | 3.99 | | | | | | | | |
| 146 G | | | | | | | | | | | | |
| 147 K | | | 4.02 | 1.81 | 1.62 | 1.68 | 178.79 | 59.13 | 32.45 | 25.27 | 29.06 | $H^{e2}$ 2.96, |
| | | | | 1.68 | 1.43 | 1.68 | | | | | | $H^{e3}$ 2.96, |
| | | | | | | | | | | | | $C^{e}$ 41.88 |
| 148 A | 7.73 | 120.87 | 3.96 | 1.38 | | | 178.28 | 55.37 | 18.45 | | | |
| 149 A | 7.63 | 119.73 | 3.91 | 1.45 | | | 180.10 | 55.02 | 18.02 | | | |
| 150 K | 7.92 | 118.29 | 4.06 | 1.71 | 1.46 | 1.65 | 179.31 | 58.72 | 32.30 | 25.07 | 29.14 | $H^{e2}$ 2.94, |
| | | | | 1.87 | 1.41 | 1.65 | | | | | | $H^{e3}$ 2.94, |
| | | | | | | | | | | | | $C^{e}$ 42.12 |
| 151 F | 7.86 | 120.03 | 4.25 | 3.07 | | 7.14 | 176.24 | 60.47 | 39.54 | | 131.44 | |
| | | | | 3.20 | | | | | | | | |
| 152 A | 7.96 | 117.76 | 3.94 | 1.41 | | | 177.34 | 53.20 | 19.31 | | | |
| 153 G | 7.58 | 104.45 | 3.85 | | | | 174.26 | 45.31 | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 154 E | 8.00 | 120.34 | 4.29 | 1.79<br>2.03 | 2.15<br>2.23 | | 175.37 | 56.28 | 30.80 | 36.30 | | |
| 155 N | 7.97 | 124.78 | 4.38 | 2.68<br>2.57 | | 6.72<br>7.40 | 179.42 | 54.87 | 40.53 | | | $N^{d2}$ 112.61 |