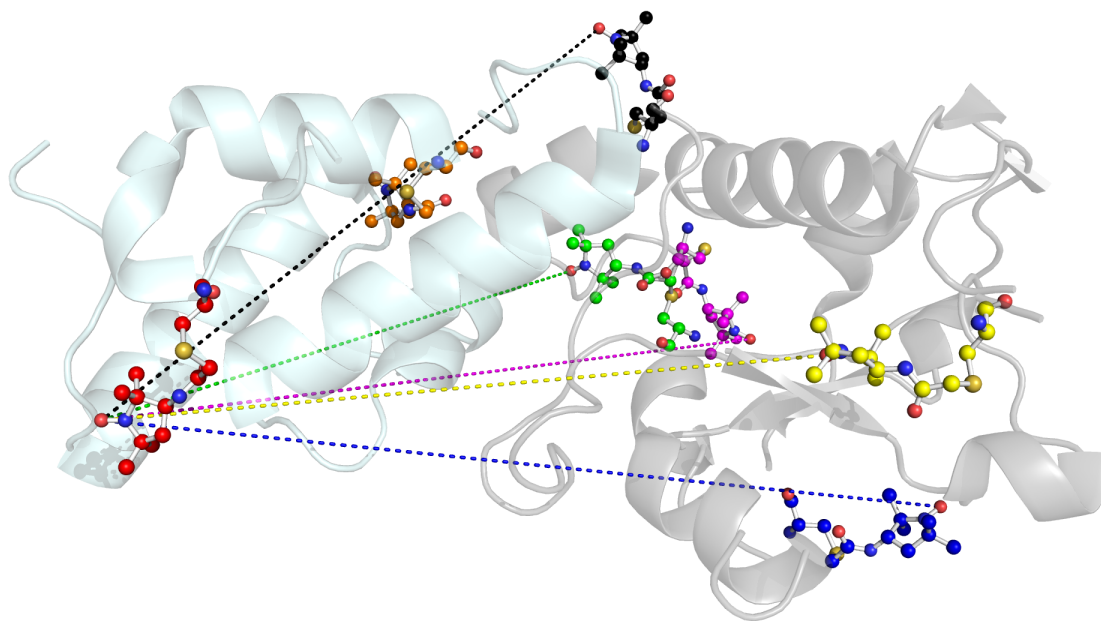


An efficient EPR method for characterising protein-protein interactions



Morgan Haydn Bye

A thesis submitted for the degree of Doctor of Philosophy.

School of Chemistry, University of East Anglia, Norwich, UK.

Submitted 2nd September 2013.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Declaration

I certify that the work contained in this thesis submitted, by me, for the degree of Doctor of Philosophy, is my own original work, except where due reference is made to other authors or co-workers and has not previously been submitted, by me, for a degree at this or any other university.

Signed:

A handwritten signature in black ink that reads "Morgan Bye". The signature is written in a cursive style with a large, prominent 'M' and 'B'.

Morgan Bye

31st August 2013

Acknowledgments

Several people have been instrumental in the completion in this work and to them I shall always be indebted.

First let me thank those that have directly had influence and guided the work. The BBSRC and University of East Anglia for supporting this work in finance and infrastructure. The collaborators Nick Housden (started University of York, now University of Oxford) and especially Mireille Vankemmelbeke (University of Nottingham). Christian Teutloff (FU Berlin) for answering many EPR questions and allowing me a week of machine time in Berlin. Colin MacDonald for introducing me to HADDOCK and the NMR servers. Christian Roghi who consulted with me upon DNA where his experience truly shone through.

The ever talented and always supportive academics, namely: Geoff Moore, Myles Cheesman, Tharin Blumenschein and Andrew Thomson.

The MacMillan lab group in the soon-to-be-mum Jess van Wonderen and the now-Peruvian Matt Bawn, as well as the now-returned-to-Germany Olly Hecht in the NMR sister group for their constant input, advice, friendship and coffee every morning.

To my then partner Liz for a sympathetic ear when I needed it. To my parents and brother for their support and effort in trying to understand what on Earth I've been doing all this time.

Thanks to my brothers-in-arms that every day have fought side-by-side through the boredom, monotony, the "HOW IS THIS STILL NOT WORKING?", the "WHAT THE HELL WAS THAT?" and much more besides. To them I owe the most. James Tolchard, Justin Bradley and Girgis Obaid

Thank you to the project students along the way including Alicia Stratford and the tirelessly working Helen Newson. But most of all Alex Morley, that gave above and beyond, working far harder than he needed to for an entire year to get this project to where it is; and somewhere along the line became a life-long friend in the process.

Finally, my thanks go to Fraser MacMillan for the opportunity. Allowing me free-roam of the lab, computers and spectrometers.

And with that, I thank you dear reader, for showing an interest in my work and to anyone else that I might have missed out along the way.

And remember:

"An expert is a man who has made all the mistakes which can be made, in a narrow field."

- Niels Bohr

Contents

I. Background	3
1. Fundamentals of electron paramagnetic resonance	5
1.1. Introduction to electron paramagnetic resonance	5
1.2. Site directed spin labelling	6
1.3. Basic EPR theory	7
1.3.1. Electron Zeeman	8
1.3.2. Hyperfine interaction	11
1.3.3. Nuclear quadrupole interaction	13
1.3.4. Zero field splitting / fine structure	13
1.3.5. Exchange coupling	14
1.3.6. Dipole-dipole coupling	15
1.4. Continuous wave versus pulsed EPR	16
1.5. Continuous wave EPR	17
1.5.1. EPR characteristics of nitroxides	17
1.5.2. Mobility	18
1.5.3. Accessibility	19
1.5.4. Multiple frequency EPR	20
1.5.5. Line broadening	22
1.6. Pulsed EPR theory	22
1.6.1. Magnetization in the laboratory frame	22
1.6.2. Magnetization in the rotating frame	24
1.6.3. Free induction decay and off resonances	25
1.7. Pulsed EPR applications	27
1.7.1. Relaxation times	27
1.7.1.1. Transverse relaxation	28
1.7.1.2. Inversion recovery	29
1.7.2. Field swept electron spin echo	31
1.7.3. Electron-electron double resonance	31
2. Cell biology and biochemistry of colicin E9	35
2.1. Organisms and cells	35
2.2. Bacteria and colicins	35
2.3. Domains and structures of E-type colicins	37
2.4. Extracellular release of colicins	40
2.5. Receptor binding and membrane translocation	41
2.5.1. Initial binding	41
2.5.2. Colicin cleavage at cell surface	42
2.5.3. Release of immunity protein	42
2.5.4. Colicin import	42
2.6. Cytotoxic activity of Colicin E9	43
II. Investigation	47
3. Aims and objectives	49
4. Protein production, purification and performance	51
4.1. DNA stocks	51
4.1.1. Naturally derived mutants	51
4.1.2. Synthetically derived mutants	53
4.2. Protein expression	53
4.3. Crude purification	54
4.3.1. Naturally derived colicin E9 DNases	54
4.3.2. Synthetically derived colicin E9 DNases and immunity protein Im9s	54
4.4. Polishing purification	55

Contents

4.5. Protein confirmation	56
4.5.1. Optical spectroscopy	56
4.5.2. Mass spectrometry	57
4.6. Yields	57
4.7. Binding activity	58
4.8. Enzymatic activity	59
5. Spin label characterisation	63
5.1. Spin label efficiency	63
5.2. Continuous wave EPR characterisation	64
5.2.1. Colicin E9 DNase S3C ^{IA1}	65
5.2.2. Colicin E9 DNase S30C ^{R1A}	65
5.2.3. Colicin E9 DNase S3C ^{IA1} / S80C ^{IA1}	66
5.2.4. Colicin E9 DNase S30C ^{IA1} / S80C ^{IA1}	66
5.2.5. Colicin E9 DNase E66C ^{IA1}	67
5.2.6. Colicin E9 DNase D20C ^{IA1} / E66 ^{IA1}	67
5.2.7. Im9 C23 ^{IA1}	68
5.2.8. Im9 C23S / S6C ^{IA1}	68
5.2.9. Im9 S6C ^{IA1} / C23 ^{IA1}	69
5.3. Spin label properties	70
5.4. Spin label accessibility	72
5.5. Summary and discussion	77
6. Echo optimization	79
6.1. Spectra fitting	79
6.2. Temperature optimization	81
6.2.1. T_2 relaxation time	82
6.2.2. T_1 relaxation time	84
6.3. pH optimization	84
6.4. Deuteration optimization	86
6.5. Glycerol optimization	88
6.6. Glycerol deuteration optimization	88
6.6.1. IA-PROXYL	90
6.6.2. Colicin E9 DNase S3C	90
6.7. Sample cooling	90
6.8. Pulse length optimization	94
6.9. Summary and discussion	95
7. Pulsed Investigation	99
7.1. T_1 and T_2 measurements	99
7.2. Distance measurement	100
7.2.1. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	102
7.2.2. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1}	104
7.2.3. Im9 S6C ^{IA1} /C23 ^{IA1}	106
7.2.4. Colicin E9 DNase S3C ^{IA1} - Im9 C23 ^{IA1}	108
7.2.5. Colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23S	110
7.2.6. Colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	112
7.2.7. Colicin E9 DNase S30C ^{R1A} - Im9 C23 ^{R1A}	114
7.2.8. Colicin E9 DNase E66C ^{IA1} - Im9 C23 ^{IA1}	116
7.2.9. Colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23S	118
7.2.10. Colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	120
7.2.11. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	122
7.2.12. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	124
7.2.13. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	126
7.2.14. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	128
7.2.15. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	130
7.2.16. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	132
7.3. Orientation selection	134
7.4. Summary	138
8. Protein - protein docking with HADDOCK	141
8.1. Docking principles	142
8.2. HADDOCK overview	142
8.2.1. Docking methodology	144
8.2.2. Colicin protein structure	145
8.2.3. HADDOCK's application to colicins	147

8.3. <i>in silico</i> spin labelling of proteins	148
8.3.1. MMM	149
8.4. Conversion of distance distributions to distance restraints	150
8.5. MISHAP	153
8.5.1. Implementation of MMM rotamers into PDB format	154
8.5.2. Acquiescence of HADDOCK to spin labels	155
8.5.2.1. Topology	155
8.5.3. Distance distributions and HADDOCK restraints	156
8.6. HADDOCK runs	157
8.6.1. Without constraints	159
8.6.2. With chemical shift perturbation constraints	160
8.6.3. With CSP constraints and spin labels	163
8.6.4. With intramolecular PELDOR restraints	163
8.6.4.1. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	163
8.6.4.2. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1}	164
8.6.4.3. Im9 S6C ^{IA1} /C23 ^{IA1}	166
8.6.5. With intermolecular PELDOR restraints from colicin E9 DNase S3C ^{IA1}	166
8.6.6. With intermolecular PELDOR restraints from colicin E9 DNase E66C ^{IA1}	169
8.6.7. With intermolecular PELDOR restraints to Im9 C23 ^{IA1}	170
8.6.8. With intermolecular PELDOR restraints to Im9 S6C ^{IA1}	171
8.6.9. With all PELDOR restraints and CSPs	173
8.6.10. With only PELDOR restraints	175
8.6.11. With a MMM derived restraint	176
8.6.12. With MMM derived restraints for all spin labels	177
8.7. Summary and conclusions	178
9. Discussion and conclusions	181
9.1. Protein production	181
9.2. cw-EPR	182
9.3. Pulsed EPR	183
9.4. Docking methodology	186
9.5. Outlook	190

III. Methods and materials

193

10. Molecular biology	195
10.1. General techniques	195
10.1.1. SDS-PAGE	195
10.1.2. UV/Vis Spectroscopy	195
10.1.3. Dialysis	196
10.2. DNA	196
10.2.1. General DNA techniques	196
10.2.1.1. Agar plates	196
10.2.1.2. Transformation of DNA plasmid into <i>E. coli</i>	196
10.2.1.3. Glycerol stocks	196
10.2.1.4. DNA sequencing	196
10.3. Plasmids	197
10.3.1. Synthetically derived plasmids	198
10.3.1.1. Colicin E9 DNase S3C / S80C / H127A - pMB1	198
10.3.1.2. Colicin E9 DNase S30C / S80C / H127A - pMB2	199
10.3.1.3. Colicin Immunity protein Im9 S6C - pMB3	199
10.3.1.4. Colicin E9 DNase S30C / H127A - pMB4	199
10.3.1.5. Colicin E9 DNase S80C / H127A - pMB5	200
10.4. Protein production	200
10.4.1. Colicin E9 DNase production	200
10.4.1.1. Large scale over-expression	200
10.4.1.2. Liberation of protein from cells	201
10.4.1.3. Immobilised metal ion affinity chromatography	201
10.4.1.4. Dialysis	201
10.4.1.5. Lyophilisation	201
10.4.1.6. Size exclusion chromatography	201
10.4.1.7. Purification and concentration	202
10.4.1.8. Protein concentration determination	202
10.4.1.9. Protein storage	202
10.4.2. Colicin E9 production	203

10.4.3. Im9 production	203
10.4.3.1. Large scale over-expression	203
10.4.3.2. Liberation and purification	203
10.4.3.3. Purification, concentration and storage	203
10.5. Spin labelling	203
10.5.1. MTSL	203
10.5.2. IA-PROXYL	204
11. EPR and other spectroscopies	205
11.1. cw-EPR	205
11.1.1. X-band	205
11.1.1.1. Room temperature continuous wave	205
11.1.1.2. Low temperature continuous wave	205
11.1.2. S- and Q-band	206
11.1.2.1. S-band	206
11.1.2.2. Q-band	206
11.1.3. W-band	206
11.2. Pulsed EPR	207
11.2.1. Sample preparation	207
11.2.1.1. Protonated samples	207
11.2.1.2. Deuterated samples	207
11.2.1.3. pH shifted samples	207
11.2.1.4. Samples with glycerol	208
11.2.1.5. Sample mixing	208
11.2.1.6. Sample freezing	208
11.2.2. General experimental setup	208
11.2.3. T_1	209
11.2.4. T_2	209
11.2.5. FSE	209
11.2.6. ELDOR nutation	209
11.2.7. PELDOR	209
11.3. Other spectroscopy	210
11.3.1. NMR	210
11.3.1.1. Chemical shift perturbations	210
11.3.2. Ultra-violet / visible spectroscopy	210
11.3.3. Circular dichroism	211
11.3.4. Fluorometry	211
12. Computational	213
12.1. EPR toolbox	213
12.1.1. cwViewer	213
12.1.2. EasyRefiner	214
12.1.3. PowerSat	215
12.2. NACCESS	216
12.2.1. Define spin labels	216
12.3. MMM	217
12.3.1. Selenomethionine	217
12.3.2. PDB file write function	218
IV. Appendices	219
A. Useful references	221
B. Closed cycle low temperature EPR	223
C. MISHAP manual	265
D. Additional experimental figures	315
D.1. PELDOR setup experiments	315
D.2. PELDOR experiments	317
D.2.1. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	317
D.2.2. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1}	318
D.2.3. Im9 S6C ^{IA1} /C23 ^{IA1}	319
D.2.4. Colicin E9 DNase S3C ^{IA1} - Im9 C23 ^{IA1}	319
D.2.5. Colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23S	320
D.2.6. Colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	320

D.2.7. Colicin E9 DNase S30C ^{R1A} - Im9 C23 ^{R1A}	321
D.2.8. Colicin E9 DNase E66C ^{IA1} - Im9 C23 ^{IA1}	322
D.2.9. Colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23S	322
D.2.10. Colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	323
D.2.11. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	323
D.2.12. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	324
D.2.13. Colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	324
D.2.14. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	325
D.2.15. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	325
D.2.16. Colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	326
E. Consulted materials	327
F. Personal	333

List of Figures

1.1. Labelling mechanisms of common nitroxides to cysteine's S_γ	7
1.2. An energy diagram representing electron Zeeman splitting in the presence of a magnetic field. The stronger the magnetic field the greater the degree of separation	10
1.3. Definition of angles in a spherical co-ordinate system used in equations 1.5 and 1.6, when z is defined as the external magnetic field, B_0 , θ is from the z -axis and ϕ the projection in the xy plane	10
1.4. The hyperfine interaction created between a paramagnetic electron and non-zero spin nucleus	12
1.5. Coupling between two magnetic moments, $\vec{\mu}_1$ and $\vec{\mu}_2$ in a general orientation to the magnetic field with angles, θ_1 , θ_2 and ϕ	15
1.6. Coupling between two isotropic magnetic moments, μ_1 and μ_2 in a single orientation to the magnetic field with angle, θ	16
1.7. Energy level diagram of electron Zeeman, nuclear Zeeman and ^{14}N hyperfine interactions in a continuous wave EPR spectrum of a nitroxide radical	18
1.8. <i>Angular momentum variations shown as a result of velocity and torque.</i>	23
1.9. The effective microwave magnetic field in the rotating frame	26
1.10. The effect of varying the pulse length in pulsed-EPR, these spectra show the same sample of perinaphthenyl radical. Shorter pulses maximise spectra resolution. Adapted from Elexsys E580 User's manual	27
1.11. Upon the application of the pulse, all magnetization lies along $-y$ (as shown left diagram). With time fanning out about xy occurs (right diagram, faster spin shown lighter), decreasing the magnetization	28
1.12. Echo detected EPR pulse sequence	29
1.13. Moving pulse 2 pulse ESEEM experiment pulse sequence	29
1.14. Inversion recovery experiment pulse sequence	30
1.15. FSE experiment pulse sequence	31
1.16. 4 pulse PELDOR pulse sequence	31
1.17. ELDOR pulse power optimisation sequence	34
1.18. ELDOR nutation experiment pulse sequence	34
2.1. Expression, entry and action of colicins from host cell to target cell	36
2.2. A schematic representation of the colicin E9 domains; the translocation T domain, receptor R domain and cytotoxic C domain as well as Im9, with included masses.	37
2.3. Dissociation constants for colicin E9 against other well studied protein-protein interactions	38
2.4. Dissociation constants comparing cognate and non-cognate colicin E9 interactions	39
2.5. Im9 bound to the E9 DNase domain, with Im9's binding helices highlighted	39
2.6. The binding event and recruitment of host cell proteins by colicin E9.	43
2.7. Colicin E9 DNase bound to dsDNA	45
2.8. Colicin E9 shown binding to DNA with key motifs and metal ions highlighted.	45
4.1. A typical large scale overexpression	53
4.2. A typical crude purification with a GdnHCl unfolding <i>in situ</i> step before the imidazole titration. Shown here colicin E9 DNase S3C.	54
4.3. A typical crude purification with a one step imidazole titration. Shown here colicin E9 DNase S3C/S80C.	55
4.4. A typical polishing purification using size exclusion chromatography. Shown here colicin E9 DNase S3C	56
4.5. A typical optical spectroscopy profile. Shown here colicin E9 DNase S3C. Blue shows the folded protein after dialysis. Red shows the unfolded protein taken directly after the crude purification. Folded protein traces before and after size exclusion are identical.	57
4.6. Typical MALDI mass spectrometry profiles for a colicin E9 DNase and immunity protein Im9.	58
4.7. Fluorescence enhancement as a result of addition of immunity protein Im9 to colicin E9 DNase	59
4.8. Growth curves of <i>E. coli</i> BL21 (DE3) cells in culture with the addition of colicin E9 DNase	60
4.9. Growth curves of <i>E. coli</i> BL21 (DE3) cells in culture with the addition of synthetic colicin E9 DNase.	60
4.10. Growth curves of <i>E. coli</i> BL21 (DE3) cells with <i>Col</i> plasmid in culture with the addition of colicin E9 DNase.	61
4.11. Growth curves of <i>E. coli</i> BL21 (DE3) cells with <i>Col</i> plasmid in culture with the addition of synthetic colicin E9 DNase.	61
5.1. cw-EPR results for colicin E9 DNase S3C ^{I A1} at S, X and Q - band	65
5.2. cw-EPR results for colicin E9 DNase S30C ^{R1A} at S, X, Q and W - band	65
5.3. cw-EPR results for colicin E9 DNase S3C ^{I A1} / S80C ^{I A1} at S, X and Q - band	66
5.4. cw-EPR results for colicin E9 DNase S30C ^{I A1} / S80C ^{I A1} at S, X and Q - band	66
5.5. cw-EPR results for colicin E9 DNase E66C ^{I A1} at S, X and Q - band	67

List of Figures

5.6. cw-EPR results for colicin E9 DNase D20C ^{IA1} / E66C ^{IA1} at S, X and Q - band	67
5.7. cw-EPR results for Im9 at S, X, Q and W - band	68
5.8. cw-EPR results for Im9 C23S / S6C ^{IA1} at S, X and Q - band	68
5.9. cw-EPR results for Im9 S6C ^{IA1} / C23 ^{IA1} at S, X and Q - band	69
5.10. Power saturation profile of colicin E9 DNase S3C ^{IA1}	75
5.11. Power saturation profiles of other labelling sites. Each shows the power saturation profiles of the sample in the presence of a nitrogen atmosphere (blue circles), atmospheric oxygen (red squares) and a nitrogen atmosphere with 1 mM NiEDDA (green triangles). Lines are coloured to each experiment and show the fit of data using equation 5.5 and PowerSat's Lorentzian derivative fitting mode. 100 μ M protein, 298 K.	77
6.1. Possible fitting methods of the T_2 relaxation time	81
6.2. T_2 temperature optimization of E9 DNase S3C in water.	83
6.3. A Fourier transformation of the 4 K experiment shown in 6.2.a	83
6.4. T_1 temperature optimization of E9 DNase S3C. a) echo decay traces showing key temperatures b) fitted T_1 times plotted against temperature c) fitted echo amplitude plotted against temperature. In b and c error bars show 95 % confidence bands of the fitting. 1 scan, 2 shots/point, 4000 points, 60496 μ s shot rep. time, 3439 G, 9.6642 GHz, 4.4 dB attenuation, 200 MHz video bandwidth, 63 dB video gain. 100 μ M protein in H ₂ O.	85
6.5. T_2 pH temperature optimization of E9 DNase S3C.	87
6.6. T_2 buffer deuteration optimization of E9 DNase S3C.	87
6.7. T_2 relaxation time determination of E9 DNase S3C ^{IA1} at 70 % deuteration of the buffer	89
6.8. T_2 glycerol concentration optimization of E9 DNase S3C	89
6.9. T_2 glycerol deuteration optimization of IA-PROXYL	91
6.10. T_2 glycerol deuteration optimization of E9 DNase S3C	91
6.11. Effect of T_1 relaxation by different sample cooling methods of E9 DNase S3C. Error bars show 95 % confidence bands of the fitting. 50 K, 2 scans, 4 shots/point, 1024 points, 50000 μ s shot rep. time, 3455 G, 9.6903 GHz, 20 MHz video bandwidth, 60 dB video gain. 100 μ M protein.	93
6.12. Effect of T_2 relaxation by different sample cooling methods of E9 DNase S3C. Error bars show 95 % confidence bands of the fitting. 50 K, 4 scans, 4 shots/point, 1024 points, 10000 μ s shot rep. time, 3455 G, 9.6903 GHz, 20 MHz video bandwidth, 51 dB video gain. 100 μ M protein.	93
6.13. T_2 $\pi/2$ pulse length optimization of E9 DNase S3C. a) echo decay traces b) fitted T_2 times plotted against $\pi/2$ pulse length c) fitted echo amplitude plotted against $\pi/2$ pulse length. In b and c error bars show 95 % confidence bands of the fitting. 50 K, 4 scans, 4 shots/point, 1024 points, 10000 μ s shot rep. time, 3454.5 G, 9.6846 GHz, 20 MHz video bandwidth, 48 dB video gain. 100 μ M protein.	94
7.1. Relaxation times of each labelled protein in the E9 DNase - Im9 bindind complex under optimised conditions from chapter 6	100
7.2. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	102
7.3. Rotamer comparison of colicin E9 DNase S3C ^{IA1} and S80C ^{IA1}	103
7.4. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1}	104
7.5. PELDOR experiment results for Im9 S6C ^{IA1} /C23C ^{IA1}	106
7.6. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 C23 ^{IA1}	108
7.7. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23S	110
7.8. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	112
7.9. PELDOR experiment results for colicin E9 DNase S30C ^{R1A} - Im9 C23 ^{R1A}	114
7.10. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 C23 ^{IA1}	116
7.11. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} / C23S	118
7.12. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} / C23 ^{IA1}	120
7.13. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	122
7.14. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	124
7.15. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	126
7.16. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	128
7.17. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	130
7.18. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} to Im9 S6C ^{IA1} /C23 ^{IA1}	132
7.19. Q-band PELDOR experiment results for colicin E9 DNase S30C ^{R1A} to Im9 C23 ^{R1A}	136
7.20. Comparison of X- and Q-band PELDOR experiment results for colicin E9 DNase S30C ^{R1A} to Im9 C23 ^{R1A}	137
7.21. A binding cross summary of distance distributions from all PELDOR experiments. Presented horizontally are the spin labelled colicin E9 mutants plotted against the spin labelled Im9 proteins. The top row and first column are blank as a result of the protein being run in isolation.	140
7.22. Graphical summary of PELDOR experiments with final distances superimposed onto the protein structure of PDB file, 2K5X.	140
8.1. Residues with significant chemical shift perturbations mapped onto the NMR structure of the colicin E9 DNase / Im9 complex (PDB ID: 2K5X), side chains are coloured by atom (oxygen red, nitrogen blue)	141
8.2. Flowchart representation of the HADDOCK modelling process, model restraints are applied at with each of the four energy optimisation steps	144
8.3. Visual overlays of various cognate partner strucutre files	146

8.4. Results of HADDOCK run using PDB entry 2K5X structures and NMR CSP data from ^1H - ^{15}N HSQC and triple resonance experiments.	148
8.5. Structures of spin labels IA-PROXYL and MTSL attached to a cysteine residue with bond angle motion highlighted.	149
8.6. Visualisation of MMM rotamers upon E9 DNase S30C	150
8.7. Graphical representation of the distance distribution fitting mechanism employed. a) defines the common terms in a Gaussian distribution, with the centre point (d), the full width at half height, and the first two standard deviations of data in red and green, representing 68 % and 95 % of the data. b) shows the symmetrical fitting of the single, largest peak to a Gaussian distribution in colicin E9 DNase S30C ^{R1A} - Im9 C23 ^{R1A} , with the black lines highlighting the first standard deviation. c) shows the symmetrical fitting in red and asymmetric fit in green of colicin E9 DNase S30C ^{IA1} - Im9 C23 ^{IA1} . d) shows the multiple peak fitting using asymmetrical fitting upon colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} , the first standard deviation for each peak is shown in the peak colour, whilst the overall d position is shown as a vertical black line with dplus and dminus highlighted as black, dashed arrows.	152
8.8. MISHAP PDB Creator user interface with 2K5X parameters loaded	154
8.9. MISHAP distance constraint generator user interface	156
8.10. HADDOCK run results with no restraints, and only minimal, randomly generated AIRs	159
8.11. HADDOCK run with NMR AIRs restraints.	161
8.12. HADDOCK run with NMR AIRs restraints upon MISHAP generated proteins with all spin labels present.	162
8.13. HADDOCK run with NMR AIRs and colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} PELDOR restraint.	165
8.14. HADDOCK run with colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} PELDOR restraint.	165
8.15. HADDOCK run with Im9 S6C ^{IA1} /C23 ^{IA1} PELDOR restraint.	167
8.16. HADDOCK run with NMR AIRs and colicin E9 DNase S3C ^{IA1} -Im9 S6C ^{IA1} PELDOR restraint	167
8.17. HADDOCK run with NMR AIRs and colicin E9 DNase E66C ^{IA1} -Im9 C23 ^{IA1} PELDOR restraint	168
8.18. HADDOCK run with colicin E9 DNase E66C ^{IA1} -Im9 S6C ^{IA1} PELDOR restraint	168
8.19. HADDOCK run with NMR AIRs and colicin E9 DNase S3C ^{IA1} /S30C ^{IA1} /E66C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1} PELDOR restraint	170
8.20. HADDOCK run with NMR AIRs and colicin E9 DNase S3C ^{IA1} /S30C ^{IA1} /E66C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} PELDOR restraint	171
8.21. HADDOCK run with AIRs and colicin E9 DNase S3C ^{IA1} /S30C ^{IA1} /E66C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1} PELDOR restraints	172
8.22. HADDOCK run without AIRs, only colicin E9 DNase S3C ^{IA1} /S30C ^{IA1} /E66C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1} PELDOR restraints	174
8.23. HADDOCK run with NMR AIRs and colicin E9 DNase S3C ^{IA1} -Im9 C23 ^{IA1} MMM restraints	176
8.24. HADDOCK run with NMR AIRs and MMM derived distance restraints between every spin label	177
9.1. Comparison of colicin E9 DNase to the homology model colicin E3	189
9.2. Colicin E9 DNase and Im9 superimposed onto the homology model colicin E3 structure	189
12.1. Screenshot taken from cwViewer, showing the room temperature X-band spectrum of colicin E9 DNase S30C ^{R1A}	213
12.2. Screenshot taken from EasyRefiner, showing a file with default nitroxide system parameters loaded	214
12.3. Screenshot taken from EasyRefiner, showing fitting of a nitroxide system	214
12.4. Screenshot taken from PowerSat, showing a Gaussian derivative fitting method	215
D.1. Initial parameter finding experiments for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	316
D.2. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1}	317
D.3. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1}	318
D.4. The resulting distance distributions for the colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} PELDOR experiment when the background subtraction time is moved.	318
D.5. PELDOR experiment results for Im9 S6C ^{IA1} /C23 ^{IA1}	319
D.6. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 C23 ^{IA1}	319
D.7. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23S	320
D.8. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	320
D.9. The resulting distance distributions for the colicin E9 DNase S3C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1} PELDOR experiment when the background subtraction time is moved.	321
D.10. PELDOR experiment results for colicin E9 DNase S30C ^{R1A} - Im9 C23 ^{R1A}	321
D.11. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 C23 ^{IA1}	322
D.12. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23S	322
D.13. PELDOR experiment results for colicin E9 DNase E66C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	323
D.14. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	323
D.15. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	324
D.16. PELDOR experiment results for colicin E9 DNase S3C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	324
D.17. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 C23 ^{IA1}	325
D.18. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23S	325
D.19. PELDOR experiment results for colicin E9 DNase S30C ^{IA1} /S80C ^{IA1} - Im9 S6C ^{IA1} /C23 ^{IA1}	326

Nomenclature and abbreviations

AIR	ambiguous interaction restraint
CAPRI	critical assessment of prediction of interactions
CNS	crystallography and NMR system
CSP	chemical shift perturbation
CV	column volume equivalents
cw	continuous wave
DEER	double electron-electron resonance
DNA	deoxyribose nucleic acid
DPPH	2,2-diphenyl-1-picrylhydrazyl
dsDNA	double stranded DNA
DTT	dithiothreitol
EDDA	ethylenediaminediacetic acid
EI	exchange interaction
ELDOR	electron-electron double resonance
ENDOR	electron-nuclear double resonance
EPR	electron paramagnetic resonance
ESE	electron spin echo
ESEEM	electron spin echo envelope modulation
ESR	electron spin resonance
EZ	electron Zeeman
FID	Fourier induced decay
FID	free induced decay
FSE	field swept ESE
FT	Fourier transform(ation)
FWHM	full width at half height
GdmCl	guanidinium chloride
HADDOCK	high ambiguity driven biomolecular docking
HF	hyperfine interaction
HSQC	heteronuclear single-quantum correlation
HYSCORE	hyperfine sub-level correlation
IA-PROXYL	3-(3-(2-iodo-acetamido)-propyl-carbamoyl)-PROXYL
IA1	PDB residue code for cysteine attached IA-PROXYL
IMAC	immobilised metal ion affinity chromatography
LCR	ligase chain reaction
MALDI-MS	matrix assisted laser desorption ionisation mass spectrometry

List of Figures

MISHAP	MMM in silico simulated spin label to HADDOCK program
MMM	multiscale modelling of macromolecular systems
MS	mass spectrometry
MSE	selenomethioine
MTSL	S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate
NiEDDA	nickel (II) ethylenediaminediacetic acid
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NQ	nuclear quadrupole
PCR	polymerase chain reaction
PDB	protein data bank
PELDOR	pulsed electron-electron double resonance
PRE	paramagnetic relaxation enhancement
PROXYL	2,2,5,5-tetramethylpyrrolidinyl-1-oxy
R1A	PDB residue code for cysteine attached MTSL
RMSD	root mean squared deviation
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SDSL	site directed spin labelling
SNR	signal-to-noise
ssDNA	single stranded DNA
T1	spin lattice relaxation time
T2	transverse relaxation time
TEMPO	2,2,6,6-tetramethylpiperidinyl-1-oxy
UEA	University of East Anglia
UV	ultra-violet
VdW	Van der Waal's
ZFS	zero field splitting

Abstract

Atomic resolution structures of protein-protein interactions were initially limited to X-ray crystallography; with unnatural buffers, pHs and forced lattice structures. NMR extended investigations to be able to determine solution based structures of small complexes, but proteins of any size or high dynamicity remain problematic, whilst *ab initio* investigation remains almost impossible.

Using site-directed spin labelling (SDSL) across both docking partners, in combination with the electron paramagnetic resonance (EPR) technique pulsed electron-electron double resonance (PELDOR) allows for the precise measurement of intermolecular distances. Producing many mutant proteins with spin label sites across the protein allows for a 3-dimensional map to be built. Each EPR distance measurements is incorporated as a distance restraint into existing docking protocols with an aim to provide better quality docking models.

The method outlined in this investigation presents EPR as a technique with immense potential. EPR only measures paramagnetic species, in this case introduced nitroxide spin labels, making it applicable to water-soluble and membrane proteins, without the protein size limitations as with NMR. Furthermore, incorporation with freeze-quench apparatus could yield future time-dependent modelling in a way unavailable to traditional biomolecular structural techniques.

Using the well-characterised system of the DNase domain of colicin E9 and its inhibitor, colicin E9 immunity protein, Im9, a combination of structural NMR models and EPR long range distance restraints were able to present a structural solution with comparable resolution to high resolution NMR or X-ray crystallography. Looking forward, this method could be extended to molecules that have previously been available to study such as large proteins, proteins with high dynamicity, membrane proteins in lipids or even to build a structure in combination with *in silico* techniques.

Part I.

Background

1. Fundamentals of electron paramagnetic resonance

1.1. Introduction to electron paramagnetic resonance

Zavoisky published the first articles on paramagnetic resonance in 1945, quickly followed by a further 2 papers in the following year [111, 113, 112, 114]. The first paper entitled "Spin-magnetic resonance in paramagnetics", described how hydrated cupric chloride in the presence of applied radiation and magnetic field gave resonance of a particular applied electromagnetic radiation / magnetic field ratio. From these results it was later possible to determine that the spin quantum number of cupric chloride as $S = 1/2$. Only one year later did the initial work on nuclear magnetic resonance (NMR) begin to appear in the literature.

The nuclei of atoms are made up of protons and neutrons. A proton in elementary particles is composed of up, up, down quarks strongly held together by gluon force, giving an overall positive charge. The slightly heavier neutron consisting of up, down, down quarks yields a positive core $\sim 0.3 \text{ fm}$ in diameter negated by a negatively charged radius of 2 fm . As a result of this quark build up, each nucleon has an intrinsic quantum property of spin. With NMR it is clear that the nuclear moment is a result of the total nuclear spin.

Electrons, however, are negatively charged subatomic point particles [21], which is to say are zero dimensional in any relevant context [24]. Despite this an electron still has mass ($\sim 1/1836$ that of a proton), energy, spin ($S = 1/2$) and electrical charge. Spin is a unique degree of freedom fundamental to all elementary particles and, in the case of electrons, it is the directional spin of the particle upon its own axis. As the electron is charged and is moving, it also has a magnetic moment, closely related to the angular momentum or spin of the particle due to the gyromagnetic ratio (ie. the faster a charged particle is spun the greater the magnetic field). This spontaneous magnetization of spinning objects is governed according to the Barnett effect [3].

In the case of the electronic magnetic moment, the orbital angular momentum of the charged electron moving through space contributes as well as the spin of the electron. As a result, the technique formerly known as electron spin resonance (ESR) has slowly been renamed to electron paramagnetic resonance (EPR).

From a methodological view the main difference between NMR and EPR is the size of these magnetic moments. The electron moment is 660 times that of a proton, which lends EPR to a much higher sensitivity at a given spin concentration. The greater magnetic moment and intrinsic time sensitivity allows for the observation of spin-spin interactions over much larger distances and to molecular motions on a much shorter time scale.

From an applied perspective almost all molecules contain magnetic nuclei and are thus accessible to NMR spectroscopy, whilst very few molecules contain unpaired electrons and

1. Fundamentals of electron paramagnetic resonance

are thus inaccessible to EPR spectroscopy. In fact biologically, nature goes to extraordinary lengths to stop the natural occurrence of unpaired electrons as these often result in highly reactive species. A naturally occurring EPR signal is thus often related to an area of enhanced reactivity such as transition metal complexes in enzyme cores or free radicals.

1.2. Site directed spin labelling

In the absence of paramagnetism, as is common in biological systems, an artificial source is required. Paramagnetic spin probes have been routinely used in NMR paramagnetic relaxation enhancement (PRE) experiments since the 1950s [85]. Commonly metal ions have been used, but for these paramagnetic sources to be of any use they must be in an oxidation state that is inert with respect to reaction with water and amino acids. Consequently, examples using Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Cu^{2+} , and most of the lanthanide series ions appear in the literature [70]. Historically, paramagnetic metalloproteins kick-started the development of protein NMR spectroscopy as the paramagnetism shifted protons in such a way that they could be resolved in (what is now considered) low fields.

Many new EPR techniques developed as researchers became more aware that a paramagnetic centre could yield information as to the atoms that surround it, harnessing the considerable increase in spectral resolution over NMR. Particularly since the development of pulsed and high field/high frequency spectrometers [89]. Whilst some estimate up to half of proteins contain a metal [88], most of these will be diamagnetic and of those that are paramagnetic, still fewer will have the paramagnetic centre near an area of interest. For these reasons it has long been commonplace to introduce paramagnetic labels into proteins for study by EPR.

The use of metal based labels is becoming more prevalent, particularly in high field applications [79], but for the most part, by far and away the most common paramagnetic labels introduced to protein systems utilise nitroxide radicals. Pioneered by the McConnell group in the early 1960s [86], the nitroxide group of compounds consists of a radical delocalised across a nitrogen-oxygen bond. Often to stabilise the radical, to temperatures over 400 K, the immediate neighbouring carbons have branching methyl groups. The nitroxide radicals do however remain susceptible to low pHs and oxidation to hydroxylamines by weak reducing agents such as ascorbic acid. The most common of these compounds; IA-PROXYL, MTSL, TOAC and TEMPO are now commercially available at moderate cost and can be obtained in various isotopically enriched forms. Typically, nitroxides have the nitrogen of the $NO\bullet$ cyclised into a 5 or 6 membered ring with the ring tethered to a sulphate or other leaving group. Through the use of sulphur chemistry a disulphide or thioester bond can be formed to cysteine sulphurs, as demonstrated in figure 1.1.

The natural abundance of cysteine in proteins is only 1.2 % (though it constitutes 5.6 % of catalytic residues) [4]. The chance that a cysteine is present in any given protein, at a site that could be spin labelled and at a site of interest, is low. Therefore, genetic engineering is used and, through modification of only a few DNA bases, cysteines can be introduced at desired positions in the protein. These provide anchoring points to which extrinsic spin labels can be attached, the combined process is known as site directed spin labelling (SDSL).

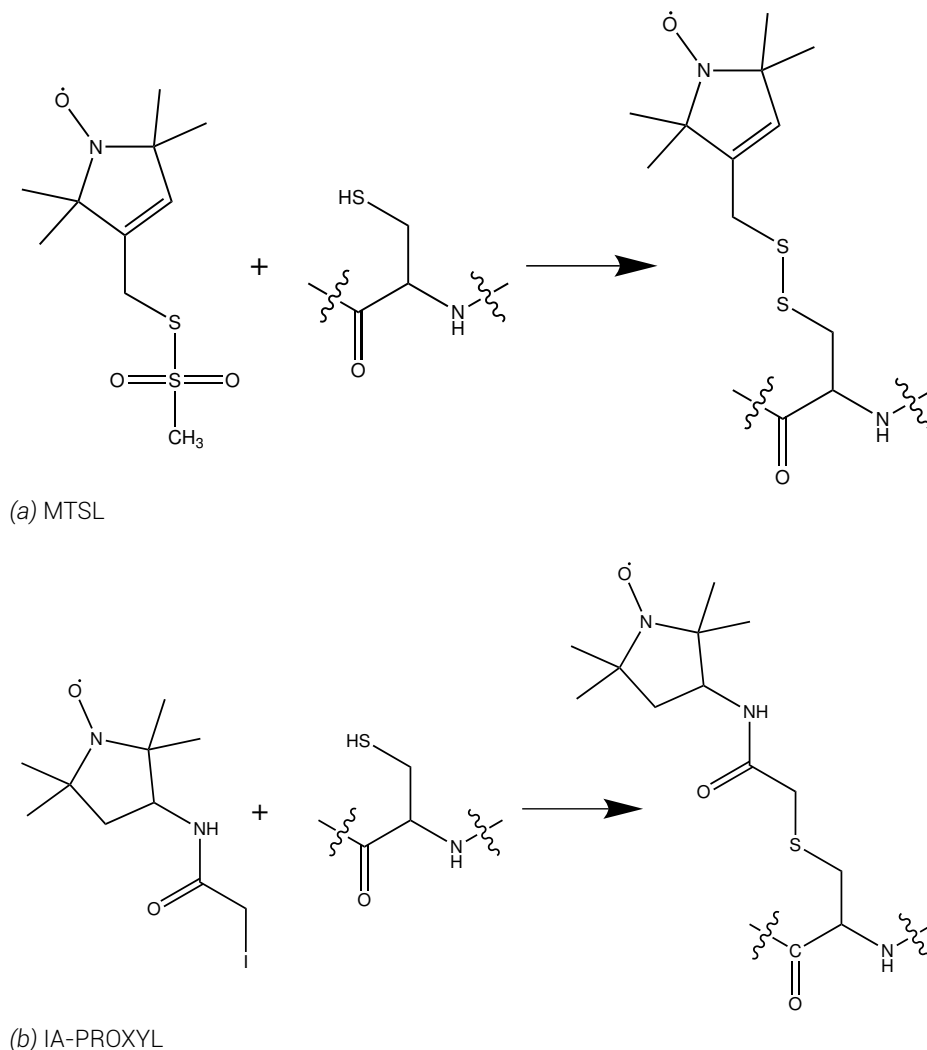


Figure 1.1 Labelling mechanisms of common nitroxides to cysteine's S_γ

In the case of poly-cysteine proteins it is common to first mutate all cysteines to alanines to present a simple model and then systematically mutate back the cysteines. Care needs to be taken when mutating away naturally occurring cysteines, as they often form disulphide bonds which can be key to protein structure, or are used at catalytic sites in metal coordination. Equally, adding cysteines to proteins that previously had none requires care. Surface exposed cysteines can easily generate intermolecular disulphide bonds forming poly-protein chains. Mutating, sometimes previously unknown, key residues to binding or activity can render the mutant inactive and sometimes the mutation will produce much reduced yields or will not express at all. After mutations have been performed full biochemical analyses are required to confirm that fully formed and functional proteins are used in the investigation.

1.3. Basic EPR theory

EPR as a technique for studying materials with unpaired electrons has much of its fundamental theory within quantum physics. In Schrödinger's equation, electrons assume pre-defined electron orbitals with certain levels of energy. Electrons will always occupy the lowest possible energy state unless an external influence is applied. Once the low energy orbitals have been filled then electrons will enter the next lowest orbital, this was known and

1. Fundamentals of electron paramagnetic resonance

used from as early as 1920 (before quantum theory) by Bohr and others in the Aufbau principle [105] and later formalised into the Pauli Exclusion Principle [74] and Hund's rules.

Electrons strive to find the minimal orbiting energy, however as a result of the Pauli Exclusion Principle (as electrons are fermions and thus must obey Fermi-Dirac dynamics), electrons cannot have the same quantum number and will often occupy the same energy level with opposite spins. The resulting spin pair, with opposing spins means that the paramagnetism on each electron is negated leaving the atom diamagnetic.

An unpaired electron, however, remains paramagnetic, that is to say it is attracted to and influenced by a magnetic field.

An electron will occupy the lowest possible energy state, however, through the application of energy, the electron can be excited from the ground state to an excited state. This is most commonly achieved by irradiation with photons (light) of a particular frequency. With time, the energy will be lost as an emission of a photon and the electron will return back to the ground state. Absorption of visible light gives rise to colours and has been exploited in science in colorimetry, luminescence and fluorescence experiments amongst many others, where absorption/emission intensity, frequency or decay times are measured.

In an EPR experiment numerous different factors will affect the electron's magnetic properties which will in result in differentiation in the total energy of the system. The total energy of a spin system can be described by a Hamiltonian, \hat{H} , where total energy the formal term for the combined kinetic and potential energies of a system. The spin Hamiltonian formalism considers the energy of an electron to be determined by its spin only. For an electron in the absence of an applied photon energy the spin Hamiltonian is given as

$$\hat{H} = \hat{H}_{EZ} + \hat{H}_{HF} + \hat{H}_{NQ} + \hat{H}_{ZFS} + \hat{H}_{EI} \quad (1.1)$$

where the Hamiltonian is the sum of the electron Zeeman (EZ), hyperfine interaction (HF), nuclear quadrupole (NQ), zero field splitting (ZFS) and exchange interaction (EI). Each term in the spin Hamiltonian will be described in greater detail in the following sections.

1.3.1. Electron Zeeman

When considering the simplest system, that of a single electron in a hydrogenic orbital, the application of an external magnetic field splits the electron ground state. With intrinsic spin and charge, the electron also has a magnetic property known as a magnetic moment (turning force). The magnetic moment creates a directionality to the electron's magnetism. As a result, when the electron is placed within a magnetic field it can align with the field in either a parallel or antiparallel manner. Opposing the external magnetic field requires more energy than aligning parallel to the field, this creates the splitting of the ground state into two distinct energy levels. The energy difference between the two energy levels creates a Boltzmann distribution between the two states in the unperturbed system. The splitting energy between states is directly proportional to the size of the applied magnetic field.

With the application of energy, in the form of electromagnetic radiation, the Boltzmann distribution is distorted and greater spin density is found in the higher energy ground state. With time, electrons in the higher energy state will relax to the lower energy ground state by

emission of a photon. Whilst the Boltzmann distribution no longer applies, a difference in populations between energy states is required to maintain a signal.

EPR is a resonance technique, in which the transitions of electrons between split ground states are measured by the absorption of photons. Detection is made by disruption of a standing wave causing reflections of microwaves from the sample cavity.

The relationship between energy spin levels and photon frequency is stated by Planck's law. Whereby the change in energy, ΔE , is related to the frequency of the absorbed electromagnetic radiation

$$\Delta E = h\nu \quad (1.2)$$

where h is Planck's constant and ν is the frequency of the radiation. Many different spectroscopies utilise the same principle using different frequencies of radiation, for example; NMR uses 10^6 Hz (radio frequencies - corresponding to nuclear resonance), terahertz spectroscopy uses 10^{12} Hz, infra-red spectroscopy uses 10^{12-14} Hz (corresponding to covalent bonds), ultra-violet/visible spectroscopy uses 10^{15-16} Hz, to name but a few. EPR uses 10^9 Hz radiation, commonly referred to as microwaves. The most common EPR spectroscopy is conducted at X-band corresponding to approximately 9 GHz (3.3 cm wavelength). In EPR, unlike other spectroscopies, for experimental reasons it is more convenient to keep the frequency static and sweep the magnetic field through resonance positions.

An EPR spectrum is a graphical representation of the absorption of energy (electromagnetic radiation, in the microwave frequency range) against the strength of the magnetic field. This ultimately can give information that can elucidate the environment of the paramagnetic electron, including interactions to other paramagnetic centres and sample dynamics reported by the orbital angular momentum that are otherwise lost in the spin Hamiltonian formalism.

The largest influence to the electron in an EPR experiment is usually that of the electron interacting with the magnetic field, this is called the *Zeeman effect*. The electron has a magnetic moment, μ , which as discussed previous results in the lowest energy state to result when the electron polarity is aligned with the field.

The two energy levels are labelled according to the direction of their spin, m_s relative to the direction of the magnetic field. For a simple single electron, spin half system the parallel state is designated $m_s = -1/2$ and the antiparallel state is designated $m_s = +1/2$. From this the primary EPR equation can be deduced:

$$E = g_e \mu_B B_0 m_s = \pm 1/2 g_e \mu_B B_0 \quad (1.3)$$

where g_e is the electron g-factor, a proportionality constant (2.002319304). μ_B is the Bohr magneton ($\sim 9.27 \times 10^{-24}$ J.T⁻¹), the natural unit of an electronic magnetic moment, the moment that would be expected if an electron was a spinning charge according to classical physics (universal constant). B_0 is the external magnetic field. In figure 1.2 it can be seen that with no magnetic field the ground state is degenerate, but as the magnetic field increases the energy difference between the two spin projections increases.

Formally however, the magnetic field is a vector giving it directionality as well as a magnitude. When considering the magnetic field as a vector the spin Hamiltonian becomes:

1. Fundamentals of electron paramagnetic resonance

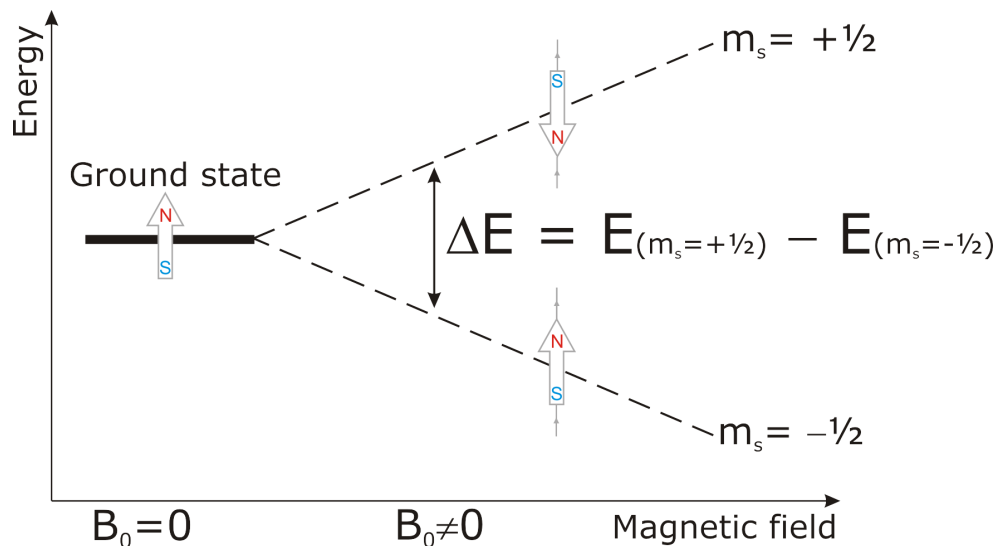


Figure 1.2 An energy diagram representing electron Zeeman splitting in the presence of a magnetic field. The stronger the magnetic field the greater the degree of separation

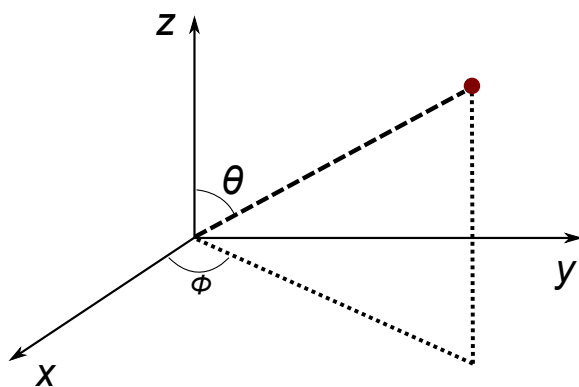


Figure 1.3 Definition of angles in a spherical co-ordinate system used in equations 1.5 and 1.6, when z is defined as the external magnetic field, B_0 , θ is from the z -axis and ϕ the projection in the xy plane

$$\hat{H} = \frac{\mu_B}{\hbar} \vec{B}_0 \underline{g} \hat{S} \quad (1.4)$$

where μ_B is the Bohr magneton, \hbar is the reduced Planck's constant ($\hbar/2\pi \approx 1.05 \times 10^{-34}$ J.s), \vec{B}_0 the transpose of the magnetic field vector $\vec{B}_0 = (B_{0x}, B_{0y}, B_{0z})$, \underline{g} the g tensor (not strictly a tensor as it connects two independent spaces; that of the spin and that of the laboratory) and \hat{S} the electron spin vector ($\hat{S}_x, \hat{S}_y, \hat{S}_z$).

When in the principal axis system of the g tensor the magnetic field can be written as:

$$\vec{B}_0 = B_0(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \quad (1.5)$$

where θ and ϕ are polar angles that characterise the direction of the magnetic field within the principal axis system (see figure 1.3). Combining equations 1.4 and 1.5 yields:

$$\hat{H}_{EZ} = \frac{\mu_B}{\hbar} B_0(g_x \sin \theta \cos \phi \hat{S}_x + g_y \sin \theta \sin \phi \hat{S}_y + g_z \cos \theta \hat{S}_z) \quad (1.6)$$

When the g tensor is anisotropic the direction of the effective magnetic field to the principal axis system is dependent upon the orientation of the molecule within the magnetic field. As the electron spin system is not necessarily parallel to the magnetic field and could have

different orientations, each with differently directed effective magnetic fields, the effective g tensor is given as:

$$g_{eff} = \sqrt{g_x^2 \sin^2 \theta \cos^2 \phi + g_y^2 \sin^2 \theta \sin^2 \phi + g_z^2 \cos^2 \theta} \quad (1.7)$$

At high magnetic fields the electron Zeeman interaction dominates over all other interactions and the contribution is proportional to the combination of the magnetic quantum number of the electron spin and the effective g tensor. The resonance condition for this orientation then becomes:

$$B_{0,res} = \frac{h\nu_{mw}}{g_{eff}\mu_B} \quad (1.8)$$

The effective g tensor will deviate slightly from that of the free electron g value, g_e , as a result of the orbital angular momentum which otherwise would not exist. Perturbation theory can resolve this deviation accurately but can be approximated as occupied electron orbitals making a negative difference (increase in field) whilst empty p -, d - or f - orbitals increase the g tensor (low field shift). The g tensor is therefore a complicated function that can act as a fingerprint to the electron orbitals of a paramagnetic centre since small nuclei elements will have a small spin-orbit coupling and the g tensor deviation from g_e will be small whilst first row transition metals with extended electron orbitals will show greater deviations from g_e .

1.3.2. Hyperfine interaction

Whilst the paramagnetic electron's g factor yields much information, it mainly reports only upon the electron, it speaks little of the environment surrounding the electron. Fortunately, the same paramagnetic electron is very sensitive to the immediate area. A large number of naturally abundant nuclei have a non-zero nuclear spin, resulting in their own magnetic momenta and thus magnetic fields. This nuclear magnetic field interacts with the electron magnetic field with the electron energy reflecting the interaction, an effect known as the hyperfine interaction.

The magnetic field resulting from the nucleus, B_1 , can either add to or oppose the external magnetic field which in turn raises or lowers the energy required to fulfil the resonance condition. In the case of a nucleus with spin $I = 1/2$, the single resonance line at B_0 is split in two representing the $B_0 - B_1$ and $B_0 + B_1$ conditions.

The hyperfine interaction term of the spin Hamiltonian for a $S = 1/2$ is given by

$$\hat{H}_{HF} = \sum_k \hat{S} \underline{A}_k \hat{I}_k \quad (1.9)$$

where k represents every magnetic nuclei ($I_k > 0$) in the molecule that has hyperfine couplings, \hat{I}_k are vector operators for each nuclei and \underline{A}_k are hyperfine tensors.

The hyperfine coupling of the electron spin to the nuclear spin comes about by a through-space dipole-dipole coupling of the two magnetic moments and by the Fermi contact interaction as a result of the non-zero probability that the electron is in the nucleus. The Fermi

1. Fundamentals of electron paramagnetic resonance

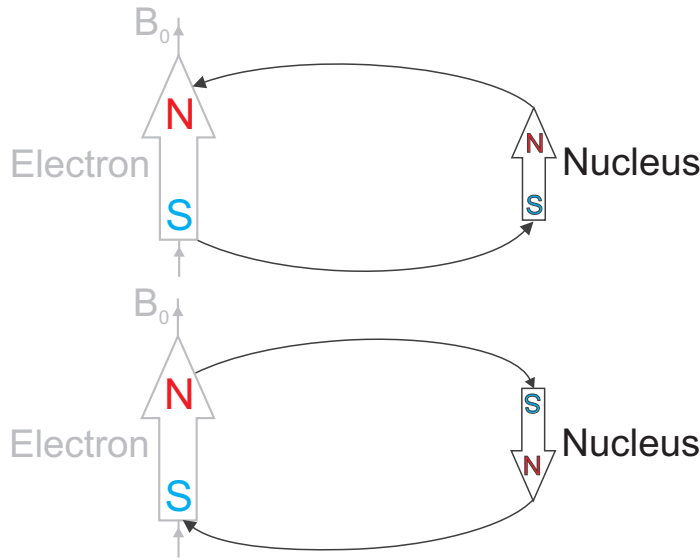


Figure 1.4 The magnetic moment of a $I \neq 0$ nucleus can also be thought of in classical physics as a small bar magnet. The resulting magnetic field, B_1 that can either add to or oppose the external magnetic field.

contact interaction can only occur when the electron is in a s -orbital and yields a purely isotropic coupling, $\hat{H}_{HF} = a_{iso} \hat{S} \hat{I}$ with

$$a_{iso} = \rho_s \cdot \frac{2}{3} \frac{2\mu_0}{\hbar} g_e \mu_B g_n \mu_n |\psi_0(0)|^2 \quad (1.10)$$

where ρ_s the spin density in the s -orbital, g_n the nuclear g value and μ_n the nuclear magneton ($\sim 5.05 \times 10^{-27} \text{ J.T}^{-1}$). $|\psi_0(0)|^2$ is the probability that an electron in the ground state (ψ_0) has a displacement of zero from the nucleus.

Paramagnetic electrons in p -, d - or f - orbitals do not have a Fermi contact interaction as these orbitals have a node at the nucleus and so the electron will never be found there. However, as these orbitals are non-spherical in shape and symmetry, couplings between these orbitals and nuclei do not average. The spin density in this case gives purely anisotropic hyperfine couplings. The total anisotropic hyperfine coupling tensor, \underline{T} of matrix, $T_{i,j}$ in this case can be calculated from the ground state wavefunction:

$$T_{i,j} = \frac{\mu_0}{4\pi\hbar} g_e \mu_B g_n \mu_n \left\langle \psi_0 \left| \frac{3r_i r_j - \partial_{ij} r^2}{r^5} \right| \psi_0 \right\rangle \quad (1.11)$$

In special cases, such as a proton or alkaline metal which have no significant spin density in p -, d - or f - orbitals the anisotropy can only occur from through-space dipole-dipole coupling to other spin density centres, though this is not significant to this investigation. For protons a good approximation is to assume only spin density at the central ion.

As with the electron Zeeman, orbital angular momentum has a small contribution to the dipole-dipole coupling in hyperfine coupling. As a result a small correction factor for the dipole-dipole hyperfine tensor is required

$$\underline{A}_{dd} = \frac{g \underline{T}}{g_e} \quad (1.12)$$

Yielding the final hyperfine tensor \underline{A}_k in equation 1.9 as

$$\underline{A}_k = a_{iso,k} \cdot 1 + \frac{g \underline{T}_k}{g_e} \quad (1.13)$$

1.3.3. Nuclear quadrupole interaction

With nuclei of spin $I_k > 1/2$ there are 2 or more centres of magnetism within the nucleus. Whilst each source will behave as a dipole and yield a dipolar field in isolation, the close proximity of the centres dampens the dipole moment and yields a quadrupole moment. Typically the nuclear quadrupole interaction does not contribute significantly to the EPR signal unless it is of the same order of the hyperfine coupling to the nucleus. Usually, this only occurs in heavy elements in the third row of the periodic table. However, in electron-nuclear double resonance (ENDOR), electron spin echo envelope modulation (ESEEM) and hyperfine sub-level correlation (HYSCORE) spectroscopy investigations the nuclear quadrupole effect needs to be considered. In EPR the nuclear quadrupole Hamiltonian for a single spin is written as

$$\hat{H}_{NQ} = \hat{I} \underline{P} \hat{I} \quad (1.14)$$

where the principle values of the nuclear quadrupole tensor are given as

$$P_z = \frac{e^2 q Q}{2I(2I-1)\hbar}, \quad (1.15)$$

$$P_x - P_y = \eta P_z, \quad (1.16)$$

and

$$P_x + P_y + P_z = 0. \quad (1.17)$$

In the high field approximation the nuclear Zeeman is much larger than the quadrupole interaction energies. This allows the evaluation of the nuclear quadrupole as the product of the square of nuclear spin quantum number and the effective nuclear quadrupole interaction.

1.3.4. Zero field splitting / fine structure

In the case of many electrons about a nucleus the electrons will distribute degenerately first singly occupying each orbit according to Hund's rule as electron pairing is a higher energy state. In the presence of a ligand field that removes orbital degeneracy to the extent that orbital separation is greater than spin pair repulsion, then spin pairing will be favoured. Depending upon the strength of the ligand field several electrons may be unpaired in high spin states or paired as far as possible in the low spin state.

In the case of transition metals all unpaired electrons stay mostly with the metal, the electrons remain so close in space and coupled such that they cannot be individually excited. In this case it is appropriate to consider the many electrons as a single collective group. For

1. Fundamentals of electron paramagnetic resonance

example Fe^{3+} with electron configuration $3d^5$, with weak ligands assumes a $S = 5/2$ high spin state (5 unpaired electrons, 3 low spin, 2 high spin), whilst in a strong ligand field with 4 electrons coupled has a low spin state of $S = 1/2$ (4 paired electrons, 1 unpaired, all low spin). Note that a transition metal does not have to be high spin to exhibit zero field splitting. In the case of a $3d^4$ system, a $S = 1$ state can occur from all electrons being low spin, with 2 electrons pairing and 2 remaining unpaired.

In the case of the high spin Fe^{3+} ($S = 5/2$) and other $S > 1/2$ cases, the d -orbitals are not symmetrical which creates a small distortion not entirely dissimilar to the nuclear quadrupole, a magnetic effect which would be present even in the absence of an external magnetic field. The spin Hamiltonian for zero field splitting is

$$\hat{H}_{ZFS} = \hat{S}^T \underline{D} \hat{S} \quad (1.18)$$

where \underline{D} is the zero field splitting tensor. In the principle axes frame of this tensor the Hamiltonian can be written as

$$\hat{H}_{ZFS} = D_x \hat{S}_x^2 + D_y \hat{S}_y^2 + D_z \hat{S}_z^2 \quad (1.19)$$

1.3.5. Exchange coupling

When two unpaired electron spins, S_1 and S_2 , of $S = 1/2$, such as two nitoxide radicals, are in close proximity up to approximately 15 \AA there is a significant overlap of the singly occupied molecular orbitals, which allows the electrons to exchange [48]. The exchange interaction allows for a splitting between the singlet ($S = 0$) and triplet ($S = 1$) states. If this splitting is small then transitions between singlet and triplet state can be excited in the EPR experiment, but it is often more convenient, especially at longer distances, to consider the system as two individual spins.

The exchange term of the spin Hamiltonian is given as

$$\hat{H}_{EI} = J \hat{S}_1 \hat{S}_2 = J(\hat{S}_{1x}\hat{S}_{2x} + \hat{S}_{1y}\hat{S}_{2y} + \hat{S}_{1z}\hat{S}_{2z}) \quad (1.20)$$

This term describes a purely isotropic coupling, whilst anisotropic contributions to the exchange coupling can occur, this only occurs for heavy elements and can be neglected for the purposes of this investigation.

The high field approximation of the energy contribution of the exchange interaction is proportional to the product of spin quantum numbers of both spins and J . However, a more practical approximation for nitroxide systems is that the exchange interaction decays exponentially with distance between the spins subject to the magnetic conductivity of the medium. In systems where J is significant, such as copper acetate, this approximation cannot be made.

1.3.6. Dipole-dipole coupling

In systems with large g anisotropy or where the sample is derived from a frozen solution, as is the case in biological systems, the magnetic coordinate frames are highly unlikely to be parallel to the external magnetic field. When 2 spins are considered then, neither will be aligned with the field or with each other. The energy between spins and the spin Hamiltonian describing that energy will therefore be a product of the angles between the spins, ϕ , the angle of each spin to the magnetic field, θ , and the distance between the spins. This interaction is shown graphically in figure 1.5 and mathematically derived as:

$$E = -\frac{\mu_0}{4\pi} \cdot \mu_1 \mu_2 \cdot \frac{1}{r^3} \cdot (2 \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \cos \phi) \quad (1.21)$$

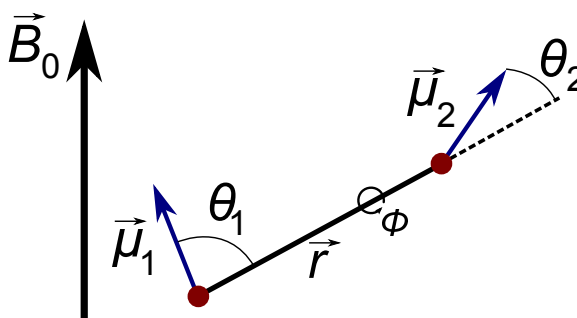


Figure 1.5 Coupling between two magnetic moments, $\vec{\mu}_1$ and $\vec{\mu}_2$ in a general orientation to the magnetic field with angles, θ_1 , θ_2 and ϕ

with the dipole-dipole coupling term in the Hamiltonian becoming:

$$\hat{H}_{dd} = \hat{S}_1 D \hat{S}_2 = \frac{1}{r^3} \cdot \frac{\mu_0}{4\pi\hbar} \cdot g_1 g_2 \mu_B \left[\hat{S}_1 \hat{S}_2 - \frac{3}{r^2} (\hat{S}_{1r}) (\hat{S}_{2r}) \right] \quad (1.22)$$

In the special case when anisotropy is so small that orientation can be neglected when calculating \hat{H}_{dd} , as in two organic radicals when $S = 1/2$ and $g_{xy} \approx g_{xz} \approx g_{yz}$, the dipole-dipole tensor \underline{D} has the principal values $-\omega_{dd}$, ω_{dd} and $2\omega_{dd}$ where

$$\omega_{dd} = \frac{1}{r^3} \cdot \frac{\mu_0}{4\pi\hbar} \cdot g_1 g_2 \mu_B^2 \quad (1.23)$$

Whereupon the orientation of the molecule can be characterized using a single angle, θ , between the magnetic field and spin-spin vector, shown in figure 1.6

$$d(\theta) = (3 \cos^2 \theta - 1) \omega_{dd} \quad (1.24)$$

When the g tensors for each spin are approximately equal to g_e an approximation of the coupling in frequency units is

$$\nu_{dd} = \frac{52.04 \text{ MHz}}{r^3 \text{ nm}} \quad (1.25)$$

when

$$\nu_{dd} = \frac{\omega_{dd}}{2\pi} \quad (1.26)$$

1. Fundamentals of electron paramagnetic resonance

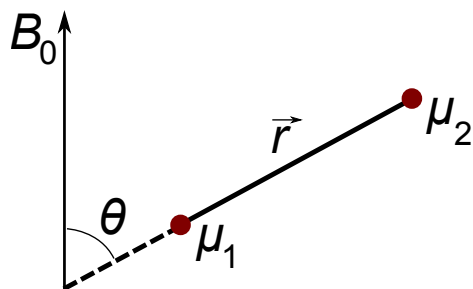


Figure 1.6 Coupling between two isotropic magnetic moments, μ_1 and μ_2 in a single orientation to the magnetic field with angle, θ

With typical transverse relaxation times in electrons being in the order of a few microseconds dipole-dipole couplings can be measured to around 100 kHz , corresponding to a distance of 8 nm [102, 103].

1.4. Continuous wave versus pulsed EPR

For a long time, due to the technological demands of such small time scales and difficult electromagnetic radiation frequencies, EPR had been a continuous wave (cw) spectroscopy. As with NMR the field has received substantial gains in sensitivity and shortened experiment time by moving from a continuous application of radiation to a pulsed Fourier transformation (FT) process. By measuring in the time domain and using multi-dimensional techniques more data than could ever have been previously imagined can be obtained. However, unlike in NMR, continuous wave EPR techniques still provide invaluable information and looks unlikely to be replaced.

In the most basic setup, pulsed-EPR involves using a short (nanosecond) and intense (hundreds of Watts) burst of microwaves to excite the sample and move the Boltzmann distribution from thermal equilibrium. After a brief delay, the subsequent relaxation back to the lower thermodynamic energy Boltzmann distributed ground state is recorded. This gives data in the time domain which can be Fourier transformed into the frequency domain.

The most common analogy for the differences between cw and pulsed EPR is that of tuning a bell. If given an unknown bell and assigned to finding the frequency spectrum of the bell, one approach would be to apply a low power frequency generator to the bell and slowly scan through all the frequencies and listening. Any sound detected is as a result of acoustic resonances in the bell; this is analogous with cw-EPR. This approach takes a long time to scan through the frequency range if a reasonable resolution is desired. The second approach would be to strike the bell with a hammer and a large amount of energy. The bell will immediately resonate to disperse the energy. The ringing noise is an overlay of all the acoustic resonances at multiple frequencies. Recording the decaying noise, digitizing the sound and performing a Fourier transformation gives the frequency spectrum of the bell. In one short experiment the whole spectrum can be acquired, this is called the multiplex advantage. This was not possible with EPR until the development of fast desktop computers as performing Fourier transformations of complex signals is computationally difficult.

By extending the metaphor a few technical problems become apparent. The detector (in this case a microphone) cannot record resonances at the point in time when the hammer

strikes the bell, otherwise all that is heard is the noise of the application of the hammer and not the bell. The hammer noise will also be much louder (single event energy transfer) than the desired resonances to be recorded, typically this would overload a detector optimised for listening to the resonances. Secondly, because frequencies are being measured the collection time needs to be as long as possible, however, with time each oscillation is dampened and amplitude is lost. By increasing the experiment lifetime the number of observed oscillations is increased, which reduces uncertainty in the wavelength, but this is not compatible with fast relaxing species. Finally, the experiment cannot be repeated until the bell has returned to equilibrium with its environment, something which is incompatible with extended lifetime experiments.

1.5. Continuous wave EPR

1.5.1. EPR characteristics of nitroxides

The unpaired electron of the nitroxide radical is located mainly in a π bond formed by the nitrogen and oxygen atoms' p_z orbitals. Some characteristic s orbital spin density can be seen from sp hybridisation of the N-O bond, but very little further delocalisation onto the carbon atoms is seen. The g tensor for nitroxides differs slightly from the g_e as a result of the spin orbit coupling of low lying excited states that have unquenched orbital angular momenta upon the oxygen atom. The greatest deviation is seen in the g_x , along the N-O bond, with some in g_y and almost none in g_z . By contrast the greatest hyperfine coupling to the ^{14}N is seen when the external magnetic field is aligned parallel with the p_z orbitals, whilst hyperfine coupling in molecular x and y directions is small. A typical nitroxide will therefore have tensors such $g_{xx} = 2.0090$, $g_{yy} = 2.0060$, $g_{zz} = 2.0030$, $A_{xx} = A_{yy} = 18 \text{ MHz}$ and $A_{zz} = 96 \text{ MHz}$. In a nitroxide with very fast motion, ie. free in solution or bound to a mobile location of a small protein then some hyperfine coupling to surrounding carbons can be seen but usually the motion is too slow and the carbons contribute only to line broadening.

The EPR spectrum of a nitroxide in a single orientation or undergoing fast rotation yields 3 equally sized lines. The electron Zeeman splitting in a magnetic field yields 2 energy levels. The ^{14}N nuclear Zeeman further splits each energy level into 3 ($I = 1$, where $2I + 1 = 3$). However, considering only electron and nuclear Zeeman contributions the allowed energy transitions are degenerate, as each is the same energy difference and would yield only a single line spectrum. The hyperfine coupling of the electron spin to the ^{14}N nuclear spin results, when spins are parallel (aligned), spin repulsion results in an increase in the energy level, as can be seen in figure 1.7. However, when spins are anti-parallel (of opposite sign) there is an energy drop. This creates three transitions, the $m_s = -1/2$, $m_I = +1$ to $m_s = +1/2$, $m_I = +1$ now has a larger difference in energy, the $m_s = -1/2$, $m_I = -1$ to $m_s = +1/2$, $m_I = -1$ is reduced in energy and the $m_s = -1/2$, $m_I = 0$ to $m_s = +1/2$, $m_I = 0$ remains unchanged. With a different energy for each of the three transitions, there are now three resonance conditions resulting in a three line spectrum. In analysis of the fast regime spectra, the centre line corresponds to isotropic g value ($g_{iso} = (g_{xx} + g_{yy} + g_{zz})/3$) with the two additional lines split according to the isotropic A tensor ($2A_{iso} = (A_{xx} + A_{yy} + A_{zz})/3$).

At the other extreme, the single crystal spectrum of a nitroxide shows the centre line at the g

1. Fundamentals of electron paramagnetic resonance

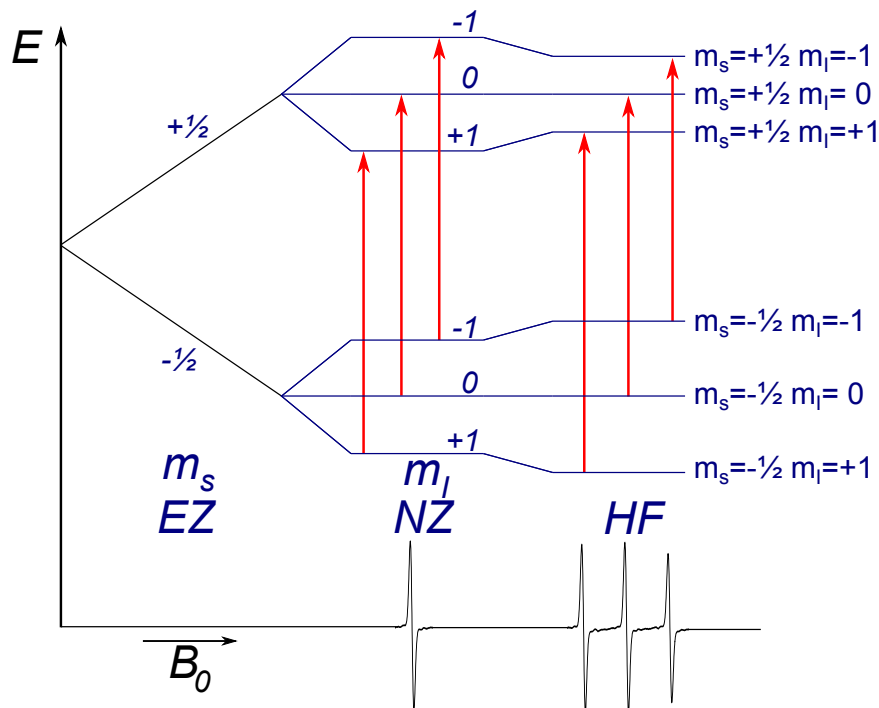


Figure 1.7 Energy level diagram of electron Zeeman, nuclear Zeeman and ^{14}N hyperfine interactions in a continuous wave EPR spectrum of a nitroxide radical

value positioned along the magnetic field direction in the molecular frame, with the splitting determined by the hyperfine coupling in this direction. In the case of a powder, or frozen solution, with multiple spins the resulting spectrum is the superposition of all orientations with proper weighting, such that there are 3 peaks corresponding to each g value and the 2 outer peaks are split by $2A_{zz}$.

1.5.2. Mobility

Rotational motion can best be thought of in nitroxides as a multi-stage process between different orientations where the exchange rate is related to the rotational diffusion time, R , through

$$R = \frac{1}{6\tau_r} \quad (1.27)$$

where τ_r is the rotational correlation time. In slow exchange the spectrum is broadened, but as the motion becomes faster, exchange between orientations results in averaging of the resonance frequencies resulting in an apparent reduction of $2A_{zz}$, visually seen experimentally as the 2 outer peaks moving closer to the central peak.

By analysing the line shapes of the spectra, information can be gathered about the mobility of nitroxide. In the fast regime, the widths of the three lines are identical as a result of averaging toward g_{iso} . In the slow regime a powder spectrum is obtained. However, in the intermediate regime the linewidths are different. The central line with $m_I = 0$ has the smallest linewidth as a result of having the smallest difference in resonant frequencies. The high field line, with $m_I = -1$, is the broadest as it has the largest range exchange interactions and therefore in resonant frequencies. Only at very fast motion ($> 10 \text{ ps}$) do the line shapes and widths become equal as other factors become the major contributors to line broadening. Using

Kivelson theory it is possible to calculate the relaxation time for the central line using analysis of line shapes and width [52].

1.5.3. Accessibility

The g and hyperfine tensors of the nitroxide depend upon the spin distribution of the electron across the bond and atoms. This distribution will depend upon the polarity of the environment. In a hypothetical polar environment the oxygen could exist in the charged O^- state and be stabilised by the polar environment whilst the electron could be located upon the positively charged nitrogen atom, $N^{\bullet+}$. In a non-polar environment, the charged polar states cannot be stabilised by the environment and thus the nitroxide atoms will preferentially sit in an uncharged state and so the electron will reside mostly upon the oxygen, O^\bullet , with an uncharged nitrogen.

In a polar environment with the electron contributing most spin density to the nitrogen, ρ_N , the isotropic hyperfine coupling from the ^{14}N increases with the increased spin density in the p_z orbital (A_{zz}). As the hyperfine tensor increases the g tensor decreases from g_e , as spin density has been withdrawn from the high energy excited states upon the oxygen, which further changes the g tensor as the energy levels between the lone electron and the lone pairs on the oxygen changes. Furthermore, hydrogen bonding to the lone pairs on the oxygen causes a small amount of delocalisation of the lone pair orbitals and an adjustment in the g_{xx} tensor.

By placing paramagnetic quenchers near a nitroxide spin label, Heisenberg exchange of unpaired electrons can occur between the nitroxide and the quencher. The quench rate will be directly proportional to the concentration of the quencher. A common laboratory quencher is diatomic oxygen that exists as a triplet in the ground state. Diatomic oxygen has two electrons in degenerate orbitals, which the exchange interaction splits into 1 singlet state ($S = 0$) and 3 triplet states ($S = 1$); Hund's rule states that the singlet state will be higher in energy than the triplet states, giving oxygen a non-zero spin in the ground state so it is paramagnetic. Diatomic oxygen is more soluble in apolar environments such as lipid bilayers than in polar environments such as water. Water soluble quenchers usually consist of transition metals stabilised in soluble ligands such as the uncharged $Ni(II)$ ethylenediaminediacetic acid (EDDA)(NiEDDA) or negatively charged $Cr(III)$ tris-oxalate ($[Cr(C_2O_4)_3]^{3-}$). At high concentrations the quenchers will cause line broadening through excessive exchange interaction. At lower concentrations the relaxation influence can be quantified using power saturation experiments.

In a saturation experiment the microwave power is increased where the amount of spins exciting and contributing to resonance increases in a linear fashion, yielding an increase in signal intensity. Approaching the saturation point, the amount of microwave energy being applied to the sample increasingly matches the rate of excitation with the rate of decay, yielding no net increase in resonance. Any further increase in microwave power begins to saturate the sample, whereby the rate of excitation exceeds the rate of decay to the ground state resulting in a loss of signal.

The rate of relaxation is purely through collisions of the quencher with the nitroxide. If the longitudinal relaxation time of the quencher is shorter than the lifetime of quencher/nitroxide complex then the exchange rate, W_{ex} , can be expressed as follows

1. Fundamentals of electron paramagnetic resonance

$$\Delta\left(\frac{1}{T_1}\right) = \Delta\left(\frac{1}{T_2}\right) = W_{ex} = k_{ex} C \quad (1.28)$$

where k_{ex} is the exchange rate constant and C is the quencher concentration. In the case that the quencher concentration is kept low the exchange rate can be assumed to be much less than $1/T_2$. W_{ex} can be determined from the saturation curves by measuring the first derivative peak-to-peak amplitude, A , of the central line of the nitroxide spectrum ($m_I = 0$) against the applied microwave power, P using

$$A(P) = \frac{I\sqrt{P}}{\left[\frac{(1+(2^{1/\epsilon}-1))P}{P_{1/2}}\right]^\epsilon} \quad (1.29)$$

with I the power independent amplitude factor, $P_{1/2}$ the power at half saturation and ϵ the inhomogeneity parameter, typically 1.5, but can be as low as 0.5 for inhomogeneous samples. More correctly, $P_{1/2}$ is the microwave power when the peak-to-peak amplitude is reduced to half of its unsaturated maximum. Given by

$$P_{1/2} = \frac{\sqrt[3]{4} - 1}{\gamma_e^2 \Lambda^2 T_1 T_2} \quad (1.30)$$

where $\Lambda = B_1/\sqrt{P}$ is the conversion efficiency of the spectrometer resonator. When quencher concentration is low, exchange broadening is small and T_2 is constant despite the presence of the quencher. Thus the change in $P_{1/2}$ can be simplified to

$$\Delta P_{1/2} = \frac{(\sqrt[3]{4} - 1) W_{ex}}{\gamma_e^2 \Lambda^2 T_2} \quad (1.31)$$

To remove the spectrometer dependency on Λ and sample dependency on T_2 , the $P_{1/2}$ is converted into a dimensionless accessibility parameter, Π , defined as:

$$\Pi = \frac{[\Delta P_{1/2}/\Delta B_{pp}]_{sample}}{[\Delta P_{1/2}/\Delta B_{pp}]_{reference}} = \alpha W_{ex} \quad (1.32)$$

where α is a constant derived experimentally by comparison of line broadening and saturation values, ΔB_{pp} is the change in peak to peak width (in the magnetic field), where the denominator is obtained by the measurement of a known standard such as a magnetically dilute diphenylpicrylhydrazyl (DPPH) solution in the same spectrometer.

The fundamentals of these experiments were extensively investigated by Altenbach of the Hubbell group on T4 lysozyme [1, 81]. Later work particularly in the implementation of power saturation experiments on transmembrane and membrane-interacting proteins has been conducted on colicin A by the Steinhoff group [72].

1.5.4. Multiple frequency EPR

The majority of nitroxide based EPR study remains at X-band (9-10 GHz, 0.3 T), however experiments conducted at high field/frequency such as at W-band (~94 GHz, 3.4 T) allows

the opportunity to offer new quantitative and qualitative data. The main transformations in line spaces occur around $1.1 T$, above this point the magnitude of Zeeman anisotropy dominates over the ^{14}N hyperfine coupling in the spin Hamiltonian. With further increases in field/frequency the g tensors are better resolved. At X-band, the nitroxide spectrum is dominated by axial hyperfine coupling and signal averaging from molecular motion. At W-band as the rhombic Zeeman term becomes dominant new information on protein structure and dynamics becomes accessible. At higher fields, the Larmor frequency is also higher which can be used to deconvolute label motion from labelled small proteins [106].

Developments in high field EPR were long hampered by technological issues, but recently has seen a dramatic increase in development thanks to improvements in quasi-optical bridges, digital electronics engineering and microwave technologies (thanks to mobile telecommunication industry demand). At the same time improvements in cooling such as cryogen-free superconducting magnets have significantly reduced the operating costs of such spectrometers.

One notable feature of nitroxide spin-label EPR is the ability to provide detailed information upon the complex rotational motion of the spin-labelled molecule. In a molecule that has been spin labelled, the paramagnetic centre is undergoing random motion and therefore is undergoing stochastic reorientation of the molecule relative to the external magnetic field. Thus, the polar θ and ϕ angles between the external magnetic field and molecular coordinate system are constantly changing in a time-dependent fashion. As a result, the spin Hamiltonian becomes time-dependent as the magnetic properties of the system vary such that $g = g(t)$ and $A = A(t)$. It is easiest to consider the time dependent spin Hamiltonian in isotropic, H_0 , and anisotropic, $H_1(t)$, parts, such that

$$H = H_0 + H_1(t) \quad (1.33)$$

Where the anisotropic part is given as

$$H_1(t) = \left[\mu_B \vec{B} \cdot [g(t) - g_{iso}] \cdot \hat{S} \right] + \left[h \hat{S} \cdot [A(t) - A_{iso}] \cdot \hat{I} \right] \quad (1.34)$$

where \hat{S} and \hat{I} are the electron and nuclear spin operators. The isotropic parameters are defined as

$$g_{iso} = \frac{g_{xx} + g_{yy} + g_{zz}}{3} \quad (1.35)$$

$$A_{iso} = \frac{A_{xx} + A_{yy} + A_{zz}}{3} \quad (1.36)$$

If the time dependency arises solely from molecular motion the isotropic part will not fluctuate with time but the anisotropic part will. Fluctuations in the anisotropic part will cause modulations in the energy levels and therefore the resonance frequencies, giving rise to a single EPR transition line occurring at multiple positions and generating line broadening.

Notice that the first term in the time-dependent Hamiltonian (equation 1.34) reflecting the electron spin is proportional to the magnetic field. Whilst the second term, reflecting the

1. Fundamentals of electron paramagnetic resonance

hyperfine interaction is field independent. By increasing the magnetic field we would therefore expect the contribution from the anisotropic Hamiltonian to increase causing protein motion to appear in the slow motion regime. The increased anisotropy also has the result of broadening the spectral width and enhancing sensitivity to molecular motion. Conversely, by decreasing the magnetic field the contribution from the time-dependent Hamiltonian is reduced, dropping the spectral width, resulting in the collapse of g tensors toward g_{iso} .

1.5.5. Line broadening

The first derivative absorbance spectrum of a nitroxide spin label at the rigid limit typically has a linewidth of 0.4 mT (11.2 MHz). A broadening of up to half this width from dipole-dipole interactions can be measured with confidence without distortion. Using equation 1.25, this corresponds to an upper detection limit of 21 \AA . In special circumstances when each of the spin labels are in near identical environmental conditions, and line broadening contributions are largely unresolved hyperfine couplings to methyl carbons, the limit can be slightly extended and with deuteration of the spin label distances up to 25 \AA can be resolved. In hydrogen bonding or polar environments, when the spin labels have different tensors, additional broadening of the signal occurs and the detection limit drops to approximately 16 \AA .

The interspin distance can only be deconvoluted when a high precision spectrum for each individual label as well as the combined spectrum exists. The doubly labelled spectrum is fitted by convolution of the sum of the singly labelled spectra. This method works best when both labelling sites are completely labelled though underlabelling can be compensated for this can be relaxed using a superposition of the dipolar broadened and unbroadened spectra, however it is less accurate and more computationally expensive. In extreme situations simulated spectra can be used in place of one of the singly labelled species.

1.6. Pulsed EPR theory

1.6.1. Magnetization in the laboratory frame

In order to describe electron position and motion accurately requires the use of an axis system. Most people are familiar with the Cartesian or laboratory frame that uses the standard x, y, z system of 3 perpendicular axes. The laboratory frame in EPR is described such that the magnetic field, B_0 is parallel to the z -axis (horizontal in the lab), the microwave magnetic field, B_1 is parallel to the x -axis (vertical in the laboratory), whilst the y -axis is orthogonal to the x and z -axis (also horizontal in the laboratory, but into space not the magnet).

According to classical physics, an electron having mass and spin has angular momentum (ie it would require energy to stop it), a product of its mass and velocity. When this electron is put into a magnetic field, a torque (turning force) is exerted which causes the electron to precess about the z -axis. Think of a wheel spinning with its axis tied to a piece of string. If there were no gravity then the wheel would be free to spin in any direction, with the inclusion of gravity the wheel hangs by the piece of string (axis pointing down, wheel spin horizontal)

aligned parallel with the field (gravity). By moving the axis by a few degrees the wheel can be seen to be rotating about the z -axis (this is always true, it is just that at very small angles this cannot be seen). This rotation about the z -axis is called precession, and is independent of the angle from z . The speed, or more correctly the frequency, of this precession is called the Larmor frequency and is related to the magnetic field by

$$\omega_L = -\gamma B_0 \quad (1.37)$$

Where ω_L is the Larmor frequency, γ is the proportionality constant known as the gyromagnetic ratio and B_0 is the magnetic field. A free electron has a gyromagnetic ratio of $\sim 2.8 \text{ MHz/G}$, which given a typical X-band experiment at 9.75 GHz , yields a magnetic field of 3480 G .

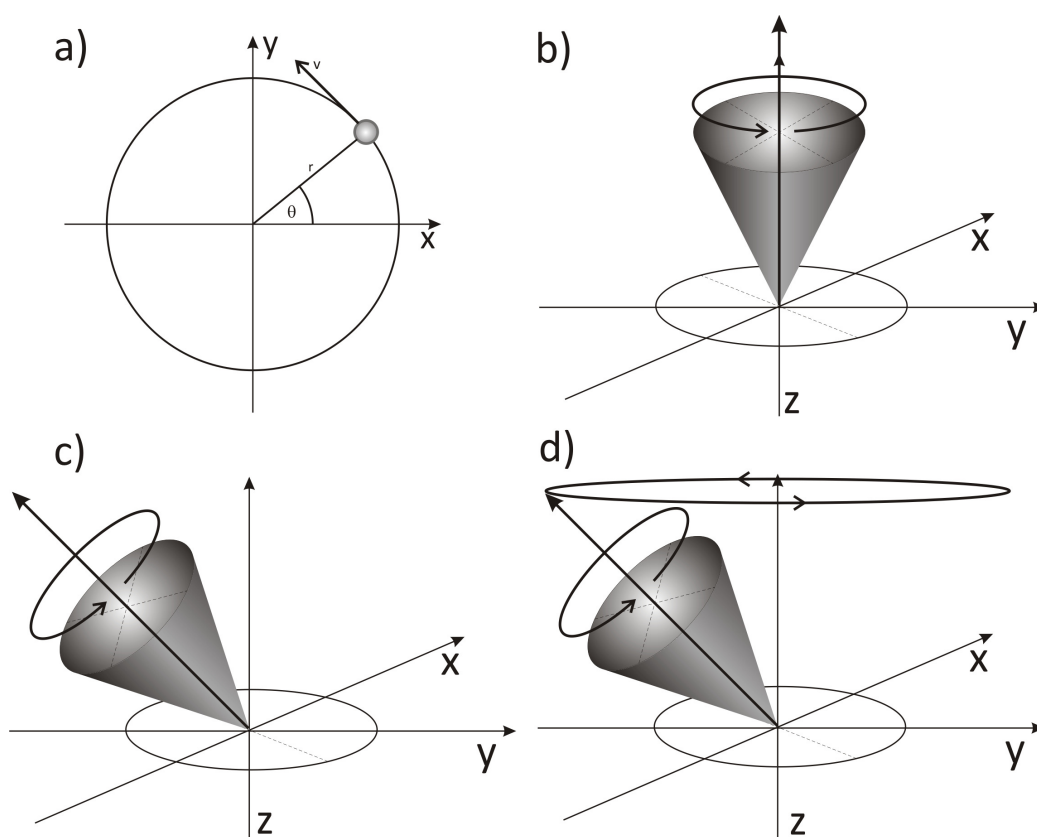


Figure 1.8 Diagram explaining the variations in angular momentum a) shows in 2 dimensions the effect of mass in a circular path; r is radius, θ , the angle from x , v the velocity b) shows that spin when parallel to the force of external magnetic field c) shows spin at an angle that is not along the z -axis d) shows that the torque (as of a result of the forces in c) cause a precession around the z -axis. The actual position of an electron can be anywhere around the edge of the top of the cone, it is simply more convenient to show this as a solid cone.

In any given sample there is not simply one molecule, there are several hundreds of thousands. In the classical model, an electron is treated as a single sphere which is not sufficient to explain all phenomena that arise from magnetic resonance. To consider the full situation a quantum model is needed in which the magnetic spins of the elementary particles can be considered to have 2 different spin axes pointing either up or down. In the absence of an external field these spins are identical in energy and are at thermal equilibrium with exactly half the population in either state. However, upon the application of a magnetic field

1. Fundamentals of electron paramagnetic resonance

the up energy state is aligned with the external magnetic field and as a result experiences a drop in energy. Alternatively, the particle can align against the external field (down) and rises in energy to an excited energy state. As this phenomenon is quantum in nature there are no possible intermediate states. The energy difference between the two states is directly proportional to the strength of the external magnetic field.

The number of electrons in either state follows Boltzmann distribution dynamics, but as the aligned, parallel state is lower in energy this will be the most populated state. The net magnetization, the sum of all the magnetic moments in the sample, will therefore be along the z -axis as the majority of electrons whilst randomly scattered about the xy plane will be precessing around the positive z -axis cancelling any net effect.

1.6.2. Magnetization in the rotating frame

Magnetization of a sample can result in incredibly complicated incurrent motion, with motions happening in three dimensions whilst rotating (think narrowing corkscrew). A useful technique to simplify the mathematics and to allow understanding is to switch from a static frame to a rotating one. The classic analogy of this effect involves a carousel and two people trying to have a conversation, whilst one is on the carousel and the other is not. In a static frame example, the observer is stationary and so meaningful conversation can only be held as only a few brief words exchanged per revolution. If the observer now walks at the same speed as the carousel is rotating then the two people can talk continuously as if they were not rotating at all. Of course this setup has more benefits, if the observer was trying to measure the riders speed (given they knew their own speed) or how often the carousel horse moved up and down this is much easier from a rotating frame (when the observers frame is approximately equal to the carousels precession and so appears static within the frame of reference).

EPR experiments use linearly polarized microwaves designed to produce a magnetic field, B_1 perpendicular to the applied magnetic field, B_0 . Linearly polarized electromagnetic radiation can be thought of in one of two ways. First it can be considered as an oscillating magnetic field (sine wave). Alternatively, and perhaps far more usefully, it can be thought of as two separate magnetic fields rotating in opposite directions to each other at the same rate. If both have a value of $1/2$ then when they both align they give a value of 1, 45° later in the sine wave the total magnetization value is $1/2$ as both fields have both dropped to $1/4$, at 90° through the sine wave each field is opposing and negates the magnetization to give a value of 0 and so on.

To simplify the mathematics of looking at a constantly moving coordinate system the coordinate system is moved at the same rate to create a rotating frame of reference when

$$\omega_L = \omega_O \tag{1.38}$$

where ω_O is the microwave frequency. By making the coordinate system rotate at the Larmor frequency, ω_L , one component of B_1 appears stationary (the other components now rotate at an apparent $2\omega_O$ and thus can be neglected). The rotating frame also makes magnetization components rotating at the Larmor frequency appear stationary.

Equation 1.37 showed that with a magnetic field the magnetization precesses around the direction of magnetic field. Thus we can extend this to B_1

$$\omega_1 = \gamma B_1 \quad (1.39)$$

Where ω_1 is also known as the Rabi frequency. If we assume that B_1 is parallel to the x -axis then the magnetic field will now rotate along the $+x$ axis for as long as microwaves are applied.

The angle with which the bulk magnetization, M_0 , is rotated (away from the z -axis) is equal to

$$\alpha = -\gamma|B_1|t_p \quad (1.40)$$

Where α is the angle from z , γ is the gyromagnetic ratio and t_p is the length of the pulse. Pulses are often labeled by their resulting tipping of the bulk magnetization vector, the most common of which are $\pi/2$ (90°) and π (180°). However, the tip angle of a pulse varies both with the length of the pulse and the magnitude of B_1 .

Note that when the bulk magnetization lies along the z -axis and a field, B_1 is applied along the x -axis, the resulting magnetization lies along the $-y$ axis (Fleming's left hand rule). Also, as B_1 is applied parallel to the x -axis it is known as a $+x$ pulse. By shifting the phase of the microwaves by 90° B_1 would lie along the $+y$ axis and the magnetization would lie upon the $+x$ axis, hence pulses are labelled by both their tip angle and which axis they lie parallel to.

1.6.3. Free induction decay and off resonances

It is important to remember that whilst in the rotating frame that after a pulse, the bulk magnetization lies along the $-y$ axis and is stationary, whilst in the laboratory frame this is actually rotating through the xy plane at the Larmor frequency. This rotation of charge creates currents and voltages in the resonator, using the same principles as a generator, and is called the free induction decay (FID)

Thus far we have only considered the conditions behind exact resonance; when the microwave frequency exactly matches the Larmor frequency. However, in any given sample there will be many different frequencies, of which not all can be tuned to be on-resonance at once, making it necessary to consider what happens to off-resonance signals.

Let us first consider the rotating frame of magnetization, with frequency ω following a 90° pulse. At the end of the pulse the magnetization will be aligned to the $-y$ axis. However, as the frequency does not equal the Larmor frequency the magnetization will appear in the xy plane, rotating faster or slower than the microwave magnetic field. The apparent rotation rate will be equal to the difference

$$\Delta\omega = \omega - \omega_0 \quad (1.41)$$

When $\Delta\omega = 0$ the rotation rate is zero appears stationary in the rotating frame, which is exactly what is expected for on resonance. If $\Delta\omega > 0$ the magnetization is gaining and will

1. Fundamentals of electron paramagnetic resonance

rotate in an anticlockwise direction, similarly if $\Delta\omega < 0$ the magnetization is losing and will rotate in a clockwise direction within the rotating frame.

This behaviour of rotating faster or slower than the reference frame gives detail toward how the free induced decay is encoded as each individual resonance frequency will appear as a magnetization component rotating in the xy plane according to $\Delta\omega$. By measuring the magnetization in the rotating frame and knowing the rotating frame frequency, the frequency components can be extracted and used to construct the EPR spectrum.

Earlier, it was stated that in an on resonance state that B_0 disappears. However, in off resonance conditions B_0 cannot be ignored and as a result the microwave magnetic field, B_1 , tips the magnetization into the xy plane differently.

$$\Delta\omega = -\gamma B_0 \quad (1.42)$$

Now the magnetization is not tipped by B_1 alone, instead by the vector sum of both B_0 and B_1 , the effective magnetic field, B_{eff} . As B_{eff} is the combination of two fields it follows that the precession frequency is the faster sum of precessions

$$\omega_{eff} = \sqrt{(\omega_1)^2 + (\Delta\omega)^2} \quad (1.43)$$

B_{eff} is a combination of vectors in the z and x -axis, resulting in the magnetization not lying in the xy plane as B_1 does but, instead, its motion is that of a cone as shown in figure 1.9.

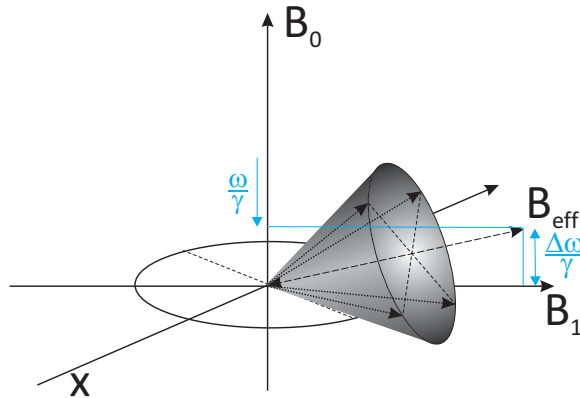


Figure 1.9 The effective microwave magnetic field in the rotating frame

The magnetization tipped toward the xy plane actually moves in an oscillating manner which decreases as the precession frequency increases (which is expected if we think back to the classical model, a spinning top that spins fast enough will always right itself). This oscillation is defined such:

$$M_{-y} = M_0 \times \frac{1}{\sqrt{1 + \left(\frac{\Delta\omega}{\omega_1}\right)^2}} \times \sin\left(\sqrt{1 + \left(\frac{\Delta\omega}{\omega_1}\right)^2} \times \frac{\pi}{2}\right) \quad (1.44)$$

In the case of a broad EPR spectrum from a wide range of frequencies ($\Delta\omega > \omega_1$) the signal is greatly diminished as the tip from the $\pi/2$ pulse will align fewer spins to be coherent in the xy plane and thus be detectable. For this reason it is important to maximise ω_1 for broad signals to maximise the detectable signal. This is experimentally achieved by using short

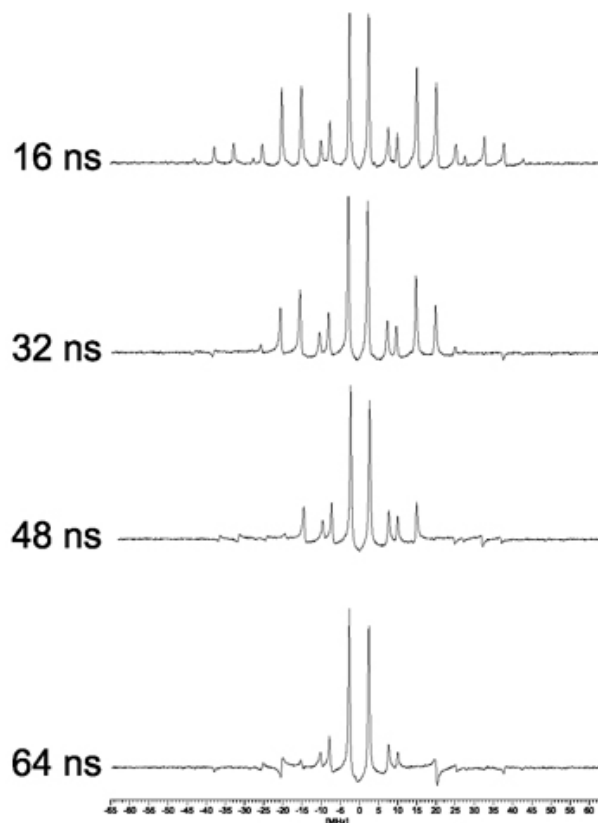


Figure 1.10 The effect of varying the pulse length in pulsed-EPR, these spectra show the same sample of perinaphthenyl radical. Shorter pulses maximise spectra resolution. Adapted from Elexsys E580 User's manual

pulse lengths (see figure 1.10) and a significant technological hurdle in the development of pulsed EPR techniques was the accurate generation of pulses of such small lengths.

1.7. Pulsed EPR applications

1.7.1. Relaxation times

Thus far it has been assumed that a $\pi/2$ pulse transfers the bulk magnetization from the z -axis into the xy plane. In reality the electrons are constantly interacting with their environment. To maintain magnetization in the xy plane is a higher energy state than to align with the laboratory magnetic field. As a result magnetization about the xy plane will decay and slowly move back toward the z -axis. This process, called relaxation, is governed by two constants. T_1 , the spin lattice relaxation time which describes how quickly the magnetization returns to the Boltzmann distribution about the z -axis and T_2 , the transverse relaxation time which describes how quickly the coherence in magnetization in the xy plane dissipates.

The result of the $\pi/2$ pulse is an immediate Fourier induced decay of the electron magnetization of frequency corresponding to the microwave pulse bandwidth. As almost all samples exhibit inhomogeneous broadening that extends beyond the pulse bandwidth a small number of spins are excited and relax very quickly.

1. Fundamentals of electron paramagnetic resonance

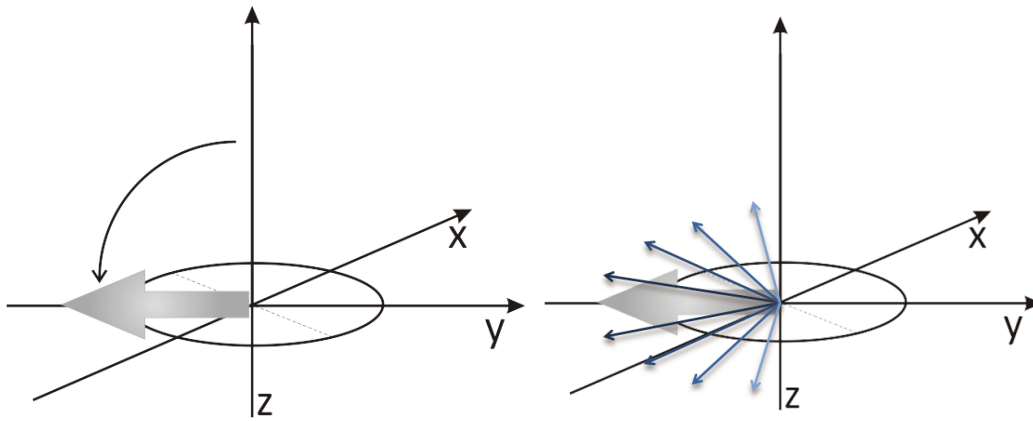


Figure 1.11 Upon the application of the pulse, all magnetization lies along $-y$ (as shown left diagram). With time fanning out about xy occurs (right diagram, faster spin shown lighter), decreasing the magnetization

1.7.1.1. Transverse relaxation

The transverse relaxation time, T_2 , corresponds to the decay of magnetization about the xy plane. Typically, this constitutes two types of broadening mechanism: homogeneous and inhomogeneous broadening.

Inhomogeneous broadening is a result of several unresolved couplings overlapping one another. Spins across a sample will experience a number of different magnetic fields arising from unresolved hyperfine interactions. The difference between each interaction is so small that each coupling overlaps and gives a Gaussian distribution of spins.

The Gaussian shape can be explained when considering that the local fields will produce a large number of electrons in the xy plane all precessing at various, non-zero frequencies. All electrons will only be approximately equal to the Larmor frequency, ω , causing a fanning of the net magnetization (as thermal equilibrium is sought); the Gaussian shape appears from the average frequency. Summing all the spin packets together results in a cancellation of terms (see figure 1.11). This characteristic time decay constant is called T_2^* (T two star)

Homogeneous broadening is a result of each spin packet experiencing the same magnetic field. The resulting spin packets interact with each other causing mutual or random flipping of one or both spins. This induced flip-flopping between spin states contributes to a faster fanning action (loss of coherence) via a normal general exponential Fourier induced decay (FID):

$$M_{-y}(t) = e^{-\frac{t}{T_2}} \quad (1.45)$$

Where T_2 is often called the spin-spin relaxation time. This form of line broadening results in Lorentzian shaped broadening as opposed to typical inhomogeneous broadening that gives Gaussian shaped broadening. This is a result of the Fourier transformation, where a quickly decaying signal in the time domain gives a broad peak in the frequency domain and slowly decaying signal in time gives a narrow frequency domain peak.

Typically an echo detected pulse sequence ($\pi/2-\tau-\pi-\tau$ -echo, figure 1.12) is used but a $2\pi/3-\tau-2\pi/3-\tau$ -echo can occasionally give better signal intensity. In this sequence the $\pi/2$ pulse excites coherence with amplitude $\cos \eta$ on the two allowed transitions and coherence with amplitude $\sin \eta$ on the two forbidden transitions. During the first time delay, τ , coherence

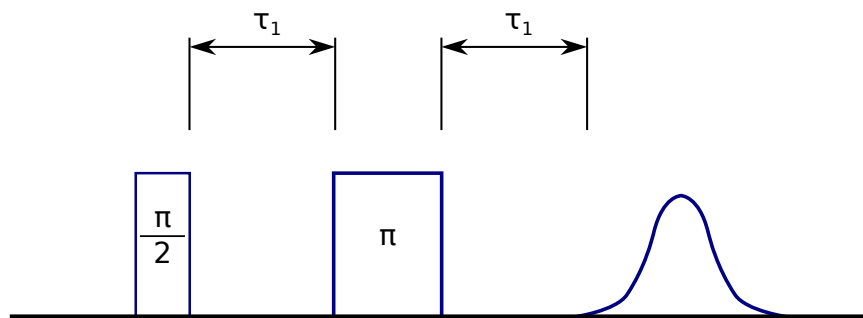


Figure 1.12 Echo detected EPR pulse sequence

evolves and decays with the transverse relaxation times T_{2a} or T_{2f} . The π pulse inverts the phase of all coherences and refocuses the magnetization such that after another time delay, τ , the spins converge to coherence to give an echo that smoothly decays with respect to the relaxation rate.

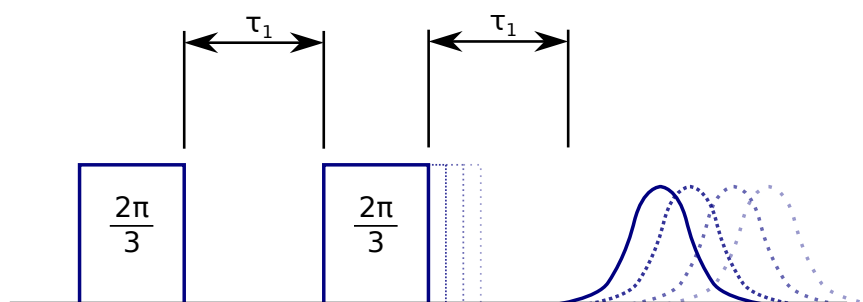


Figure 1.13 Moving pulse 2 pulse ESEEM experiment pulse sequence

To experimentally measure the transverse relaxation, a modified 2 pulse ESEEM pulse sequence, shown in figure 1.13, is used. Stepping the second pulse, as well as the detection point allows the decay to be recorded and analysed.

The π pulse changes the nuclear spin state with a probability of $\sin^2 \eta$ corresponding to a transfer in coherence between the different phase transitions. Each transferred coherence will experience a phase inversion and will oscillate with the difference of the transition frequency, allowing for the nuclear frequency to be seen in the free induced decay. A simple Fourier transformation of the decay trace will reveal the nuclear frequency of the atom on which the electron is located.

1.7.1.2. Inversion recovery

When electrons are placed inside a magnetic field, quantum mechanics allows for two states to be occupied; aligned parallel or anti-parallel to the field. The moments of the electrons will be distributed randomly with a slight majority in the lower energy, parallel to the magnetic field. The system obeys Boltzmann distribution statistics; whereby the system will find the lowest energy state of thermal equilibrium through a mix of enthalpy and entropy. The ratio of the two states can thus be defined as

$$\frac{n_{anti-parallel}}{n_{parallel}} = e^{-\frac{\Delta E}{kT}} \quad (1.46)$$

Where n represents the populations of the two states, ΔE is the energy difference between the two energy states, k the Boltzmann's constant and T the temperature (in Kelvin).

1. Fundamentals of electron paramagnetic resonance

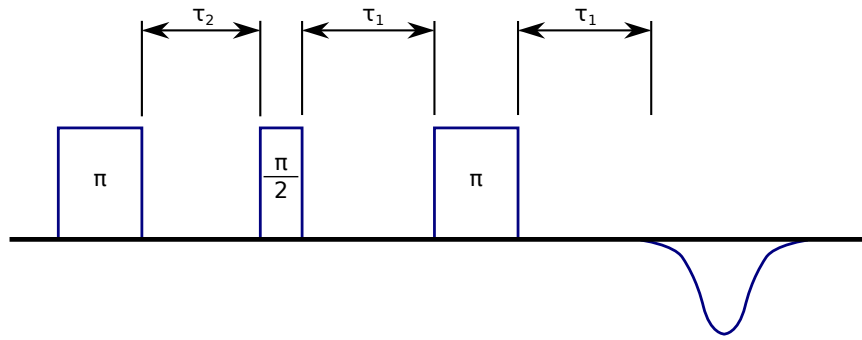


Figure 1.14 Inversion recovery experiment pulse sequence

The bulk magnetization so far discussed is actually the net vector of parallel and anti-parallel states. At equilibrium more electrons will be aligned parallel, the bulk or average magnetic moment of the sample is therefore parallel with the field. The magnetic moment is proportional to the populations of both states; with greater numbers of electrons aligned parallel yields a larger magnetic moment. Thus, a more concentrated sample (a sample with a higher number of total spins) provides more signal, so long as it remains magnetically dilute. However, achieving a larger magnetic moment requires distorting the Boltzmann distribution and requires the application of energy.

The difference in the physical number of electrons between states is not required to be very high as spectroscopy measures the transitions between states and not the population of the states. A typical X-band EPR experiment at room temperature (300 K) has 50.04% of the electrons aligned parallel and 49.96% anti-parallel, a difference of 1 in 125.

By applying a $\pi/2$ pulse, what actually occurs is that the populations in the two energy states (parallel and anti-parallel) are equalled giving no net magnetization in $+z$ or $-z$. However, the movement of electrons between states creates a magnetic moment in the xy plane, such that the bulk magnetization is considered to be in the xy plane. Using equation 1.46 we could find that $T = \infty$, which whilst counter-intuitive is thermodynamically allowed as the system is in a non-equilibrium state. Through interactions with the environment the system will eventually return to thermal equilibrium about the z axis through a process called the spin-lattice relaxation, T_1 .

By applying a π pulse the Boltzmann distribution is inverted with more electrons now lying anti-parallel (commonly referred to as an "inversion pulse"). Equation 1.46 in this instance reports a negative temperature, which is thermodynamically allowed but unfavourable which drives the system back to equilibrium. The rate constant at which M_z recovers to thermal equilibrium is the spin-lattice relaxation time, T_1 . The magnetization vector behaves for a π pulse according to:

$$M_z(t) = M_0 \times \left[1 - e^{-\frac{t}{T_1}} \right] \quad (1.47)$$

or for a $\pi/2$ pulse:

$$M_z(t) = M_0 \times \left[1 - 2e^{-\frac{t}{T_1}} \right] \quad (1.48)$$

A π pulse would generate an inverted Fourier induced decay as magnetization returns to equilibrium. Experimentally, the inversion recovery time from a single π pulse is very difficult

to record, due to pulse imperfections and a delay between pulse application and ability to turn on the receiver amplifier. To overcome this experimental restraint, a 2 pulse ESEEM pulse sequence is applied to the decaying signal as a reporter sequence, shown in figure 1.14. The reporter sequence allows the detection of a Hahn echo well away from the pulses so that the entire decay can be accurately recorded.

1.7.2. Field swept electron spin echo

By using the 2 pulse ESEEM sequence introduced in figure 1.12 and scanning through the magnetic field, a field dependent response of the sample can be detected. This is analogous to an absorbance scan in continuous wave EPR, but with detection being made using an electron spin echo (ESE). This technique is frequently referred to as a field swept ESE or field swept echo (FSE).

As with the 2 pulse ESEEM experiment for some samples, a $2\pi/3 - 2\pi/3$ pulse sequence can yield greater signal intensity than a $\pi/2 - \pi$.

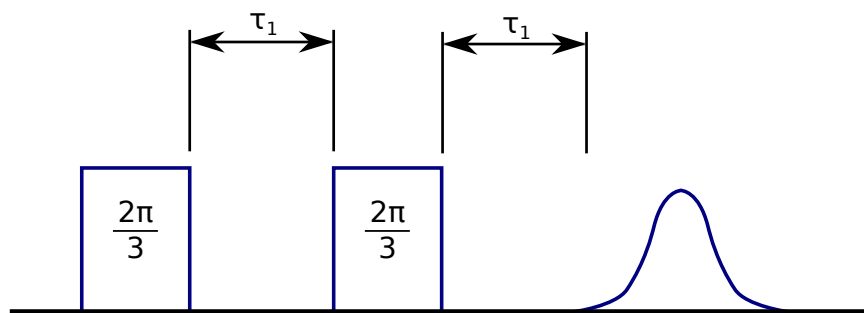


Figure 1.15 FSE experiment pulse sequence

1.7.3. Electron-electron double resonance

Pulsed electron-electron double resonance (PELDOR) also known as double electron-electron resonance (DEER) is a pulsed EPR technique for measuring the dipole-dipole interaction between spins across distances above 15 \AA . By utilising 2 different microwave frequencies the PELDOR experiment seeks to excite each spin individually. This is possible, even in the case of 2 nitroxides, as the excitation bandwidth of each microwave pulse is significantly smaller than the width of the nitroxide spectrum. Each frequency therefore can excite a different orientation or different spin state of the ^{14}N in each nitroxide, as the probability of both being in identical conditions is extremely low.

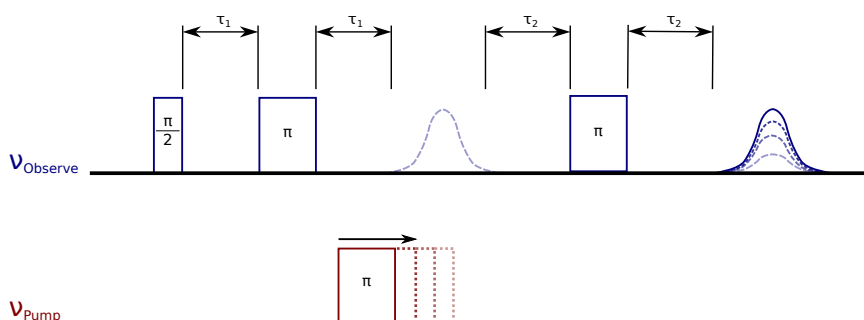


Figure 1.16 4 pulse PELDOR pulse sequence

1. Fundamentals of electron paramagnetic resonance

The PELDOR pulse sequence consists of several parts. One of the spins is selected as the observer spin and a 2 pulse ESEEM sequence is used to generate a reporter Hahn echo - $\pi/2-\tau_1-\pi-\tau_1$ -echo (where τ_1 is a delay time). To this a second part is added $\tau_2-\pi-\tau_2$ -echo, where after a second delay, τ_2 an additional π pulse is added to refocus the Hahn echo into a second echo. The observer π pulses excite only the subpopulation of A spins with linearly orientated operators (electron Zeeman and hyperfine interactions) as well as the dipole-dipole interaction between spin A and spin B.

The second microwave frequency applies only a single π pulse which exclusively excites B spins. The excitation of the B spins by the "pump pulse" inverts the state of the B spins and yields a change in the resonant frequency of the nearby A spins through the dipole-dipole interaction dependent on distance.

The time delays of the observer sequence (τ_1 and τ_2) are kept constant whilst the pump pulse is slowly stepped by variable t through the first Hahn echo and echo decay. As the pump pulse moves through the echo signal, the echo is modulated by the effect of B spin inversion with ω_{dd} as a function of time. Using ω_{dd} and equations 1.25 and 1.26, the distance between the spins can be calculated. The modulation itself is not dampened by relaxation as the total time of the experiment never changes. However, the number of observable modulations is dependent upon the detectable time of the echo, as is the maximum detectable distance. Thus the maximum observable time, t_{max} , is approximately equal to the T_2 relaxation time.

In classical physics an analogy can be made to two bar magnets aligned side by side so that their magnetic fields are coupled. If one of the bar magnets is moved with an external force we would expect the movement to be induced in the other. The degree of movement will be proportional to the strength of the coupling, in this case primarily the distance between them. By analysing the induced movement the distance between the magnets could be calculated.

To give observable echo times experiments are routinely conducted at cryogenic temperatures so as to increase the transverse relaxation time to that detectable by the spectrometer. At these temperatures the transversal relaxation is mainly driven by fluctuations in the hyperfine fields from weakly coupled neighbouring protons, hence the dropping of proton concentration, typically by deuteration of the solvent, improves signal resolution and the observation window.

As the pump pulse has only a limited excitation bandwidth, it will excite only a small fraction of the B spins, λ . It is only these spins that will contribute to echo intensity modulation, the remaining B spins ($1 - \lambda$) remain independent and not contributing to the spin pair. The PELDOR signal form factor for a given time at a spin vector angle from the magnetic field, $F(t, \theta)$ can be given as

$$F(t, \theta) = F(0) \{1 - \lambda(\theta) [1 - \cos(d(\theta)t)]\} \quad (1.49)$$

With the final form factor, for a disordered sample being the sum of powder average, such that

$$F(t) = \int_0^{\pi/2} V(t, \theta) \sin \theta d\theta \quad (1.50)$$

A small proportion of A spins will also be inverted by the pump pulse as a result of their interactions with neighbouring molecules. Assuming that these neighbours are distributed homogeneously in space then they can be subtracted using a background factor:

$$B(t) = \exp\left(\frac{2\pi g^2 \mu_B^2 \mu_0 N_A}{9\sqrt{3}\hbar} \lambda ct\right) \quad (1.51)$$

where the average modulation depth λ is the fraction of spins excited by the pump pulse, g is the average g value and c is the concentration of spins. When combined this gives the final PELDOR signal as:

$$V(t) = F(t)B(t) \quad (1.52)$$

By combination of the Fourier induced decay and the PELDOR signal form factor, $V(t)$ the experimental trace is given.

Due to the flexibility inherent in biological molecules and nitroxide linker chains, the spin-spin distance is not a sharply defined distance but rather a distribution. Extraction of the distribution of distances, $p(r)$, from the form factor, $F(t)$, is an ill-posed problem where small errors in fitting the form factor can generate large errors in the distance distribution. Errors occurring from noise in the raw data and inaccurate background subtraction cannot be avoided. One method to stabilise the solution is with additional constraints. These are defined to be non-negative $p(r)$ (as a distance cannot be negative) and must give a smooth distribution. The deviation of this new constraint from the experiment can be evaluated to give a second constraint. Using matrix algebra and a Tikhonov regularisation the subspace can be scanned to give data that is smooth, without being over-smoothed which still yields a good fit to the original data. This method is implemented in the software package Deer-Analysis developed by Prof. Dr. G. Jeschke of ETH Zurich (Switzerland), which plots the log of each parameter in the result of the Tikhonov regularisation to form a characteristically shaped L-curve. At the steep section of the curve a small increase in the smoothing parameter yields a large increase in the quality of the fit. In the flat section of the curve the data is over-smoothed and results in a poor fit to the original data. The best compromise then is at the discontinuity of the L-curve [49].

The setup of a PELDOR experiment requires careful optimisation of parameters. Typically a field swept echo is first performed to select optimum positions for the pump and observation pulse sequences. Following the field swept echo, a transverse relaxation time experiment is performed to determine the maximum observation window for the echo. An inversion recovery experiment shows the time required for the system to return to thermal equilibrium after a pulse has been applied, this allows for the setting of the pulse shot repetition time - the time between experiments in signal averaging. The spectrometer is then calibrated such that the 2 pulse ESEEM reporter sequence gives the maximum signal intensity without overloading the detector.

Spectrometer design is such that the second microwave channel is completely independent of the first in frequency, power and phase. Each of these need to be adjusted to the sample to give maximum inversion in the PELDOR experiment. First, a modified inversion recovery pulse sequence, shown in figure 1.17, is applied where the ELDOR channel is set to the same frequency as the normal channels and initiates the sequence with a π pulse. This allows for

1. Fundamentals of electron paramagnetic resonance

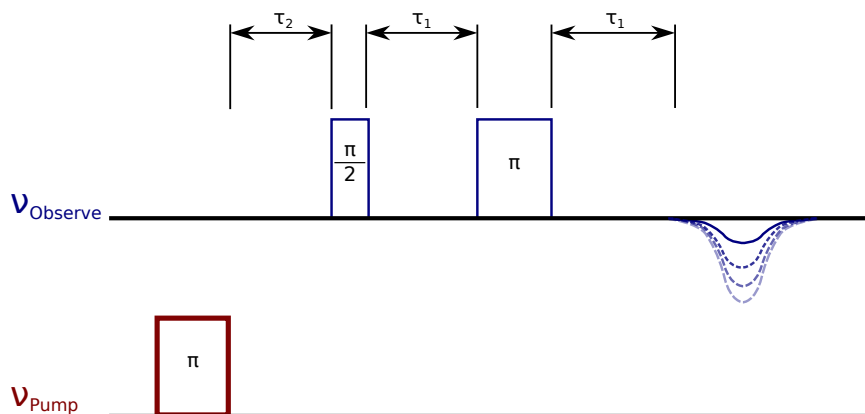


Figure 1.17 ELDOR pulse power optimisation sequence

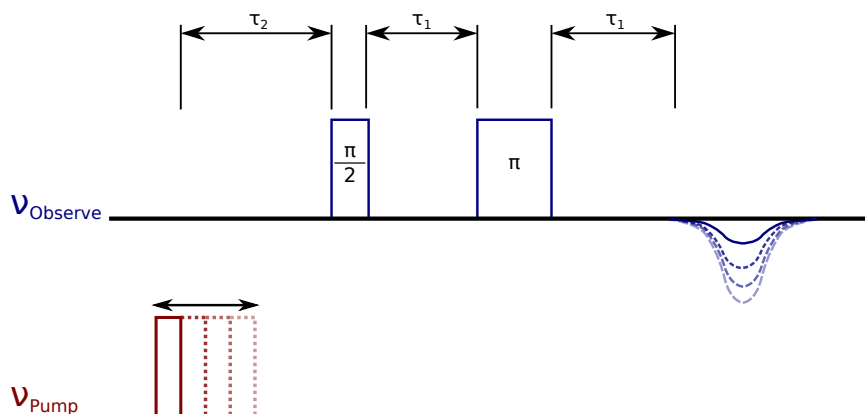


Figure 1.18 ELDOR nutation experiment pulse sequence

the optimisation of the ELDOR pulse power (with zero power the echo should be entirely positive) to achieve maximum inversion.

As the ELDOR channel uses different electronics, whilst the frequency and power are matched as far as possible there may still be some variation in channels. To ensure equal inversion of π pulses in both channels an ELDOR nutation experiment is performed, shown in figure 1.18. With the ELDOR channel power optimised and frequency matched, the ELDOR pulse begins very short and is incremented up in size whilst the inversion of the echo is monitored. Again maximum inversion is sought and should be approximately equal to the π pulse in the observer channel. When the pulse is too short, not enough energy is supplied to sufficiently move the bulk magnetisation into the xy plane; when the pulses are too long the magnetisation will be moved past the xy plane.

2. Cell biology and biochemistry of colicin E9

2.1. Organisms and cells

From the 7 characteristics of life; homeostasis, organisation, metabolism, growth, adaption, response to stimuli and reproduction [66], Linnaean taxonomy defines 3 distinct domains that of archaea, bacteria and eukarya. Traditionally, these domains are split into the 6 kingdoms of eubacteria, archaebacteria, protista, fungi, planae and animalia [42].

The cell is the basic biological, structural and functional unit of life, common to all living organisms and often referred to as the building block of life. The number and complexity of cells vary greatly between species from unicellular *Amoeba protea* (290 billion base pair genome) up to species such as *Homo sapiens* having an estimated 100 trillion cells (from a 2.9 billion base pair genome)¹. The word cell comes from the Latin *cellula*, meaning small room and first coined by Robert Hooke in 1665 when he observed that cork under a microscope looked similar to the small rooms that monks of the time lived in [61]. However, it was not until 1839 when it was hypothesised that cells were produced by pre-existing cells [57].

Cells have since been classified into 2 classes. The first, prokaryotic cells, are usually independent, smaller, simpler and typically classed by their cytoplasmic DNA (no membrane encapsulated nucleus) and flagella for motility (though not always required). The second, eukaryotic cells, are typically part of a multicellular organism, approximately 15x the size of a typical prokaryotic cell. Characteristically, with membrane bound organelles such as mitochondria, and DNA is formed into linear chromosomes inside a nuclear envelope.

Bacteria are further divided into Gram-positive and Gram-negative historically dependent upon their reaction to Gram staining, named after its inventor, Hans Christian Gram [78]. Gram staining differentiates bacteria by the chemical and physical characteristics of their cell walls by detecting the presense of peptidoglycan. Gram-positive bacteria have a single cellular membrane surrounded by a thick peptidoglycan layer. Gram-negative bacteria have a thin peptidoglycan layer that is sandwiched between two cellular membranes, having 2 membranes requires more elaborate transport systems with the outer membrane having many different porins for transport of materials.

2.2. Bacteria and colicins

Bacteria such as *Escherichia coli*, amongst other Gram-negative enterobacteriaceae often find themselves in less than favourable environmental conditions. At this time essential nutritional factors quickly become scarce.

¹Database of Genomic Sizes, the Growth of GenBank <http://www.cbs.dtu.dk/databases/DOGS/GBgrowth.php>

2. Cell biology and biochemistry of colicin E9

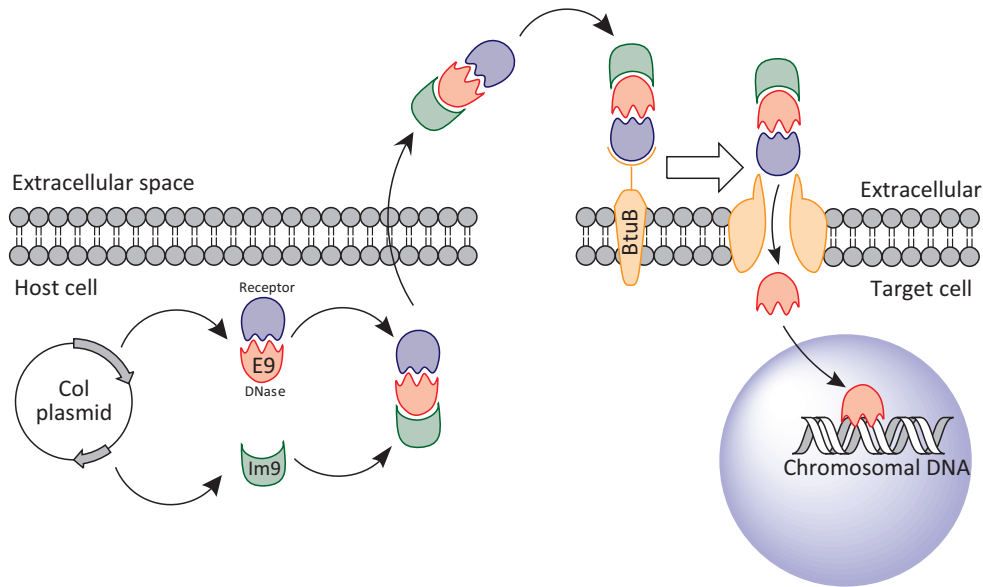


Figure 2.1 Expression, entry and action of colicin E9 from host cell to target cell. The ColE9 plasmid starts expression in response to SOS signalling and expresses the colicin E9; subdivided into the 15 kDa DNase domain (red) and the 45 kDa receptor binding and translocation domain (blue). The ColE9 Plasmid also expresses the 9 kDa colicin immunity protein, Im9 to stop the DNase activity inside the host cell. The E9/Im9 complex is released from the host and binds to sensitive bacteria via the BtuB receptor.

One response to biological stress is that of the SOS response pathway which allows the DNA repair protein RecA to bypass lesions and base pair mismatches in the DNA. RecA and other SOS response proteins are negatively regulated by LexA protein dimers binding to DNA at SOS box sequences (~ 20 bp in length). DNA damage increases in the amount of single stranded DNA (ssDNA) at replication forks where DNA polymerase has been blocked. RecA forms a filament around the ssDNA region, while with activation by ATP, allows RecA to interact with LexA's repressor domain and trigger self-cleavage from SOS boxes, allowing the expression of other SOS response system proteins [59]. Colicins are plasmid encoded antibacterial toxins (expressed by *E. coli* with the appropriate *Col* plasmid) that are one group of proteins that are expressed as a result of the SOS response. By delivering them into the extracellular environment gives them a competitive edge in nutritionally restrictive times [32].

Colicins are composed of three protein domains [15]. The first domain is responsible for the recognition of specific receptors on the target cell, such as the BtuB receptor for the E-group colicins [108]. The second domain is responsible for the translocation of the colicin across the cell membrane after binding by co-opting the machinery of the target cell. The third domain is the killing domain with either cytotoxic or pore formation activity which ultimately leads to death of the target cell. Colicins usually exhibit a "1-hit" kill kinetic by which only a very small number of molecules are required to successfully kill a target cell. Based upon the cytotoxicity method, colicins are divided into 4 groups:

- ◇ The ion channel formers; colicins E1 (reviewed by Cramer *et al.* [23]), A, B Ia [58], Ib and N - these form voltage dependant ionic channels in the cell membrane causing membrane depolarisation and stopping the cell producing energy.
- ◇ The cell wall inhibitor, colicin M - that inhibits the synthesis of the cell wall, eventually causes cell lysis, seen typically through initial cell blebbing [41]
- ◇ The RNases, colicins E3 [12], E4, E6 - typically cleaving the 16S ribosomal RNA, making RNA-to-protein translation impossible, colicin E5 is a RNase but instead targets

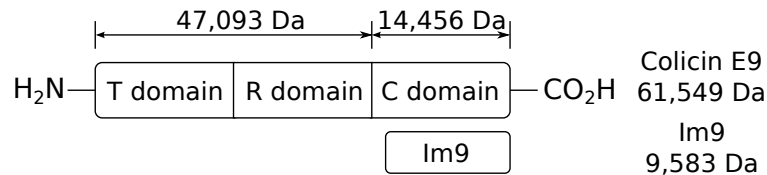


Figure 2.2 A schematic representation of the colicin E9 domains; the translocation T domain, receptor R domain and cytotoxic C domain as well as Im9, with included masses.

transfer RNA [68].

- ◇ The DNases, colicins E2, E7, E8, E9 [97] - cleave the DNA in both singly and doubly stranded fashion.

2.3. Domains and structures of E-type colicins

E-type colicins (molecular weight ~ 61 kDa) are composed of three domains, each designed for a specific function in assisting to the overall lethal toxicity. Due to colicins being such potent antibacterial agents their structure and activity has been well studied. Proteolytic digestion and deletion analysis revealed that there is linearity to the domains in the amino acid sequence, giving common structural function to all colicins [73].

The first colicin domain, the N-terminal T domain is involved in the translocation of the colicin across the target's outer cell membrane. The second domain, the R domain, is responsible for receptor recognition; namely the BtuB cell surface receptor protein. The cytotoxic activity is carried by the third C-terminal cytotoxic "killing" domain.

Colicins within a class (such as the E group colicins) have been shown to have high sequence homology within their T and R domains. The difference between each colicin in a class lies in their C domains, where only those colicins with similar cytotoxic action show any sequence homology [20, 90, 25, 30]. For example within the DNase E colicins (E2, E7, E8 and E9), an 80% sequence homology in the C domain is seen [46].

Colicin production is executed with simultaneous coexpression of colicin specific immunity proteins (molecular weight ~ 10 kDa). Amongst the E group, DNase colicin immunity proteins there is a sequence homology of approximately 50% [46]. Each immunity protein specifically binds its partner colicin, referred to as its cognate partner. The immunity protein tightly binds to the colicin C-terminal DNase domain thereby blocking the DNA binding region and thus protecting the host cell from the affect of DNase activity.

This pairing means that each immunity protein (Im) is named according to its cognate partner, for example E9-Im9, E7-Im7, etc. Sequence homologies between the colicin E activity domains as well as the immunity proteins is high. This allows for some amount of cross protection of colicin expressing cells from other colicins, even if the attack does not have explicit immunity to that colicin. This is a complex interaction that is as yet not fully understood as cross protection works well *in vitro* yet is rarely seen *in vivo*.

Since the first discovery of colicin DNases [82] and RNases [66] it has been found that the corresponding immunity proteins form a very tightly bound complex with the colicin, that is only chemically breakable through the use of strong denaturation agents.

2. Cell biology and biochemistry of colicin E9

Wallis *et al* [96, 97, 98] were the first to measure the binding affinities for colicin E9 to Im9; as well as a number of other noncognate immunity proteins. The dissociation constant (K_d) for the cognate pairing was found to be one of the highest measured affinities for a protein-protein interaction even to this date (figure 2.3). The outcome of these experiments both *in vivo* and *in vitro* gave an order to cross reactivity affinity for colicin E9 as $\text{Im9} \gg \text{Im2} > \text{Im8}$, with Im7 offering no detectable binding to colicin E9 in the methods employed by Wallis *et al* (figure 2.4), but later using other methods found to have a K_d of $1 \times 10^{-4} M$. The non-cognate immunity proteins, whilst binding colicin E9 showed binding affinities (concentration of immunity protein required for half the population to be complexed) in the range of 10^{-8} to $10^{-4} M$. The kinetic experiments however, show that the association rate constants (the speed at which the complex forms) are similar amongst cognate and non-cognate pairs. The differences are only really pronounced in the dissociation rate constants (strength of binding, ie how difficult it is to separate the complex) where the cognate pair is over six orders of magnitude slower than non-cognates. Most interestingly, these experiments showed that whilst non-cognate of different immunity proteins to the same colicin varied by four orders of magnitude, the non-cognate pairing of the same immunity protein to different E group colicins varies little.

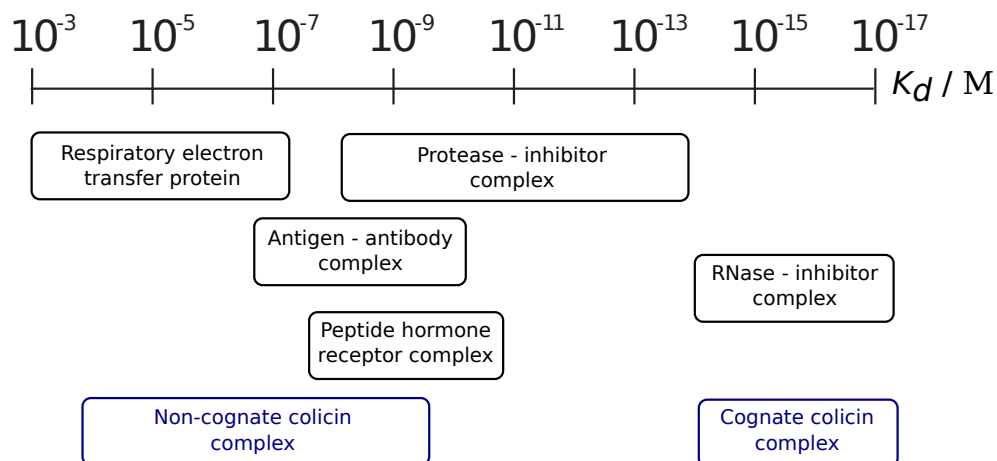


Figure 2.3 A comparison of dissociation constants for E9 DNase-Im protein complexes with a number of other well studied protein-protein interactions. Exact figures are not given but instead an approximate range as colicin E9-Im protein interaction is salt concentration dependant. Figure adapted from Figure 6 of [96].

ot

Wallis *et al* additionally showed that the dissociation constant for Im9 binding to full length colicin E9 or the E9 DNase domain was near identical, suggesting that the immunity protein makes no contacts with the T and R domains of any energetic significance.

Homologue-scanning mutagenesis was conducted whereby the key secondary structures of Im9 were systematically mutated to those of Im2 allowing for the elucidation of the secondary structures responsible for binding to colicin E9 [55]. With further investigation using nuclear magnetic resonance (NMR) [69] it was found that a single, 14 residue helix (helix II) was mostly responsible for colicin specificity. The neighbouring 6 residue helix III was found to be important in binding (see figure 2.5). Helix III remains conserved amongst the immunity protein family. Later investigation by alanine scanning mutations (where the 6 residues of helix III are systematically mutated to alanine) confirmed that helix III was indeed the main DNase binding site and contributed over $2/3$ of the binding energy to colicin E9 [96].

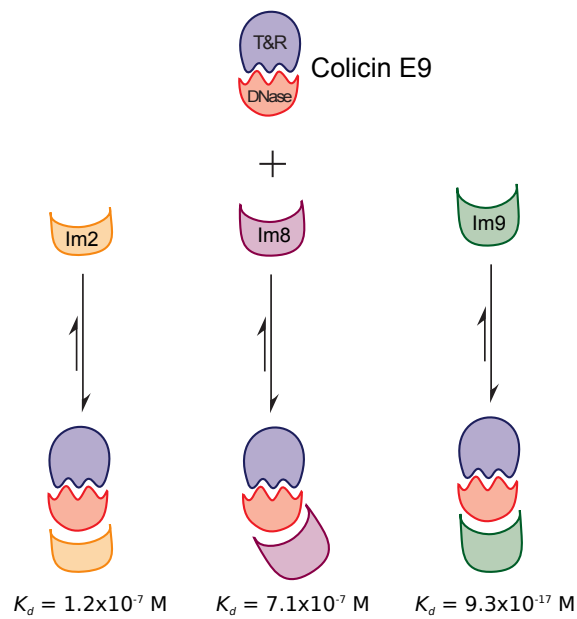


Figure 2.4 A diagram comparing the affinity of 3 different immunity proteins with the E9 colicin, the cognate Im9 and the non-cognates Im2 and Im8. As each immunity protein is different they bind the E9 in slightly different ways giving rise to different binding affinities.

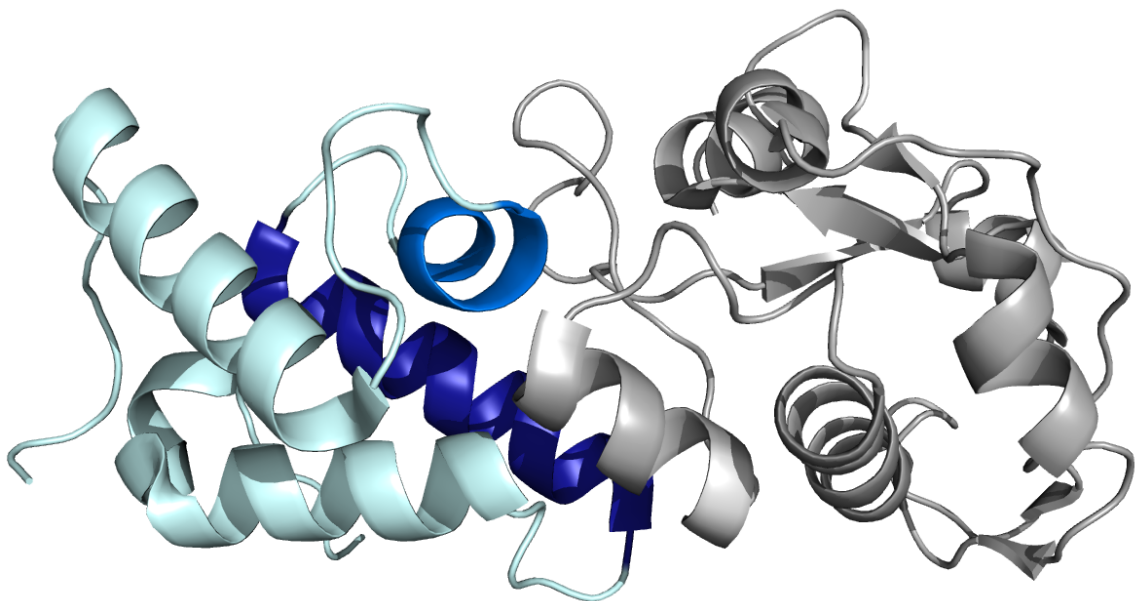


Figure 2.5 A cartoon representation of Im9 bound to the E9 DNase domain, with Im9's helices highlighted from PDB 2K5X. E9 is shown in grey, Im9 in cyan with helix II in dark blue and helix III in blue

2.4. Extracellular release of colicins

During times of environmental stress LexA dimers are removed from SOS boxes as a result of ssDNA replication forks, a classic example of DNA damage. The removal of inhibition allows for the resulting translated SOS response proteins to activate SOS boxes and start transcription of downstream genes. However, the LexA inhibition system is not completely effective, as there is always a small amount of the downstream genes transcribed into protein. This results in a very small background level of colicins in any cell containing the *Col* vector. Upon SOS box activation colicin production is positively upregulated but the activation system results in a lag time of approximately 60 to 90 minutes between stress and observable increases in colicin production. After the initial lag colicin levels increase exponentially to around 1000 times that of pre-induction [96] as a result of a very strong promoter and multicopy plasmids.

The release of colicins into the extracellular environment is very different to other bacterial extracellular proteins and is still thought to be largely unique [96]. Most extracellularly released proteins use a specific signal sequence, typically an encoded amino acid sequence (in much the same way as eukaryotes) or posttranslational modification (such as glycosylation), to bind the chaperone SecB. SecB binds the membrane bound SecA, an ATPase that delivers the protein to the translocation complex consisting of SecY, E and G. This SecA and SecYEG complex translocates the protein stepwise, 20 residues at a time, with each step powered by the hydrolysis of an ATP molecule [66]. There are 4 other, less common, methods of extracellular release by bacteria reviewed by Nikaido [67] and Thanassi and Hultgren [87].

Instead, colicins are secreted by a method that requires no signal sequence but rather the expression of only one gene which encodes the colicin lysis protein, also known as the killing (*kil*) protein or bacteriocin release protein (BRP). Colicins are released non-specifically after an arbitrary amount of time after synthesis with an amount of additional cytoplasmic and periplasmic proteins. The reason for such erratic release comes from the colicin lysis protein gene being located proximally to the immunity protein gene with a weak transcription terminator between them. Transcription then is reliant upon transcriptional read-through from the colicin promoter across the entire operon. Under normal (non-SOS) conditions very little lysis protein is produced as the chance of read-through on an inhibited gene is very low. With SOS activation, inhibitors are removed and there is increased reading of the entire operon, which results in greater colicin lysis protein production, but still at a low and random rate.

Thus when a cell activates colicin production, the process is usually irreversible. With *in vitro* studies artificial colicin induction immediately causes a drop in the number of viable cells within a colony. This cell death is as a result of lysis protein coexpression with colicin and not from the colicins themselves which the cell is protected from with the immunity proteins. The exact method by which cell death is caused still remains unknown but it seems to be linked to the shutting down of chromosomal protein synthesis [60].

Many studies have been conducted upon the effectiveness of the colicin lysis protein and the operators surrounding the colicin genes. In the event of the lysis gene being removed or having base mutations, insertions or deletions, release of colicins from the cell is inhibited.

Whilst cloning of the lysis gene and induction provokes protein release and cell death even in non-colicinogenic cells [15].

Ultimately, cells that underproduce or are slower to initiate a response are favoured for survival in a colony environment. However, switching the mechanism off or having it respond too slowly would also be detrimental to the population in a mixed environment. As a constant turnover of immunity protein is required to protect the host cell from its neighbour's colicin production.

In vivo experiments have shown that colicin producing strains of *E. coli* actively promote microbial diversity and *E. coli* colonies exist in a constant state of rock-paper-scissors between colicin producing, colicin sensitive and colicin resistant. As resistant outcompetes producing, producing kills sensitive and sensitive outcompetes resistant strains [37, 51].

2.5. Receptor binding and membrane translocation

2.5.1. Initial binding

Whilst once thought to be a simple method of the colicin binding a transmembrane protein and then a simple translocation across the cell membrane. It was found not to be the case with the characterization of several colicin insensitive mutants. Initially colicin resistant mutants, which did not bind the colicins were found more recently to be "tolerant" mutants, which bind the colicins but were unaffected were discovered. Following experimentation the colicin targetted transmembrane proteins were called "Tol" for their tolerance of colicin, with the most abundant protein being TolC. These mutants were found to inhibit colicin activity via one of two methods. Classification of tolerant cells to colicins into groups A or B was made by whether the colicin was translocated via the Tol or Ton systems respectively.

Both A and B group colicins target *E. coli* cells by interacting with specific outer membrane proteins. Group B colicins such as colicin B, D, M, Ia, Ib use the Ton system; almost exclusively a single TonB-dependant transporters in the outer membrane, only group B colicins 5 and 10 have been shown to use two outer membrane proteins [13].

Group A colicins largely target the TonB-dependent vitamin B₁₂ transporter BtuB for binding (including colicins A and E1 to 9, whilst colicin K binds Tsx and U binding OmpA) These colicins also require a second outer membrane protein for translocation (OmpF in all cases except colicin E1 which uses TolC, note that the TolC pore is twice the size of OmpF's). This two step translocation has lead many researchers to believe that the initial binding event serves to orientate the colicin into position on the cell surface before translocation. This is not necessarily a requirement as BtuB can be bypassed by the pore forming colicin A proteins when cells are incubated in low osmotic conditions suggesting that OmpF can be used as a sole receptor [18]. Cramer *et al.* have recently proposed a controversial fishing pole mechanism where the strong binding event of colicin to BtuB allows the colicin T domain to search the outer membrane whilst still being tethered to the membrane [109].

The actual binding event between colicins (both groups A and B) and receptor is now quite well characterised through genetic, biochemical and biophysical approaches, as well as extensive X-ray crystallography and NMR investigation (reviewed extensively by Cascales *et al.*

2. Cell biology and biochemistry of colicin E9

[15]). Interestingly, colicins despite their size exhibit the same type of receptor specificity to that of the small molecule ligand of the receptor. Immunity to colicin attack can be given to cells if a surplus of ligand is added to the extracellular environment, this competition differs greatly depending upon the natural ligand the individual colicin [17].

2.5.2. Colicin cleavage at cell surface

Many different laboratories working with different colicins (A, E1 to E4, Ia and Ib) have characterised cleavage of the killing domain at the outer membrane. Most groups attributed this cleavage to the outer membrane protease OmpT fragmenting the colicin into smaller products at the cell surface. Further investigation using *ompT* deletion mutants showed no fragmentation of colicins but remained sensitive to colicin kill domains and died [16].

This proteolytic cleavage seems to be essential for the effectiveness of some nuclease acting colicins [18], whereas *ompT* deletion mutants are killed with the same efficiency as wild type by colicin Ia.

Recently at the University of Oxford using a combination of cysteine trapping, mass spectrometry and X-ray crystallography, colicin E9 has been shown to move across the outer membrane using OmpF [43]. The cleavage question of E group colicins remains hotly debated, with some claiming that the Oxford trapping method artificially creates the results they want to see.

2.5.3. Release of immunity protein

Colicin expression in cell requires coexpression with an immunity protein to prevent autoinduced apoptosis. Due to the importance of this inhibition, the immunity protein is bound very tightly. In the case of colicin E9 with its immunity protein, Im9 the complex displays a K_d of 10^{-17} M. This immunity protein needs to be removed at the target cell in order to be an effective cytotoxic protein, but such a tight binding presents a problem.

Studies have shown that the colicin is unfolded or at least goes through a large conformational change during the translocation process. This could explain the release of the immunity protein; however, isothermal titration calorimetry combined with structural studies have shown that colicin E9 does not unfold on binding. Instead the colicin E9-Im9 complex is able to bind BtuB and recruit OmpF forming the BtuB-OmpF-colicin E9 translocon without displacing Im9 [44]. The exact mechanism by which the immunity protein is shed remains unknown and is likely different for each colicin-immunity protein complex.

2.5.4. Colicin import

Colicin import is the term used for the movement by which parts of the colicin (N and C domains) reach their final target including recognition and binding to the outer membrane of target cells. To avoid complication only the E colicins (group A)(nucleases) shall be discussed further.

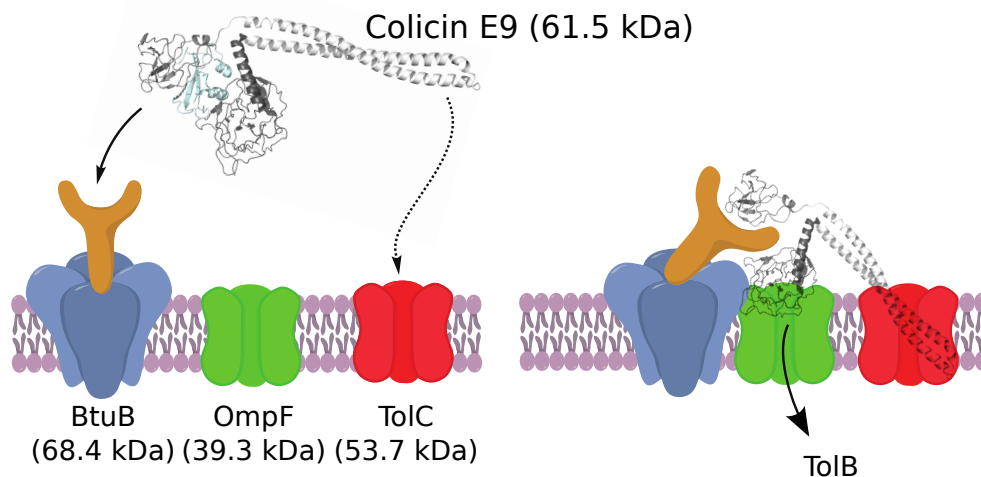


Figure 2.6 The binding event and recruitment of host cell proteins by colicin E9. The multi-domained colicin initially binds to the BtuB cell surface receptor. The translocation domain's tail (shown as a red and blue twist) binds TolC which further recruits OmpF. Together BtuB, TolC, OmpF and colicin form a translocon. The colicin tail is fed through TolC which brings the colicin C domain into close proximity to OmpF allowing for transport across the membrane and binding to TolB. Colicin E9 here shown using the colicin E3 homology model.

Whilst the exact translocation mechanism of colicins across the outer membrane remains poorly understood, it is plain that two proteins are required; a cell surface receptor (BtuB) and a cell penetrator (TolC) [44]. The latter also requires a co-transporter usually in the form of OmpF (an outer membrane tri-pored protein). The naturally disordered tail of the colicin's translocation domain binds the complex and forms a BtuB-OmpF-colicin translocon. The tail of the translocation domain is then fed into the cell via one of the OmpF's pores and then recruits TolB (a periplasmic protein). The colicin's cytotoxic domain is now in close proximity to another of OmpF's pores and is thought to probably enter this way [110].

Transport across the periplasmic space and final import, whilst being widely researched for many years remains speculative. What is clear is that TolA, B, Q and R are required for the translocation of colicins A, E2 to E9 and K (colicin E1 does not require TolB) [5]. Further investigation has shown that each Tol protein level increased proportionally to the amount of colicin that the cell was exposed to [38] and that the number of translocation events was similar to the number of TolA molecules on the cell surface [54].

Despite all of this research the exact release mechanism of the immunity protein from the colicin during binding and translocation remains unknown. It has been shown with colicin E2-Im2 that the binding event with receptor did not provide enough energy to separate the cognate complex [29]. Through fluorescence assays it has been shown that the energy for separating the colicin E9-Im9 complex must come from the cytoplasmic proton motive force, using the *tol* system. A crucial role in energy transfer is played by TolA, whilst separation can only occur when the first few residues of the colicin E9 translocation domain have bound TolB [93].

2.6. Cytotoxic activity of Colicin E9

Colicin E9 is known to be an enzymatic colicin that targets the phosphodiester bonds of DNA. Colicin E9 must first cross the outer cell membrane. The inner cell membrane and the

2. Cell biology and biochemistry of colicin E9

nuclear envelope, before acting upon the DNA. Debate still remains whether the transport within the cell is actively mediated ("self-propulsion") or by a system possibly at the end of the Tol/Ton system yet to be identified ("retrotranslocation"); or by a combination of both.

What is known from planar lipid bilayer experiments is that colicin E9 binds membrane surfaces (such as the nuclear envelope) and creates specific channels. The E9 DNase channels are short lived with lifetimes in the order of milliseconds and are not voltage dependant. Whilst these channels are not affected by DNA or metal ion binding, they are completely blocked when the E9 DNase is complexed with Im9 [62].

Colicin E9, as well as three other DNase colicins (E2, E7 and E8), have all had their structures elucidated, with the DNase domains holding an ~65% sequence homology. The DNase domain of colicin E9 consists of 3 antiparallel β -sheets surrounded by helices, with a concave active-site cleft for DNA binding (see figure 2.7). Interestingly most of the differences between the family of DNase colicins is confined to the immunity protein binding site, adjacent to the active site [53].

The core of the colicin DNase activity comes from the H-N-H motif, a ~34 amino acid sequence at the C-terminus of the DNase domain (and thus the protein) constructed of 2 β -strands and an α -helix sandwiched between a metal co-factor, which in all colicins resembles a distorted zinc finger structure.

The motif contains 4 histidine residues, conserved across all colicin DNases. The most important of these is the N-terminal histidine of the H-N-H motif which serves as a general base for the hydrolysis of the DNA backbone. The central asparagine (H-N-H) plays a structural role stabilizing an hydrogen bond across the motif, whilst the C-terminal histidine (H-N-H) is a metal binding residue.

Colicin DNases have been shown to destroy bacterial genomes through producing random nicks in dsDNA. Since nicks are a common occurrence in ordinary cell biology, repair mechanisms exist to protect the cell. However, cell death is likely to occur as a result of a double strand break caused by the repeated cutting of DNA, preferentially at thymine bases [76]. It has also been shown that colicin E9, like most H-N-H motif proteins, shows a dependence for high levels of Mg^{2+} or Ca^{+2} , a high optimum pH (pH 8-9) and is inhibited by monovalent cations. Colicin E9 can readily bind and use many first row transition metals such as Ni^{2+} , Co^{2+} , Cu^{2+} and Zn^{2+} although these metals influence substrate specificity and typically target ssDNA instead of dsDNA [50].

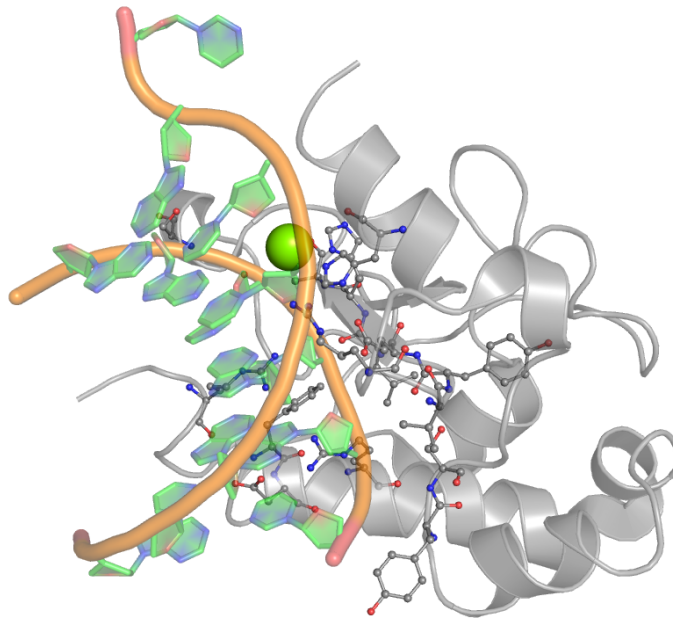


Figure 2.7 Colicin E9 DNase bound to dsDNA. Image manipulated from X-ray crystallography structure (PDB 1V14). E9 DNase shown in grey with binding residue side chains highlighted in ball and stick, DNA backbone in orange, DNA bases in green.

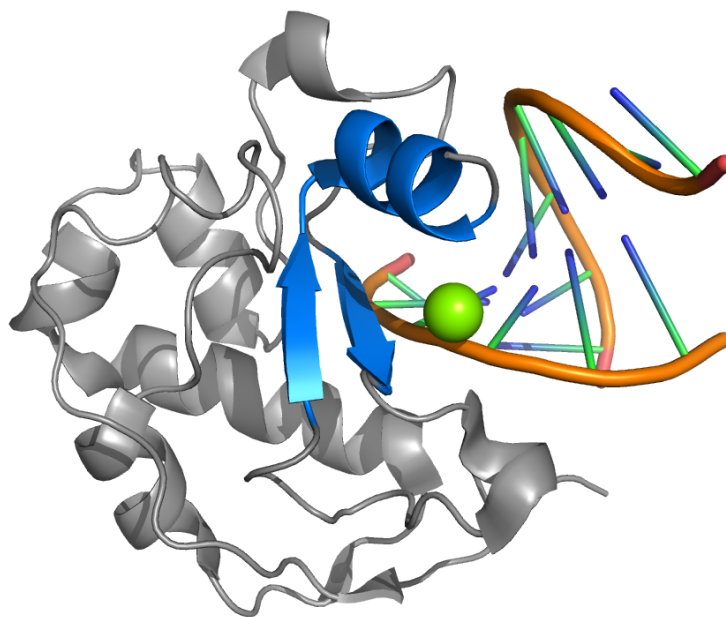


Figure 2.8 Colicin E9 DNase active site with H-N-H motif and metal. The H-N-H motif's α -helix and β -sheets are highlighted in blue, with bound magnesium ion in green, and dsDNA (backbone in orange). Image manipulated from X-ray crystallography structure (PDB 1V14)

Part II.

Investigation

3. Aims and objectives

Atomic resolution structures of protein-protein interactions were initially limited to x-ray crystallography; with unnatural buffers, pHs and forced lattice structures. NMR extended investigations to be able to determine solution based structures of small complexes, but *ab initio* information remains a constantly developing field.

Here it is hoped to express several mutant proteins of the colicin E9 DNase - Im9 complex. Each mutant will contain at least one cysteine which will be targeted by a nitroxide spin label in order to make it paramagnetic and detectable by EPR.

Each spin labelled protein will be fully characterised by various EPR techniques to determine the labelling efficiency, spin label motion, spin label accessibility and position in space relative to other labelling positions.

Using the EPR technique, pulsed electron-electron double resonance (PELDOR) it is hoped to be able to measure the interspin distances between spin labels across the colicin E9 DNase - Im9 complex.

Current docking models using NMR ambiguous interaction restraints based upon chemical shift perturbations from HSQC titrations give two potential binding sites of Im9 upon the colicin E9 DNase surface. The EPR experimentally derived interspin distances will be added as restraints to the protein-protein docking simulation to see if the quality of the model can be enhanced.

To introduce spin label to spin label distance constraints into the docking software will require:

- ◇ simulation of spin labels upon the protein structure
- ◇ generation of the protein coordinate file inclusive of spin labels
- ◇ creation of parameters for spin labelled cysteine residues, such that the software considers the residue to be an unnatural but allowed amino acid residue
- ◇ creation of restraint files based upon EPR data

4. Protein production, purification and performance

The purpose of this chapter is not to extensively describe every technique and step involved in the production of samples, this information can be found in the molecular biology methods chapter, chapter 10 (page 195). Instead this chapter seeks to highlight the typical output of a protein grow up and purification for each different production method.

Initial plans were to use genes previously generated at the University of East Anglia (UEA) by previous members of the colicin research group. However, it quickly became apparent that there were problems with these, detailed below, and other arrangements had to be sought. As a result, this investigation uses 3 basic types of protein grow up procedures, dependent upon the gene source.

First, the immunity protein Im9, which since the very earliest work has been expressed with a terminal His₆ tag; allowing for a single step crude purification method using immobilised metal affinity chromatography (IMAC) as poly-histidine sequences are known to favourably bind metal ²⁺ cations.

Second, in existing DNA stocks of colicin E9 DNases the DNase activity remains requiring co-expression with the immunity protein. Since in the *in natura* vector the immunity protein overlaps the E9 DNase domain, but in a different reading frame, this is not a problem and the 2 genes are co-expressed with the immunity protein (coming second in the gene) labelled with an His₆ tag. The crude purification then consists of binding the protein to an IMAC column then washing the column with buffer to remove weakly associated proteins. A wash with a denaturant to remove the colicin E9 DNase from the immunity protein Im9 follows, before finally removing the Im9 from the column.

Third, for newly created mutants where the existing DNA stocks had irrevocably been damaged or for new mutants, recent developments could be exploited whereby introducing a H127A mutation allowing the binding of colicin E9 DNase to DNA but inhibits the cleaving ability. These colicin mutants could be safely overexpressed in *E. coli* cells without the need for an immunity protein and could also be expressed with an His₆ tag allowing purification in a similar manner to that of the immunity protein Im9.

4.1. DNA stocks

4.1.1. Naturally derived mutants

At the start of the investigation an extensive library of colicin E9 DNase mutants existed from previous NMR and X-ray crystallography studies. These libraries were cloned copies of

4. Protein production, purification and performance

the colicin gene from the wild-type Col plasmid, with mutants being produced using primer based site directed mutagenesis. The libraries were stored in plasmid DNA stocks as well as cell glycerol stocks, where the plasmid had been transformed into *E. coli* BL21 (DE3) cells.

Test expressions were initially made for mutants colicin E9 DNase S3C, S30C, S80C, E66C and D20C/E66C as well as immunity protein Im9 from the glycerol stocks. The results of which showed that none expressed protein at a suitable molecular weight.

As no record of when the glycerol stocks had been made, a fresh transformation from the plasmid library for each was made and test expression repeated. Here the colicin E9 DNase E66C, D20C/E66C and immunity protein Im9 showed good expression. Whilst the cells with plasmids containing the colicin E9 DNase S3C, S30C and S80C mutations showed weakened antibiotic resistance and no protein expression.

The freshly transformed cells were split and mini-prepped to yield a large enough plasmid amount for DNA sequencing. DNA sequencing was conducted by The Genome Analysis Centre, Norwich, UK.

Mutant	Correct amino acid sequence
E9 DNase S3C	MECKRNKPGK ATGKGKPVGD KWLDDAGKDS GAPIPDRIAD KLRDKEFKSF DDFRKAVWEE VSKDPELSKN LNPSNKSSVS KGYSPTPKN QQVGGGRKVE LHHDKPISQG GEVYMDNIR VTPKRHIDI HRGK HHHHHH
E9 DNase S30C	MESKRNKPGK ATGKGKPVGD KWLDDAGKDC GAPIPDRIAD KLRDKEFKSF DDFRKAVWEE VSKDPELSKN LNPSNKSSVS KGYSPTPKN QQVGGGRKVE LHHDKPISQG GEVYMDNIR VTPKRHIDI HRGK HHHHHH
E9 DNase S80C	MESKRNKPGK ATGKGKPVGD KWLDDAGKDS GAPIPDRIAD KLRDKEFKSF DDFRKAVWEE VSKDPELSKN LNPSNKSSVC KGYSPTPKN QQVGGGRKVE LHHDKPISQG GEVYMDNIR VTPKRHIDI HRGK HHHHHH

Table 4.1. Sequencing results from the UEA non-yielding mutants. Black text shows residues in the correct reading frame. Green text indicates a 1 base frame shift. Blue text indicates a 2 base frame shift. Red text indicates a 3 base frame shift.

The results of sequencing (shown in table 4.1) showed that each non expressing mutant had several frame shifts in the DNA within the protein encoding region, each is usually a very rare occurrence. At best these shifts made nonsensical amino acid sequences or simply halted transcription through stop sequences. Upon further investigation it was found that these plasmids were constructed before at least 1992-1994, where they first appeared in the literature [100, 99, 101], with no record of their maintenance. Whilst holding DNA at cryogenic temperatures should preserve it indefinitely, even at these temperatures enzymes can turn over, only at a much reduced rate. It is common practice to not hold DNA in cell glycerol stocks for more than a year, and sequence DNA stocks every 5 years to ensure quality.

Communication with colicin research groups at the Universities of Nottingham and York sought to find replacement DNA for the colicin E9 DNase S3C, S30C and S80C mutants. Dr. M. Vankemmelbeke, University of Nottingham, UK was able to provide DNA for each, but confessed that these were old clones of UEA stocks and had no recent experience with them and had not had them sequenced.

The Nottingham DNA stocks were transformed into *E. coli* and split. Half going to a small scale test expression and half to mini-prepping for sequencing. E9 DNase S3C yielded good amounts of protein in the test expression and was confirmed by sequencing. E9 DNase S30C and S80C performed poorly in the test expressions and the sequencing showed frame shifts

in the protein encoding region, albeit to a lesser extent than the DNA stocks from UEA.

4.1.2. Synthetically derived mutants

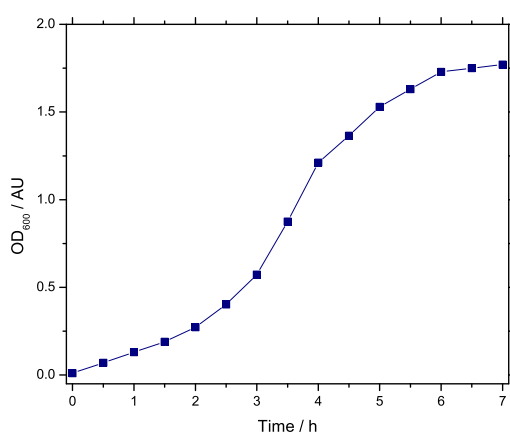
As a result of poor sequencing results and test expressions it was decided that the mutants of colicin E9 DNase S30C and S80C were unusable. Synthetic gene companies, such as Eurofins MWG Operon, offer synthetic genes cloned into vectors at prices comparable to that of cloning in the laboratory.

Whilst creating genes it was decided that investigating double mutants (mutants with 2 spin label sites) would give the investigation an added dimension allowing for triangulation of three distances within one complex. Through *in silico* analysis the possibility of a colicin E9 S3C/S30C was ruled out as the spin labels would be too close in space to give effective PELDOR distances. However the double mutants colicin E9 S3C/S80C, S30C/S80C and Im9 S6C (with a natural cysteine at position 23) could also be commissioned.

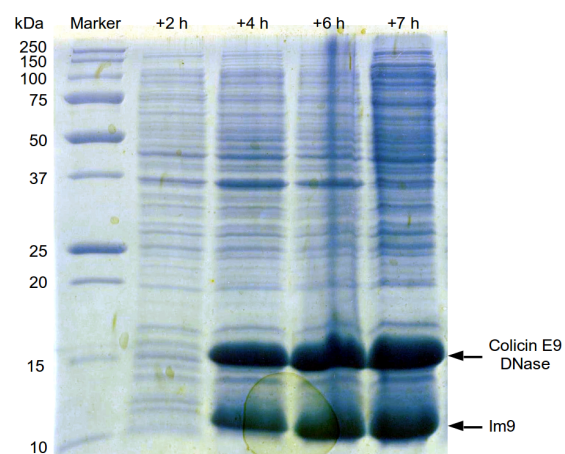
4.2. Protein expression

All mutants were grown in *E. coli* BL21 (DE3) cells transfected with the plasmid of interest. These cells were grown from a single colony overnight in antibiotic loaded growth media as a selection pressure (as per section § 10.4.1 on page 200). A proportion of the overnight growth was used to propagate a large scale growth medium.

Every half an hour the optical density of the growth media (Luria broth) was measured to monitor cell growth. At an OD_{600nm} of 0.6 an *E. coli* population is considered to be in the middle of its log growth phase and was induced such that the cells are converted from cell growth to protein expression.



(a) A plot of optical density at 600 nm



(b) The SDS-PAGE gel

Figure 4.1 A typical large scale overexpression. Shown here, a growth profile, started from 20 ml taken from the overnight culture of *E. coli* BL21 (DE3) cells transformed with the colicin E9 DNase S3C plasmid. Here the cells were added to LB broth (500 ml) with ampicillin (500 μ l, 100 mg/ml) and incubated with shaking (37 °C, 180 rpm) until reaching an optical density (less LB) at 595 nm of approximately 0.6, at which time induction occurred (3 h) with IPTG (5 ml, 100 mM). Samples were taken regularly and run upon a SDS-PAGE gel to confirm the presence or absence of colicin E9 DNase.

4.3. Crude purification

The aim of crude purification is to separate the desired, over-expressed protein from the vast majority of other soluble cytoplasmic proteins contained in the cell lysate. To this end, an IMAC column charged with Ni^{2+} was used, as it is routinely used for the purification of proteins containing a poly-histidine tail. The poly-histidine tail reversibly coordinates around the metal cation and can easily be dissociated from the column with the use of imidazole.

The crude soluble, filtered, cell lysate is passed over the column bed with a low level (50 *mM*) of imidazole to prevent non-specific binding. A long wash of several column volumes precedes the removal of the desired protein from the column.

4.3.1. Naturally derived colicin E9 DNases

Colicin E9 DNase mutants S3C, E66C and D20C/E66 are functionally active as DNases and required co-expression with the immunity protein Im9, which contains the poly-histidine tag. Due to the high affinity between the complex it was practical to separate the complex in place on the column using guanidine hydrochloride before titrating in the imidazole to remove the immunity protein Im9 from the column.

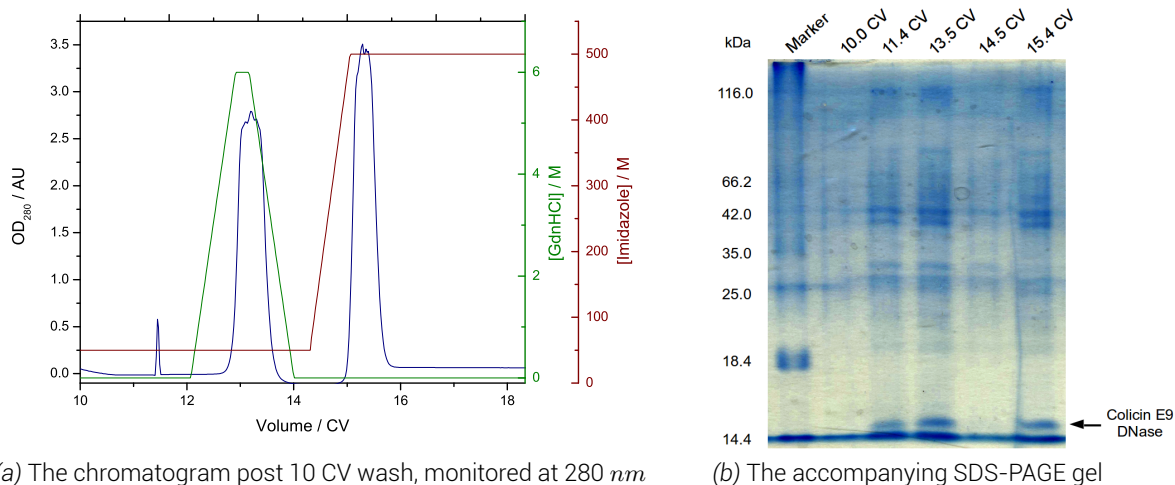


Figure 4.2 A typical crude purification with a GdnHCl unfolding in situ step before the imidazole titration. Shown here the supernatant of the spun down and sonicated cell lysate of the colicin E9 DNase S3C grow up was added to a 5 ml GE Healthcare Life Sciences HisTrap HP column activated with Ni_2SO_4 (50 μM). The protein was directly loaded onto the column with low concentration imidazole (50 *mM*) to stop non-specific binding. After a 10 column volume wash, GdnHCl was titrated in to a final concentration of 6 M to unfold the colicin E9 DNase from the immunity protein Im9. The GdnHCl was titrated away, before the imidazole was increased to a final concentration of 500 *mM* to dissociate the immunity protein Im9 from the column. Samples were taken at each stage and run upon a SDS-PAGE gel to confirm the presence or absence of colicin E9 DNase.

4.3.2. Synthetically derived colicin E9 DNases and immunity protein Im9s

The synthetically derived colicin E9 DNases of S30C, S80C, S3C/S80C and S30C/S80C were all synthesised with the additional H127A mutation making them DNA binding but not DNA

cutting and thus not a threat to the host cell during expression. This meant that these mutants could also be synthesised directly with the poly-histidine tag, allowing for a one step crude purification.

The synthetic colicin mutants cell lysate, as with the immunity protein mutants, were passed onto a Ni^{2+} IMAC column with low level imidazole to prevent non-specific binding. The column was then extensively washed before the imidazole concentration raised to dissociate the protein from the column bed.

Often there would be such a large quantity of protein that the SDS-PAGE gels showed leaking across lanes and the flow through from the column also shows the desired protein as a result of there being too much protein to bind to the column. In this case the flow through was run again over the column and the excess protein collected.

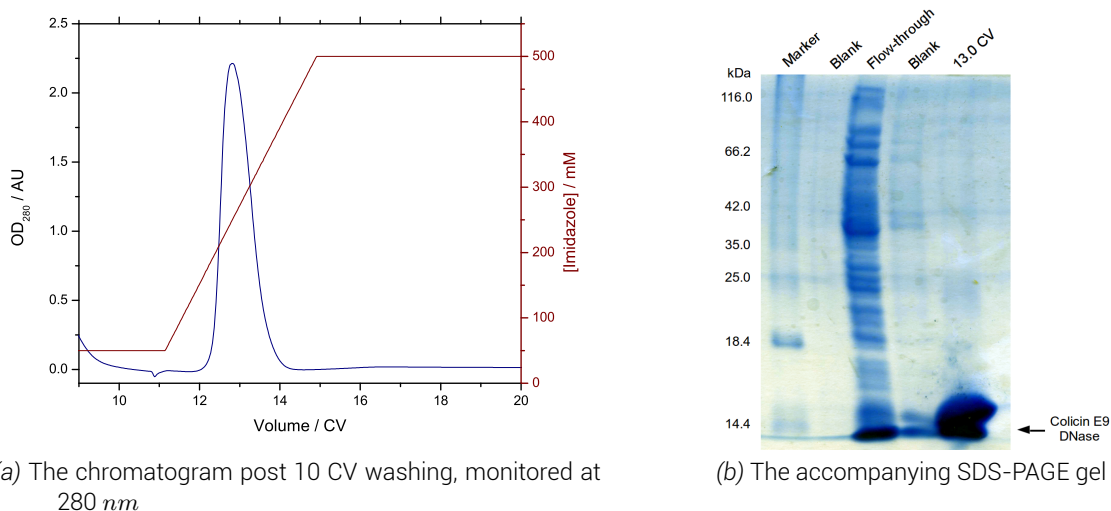


Figure 4.3 A typical crude purification with a one step imidazole titration. Shown here the supernatant of the spun down and sonicated cell lysate of the colicin E9 DNase S3C/S80C grow up was added to a 5 ml GE Healthcare Life Sciences HisTrap HP column activated with Ni_2SO_4 (50 μM). The protein was directly loaded onto the column with low concentration imidazole (50 mM) to stop non-specific binding. After a 10 column volume wash, the imidazole was increased to a final concentration of 500 mM to dissociate the protein from the column. Samples were taken at each stage and run upon a SDS-PAGE gel to confirm the presence or absence of the protein of interest. a) shows the chromatogram with only a single, large peak b) shows the associated SDS-PAGE gel showing a huge amount of colicin E9 DNase.

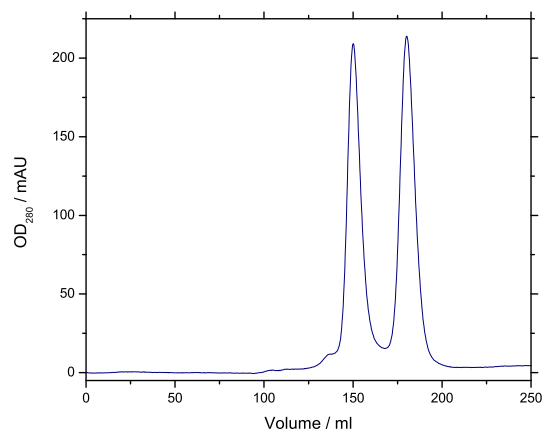
4.4. Polishing purification

After the crude purification almost all contaminants are removed from the sample, but to ensure as high purity as possible a second purification step was used. In this investigation the second purification used was a gel filtration, size exclusion chromatography step. The assumption that the probability that any naturally occurring, soluble protein within the *E. coli* cell lysate that showed affinity to the IMAC column also having a similar molecular weight would be very small.

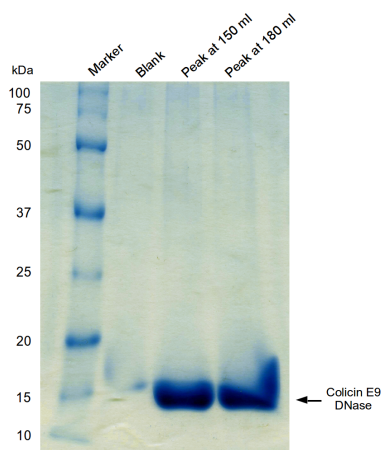
There was no difference in method at this stage between proteins or between the origins of the DNA. Most of the proteins here did however show 2 peaks during the size exclusion chromatography. When samples were taken of the 2 peaks and tested by SDS-PAGE they gave 2 bands at the same molecular weight (as seen in figure 4.4). When a native SDS-PAGE test

4. Protein production, purification and performance

was conducted (one in which no reducing agent is present), the first peak was found to be twice the apparent molecular weight of the monomer protein suggesting exposed cysteines were producing homo-dimers.



(a) The size exclusion chromatogram, monitored at 280 nm, column volume 230 ml



(b) The accompanying SDS-PAGE gel

Figure 4.4 A typical polishing purification using size exclusion chromatography. After dialysis and concentration the protein in question, shown here colicin E9 DNase S3C, would be loaded into the injection loop of a GE Health Life Sciences Äkta Explorer FPLC system, connected to a 350 ml bed volume, GE Health Life Sciences Superdex 200, 26/60 column. Samples were taken at each stage and run upon a SDS-PAGE gel to confirm the presence or absence of the protein of interest. a) shows a characteristic size exclusion chromatogram showing two distinct peaks b) shows the accompanying SDS-PAGE gel showing that under normal reducing conditions used in SDS-PAGE that both peaks contain a protein band associated with monomer colicin E9 DNase.

4.5. Protein confirmation

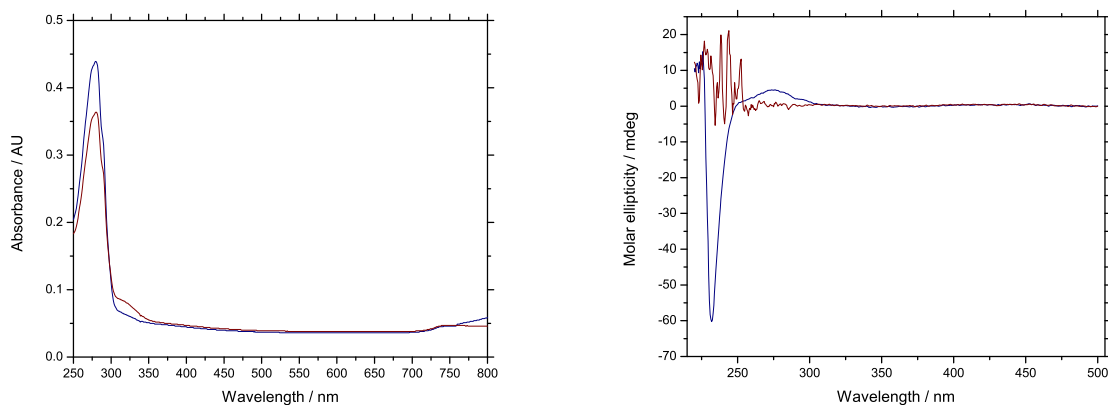
4.5.1. Optical spectroscopy

All samples after purification were tested using a wide spectrum UV/visible absorbance spectroscopy. As none of the proteins here contain metal cofactors, a single peak at 280 nm was expected in the UV/visible spectrum resulting mainly from tryptophan, tyrosine and cysteine residues.

All samples also had their circular dichroism spectrum recorded. The naturally derived colicin E9 mutants which were unfolded in the crude purification step were recorded before and after dialysis to ensure that correct folding was achieved. It was expected that for unfolded proteins a random noise like signal would be seen before dialysis and for the folded protein there would be a strong negative peak in the far-UV range (<250 nm). In the example given in figure 4.5.b performing estimation of protein secondary structure from CD spectra using the online web tool K2D2 [75], colicin E9 DNase S3C in the presence of a denaturant shows 5.4 % predicted α -helical structure and 4.3 % predicted β -strand structure. After dialysis, K2D2 gives 60.4 % predicted α -helical structure and 4.7 % predicted β -strand structure. This compares well with the published X-ray crystallography structure (PDB entry 1FSJ) that reports 8 α -helices (65 residues, 48 %) and 4 short β -strands (10 residues, 7 %). CD is only used as a prediction of secondary structure and that with the facilities measurement only to 220 nm

was possible, whereas for more accurate predictions of secondary structure require further measurement into the far-UV to 190 *nm*.

All proteins behaved as expected in this tests. All mutants, even those of synthetically derived mutants gave comparable traces.



(a) The UV/visible spectrum

(b) The circular dichroism spectrum

Figure 4.5 A typical optical spectroscopy profile. Shown here colicin E9 DNase S3C. Blue shows the folded protein after dialysis. Red shows the unfolded protein taken directly after the crude purification. Folded protein traces before and after size exclusion are identical.

4.5.2. Mass spectrometry

All samples, after purification and lypophilisation, had a small amount taken (~ 1 *mg*, far more than required for a), desalted and sent for MALDI-MS (matrix assisted laser desorption ionisation mass spectrometry) analysis at the National Mass Spectroscopy Centre, Swansea, UK. 1 *mg* of protein is far more than required for a typical MS sample, but allowed the service to properly optimise experimental conditions for clean spectra. MALDI-MS was selected over ESI-MS (electrospray ionisation mass spectrometry) as it uses a much softer ionisation process, resulting in lower charge ions which evolve at lower energies, ideal for proteins with untested stabilities. MALDI-MS is also more robust to the salt concentrations expected in biological buffers.

All samples showed a strong intensity at a mass to charge ratio (m/z) corresponding to the molecular weight of the protein with a +1 charge. A second, weaker peak could be seen at exactly half the m/z corresponding to the mutant with a +2 charge.

4.6. Yields

Yields are given as an average *mg* of protein yielded per litre of growth media. As mutants are of approximate equal weight direct comparison is possible. The average yield also reflects that lower yielding mutants were grown in multiple batches to yield a total of greater than 100 *mg* required for all further experiments.

4. Protein production, purification and performance

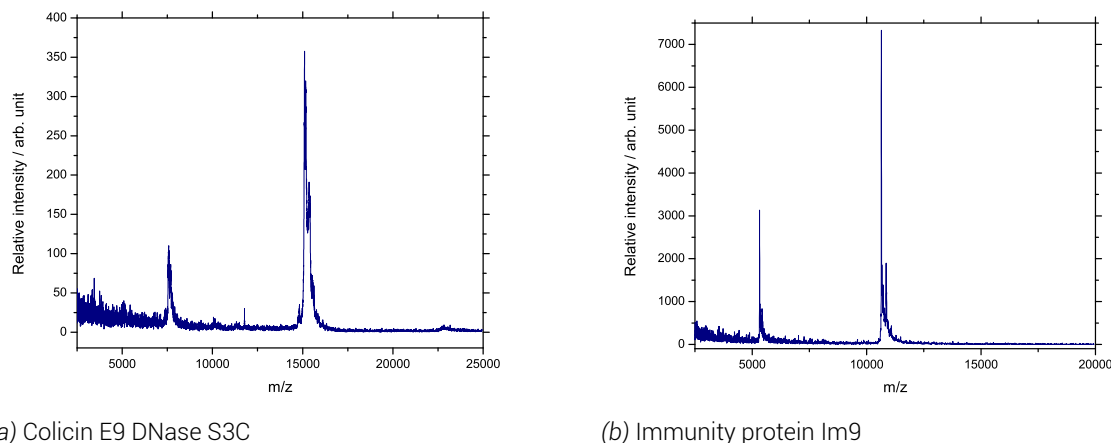


Figure 4.6 Typical MALDI mass spectrometry profiles for a colicin E9 DNase and immunity protein Im9.

Protein, labelling position	Yield / $mg.l^{-1}$
E9 DNase S3C	64.2
E9 DNase S30C	2.7
E9 DNase S80C	-
E9 DNase S3C / S80C	34.3
E9 DNase S30C / S80C	31.2
E9 DNase E66C	32.4
E9 DNase D20C / E66C	19.9
Im9 C23	36.1
Im9 C23S / S6C	11.6
Im9 S6C	12.8

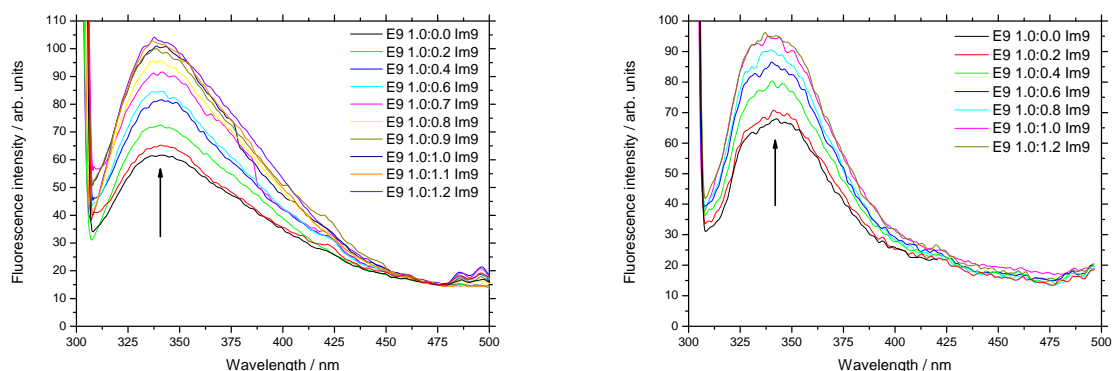
Table 4.2. The yields of each protein's grow up and purification.

4.7. Binding activity

Many of proteins in this investigation are synthetically derived such that they ultimately have the same amino acid sequence. This however does not mean that each mutant uses the same DNA base sequence, as each mutant underwent a codon optimisation process based upon the NCBI-GenBank Codon Usage Database [65]. This process both ensures that only the most common codons for each amino acid for the selected species are used whilst avoiding excessive use of similar codons in close proximity, as well as avoiding bad motifs - such as restriction enzyme target sites, hairpin motifs, etc. It was therefore important to test each synthetically derived mutants' biological activity matched that of the the naturally derived mutants, thereby ensuring that each mutation was structurally and functionally active as desired.

Free tryptophan in solution shows a fluorescence emission at 350 nm when irradiated with radiation at 295 nm . In unfolded proteins where each amino acid has little interaction with others the fluorescence of tryptophan residues reflect that of the free state. As a protein folds the amino acid side chains of other residues surround the typtophan residue and their interaction influences the fluorescene behaviour.

Pommer [77] (a more in depth discussion of Wallis *et al.* [100]) showed that the 2 tryptophans (W22 and W58) within colicin E9 DNase domain show an emission maximum at 332 nm



(a) Naturally derived mutant colicin E9 DNase S3C titrated with Im9

(b) Synthetically derived mutant colicin E9 DNase S30C/S80C titrated with Im9

Figure 4.7 Fluorescence enhancement as a result of addition of immunity protein Im9 to colicin E9 DNase. 10 μM colicin E9 DNase was added to a $10 \times 10 \text{ mm}$ quartz cuvette at 298 K. Scans were conducted using continuous excitation mode with excitation at 295 nm, with detection between 300 and 500 nm with a scan rate of 100 nm/min. For each titration a small amount of 100 μM immunity protein Im9 was added to minimise dilution effects.

indicative of at least 1 of these tryptophans being buried. When the colicin E9 DNase is denatured by pH, the fluorescence emission shifts to 354 nm. The immunity protein Im9, with no natural tryptophans shows no significant fluorescence from excitation at 295 nm. However, as Im9 is titrated into a colicin E9 DNase solution a fluorescence enhancement is seen until an equimolar stoichiometry is reached. The fluorescence emission maximum upon excitation at 295 nm shifts to 334 nm, as both tryptophans are very close to the binding interface of the immunity protein Im9.

Thus, if a fluorescence enhancement and shift is seen when titrating in immunity protein Im9 to a colicin E9 DNase solution then it can be shown that the synthetically derived mutants are forming a complex comparable to the naturally derived mutants.

Figure 4.7 is typical for all the mutants tested and shows that the synthetically derived proteins behave experimentally as the naturally derived mutants, which in turn match the results seen by Wallis *et al.* [100, 101].

4.8. Enzymatic activity

As stated at the start of this chapter, naturally derived colicin E9 DNase mutants still possess DNase activity and required co-expression with the immunity protein Im9, to stop the killing of the host cell during expression. The synthetically derived mutants were all expressed with an additional H127A mutation following the work of Garinot-Schneider *et al.* [34]. This work showed that on full length colicin E9 a H575A mutation knocked out DNase activity whilst still being able to enter target cells. However, the activity of isolated DNase domain with this mutation was previously untested. Whilst the first indication that DNase activity had been at least inhibited was given when cells transformed with the new plasmid did not immediately die; the extent of DNase activity inhibition required testing.

Likewise, the synthetically derived immunity protein Im9 (S6C) needed its biological relevance to be tested to ensure that it inhibits the enzymatic activity of the colicin E9 DNase domain.

4. Protein production, purification and performance

For this experiment 2 types of cell were used, non-transformed *E. coli* BL21 (DE3) cells and *E. coli* BL21 (DE3) cells transformed with the Col plasmid for colicin E9 DNase S3C. The cells without the plasmid should be sensitive to the cytotoxic activity of colicin E9 whilst those containing the plasmid contain a gene encoding for the immunity protein Im9. However, on the addition of colicin E9 DNase complexed with the immunity protein Im9 neither cell line should be adversely affected.

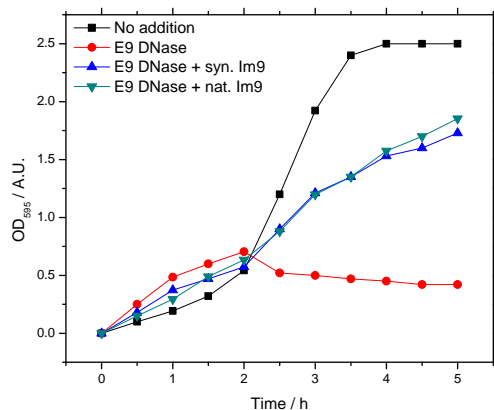
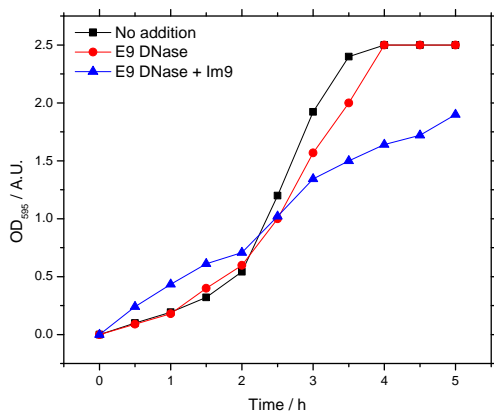
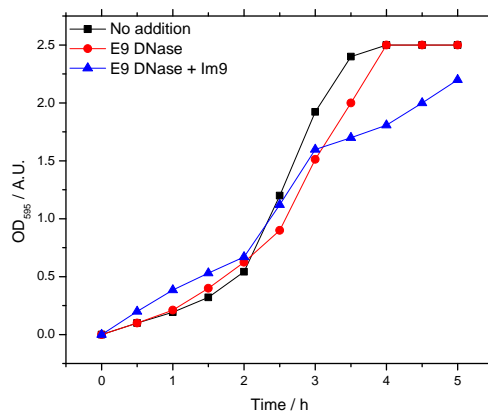


Figure 4.8 Growth curves of *E. coli* BL21 (DE3) cells in culture with the addition of colicin E9 DNase. Black squares show normal cell growth. Red circles, cells with colicin E9 DNase S3C added at 2 hours. Blue upward triangles, cells colicin E9 DNase S3C and Im9 added at 2 hours. Dark cyan downward triangles, colicin E9 DNase S3C and Im9 S6C added at 2 hours.

Figure 4.8 shows that cells lacking the Col plasmid when exposed to colicin E9 DNase without the H127A mutation show a drop in optical density and no further growth. If however the colicin is added complexed with Im9 irrespective of whether it is naturally derived or the synthetically produced Im9^{S6C} then cells survive. There is a distinct slowing of population growth as inevitably some cells will not respond in time to prevent death, but both Im9s show near identical results.



(a) With colicin E9 DNase S3C/S80C/H127A



(b) With colicin E9 DNase S30C/S80C/H127A

Figure 4.9 Growth curves of *E. coli* BL21 (DE3) cells in culture with the addition of synthetic colicin E9 DNase. Black squares show normal cell growth. Red circles, cells with colicin E9 DNase added at 2 hours. Blue upward triangles, cells colicin E9 DNase and Im9 added at 2 hours.

In figures 4.9.a and 4.9.b the addition of synthetic colicin E9 DNases with the additional H127A mutation show no distinct difference to population growth. If wild type colicin E9 DNase with DNase activity is complexed with Im9, when administered there appears to be some minor slowing, presumably from a few free molecules of colicin E9.

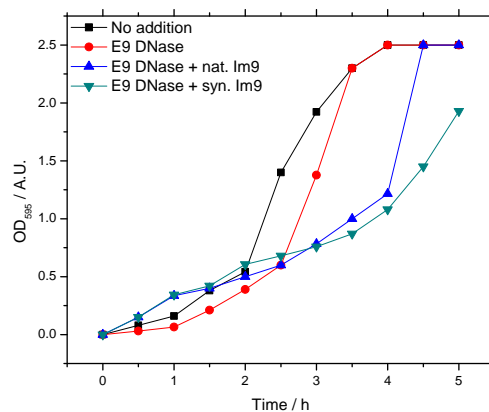
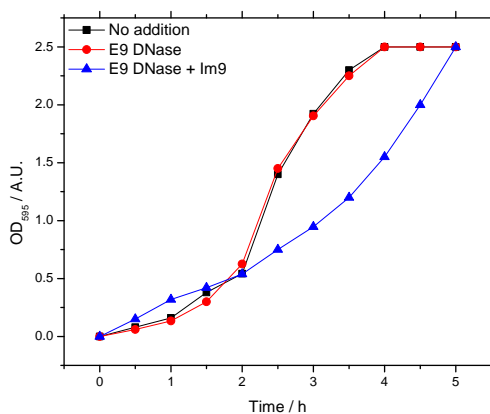
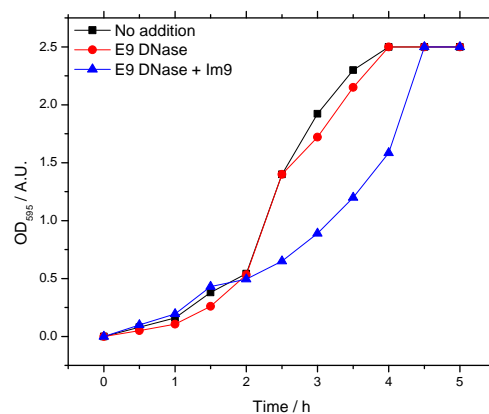


Figure 4.10 Growth curves of *E. coli* BL21 (DE3) cells with Col plasmid in culture with the addition of colicin E9 DNase. Black squares show normal cell growth. Red circles, cells with colicin E9 DNase S3C added at 2 hours. Blue upward triangles, cells colicin E9 DNase S3C and Im9 added at 2 hours. Dark cyan downward triangles, colicin E9 DNase S3C and Im9 S6C added at 2 hours.



(a) With colicin E9 DNase S3C/S80C/H127A



(b) With colicin E9 DNase S30C/S80C/H127A

Figure 4.11 Growth curves of *E. coli* BL21 (DE3) cells with Col plasmid in culture with the addition of synthetic colicin E9 DNase. Black squares show normal cell growth. Red circles, cells with colicin E9 DNase added at 2 hours. Blue upward triangles, cells colicin E9 DNase and Im9 added at 2 hours.

Cells with the Col plasmid upon the addition of colicin E9 DNase from either source does not kill population growth, but, slows it slightly. Again, the presense of Im9 with colicin E9 DNase slows population growth more than simply colicin E9 DNase alone. This could be due to some unknown positive feedback mechanism whereby the presense of Im9 in the cell makes the cell produce more Im9 to protect itself, as naturally the only time that Im9 would be produced is during co-expression with colicin E9 and free Im9 suggests free colicin within the cell.

5. Spin label characterisation

With proteins successfully expressed and verified, next was to attach a spin label so an EPR signal was produced. A full cw-EPR characterisation was then conducted to analyse the spin label's properties before moving forward to the more technical pulsed EPR methods. Spin labelled proteins are denoted by the PDB modified residue ID assignment, in superscript to the residue, where IA1 is IA-PROXYL and R1A is MTSL. The spin labelling procedure is discussed on §10.5.

5.1. Spin label efficiency

Spin label efficiency was calculated by deriving the double integral of the room temperature, X-band cw-EPR spectrum (the spin count). This was compared against a standard curve constructed of spin counts of free spin label at concentrations around the estimated protein concentration. The attached spin label concentration by spin count determination was compared to the protein concentration by UV/visible absorbance to give a labelling efficiency. Labelling efficiency is stated as a percentage for the total label attachment sites. Protein concentration was measured using absorbance at 280 nm (detailed in §10.4.1.8) and the Beer-Lambert Law (eq. 10.1). Where more than one cell culture was used the average labelling efficiency is given.

Protein, labelling position	Labelling efficiency / %
E9 DNase S3C	33.6 ^{IA1} / 1.6 ^{R1A}
E9 DNase S30C ^{R1A}	19.2
E9 DNase S80C ^{IA1}	N/A
E9 DNase S3C ^{IA1} / S80C ^{IA1}	84.7
E9 DNase S30C ^{IA1} / S80C ^{IA1}	95.9
E9 DNase E66C ^{IA1}	38.9
E9 DNase D20C ^{IA1} / E66C ^{IA1}	1.4
Im9 C23 ^{IA1}	5.3
Im9 C23S / S6C ^{IA1}	14.1
Im9 S6C ^{IA1} / C23 ^{IA1}	16.5

Table 5.1. Spin labelling efficiency summary

In this investigation, a large amount of time was spent fully optimising labelling efficiency to the highest possible level. It was known from the outset that low labelling efficiencies would be punishing in the later PELDOR investigation, for example if two proteins with labelling efficiency of 10 % are complexed then only 1 % of the sample will contain a complex where both proteins are labelled and can contribute to the PELDOR signal.

In the first instance, a thorough literature search was conducted. As some colicin E9 mutants had been previously reported [106] this protocol was replicated. This method (detailed

5. Spin label characterisation

in §10.5.1) involved in a low oxygen environment (glove-box) adding a small amount of DTT to the protein sample to reduce any protein homodimers and ensure that cysteines were in the S^- state. Excess DTT was removed using a PD-10 desalting column, such that the newly formed label-protein bond would not be reduced. The protein was mixed with a 5 fold molar excess solution of MTSL, wrapped in foil (as the reaction is light inhibited) and left for 3 *h* with stirring. In the paper, labelling efficiencies of >95 % were reported [106], compared to the obtained results of <5 %. After personal communications with Dr. G White, a labelling procedure was attempted together in identical conditions to those reported previously. Again, the resulting labelling efficiency was <5 %.

So began a spin label trial in which many conditions were investigated. Namely, oxygenation of the environment (atmospheric versus low oxygen environment), with and without DTT, labelling time (3, 16 and 56 *h*), labelling excess (2, 5, 10, 20, 100 fold molar excess of label), pH (5.0, 7.0 and 9.0) and labelling temperature (277 and 298 *K*). With optimisations in all of these variables typical labelling efficiencies were increased to low teen percentages for colicin E9 mutants, with Im9 samples not reaching over 8 %. Note that samples requiring unfolding showed an inhibited labelling efficiency by the denaturant and typically required 10× the labelling excess.

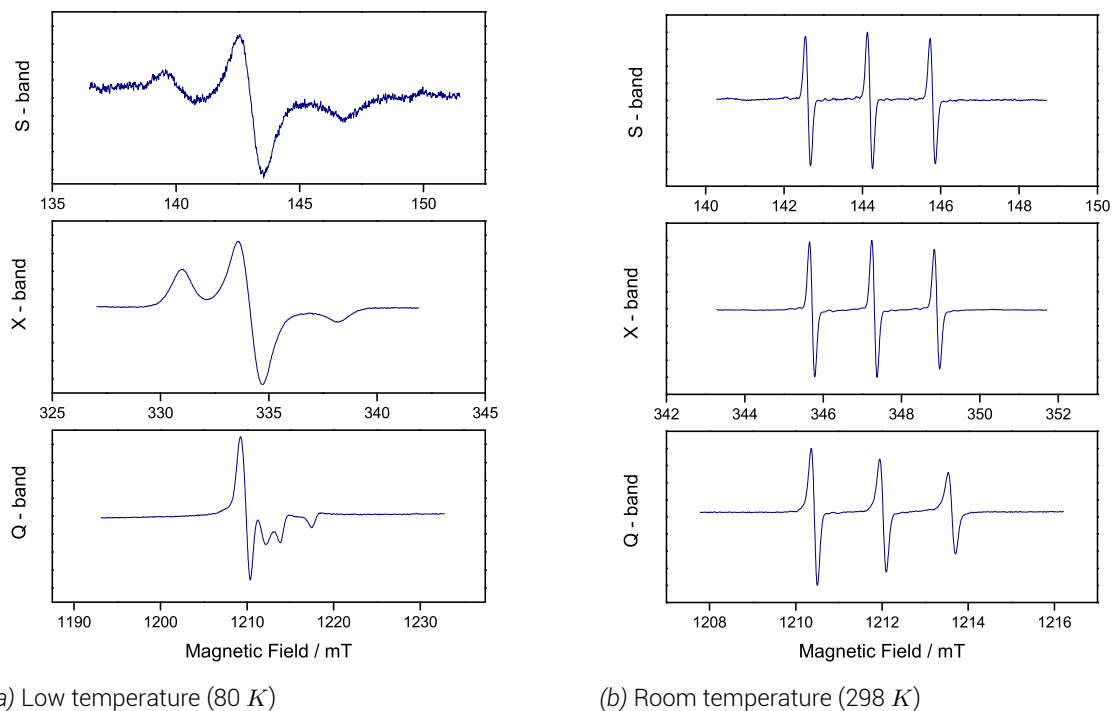
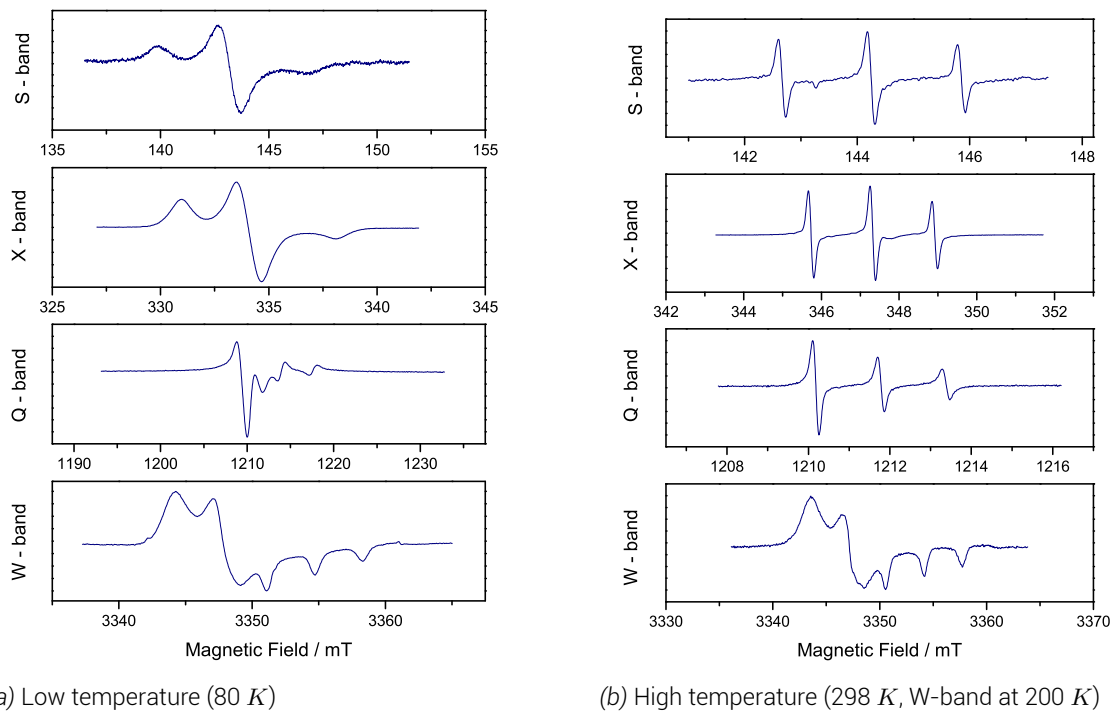
It was at this point, the newly reported IA-PROXYL was tested, as its reaction chemistry is irreversible and uses a good iodo leaving group instead of weak sulphur chemistry. The results listed in table 5.1 are the result of a small spin labelling trial with the IA-PROXYL label, in which it was found that the best labelling conditions could be obtained using a 5× molar excess, without DTT, pH 7.0, at 277 *K* with incubation overnight with stirring (16 *h*, 30 *rpm*).

Using IA-PROXYL the synthetically derived colicin E9 DNase mutants immediately jumped from 10-15 % to 84-96 %. Some naturally derived colicin E9 DNase mutants (S3C and E66C) saw an increase of around 30 % in labelling efficiency. All Im9 labelling sites remained very low efficiency, but approximately double that seen with MTSL.

5.2. Continuous wave EPR characterisation

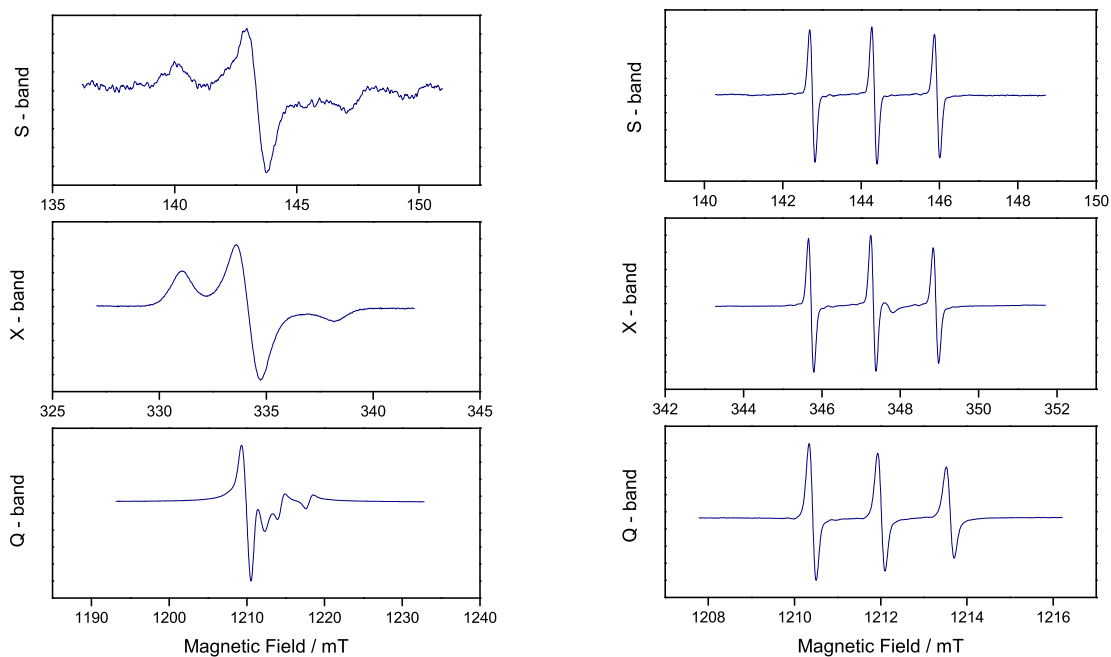
Each spin label site was characterised at room (298 *K*) and cryogenic (80 *K*) temperatures, at S, X and Q-band, with some also characterised at W-band. Each sample was run under identical conditions for each temperature and frequency (§11.1), allowing direct comparison between samples.

Q-band spectra, particularly at low temperatures are not phased well, due to the spectrometer in question having no feedback regarding phasing and thus each spectrum had to be phased by hand using a trial-and-error process.

5.2.1. Colicin E9 DNase S3C^{IA1}Figure 5.1 cw-EPR results for colicin E9 DNase S3C^{IA1} at S, X and Q - band5.2.2. Colicin E9 DNase S30C^{R1A}Figure 5.2 cw-EPR results for colicin E9 DNase S30C^{R1A} at S, X, Q and W - band

5. Spin label characterisation

5.2.3. Colicin E9 DNase S3C^{IA1} / S80C^{IA1}

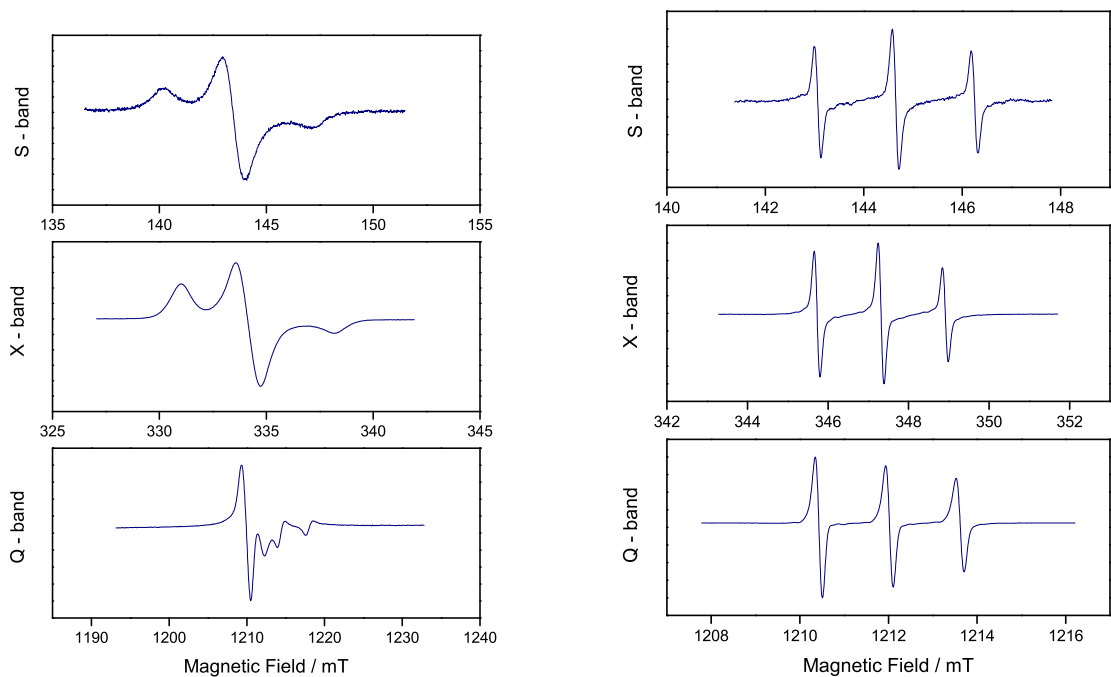


(a) Low temperature (80 K)

(b) Room temperature (298 K)

Figure 5.3 cw-EPR results for colicin E9 DNase S3C^{IA1} / S80C^{IA1} at S, X and Q - band

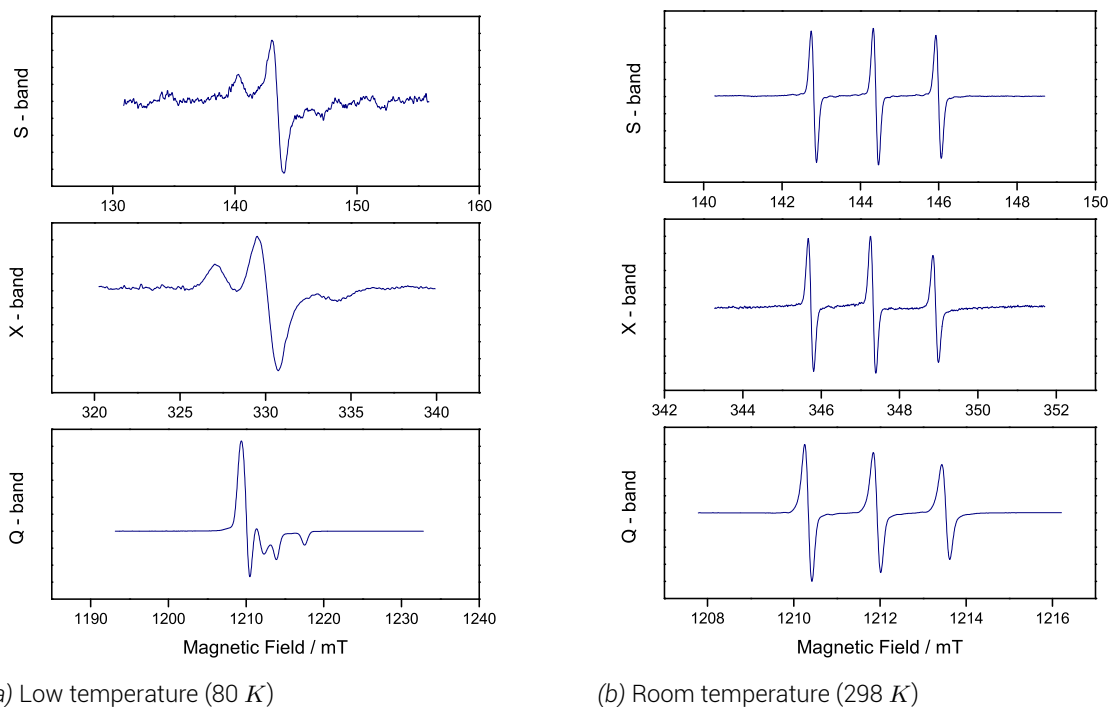
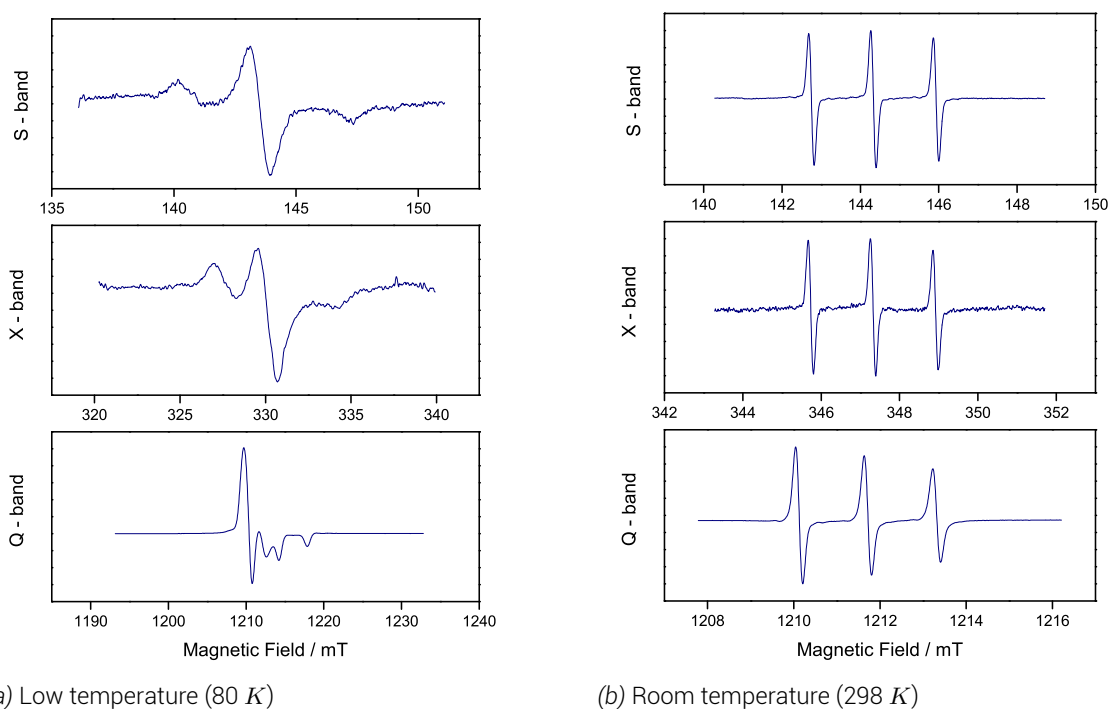
5.2.4. Colicin E9 DNase S30C^{IA1} / S80C^{IA1}



(a) Low temperature (80 K)

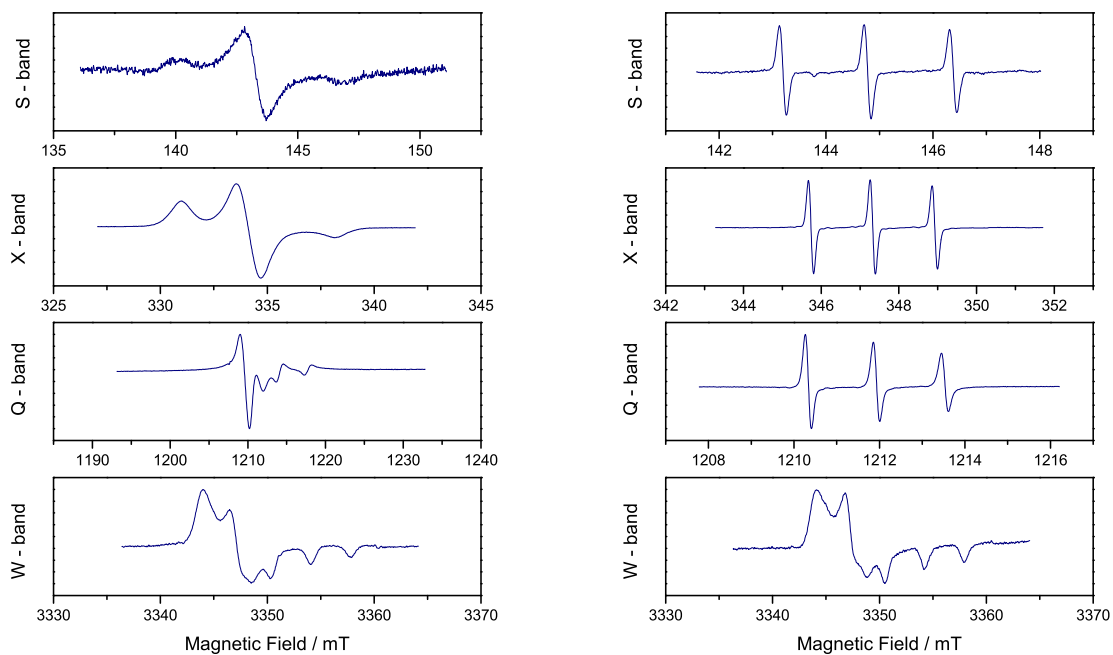
(b) Room temperature (298 K)

Figure 5.4 cw-EPR results for colicin E9 DNase S30C^{IA1} / S80C^{IA1} at S, X and Q - band

5.2.5. Colicin E9 DNase E66C^{IA1}Figure 5.5 cw-EPR results for colicin E9 DNase E66C^{IA1} at S, X and Q - band5.2.6. Colicin E9 DNase D20C^{IA1} / E66^{IA1}Figure 5.6 cw-EPR results for colicin E9 DNase D20C^{IA1} / E66^{IA1} at S, X and Q - band

5. Spin label characterisation

5.2.7. Im9 C23^{IA1}

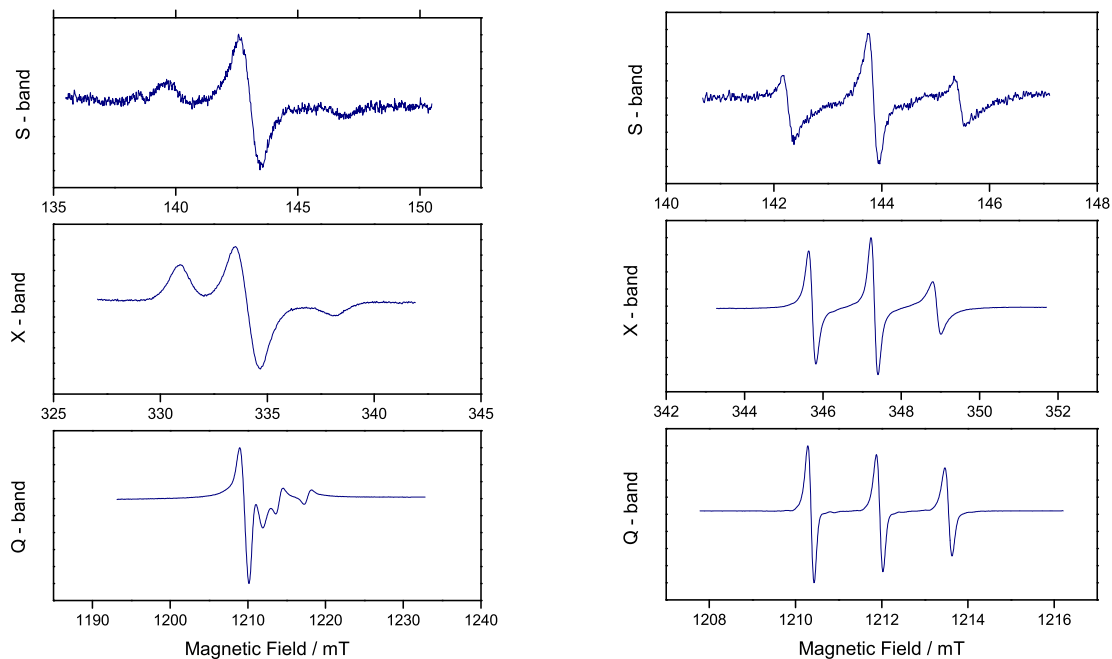


(a) Low temperature (80 K)

(b) High temperature (298 K, W-band at 200 K)

Figure 5.7 cw-EPR results for Im9 at S, X, Q and W - band

5.2.8. Im9 C23S / S6C^{IA1}



(a) Low temperature (80 K)

(b) Room temperature (298 K)

Figure 5.8 cw-EPR results for Im9 C23S / S6C^{IA1} at S, X and Q - band

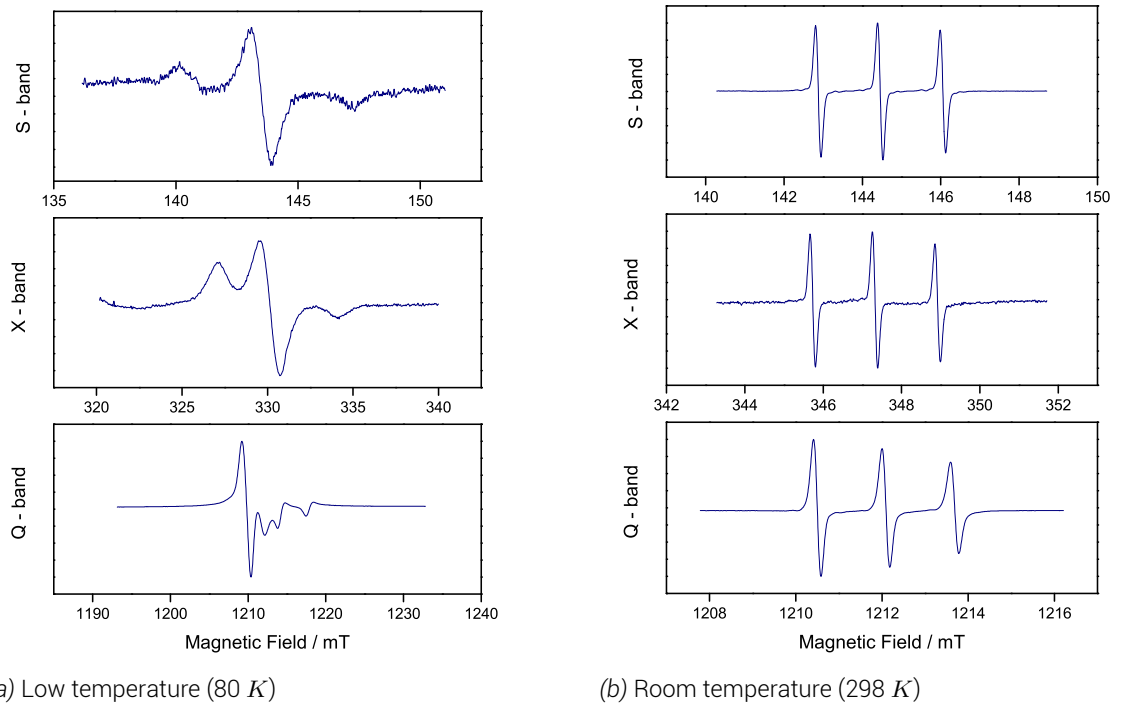
5.2.9. $\text{Im9 S6C}^{IA1} / \text{C23}^{IA1}$ 

Figure 5.9 cw-EPR results for $\text{Im9 S6C}^{IA1} / \text{C23}^{IA1}$ at S, X and Q - band

5.3. Spin label properties

Having collected each cw-EPR spectrum, each experiment was loaded into MatLab and fitted with the EasySpin package to extract its spin properties. Normal nitroxide spin label conditions were assumed as the starting point of the fitting and to optimise this fitting process the MatLab package EasyRefiner was written which allowed for the queuing of experiments and through an iterative refining of parameters process minimised fitting time whilst maintaining a high degree of accuracy (§ 12.1.2 on page 214) .

Table 5.2 on the facing page shows the resulting g -factors and hyperfine constants for all colicin E9 DNase and Im9 samples from the EasySpin/EasyRefiner process. It can be seen that most spin label sites give very similar results, typical of that for nitroxide spin labels. Moreover, there is even less variation in g or A values for the same sample at different frequencies, reinforcing that both experimental and fitting methods are sound.

Protein, labelling position	Frequency / GH_z	g			A / MH_z		
		g_x	g_y	g_z	A_{xx}	A_{yy}	A_{zz}
E9 DNase S3C ^I A1	4	2.0220	2.0166	2.0088	22.500	22.500	87.501
	9	2.0219	2.0039	2.0000	22.499	22.209	86.715
	34	2.0223	2.0195	2.0023	24.697	23.254	86.443
E9 DNase S30C ^{RLA}	4	2.0228	2.0077	2.0046	22.450	22.324	87.054
	9	2.0228	1.9991	1.9921	22.036	22.343	87.222
	34	2.0218	2.0240	2.0067	25.031	22.000	87.883
E9 DNase S3C ^I A1 / S80C ^I A1	4	2.0230	2.0055	1.9959	22.418	22.396	87.499
	9	2.0218	2.0030	1.9820	21.406	21.792	86.537
	34	2.0224	2.0085	2.0084	24.752	24.387	85.362
E9 DNase S30C ^I A1 / S80C ^I A1	4	2.0219	2.0070	2.0033	22.451	22.500	87.499
	9	2.0228	2.0053	1.9891	22.500	21.717	87.501
	34	2.0221	2.0209	2.0065	23.117	23.492	87.044
E9 DNase E66C ^I A1	4	2.0222	2.0144	2.0065	21.977	22.500	87.496
	9	2.0222	2.0103	1.9881	22.254	22.468	87.425
	34	2.0221	2.0126	2.0058	24.744	21.961	87.766
E9 DNase D20C ^I A1 / E66C ^I A1	4	2.0224	2.0129	2.0021	22.302	19.776	85.866
	9	2.0220	2.0094	1.9922	22.500	22.500	87.379
	34	2.0224	2.0026	1.9962	24.844	22.140	86.883
Im9 C23 ^I A1	4	2.0219	2.0085	1.9998	21.861	22.489	87.470
	9	2.0223	2.0044	2.0097	22.286	21.172	87.491
	34	2.0228	2.0124	2.0003	21.294	24.608	88.322
Im9 C23S / S6C ^I A1	4	2.0228	2.0158	2.0009	22.499	22.500	86.777
	9	2.0223	2.0222	2.0133	22.435	22.478	87.500
	34	2.0226	2.0103	2.0006	22.962	23.440	88.225
Im9 S6C ^I A1 / C23 ^I A1	4	2.0224	2.0175	2.0111	22.500	21.995	87.313
	9	2.0229	1.9992	1.9967	22.490	22.501	86.933
	34	2.0222	2.0036	1.9983	23.724	24.295	86.300

Table 5.2. Low temperature fitting of g and A tensor for all colicin E9 DNase and Im9 mutants

5.4. Spin label accessibility

The g and hyperfine tensors of the nitroxide depend upon the spin distribution of the electron across the bond and atoms. This distribution will depend upon the polarity of the environment. In a hypothetical polar environment the oxygen could exist in the charged O^- state and be stabilised by the polar environment whilst the electron could be located upon the positively charged nitrogen atom, $N^{\bullet+}$. In a non-polar environment, the charged polar states cannot be stabilised by the environment and thus the nitroxide atoms will preferentially sit in an uncharged state and so the electron will reside mostly upon the oxygen, O^\bullet , with an uncharged nitrogen.

In a polar environment with the electron contributing most spin density to the nitrogen, ρ_N , the isotropic hyperfine coupling from the ^{14}N increases with the increased spin density in the p_z orbital (A_{zz}). As the hyperfine tensor increases the g tensor decreases from g_e , as spin density has been withdrawn from the high energy excited states upon the oxygen, which further changes the g tensor as the energy levels between the lone electron and the lone pairs on the oxygen changes. Furthermore, hydrogen bonding to the lone pairs on the oxygen causes a small amount of delocalisation of the lone pair orbitals and an adjustment in the g_{xx} tensor.

At equilibrium within a magnetic field there exists a Boltzmann distribution between spins aligned with the field and those against the field. At room temperature and magnetic fields associated with X-band frequencies, the distribution lies such that 50.04 % of the spin population is parallel with the external magnetic field. 49.96 % of the spin population is aligned anti-parallel to the field. Without this distribution of spins magnetic resonance would not be possible. With applied radiation of the correct frequency transitions between the two energy levels occur. The absorption from the lower energy state of photons is typically detected in EPR by disturbance in a standing wave.

At low microwave powers spins the probability of a transition between energy levels is low. As the microwave power is increased the probability of transition increases in a linear fashion, thus the amount of spins absorbing photons at any given point increases yielding an increase in signal. That is until the saturation point, at which time the spin populations of both levels and the transition rate between each is equalled resulting and there is no net change in of microwave power being applied to the sample matches the rate of excitation with the rate of decay. Any further increase in microwave power "saturates" the sample whereby the rate of decay exceeds the rate of excitation and results in a loss of signal.

By placing paramagnetic quenchers near a nitroxide spin label, Heisenberg exchange of unpaired electrons can occur between the nitroxide and the quencher. The quench rate will be directly proportional to the concentration of the quencher. Thus by introducing different forms of relaxation agent that preferentially adopt hydrophobic or hydrophilic environments and observing the change nitroxide spin label saturates can yield information as to its environment.

Typically a control experiment is performed whereby the sample is tested in a nitrogen atmosphere. Then the experiment is repeated again in the presense of oxygen, air is sufficiently oxygenated. Diatomic oxygen, with its 3 triplet states, preferentially sits inside of a lipophilic environments such as a phospholipid bilayer of a cell wall and will quench nitroxides sitting

in or near membranes. A third experiment with aqueous relaxation agents such as consist of transition metals stabilised in soluble ligands such as the uncharged $Ni(II)$ ethylenediaminediacetic acid (NiEDDA) or negatively charged $Cr(III)$ tris-oxalate ($[Cr(C_2O_4)_3]^{3-}$). At high concentrations the quenchers will cause line broadening through excessive exchange interaction. At lower concentrations the relaxation influence can be quantified using power saturation experiments. By comparing the peak-to-peak linewidth broadening and intensity of the centre peak in the 3 experiments against a standard such as DPPH (2,2-diphenyl-1-picrylhydrazyl) the accessibility of the spin label to an aqueous or lipid environment can be calculated.

Briefly, the rate of relaxation is purely through collisions of the quencher with the nitroxide. If the longitudinal relaxation time of the quencher is shorter than the lifetime of quencher/nitroxide complex then the exchange rate, W_{ex} , can be expressed as follows

$$\Delta \left(\frac{1}{T_1} \right) = \Delta \left(\frac{1}{T_2} \right) = W_{ex} = k_{ex} C \quad (5.1)$$

where k_{ex} is the exchange rate constant and C is the quencher concentration. In the case that the quencher concentration is kept low the exchange rate can be assumed to be much less than $1/T_2$. W_{ex} can be determined from the saturation curves by measuring the first derivative peak-to-peak amplitude, A , of the central line of the nitroxide spectrum ($m_I = 0$) against the applied microwave power, P using

$$A(P) = \frac{I\sqrt{P}}{\left[\frac{(1+(2^{1/\epsilon}-1))P}{P_{1/2}} \right]^\epsilon} \quad (5.2)$$

with I the power independent amplitude factor, $P_{1/2}$ the power at half saturation and ϵ the inhomogeneity parameter, typically 1.5, but can be as low as 0.5 for inhomogeneous samples. More correctly, $P_{1/2}$ is the microwave power when the peak-to-peak amplitude is reduced to half of its unsaturated maximum. Given by

$$P_{1/2} = \frac{\sqrt[3]{4} - 1}{\gamma_e^2 \Lambda^2 T_1 T_2} \quad (5.3)$$

where $\Lambda = B_1/\sqrt{P}$ is the conversion efficiency of the spectrometer resonator. When quencher concentration is low, exchange broadening is small and T_2 is constant despite the presence of the quencher. Thus the change in $P_{1/2}$ can be simplified to

$$\Delta P_{1/2} = \frac{(\sqrt[3]{4} - 1) W_{ex}}{\gamma_e^2 \Lambda^2 T_2} \quad (5.4)$$

Experimentally, this is simplified by plotting the square root of microwave power against the peak-to-peak amplitude of the central line of the nitroxide spectrum ($m_I = 0$) at that power. By replicating this across a range of microwave powers a saturation curve can be constructed. At low powers, an increase in microwave power yields a linear increase in peak-to-peak amplitude. Nearing saturation the linear relationship tails off to a peak, after which any further increase in microwave power results in an exponential decay in peak-to-peak amplitude. This gives a characteristic power saturation profile which can easily be fitted using

5. Spin label characterisation

$$Y' = \left[a \cdot \sqrt{P} \cdot \frac{1 + ((2^{1/b} - 1) \cdot P^2)}{P_{1/2}} \right]^{-b} \quad (5.5)$$

Where Y' is the peak-to-peak amplitude of the central line of the nitroxide spectrum ($m_I = 0$), a and b are fitting variables, P the microwave power and $P_{1/2}$ the power at half saturation.

$P_{1/2}$ values are obtained for the sample under control, lipophilic and hydrophilic quencher conditions as well as for the reference sample (DPPH) under the same conditions.

To remove the spectrometer dependency on Λ in equation 5.4 and sample dependency on T_2 , the $P_{1/2}$ is converted into a dimensionless accessibility parameter, Π , by comparison to the reference sample (DPPH), such that

$$\Pi = \frac{[\Delta P_{1/2}/\Delta B_{pp}]_{sample}}{[\Delta P_{1/2}/\Delta B_{pp}]_{reference}} = \alpha W_{ex} \quad (5.6)$$

Where Π is a dimensionless accessibility parameter (easily converted to a percentage) and ΔB_{pp} is the peak-to-peak broadening effect of saturation.

This technique was extensively developed by Dr. C. Altenbach of the Hubbell group on T4 lysozyme [81] and further extensively tested and implemented by the Steinhoff group with particular focus on transmembrane and membrane-interacting proteins [72].

Automated analysis of power saturation experiment data through to accessibility parameters using the equations and methods detailed by Altenbach [81] were implemented into a graphical user interface MatLab package, PowerSat (§ 12.1.3 on page 215). PowerSat whilst automatically calculating $P_{1/2}$ and Π values also generates and exports a visual representation of the peak-to-peak intensity against the square root of microwave power - the power saturation profile.

The numerical accessibility of each labelling position is presented in tables 5.3 and 5.4 for aqueous and lipid accessibility respectively. In each table addition columns are given showing the results of an MMM site scan analysis of the labelling position upon the PDB entry 2K5X model. MMM takes a molecular dynamics derived rotamer library based upon allowed bond angle and energies and calculates how many of these rotamers would be allowed without substantial steric hinderance from the protein structure. The higher the number of rotamers, then the greater the amount of freedom to the site, which in the case of a surface residue means more solvent accessibility. Obviously an internalised residue may have freedom to move within a pocket but not necessarily have solvent accessibility. Furthermore, MMM then calculates the distance that the nitroxide N-O bond covers amongst all of the allowed rotamers. A small value here shows that the nitroxide bond is restricted and is predicted to move very little, whilst a large distance corresponds to high flexibility of the nitroxide corresponding to high solvent accessibility.

In figure 5.10.a the centre line of the nitroxide spectrum has been plotted in 3 dimensional space with increasing microwave power shown in the z direction. An increase in microwave power can be seen to yield a large increase in signal intensity, up until saturation at 5.0 mW . The increase in signal intensity also allows the satellite peaks of hyperfine coupling to ^{13}C to be seen on the shoulders.

Protein, labelling position	Accessibility		MMM	
	$P_{1/2}/mW^{1/2}$	Aqueous / %	No. rotamers	Distribution size / nm
E9 DNase S3C ^{IA1}	27.6	63.6	50	0.69
E9 DNase S30C ^{R1A}	14.8	79.4	46	0.49
E9 DNase E66C ^{IA1}	15.9	5.9	37	0.51
Im9 C23 ^{IA1}	33.6	60.2	3	0.18
Im9 C23S / S6C ^{IA1}	12.6	2.8	46	0.57

Table 5.3. Spin labelling aqueous accessibility summary from power saturation experiments for colicin E9 DNase and Im9 labelling positions with comparison to their MMM predicted labelling statistics.

Protein, labelling position	Accessibility		MMM	
	$P_{1/2}/mW^{1/2}$	Lipid / %	No. rotamers	Distribution size / nm
E9 DNase S3C ^{IA1}	8.8	16.6	50	0.69
E9 DNase S30C ^{R1A}	41.3	14.5	46	0.49
E9 DNase E66C ^{IA1}	15.4	4.5	37	0.51
Im9 C23 ^{IA1}	12.6	0.1	3	0.18
Im9 C23S / S6C ^{IA1}	12.5	2.5	46	0.57

Table 5.4. Spin labelling lipid accessibility summary from power saturation experiments for colicin E9 DNase and Im9 labelling positions with comparison to their MMM predicted labelling statistics.

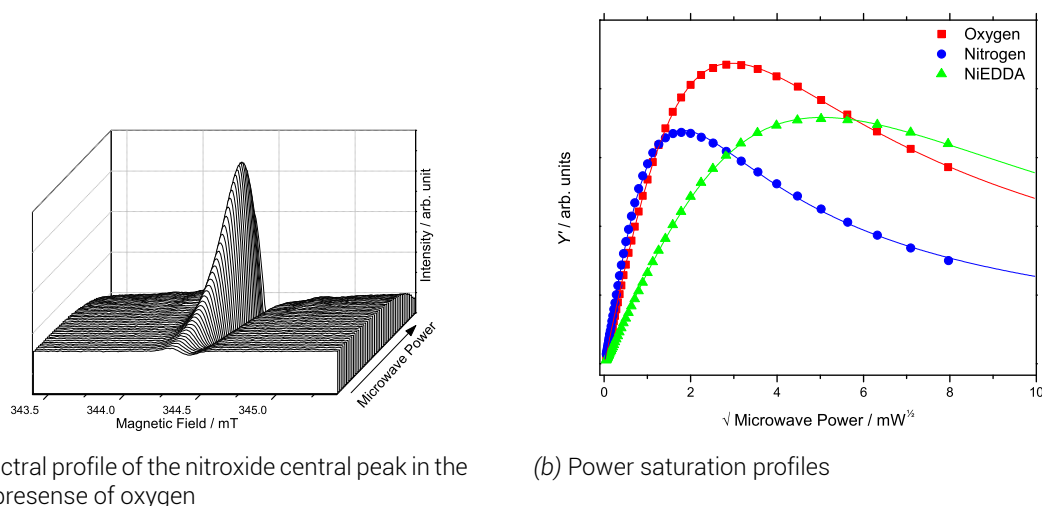


Figure 5.10 Power saturation profile of colicin E9 DNase S3C^{IA1} a) shows each microwave power slice of the central nitroxide peak in the oxygen experiment b) shows the power saturation profiles of samples of colicin E9 DNase S3C^{IA1} in the presence of a nitrogen atmosphere (blue circles), atmospheric oxygen (red squares) and a nitrogen atmosphere with 1 mM NiEDDA (green triangles). Lines are coloured to each experiment and show the fit of data using equation 5.5 and PowerSat's Lorentzian derivative fitting mode. 100 μ M protein, 298 K.

5. Spin label characterisation

As this spin label position is very much solvent exposed the NiEDDA curve in figure 5.10.b can be seen to be of less intensity due to the presence of a strong relaxant in solution stopping the maximum signal increase from microwave power. But, as the nitroxide spins relax that much faster the amount of power required to achieve saturation is much higher at 25.2 mW . This change can be seen in the $P_{1/2}$ values as when oxygen is used saturation occurs at a low value of 8.8 $mW^{1/2}$, but with NiEDDA the $P_{1/2}$ vastly increases to 27.6 $mW^{1/2}$. This corresponds to 16.6 % accessibility to a lipophilic environment and 63.6 % accessibility to a aqueous environment. These numbers align well with the large numbers of expected MMM rotamers and the large sample space that these rotamers cover.

In figure 5.11 the other single cysteine spin labelled mutants are shown. Figure 5.11.a shows colicin E9 DNase S30C^{R1A}, another solvent exposed site and as with the colicin E9 DNase S3C^{IA1} the oxygen and NiEDDA curves are above the nitrogen control, with the NiEDDA curve never becoming as intense as the oxygen, but saturating at a higher power with a $P_{1/2}$ of 41.3 $mW^{1/2}$.

It is interesting to note that the aqueous accessibility of colicin E9 DNase S30C^{R1A} is higher than that of colicin E9 DNase S3C^{IA1}. This is despite the MTSL label having one less atom in the linker chain between nitroxide and protein backbone, and the nitroxide ring being more rigid than IA-PROXYL due to the double bond between C_3 and C_4 (see figure 8.5 on page 149). The MMM assessment of this position also predicts for their to be fewer rotamers and for those rotamers to have a smaller spread of the NO position than the S3C^{IA1} label.

Figure 5.11.c shows Im9 C23^{IA1}, again a solvent exposed site and the curves show a similar tale but are not as far removed from the nitrogen control curve. This suggests that whilst being solvent exposed it is not as accessible as the colicin E9 DNase S30C^{R1A} position. This is reflected in the calculated aqueous accessibility being only 60.2 % for Im9 compared to the colicin E9 DNase S30C^{R1A} position's 79.4 % accessibility (table 5.3). Perhaps most shockingly, MMM predicts this to be a very restricted site with only 3 rotamers being viable, these having a very small 0.18 nm distribution of the NO bond. This predicted restriction could explain the low (5.3 %) labelling efficiency seen previously.

Meanwhile, in figure 5.11.b the saturation profiles for colicin E9 DNase E66C^{IA1} can be seen. This site from X-ray crystallography and NMR structures is an internal residue and is expected to have low aqueous accessibility. This was reflected during labelling when this protein could only be successfully spin labelled in strong denaturing conditions. As a result the saturation curves look almost identical in the presense of either relaxant to the nitrogen control, implying a low accessibility to either, which when calculated was less than 6 %. The MMM analysis for this site does show a reasonable number of rotamers and NO distance distribution. This is as a result of a small pore inside the protein in which the label could reside, open inspection the NO bond distribution is artifically as a result of the label being in close proximity to the other side of the pore, forcing the label to sit in two opposing positions.

Figure 5.11.d shows the Im9 S6C^{IA1} labelling position. This position was chosen as it is near the flexible end of protein and from published structures should be solvent accessible to provide good labelling efficiency. As a result, this position shows a large rotamer distribution in MMM analysis with a good spread of distances of the NO bond. However, the power saturation profile would suggest otherwise. Each experimental curve can almost be overlaid, corresponding to calculated accessibilities of less than 3 %. Whilst, this site was originally

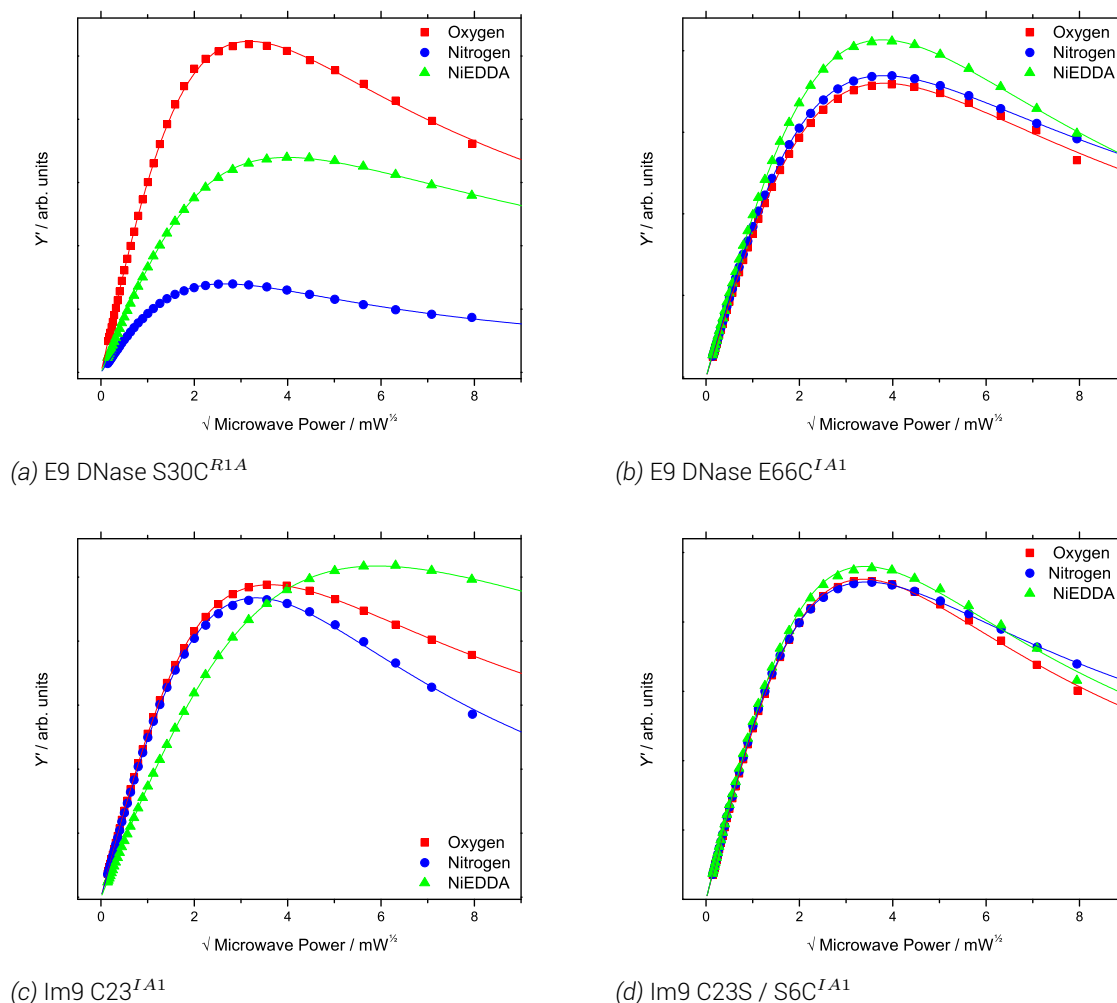


Figure 5.11 Power saturation profiles of other labelling sites. Each shows the power saturation profiles of the sample in the presence of a nitrogen atmosphere (blue circles), atmospheric oxygen (red squares) and a nitrogen atmosphere with 1 mM NiEDDA (green triangles). Lines are coloured to each experiment and show the fit of data using equation 5.5 and PowerSat's Lorentzian derivative fitting mode. 100 μ M protein, 298 K.

selected as a predicted good site for labelling the accessibility measurements suggest a site with little accessibility. If this is the case, it could explain the low (14.1 %) labelling efficiency of this site.

In addition to lone protein power saturation experiments, each experiment was repeated with with unlabelled cognate binding partner complexed. None of the experiments showed any significant change in accessibility suggesting that each labelling position is away from the docking interface.

5.5. Summary and discussion

From this investigation it has been found that several sites across the colicin E9 DNase - Im9 complex can be successfully spin labelled with nitroxide spin labels. Most of the labelling sites in the proteins expressed in the previous chapter were found to be able to be labelled with either MTSL or IA-PROXYL. However, the majority of this investigation quickly moved from using the more traditional spin label, MTSL to the IA-PROXYL label. The reason for the change in label primarily came only after exhaustive testing of MTSL and the very poor

5. Spin label characterisation

labelling efficiency of the proteins. In most cases with MTSL, the spin labelling efficiency was in the range of 10-20 %. By moving to the IA-PROXYL label in for all sites, except the colicin E9 DNase S30C, the labelling efficiency drastically improved. In the most dramatic cases such as the colicin E9 DNase double mutants (S3C/S80C and S30C/S80C) the spin label efficiency increased from around 10 - 20 % to 80 - 90 %. The labelling efficiencies using MTSL simply were not suitable for most sites to continue with in this investigation.

The notable exception was E9 DNase S30C mutant, which preferentially labelled with MTSL giving a spin labelling efficiency of 19.2 % with MTSL as opposed to being unable to give reliable results at all with IA-PROXYL. Given the poor yield and resulting very limited amounts of this protein available, it was decided to continue with MTSL for this mutant.

As most of these sites were selected for their solvent accessibility, and resulting expected high labelling efficiency it is unsurprising that these cw-EPR experiments at multiple frequencies and temperatures show little variation, due in large part to the short correlation times. Some sites, particularly on the Im9, show some slight restriction and a slightly longer correlation time with the high field peak broadening slightly. The high solvent accessibility was confirmed by power saturation experiments, with some labelling positions as high as 80 % water accessible.

The surprises here were the colicin E9 DNase E66C^{I^{A1}}, which despite the difficulty in labelling and predicted internalised position, showed low solvent accessibility but retained a low correlation time associated with high mobility. The other surprise was the Im9 S6C position which, whilst predicted to be very solvent exposed and a low correlation time associated with high label motility, showed very low solvent accessibility in the power saturation experiment. Though this could merely be representing the fact that colicin E9 DNase in solution shows some amount of intrinsically disordered protein characteristics, this molecular movement artificially decreases the rotational correlation time.

The largest problem that was encountered in spin labelling was the proteins themselves. Many of the spin labelling positions were that of solvent exposed positions. These positions whilst designed to allow access of free spin label to react with the exposed cysteine side chains, in such small, globular proteins also allowed the formation of homo-dimers. The only way to remove the disulphide linked homo-dimers was with the use of a reducing agent, but this reducing agent would indiscriminately reduce the cysteine - spin label bond we were trying to form. Alternative strategies such as low protein concentrations, anaerobic environments, pH shifts, reaction time did little to increase the labelling efficiency with MTSL. IA-PROXYL with its iodo leaving group however, reacts with the exposed cysteine sulphur atom in a non-reversible manner and greatly increased the labelling efficiency.

6. Echo optimization

It is possible to use pulsed EPR methods to elucidate distances between spin labels by exploiting the dipolar coupling between them. In order to do this it is required to pulse one of the spin labels and observe the induced reaction on the other. Due to the rapid relaxation, sample size and small difference in spin populations it is necessary to repeat the experiment many times and average the results to ensure high resolution data. Furthermore, as the electron is highly sensitive to its environment, the quality of the data obtained relies upon the correct experimental conditions.

For this reason it is necessary to optimise the conditions to maximise the transverse (T_2) relaxation time, the main contribution to echo longevity. Echoes which last longer before decoherence yield more observable data with each scan, which in turn, greatly reduces experiment time. Too short a transverse spin relaxation time yields echoes shorter than the periodicity of electron-electron dipolar oscillations and render the PELDOR experiment useless. This is not to say that spin-lattice (T_1) relaxation times and other factors were ignored, but they were not the primary focus of this investigation.

In this investigation the assumption was made that the optimisation of conditions and experimental parameters for one spin label site of colicin E9 DNase would be applicable to all the spin label sites due to their high solvent accessibility. Therefore optimum conditions were found using colicin E9 DNase S3C^{IA1}.

6.1. Spectra fitting

Using particular techniques and in this case, pulse sequences allows for the collection of raw data that contains information regarding to the relaxation of the system. To extract the relaxation times from each raw spectrum requires fitting the raw data to idealised relaxation models. As a relaxation is being measured it is important to ensure that the experiment is sufficiently long such that the system is completely relaxed, fitting data that has yet to reach thermal equilibrium could artificially generate false fitting results. To ensure accuracy a zero-filling method can be employed where the fit is compared to a second fitting of a dataset with an addition data points at later time points all with zero intensity.

In the case of transverse relaxation times, relaxation follows an exponential decay. By fitting the data to a single parameter exponential decay:

$$y = a(1 - b)^x \quad (6.1)$$

Where, y , is the decaying quantity, a , the original signal intensity before decay, b , the decay rate and x , time. In the case, of a transverse relaxation experiment the decay rate reflects the transverse relaxation.

6. Echo optimization

Complexity arises in these experiments as the electron is not isolated from the environment. Commonly in nitroxide samples a strong hyperfine interaction is seen with the solvent. The hyperfine coupling results in an oscillation of the relaxation with a periodicity related to the Larmor frequency of the nuclei. With this modulation in the relaxation fitting can be conducted using different approaches.

In the first approach the spectrum is considered as a single dataset and fitting is conducted with all data points using the assumption that the integral of positive modulations from the fit will equal the integral of negative modulations from the fit.

The second approach uses all of the dataset towards the end of the experiment, but uses only the maxima peaks of the hyperfine modulations to calculate the decay. The benefit of this method is that it will underestimate the relaxation time as a result of the rate of decay being effectively higher. This can be beneficial when later experiments rely upon the experimentally derived relaxation time as it gives the shortest usable time.

The third method, advocated by Eaton and Eaton [REF] is fit an exponential decay to only the data after the hyperfine modulations. Hyperfine oscillations typically relax away before the sample reaches thermal equilibrium. This gives the advantage that the hyperfine interaction does not interfere with the data analysis. However, in order to detect hyperfine interaction free signal requires strong, long lived signals but also allows for larger margins for error as the signal intensity used in the fitting process is only a small percentage of the total signal of the receiver.

When these three different methods are used on experimental data the difference between them quickly becomes apparent as shown in figure 6.1. Here the T_2 relaxation experiment of colicin E9 S3C^{I A1} in water at 4 K is shown. The data, as would be expected, shows obvious hyperfine modulations with a periodicity corresponding to the proton Larmor frequency. The hyperfine modulations decay faster than the overall signal and as a result the signal without detectable hyperfine modulations ($>2 \mu s$) can be fitted. Fitting a single parameter exponential decay to this data (shown in red) results in a transverse relaxation time of 471.6 ns (± 20.2 ns) and a relative signal amplitude of 179000. Clearly, whilst this fit corresponds well to data at later times it in no way reflects the experimental situation at the start, when the data is most significant. Using the assumption that hyperfine modulations average to give a consistent signal yields a much better fit to the experimental data (shown in green). Here the transverse relaxation time is calculated to be 511.6 ns (± 7.3 ns) with a relative signal amplitude of 511000. Finally, the method in which only the maxima of the hyperfine modulations are used for fitting initially and once they have decayed the rest of the data is used (shown in blue) results in a transverse relaxation time is calculated to be 507.7 ns (± 2.7 ns) with a relative signal amplitude of 570000. Here the fit is good to the data (that is being considered) and yields a relaxation time that is both reasonable as it reflects accurately the data, but it also gives the lowest possible value reflecting the fastest relaxing species within the sample.

As a result, this investigation uses the hyperfine maxima method of fitting for transverse relaxation time measurement, where only the maxima data points of each hyperfine modulation are considered. In doing so the obtained relaxation times give a lower limit to the real transverse relaxation time which is more beneficial when setting the observation window in the later PELDOR experiments.

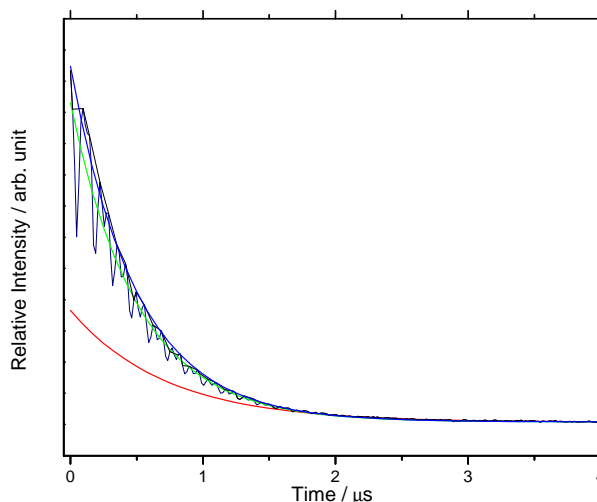


Figure 6.1 Possible fitting methods of the T_2 relaxation time, shown here E9 DNase S3C in water at 4 K. For visual clarity, this figure has been cropped at 4.0 μs but fitting was conducted with data up to 8.5 μs . Shown in navy is the raw data, in red the fitting line associated with only fitting data free from hyperfine modulations (data after 2.0 μs), in green the fitting line associated with considering all data points and in blue the fitting line associated with only taking the maxima of each hyperfine modulation as well as all data after the hyperfine modulations cannot be detected. 32 scans, 4 shots/point, 1024 points, 15000 μs shot rep. time, 3438 G, 9.6635 GHz, 5.8 dB attenuation, 200 MHz video bandwidth, 66 dB video gain. 100 μM protein in H_2O .

In the case of longitudinal relaxation experiments, each spectrum is considered as a complete data set. As an inversion recovery methodology was used to observe the relaxation time is fitted to an inverse exponential decay where the data set is inverted by multiplication of -1 before being fitted to a single parameter exponential decay, again of form:

$$y = a(1 - b)^x \quad (6.2)$$

Fitting of spectra results in the determination of two variables; the relaxation time (the rate of change with respect to time) and the relaxation intensity (the rate of change with respect to the initial intensity at time zero). The relaxation time for each spectrum is plotted with the relaxation times of the other spectra in a single plot. The relaxation intensity measurement for each spectra, being of arbitrary unit, is normalised against the largest signal amplitude before being plotted. In all cases the fitting procedure records the 95% confidence bands and these are present as error bars on associated plots.

By showing both the relaxation time as well as the fitted and normalised signal intensity allows for the determination of not only ideal relaxation times but signal. A desirable relaxation time, with little signal is not preferential to a less desired relaxation time with a larger intensity as the required extended signal averaging time to obtain a sufficient signal-to-noise ratio, far exceeds the benefit of the relaxation time.

6.2. Temperature optimization

Temperature is a key component to optimise. As the sample is cooled, energy is removed which reduces the molecular motion, of not only the bonds of the spin label, but the protein backbone as well. As motion is reduced an increase in the relaxation time would be expected

6. Echo optimization

as energy is not dissipated through motion. At low temperatures, certain conformations of the folding landscape begin to be eliminated as their energy level is no longer supported. Approaching absolute zero it is expected that all other conformations are eliminated leaving only the lowest possible energy state, which may or may not reflect the biological solution. Further it is known that the signal-to-noise ratio in echo experiments varies with temperature as a result of the temperature dependence to the polarisation of the spins. At temperatures when $k_B T > h\nu_{mw}$ polarisation is inversely proportional to temperature [49].

6.2.1. T_2 relaxation time

The experiment was started at 4 K , as the lowest feasible stable temperature to absolute zero with the experimental setup. The temperature was then raised to 10 K with further 10 K increments until an echo could not be measured.

As expected, the lowest temperature with the lowest energy gave the longest transverse relaxation time as all relaxation mechanisms slow. This had the additional advantage of giving the greatest amplitude signal. At 4 K , with no other experimental optimization, a T_2 of 507.8 ns was achieved with a large intensity, and very low signal-to-noise (SNR) after 32 scans. Subtraction of the background and Fourier transformation of the peaks of this experiment (figure 6.3), reveals two distinct peaks. The peak at 14.7 MHz matches that of the proton Larmour frequency ($42.6 MHz/T$ [104, 6]) at the experimental magnetic field. The second peak at 7.3 MHz is exactly twice the Larmour frequency and is largely a result of the background subtraction method.

Between 4 and 80 K there is a near linear relationship between diminishing transverse relaxation time and increased temperature. The amplitude of the echoes vary little between 10 and 80 K before dramatically increasing at 4 K . Suggesting a shift in the Boltzmann distribution of spins approaching very cold temperatures, allowing for the excitation of many more electron spins, seen experimentally as an increase in signal.

The 90 K data point at first glance appears erroneous. After the data point was repeated in triplicate it was concluded to be significant. In this instance the fitting gives values that are artificially distinct from the others in the temperature assay. The transverse relaxation time at this temperature is much higher than the series would expect, at 507.7 ns and is only 0.1 ns lower than the relaxation time shown at 4 K . However, the intensity is very low as displayed in figures 6.2.a and 6.2.c. The very shallow decay in the first 500 ns of this curve gives a very low signal amplitude (ie. rate of decay), but this artificially creates an incredibly long relaxation time. Realistically, despite the long calculated relaxation time this would be a terrible experimental to select as the signal intensity is so low that the experiment would require huge experimental times to achieve a reasonable signal-to-noise ratio.

Upon further investigation of the 90 K data point, it was found that this temperature at atmospheric pressure coincides with a phase transition of ice. Whereby, as the ice warms it moves from ice XI (an orthorhombic structure) to ice I_c (a metastable cubic crystalline structure) [2, 33, 64]. It is likely that the coexistence of two phases at equilibrium negates the signal from the echo as any energy put into the sample at this point shifts the equilibrium and the energy is used in the phase transition, rather than in emission of photons to the receiver.

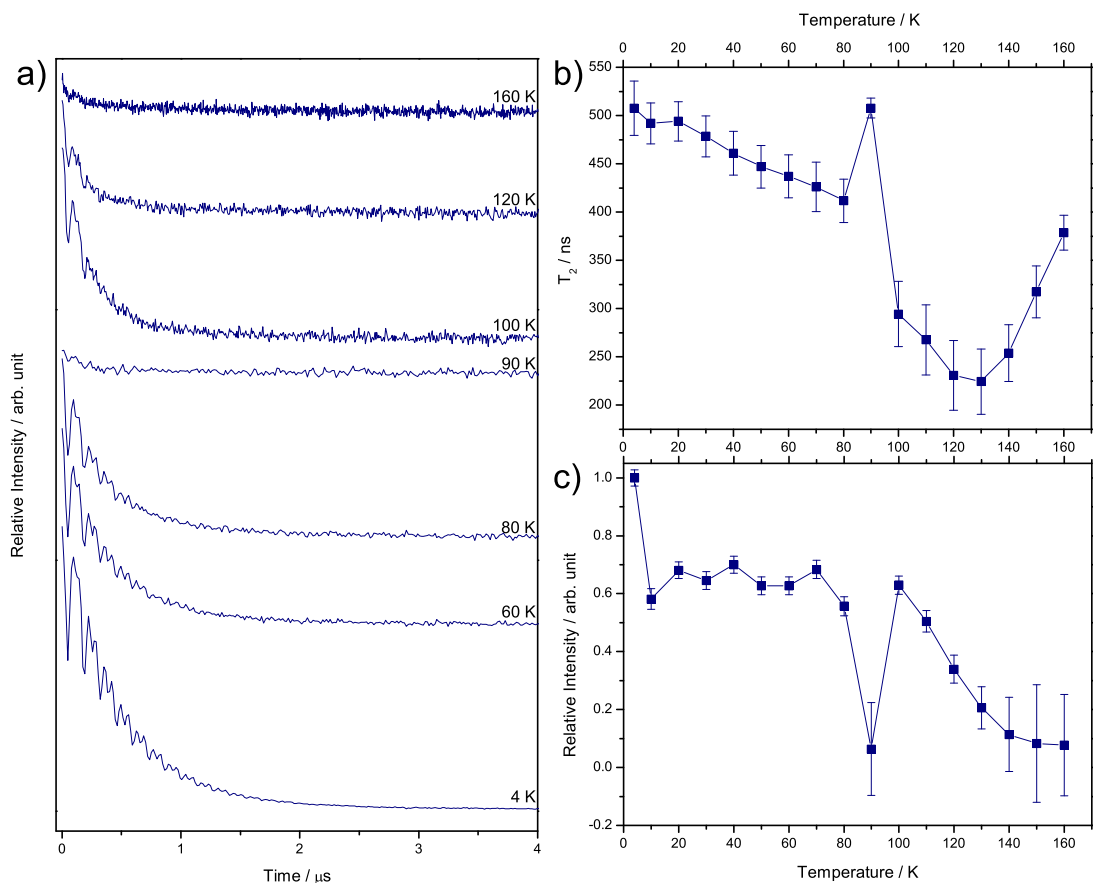


Figure 6.2 T_2 temperature optimization of E9 DNase S3C in water. a) echo decay traces showing key temperatures b) fitted T_2 times plotted against temperature c) fitted echo amplitude plotted against temperature. In b and c error bars show 95 % confidence bands of the fitting. 32 scans, 4 shots/point, 1024 points, 15000 μ s shot rep. time, 3438 G, 9.6635 GHz, 5.8 dB attenuation, 200 MHz video bandwidth, 66 dB video gain. 100 μ M protein in H_2O .

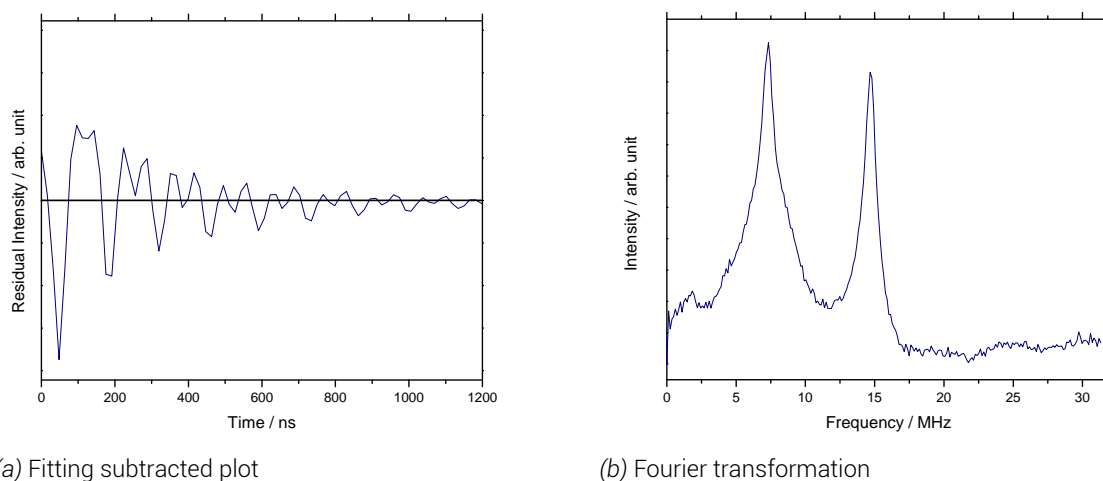


Figure 6.3 A Fourier transformation of the 4 K experiment shown in 6.2.a. 100 μ M colicin E9 DNase S3C in H_2O . a) the fitting subtracted data centred about the fitting line (zero). b) the Fourier transformation of the relaxation experiment.

6. Echo optimization

After the 90 K data point the T_2 relaxation time continues to drop as the temperature is increased. However both the relaxation time and signal intensity drop away faster than the experiments below 90 K .

At 130 K the signal intensity begins to level, but the transverse relaxation time approaches a minimum and begins to increase again with increased temperature. This can again be explained with phase transition chemistry of ice. At approximately 130 K and atmospheric pressure, ice transitions from ice I_c (metastable cubic crystalline) to ice I_h (hexagonal crystalline - common ice). The new phase of ice generates longer relaxation times as the rate of decay in the signal decreases, but this comes at the cost of less and less signal intensity, reflected in the signal-to-noise ratio and increasing error bars on the signal intensity plot.

The final data point shown at 160 K (figure 6.2.a), shows little intensity. This intensity artificially gives a high T_2 relaxation time when fitted. At 170 and 180 K no echo could be detected.

As a result of this experiment, it is clear that the lower the temperature that the experiment is conducted the better in terms of both the observed T_2 relaxation time and signal intensity, though good data is achievable at 80 K or below.

6.2.2. T_1 relaxation time

Again beginning at 4 K , as the nearest feasible temperature to absolute zero the T_1 relaxation time was measured. The temperature was then raised to 10 K and further increased in 10 K increments until an echo could not be measured.

The T_1 relaxation time is more sensitive to temperature increases than the T_2 relaxation times and it proved only possible to measure up to 80 K , though measurements at 90 and 100 K were attempted in light of the T_2 relaxation time data.

In figures 6.4.b and 6.4.c a near exponential increase in both the T_1 relaxation time and echo intensity as the temperature is dropped. At temperatures of 60 K and above it can be seen in figure 6.4.a that there is very little signal intensity from the inversion above that of the noise. This creates some uncertainty in the fitting seen in the error bars of figures 6.4.b and 6.4.c. However, having short longitudinal relaxation is preferential in that it means that the system returns to thermal equilibrium quickly and allows for faster experiment repetition. If long longitudinal relaxations are desired then these experiments again show that colder temperatures are preferential.

From this experiment the ideal temperature for further investigation was concluded to be around 40-50 K . At these temperatures the T_1 relaxation time is significantly reduced when compared to lower temperatures, for instance it is less than than a fifth of the relaxation time and intensity of that seen at 4 K .

6.3. pH optimization

Proteins have a biologically finite range of operation often within a narrow pH spectrum related to their function. Moving outside of the natural pH range can result in the loss of enzymatic function followed by a loss in secondary structure. Whilst most proteins will function

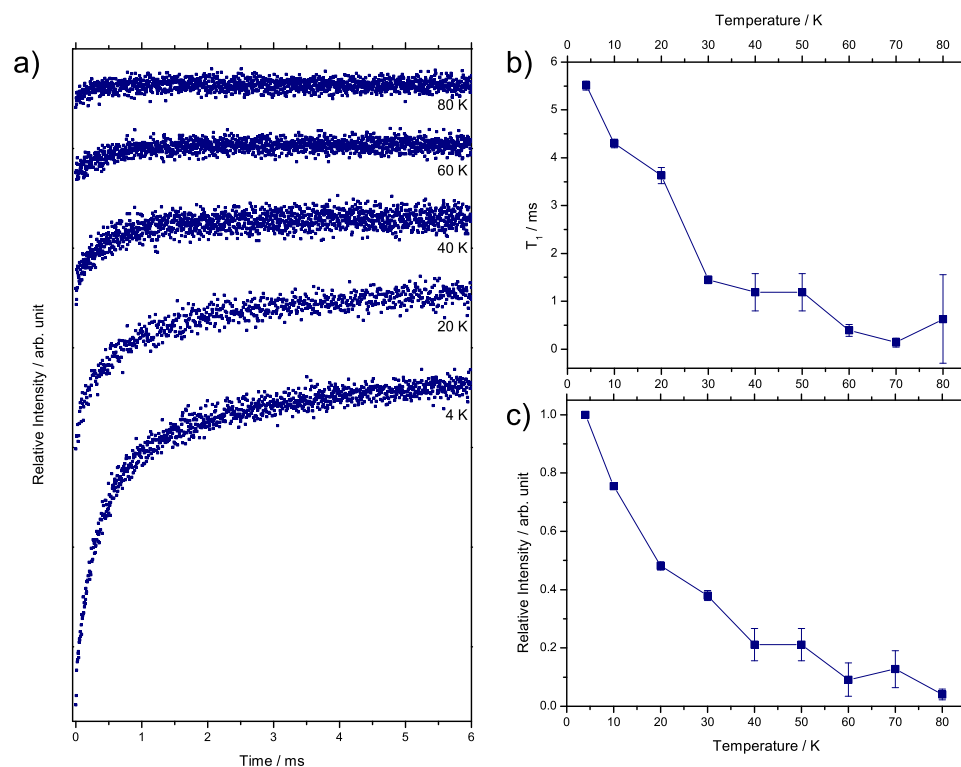


Figure 6.4 T_1 temperature optimization of E9 DNase S3C. a) echo decay traces showing key temperatures b) fitted T_1 times plotted against temperature c) fitted echo amplitude plotted against temperature. In b and c error bars show 95 % confidence bands of the fitting. 1 scan, 2 shots/point, 4000 points, 60496 μ s shot rep. time, 3439 G, 9.6642 GHz, 4.4 dB attenuation, 200 MHz video bandwidth, 63 dB video gain. 100 μ M protein in H_2O .

across a range of pHs there is usually an ideal pH at which rate and binding coefficients are much higher. For this reason most experiments with proteins are carefully buffered to ensure the sample does not move far from the desired pH thereby remaining in the most active state.

Previous studies [91] have shown that colicin E9 is typically biologically active at pH 7.4 (*in vivo*) but still has activity as low as pH 4.9 (*in vitro*). Obviously, whilst in this case the protein may continue to show activity at extreme pHs the spin label can be reduced to the O^- state or oxidised to the OH state, rendering it EPR silent. For this reason it was essential to test the bound spin label across the biological pH spectrum to ensure that an EPR signal could be seen, and that no particular pH gave a marked increase in signal.

Figure 6.5.a shows that there is no significant change in the echo decays between pH 5.0 and 9.0. This suggests that the pH change has little effect upon the environment around the spin label. As previous studies have shown there is still biological activity in this range, we would expect no loss or significant change in the colicin E9 DNase domain secondary structure.

When the echo decays are fitted (as in figure 6.5.b) a minimal change in the T_2 relaxation times is seen across the *in vitro* pH range with a difference of only 40 ns between the longest (at pH 7.5) and shortest (at pH 5.0). There is a definite optimum between pH 7.0 (494.0 ns) and pH 7.5 (494.5 ns), matching the *in vivo* optimal pH with a slight deterioration to the acidic and a faster deterioration to the basic pH regimes. The signal amplitude matches the T_2 relaxation times but shows that pH 7.0 is preferential to pH 7.5.

6.4. Deuteration optimization

Due to the non-linear relationship between the frequency of the observed signal derived from the spin-spin coupling and the spin-spin distance, the maximum range of distance measurements is often quoted as 80 \AA requiring up to $10 \mu\text{s}$ to obtain a full dipolar oscillation [49], whilst a distance of 100 \AA would require at least $20 \mu\text{s}$ and remains unobtainable in biological systems [103].

Echo dephasing in spin labelled protein systems is largely caused by the hyperfine coupling of the electron to protons in the surrounding environment and the consequent spin diffusion. It has been shown that by removing protons in solution and methyl groups within the protein via deuteration vastly increases the T_2 relaxation time [56].

It was therefore expected that as the sample became more deuterated that the T_2 relaxation time would increase whilst the ^1H hyperfine modulations seen in the oscillations would be replaced with ^2H ESEEM modulations as a result of the much larger quadrupole coupling in a deuteron. It is hypothesised that complete deuteration of sample and buffer may be able to extend the maximum PELDOR detection range to $100 - 130 \text{ \AA}$ [11].

In figure 6.6 colicin E9 DNase without deuteration shows a normal echo decay as seen in all of the previous experiments with hyperfine oscillations inside each echo oscillation. At 40 % deuteration the ESEEM effect can be seen to be appearing, with the third and fourth oscillations being higher than expected for an exponential decay. As the deuteration is further increased the ^2H ESEEM effect becomes more pronounced, but the ^1H hyperfine splitting of oscillations remains until 80 % deuteration; showing just how strong this effect is.

With every increase in deuteration level from 0 % an almost exponential increase in T_2 relaxation time is seen at the cost of some signal intensity. That is until 70% deuteration which shows an interesting, and repeatable, effect whereby the majority of signal intensity is lost which when fitted gives a high T_2 relaxation time in keeping with the trend, but could be artificially high as a result of fitting. It is likely that at 70 % there is a situation of where there is the highest level of both hyperfine and ESEEM relaxation resulting in relaxing the spin population so quickly as to remove most of the signal.

With further investigation of this experiment (shown in figure 6.7) it can be seen that the way the data is handled can make a large difference to calculated results. In red, is the single parameter exponential decay fit of the entire dataset, yielding a T_2 relaxation time of $1376.1 \pm 19.8 \text{ ns}$ with signal amplitude 87500. In green, is the fitting curve used in figures 6.6.b and 6.6.c where only the maxima of the proton hyperfine modulations are used in the fitting, where hyperfine modulations still exist in the signal. This yields a T_2 relaxation time of $1116.1 \pm 32.7 \text{ ns}$ with signal amplitude 102000, this fitting whilst having a shorter relaxation time has an increased amplitude which means that it appears closer to the experimental data even if the confidence range is larger. Finally, the line in blue shows the fitting when only the maxima of proton and deuteron modulations are included, in this instance no data points below 380 ns are used. The result of this fitting is a T_2 relaxation time of $978.3 \pm 26.6 \text{ ns}$ with signal amplitude 198000. Using the fitting parameters that consider both relaxation methods results in the large spike in figures 6.6.b and 6.6.c being reduced to little more than the 60 and 80 % points on either side.

Above 70 % deuteration the main contribution to the T_2 relaxation time is ESEEM and thus the

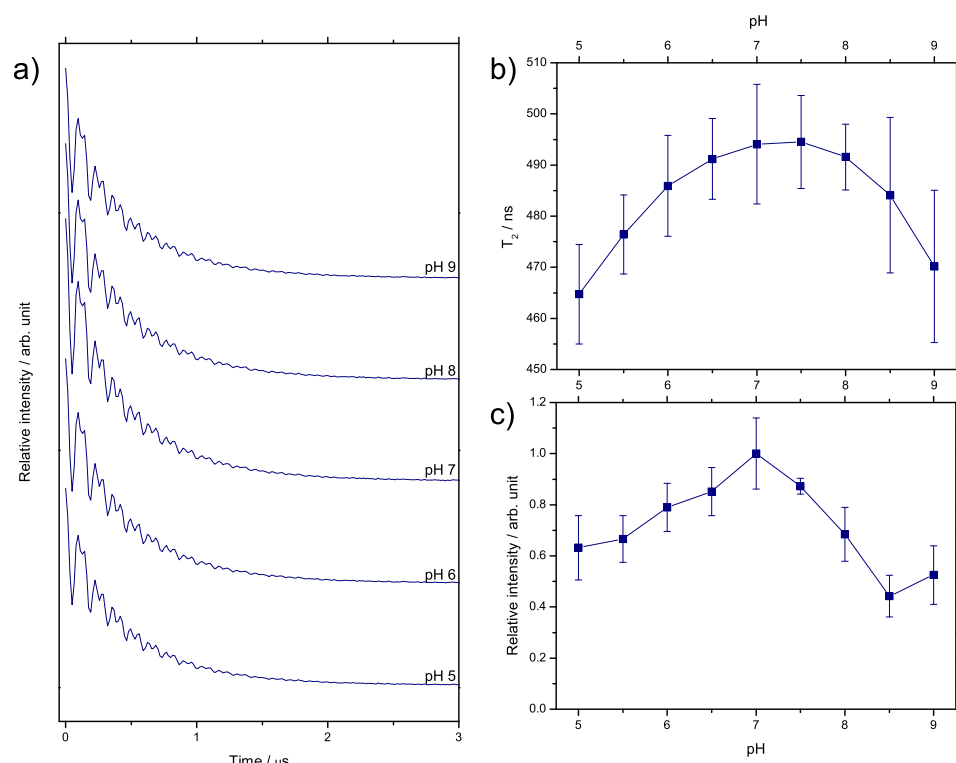


Figure 6.5 T_2 pH temperature optimization of E9 DNase S3C. a) echo decay traces showing every pH unit b) fitted T_2 times plotted against pH c) relative echo amplitude plotted against pH. In b and c error bars show 95 % confidence bands of the fitting. 15 K, 32 scans, 4 shots/point, 512 points, 15034 μs shot rep. time, 3438 G, 9.6645 GHz, 5.5 dB attenuation, 200 MHz video bandwidth, 42 dB video gain. 100 μM protein in H_2O .

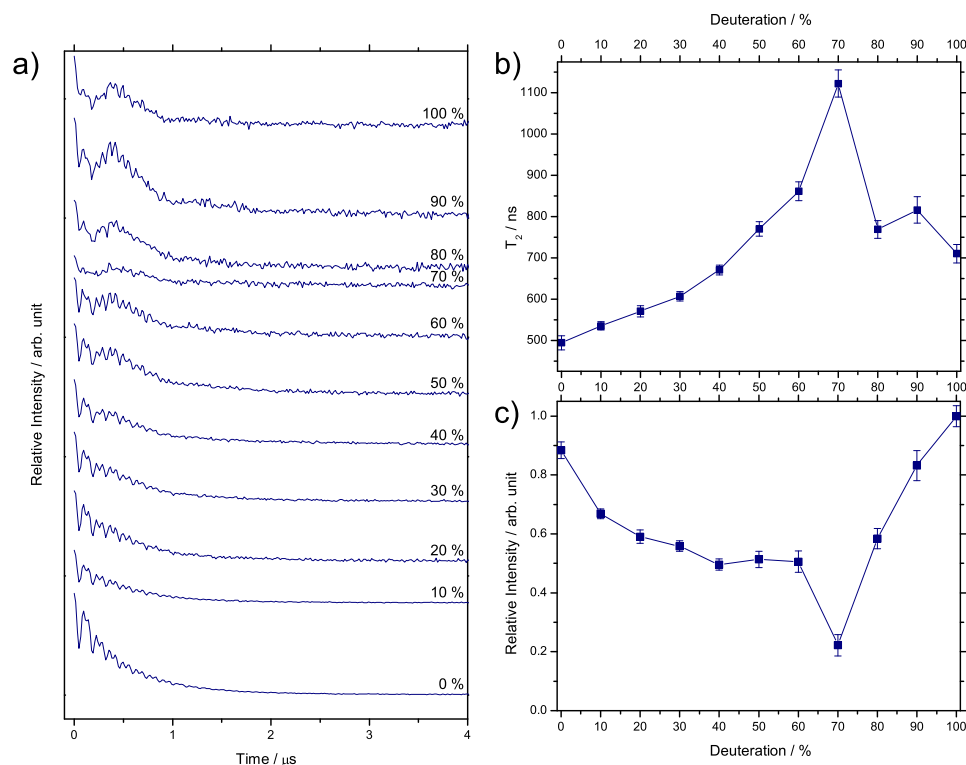


Figure 6.6 T_2 buffer deuteration optimization of E9 DNase S3C. a) echo decay traces showing buffer deuteration b) fitted T_2 times plotted against buffer deuteration c) fitted echo amplitude plotted against buffer deuteration. In b and c error bars show 95 % confidence bands of the fitting. 15 K, 32 scans, 4 shots/point, 512 points, 15034 μs shot rep. time, 3438 G, 9.6648 GHz, 5.5 dB attenuation, 200 MHz video bandwidth, 48 dB video gain. 100 μM protein in $\text{H}_2\text{O}/\text{D}_2\text{O}$.

6. Echo optimization

fitting is conducted to the ESEEM modulations only. As a result the relaxation time remains constant around 800 ns, but with further deuteration of the buffer shows a marked increase in signal intensity.

From this it can be seen that, conforming with the literature, increased levels of deuteration yields a longer transverse relaxation time. With increased echo coherence comes greater intensity echoes, so long as 70 % deuteration is avoided.

6.5. Glycerol optimization

Glycerol increases the viscosity of the sample which, given constant temperature, has been shown to linearly increase the molecular tumbling correlation time in proteins [39, 80]. The correlation time is a key factor to the relaxation time with fast molecular tumbling causing an averaging in the g and A anisotropies making it difficult to detect between the pulsed spin and the observed induced spin.

Glycerol slows the molecular motion of neighbouring methyl groups to which the unpaired electron is coupled, therefore further reducing methods by which the echo can rapidly decay.

It was expected, prior to the investigation, that even a small addition of glycerol would vastly increase the relaxation time and that it would continue to increase with more glycerol until a maximum was reached.

As figure 6.8 shows there is a large increase of 33.9 % in the T_2 relaxation time on the addition of 10 % glycerol from 2.77 to 3.71 μs . With further additions of glycerol the T_2 relaxation time continues to increase but at a much slower rate, until 70 % glycerol where the T_2 relaxation time levels out and reaches a maximum of 4.11 μs at 80 % glycerol.

The amplitude of the signal appears to vary wildly in figure 6.8.c but this is as a result of the fitting and presentation. All glycerol concentrations gave approximately equal intensities which when normalized to the highest and lowest decay, gives the spread an appearance of being much more diverse than when compared to the actual curves seen in figure 6.8.a.

A final sample preparation of 50 % glycerol was decided upon. Whilst, even the smallest addition of glycerol showed a large increase in the T_2 relaxation time, it is clear that as much glycerol as is possible is optimal. However, high glycerol sample preparations have a very high viscosity making them almost unmanageable to work with. It also requires a very concentrated aqueous protein sample to achieve a suitable protein concentration after the addition of glycerol. For this reason 50 % glycerol gave the optimum compromise between sample preparation and experimental result.

6.6. Glycerol deuteration optimization

In previous sections it has been shown that for optimal T_2 relaxation times, high levels of deuteration (§6.4) and glycerol concentration (§6.5) are required. However, glycerol (CH₂(OH)CH(OH)CH₂OH) as a simple polyol can be deuterated to different levels, completely (glycerol-d₈), through non-labile protons (glycerol-d₅) or through the labile protons of the alcohols (glycerol-d₃).

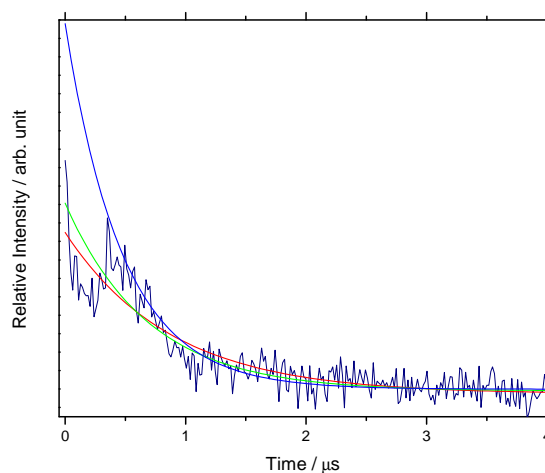


Figure 6.7 T_2 relaxation time determination of E9 DNase S3C^{I A1} at 70 % deuteration of the buffer. In navy is the raw data trace, cropped to show only the first 4.0 μ s. The red line is the fitting line obtained when the entire dataset is considered. The green line is the fitting line obtained if all of the dataset is considered but only the maxima peaks from proton hyperfine modulation are used in the fitting. The blue line shows the fitting line obtained when only the maxima hyperfine modulations of both proton and deuteron are used. 15 K, 32 scans, 4 shots/point, 512 points, 15034 μ s shot rep. time, 3438 G, 9.6648 GHz, 5.5 dB attenuation, 200 MHz video bandwidth, 48 dB video gain. 100 μ M protein in H_2O/D_2O .

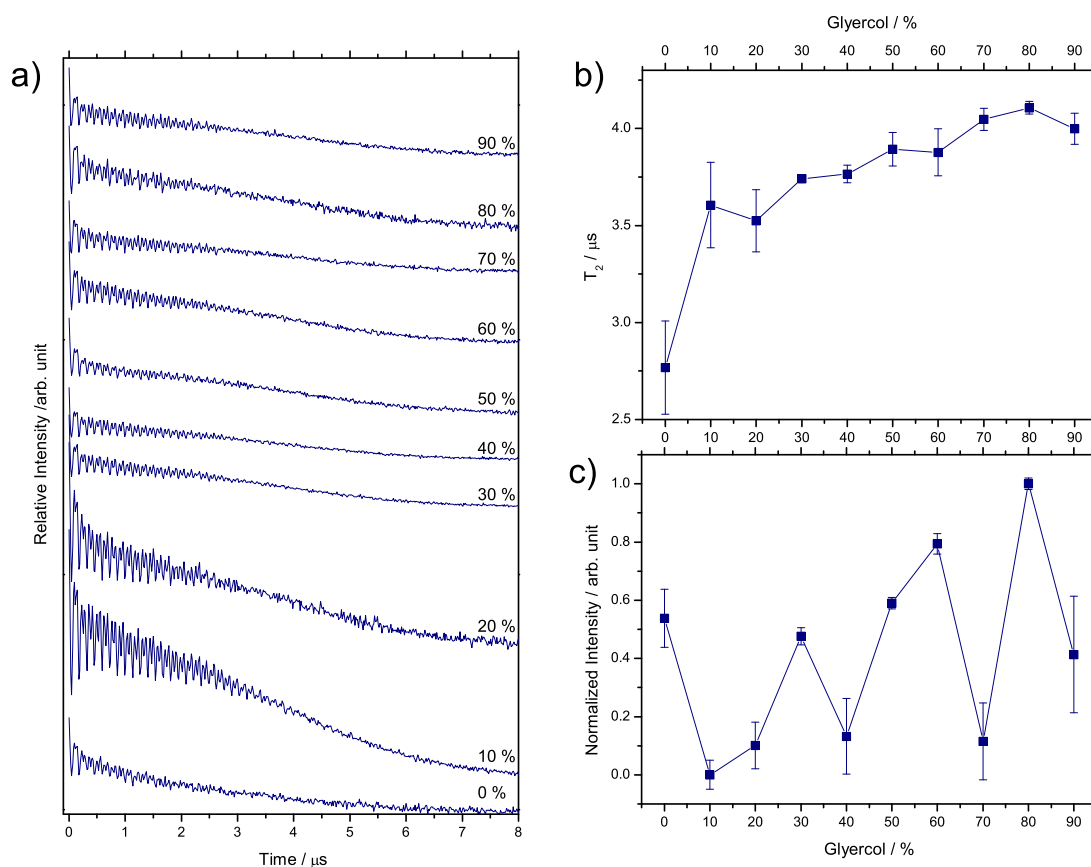


Figure 6.8 T_2 glycerol concentration optimization of E9 DNase S3C. a) echo decay traces showing key temperatures b) fitted T_2 times plotted against glycerol concentration c) fitted echo amplitude plotted against glycerol concentration. In b and c error bars show 95 % confidence bands of the fitting. 15 K, 32 scans, 4 shots/point, 1024 points, 15031 μ s shot rep. time, 3445 G, 9.7070 GHz, 7.5 dB attenuation, 20 MHz video bandwidth, 68 dB video gain. Sample made from 200 μ M protein in H_2O to give final sample as 100 μ M protein.

6. Echo optimization

As discussed in §6.4 proton concentration is inversely related to T_2 , namely, $[H^+] \propto 1/T_2$ [56]. Thus, if a sample were made of 50:50 D_2O :glycerol then, as the labile glycerol hydrogens exchange, then the concentration of free protons would increase.

To show the effect of glycerol and deuteration upon the sample and to highlight the contributions from the spin label and the protein, this investigation was conducted first upon free spin label before the label was attached to colicin E9 DNase S3C.

6.6.1. IA-PROXYL

Figure 6.9 confirms that the addition of glycerol is of paramount importance to achieve long T_2 relaxation times. Figure 6.9.b clearly shows that the addition of glycerol increases the T_2 relaxation time from $0.40 \mu s$ to $2.10 \mu s$ to $2.54 \mu s$ with glycerol- d_8 . When the proton concentration is further reduced by using a deuterated solvent (D_2O) then the T_2 relaxation time sharply increases from $0.40 \mu s$ to $3.32 \mu s$ with glycerol- d_3 and $8.34 \mu s$ with glycerol- d_8 . The combination of deuteration and glycerolisation yields a $17.5\times$ increase in T_2 relaxation time $8.7\times$ increase in modulation signal intensity.

Figure 6.9.a perhaps best shows the difference between 1H hyperfine coupling (best seen in Glycerol in H_2O) and 2H ESEEM oscillations (best seen in glycerol- d_8 in D_2O). On close inspection of figure 6.9.a the oscillations in the first few μs of the D_2O curve appears erroneous as it contains no ESEEM modulations, only some hyperfine. Whilst, the hyperfine modulations are understandable as the protein is still protonated, the ESEEM modulations simply cannot be seen as without the presence of glycerol the relaxation is so fast that the system has all but relaxed before the first ESEEM modulation.

Note that the glycerol- d_3 in D_2O echo despite a very low proton concentration still shows hyperfine oscillations upon the much longer ESEEM oscillations.

6.6.2. Colicin E9 DNase S3C

Given the results of the IA-PROXYL experiment, the spin labelled protein experiment was extended so that each solvent (H_2O/D_2O) saw every level of glycerol deuteration. Figure 6.10 shows a that combined deuterated solvent and glycerol medium yields longer T_2 relaxation times and signal intensities, but to a lesser extent than in the case of the free label ($7.8\times T_2$ relaxation time, $3.0\times$ intensity).

Figure 6.10.a whilst initially appear noisier than previous spectra, note that the abscissa is now $8 \mu s$, double that shown at the beginning of this chapter. Again the main contribution to relaxation from hyperfine and ESEEM effects as the sample becomes more deuterated is apparent, with the reduction in hyperfine relaxation lengthening the T_2 relaxation times (figure 6.10.b).

6.7. Sample cooling

EPR samples want to be in a frozen solution, a glass, not a solid. In order to achieve a good glass it is common practice for a sample to be mixed with glycerol and flash frozen in liquid

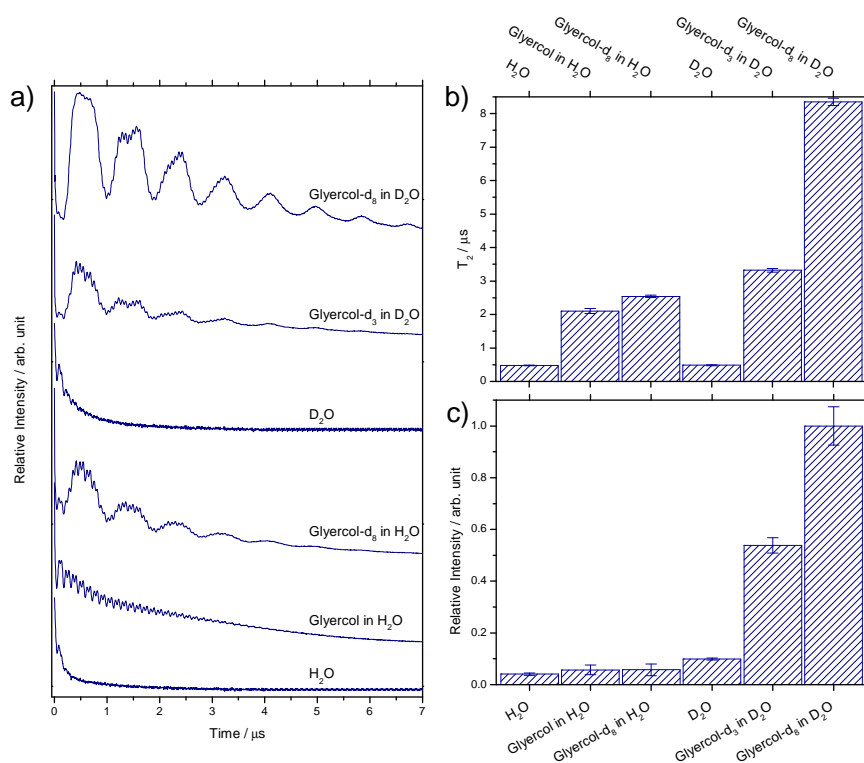


Figure 6.9 T_2 glycerol deuteration optimization of IA-PROXYL. a) echo decay traces b) fitted T_2 times plotted against glycerol deuteration c) fitted echo amplitude plotted against glycerol deuteration. In b and c error bars show 95 % confidence bands of the fitting. 50 K, 32 scans, 4 shots/point, 1024 points, 499.8 μ s shot rep. time, 3464 G, 9.7389 GHz, 2.8 dB attenuation, 20 MHz video bandwidth, 63 dB video gain. 100 μ M IA-PROXYL.

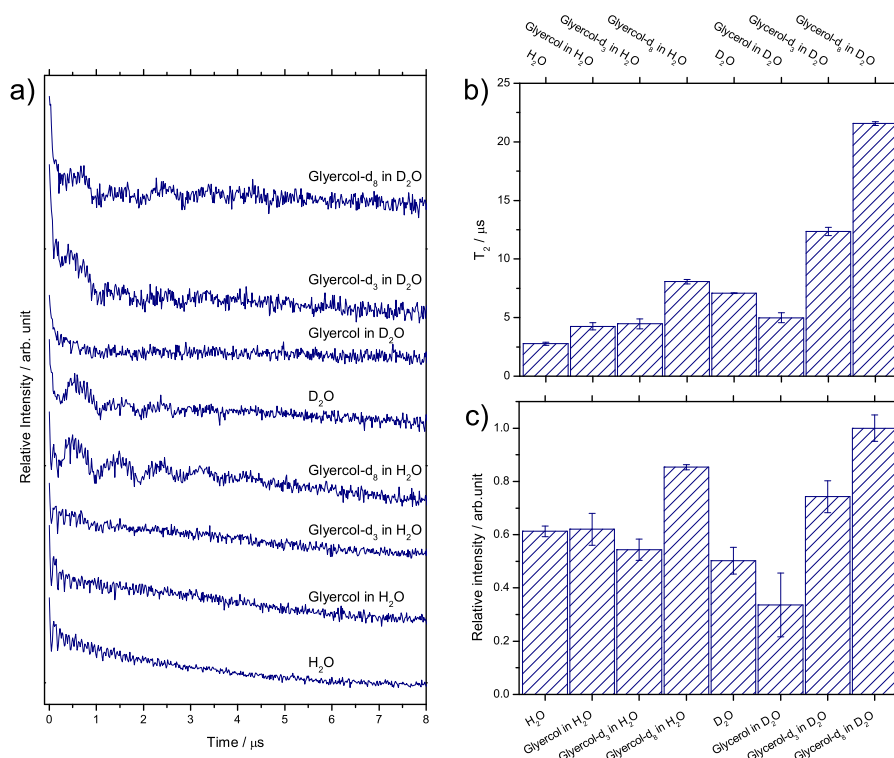


Figure 6.10 T_2 glycerol deuteration optimization of E9 DNase S3C. a) echo decay traces b) fitted T_2 times plotted against glycerol deuteration c) fitted echo amplitude plotted against glycerol deuteration. In b and c error bars show 95 % confidence bands of the fitting. 50 K, 64 scans, 4 shots/point, 1024 points, 15031 μ s shot rep. time, 3450 G, 9.7120 GHz, 5.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain. 100 μ M protein at 33 % labelling efficiency.

6. Echo optimization

nitrogen. This serves to both capture a "snapshot in time" of the sample and by creating a glass, there is low order in the lattice which minimising lattice relaxation pathways. In contrast, a slowly frozen sample in water should have a near perfect lattice with a high level of order, allowing for higher loss of magnetization via lattice relaxation. Additionally, as water freezes and forms the crystal lattice its volume increases which can create distortions within the protein structure.

The work of Dr. G. Smith and co-workers (University of St. Andrews, UK) using a specialised spectrometer, HIPER, where microwaves approach the sample in a top-down fashion have found that the quality of the sample glass surface greatly affects their signal-to-noise (SNR) ratio.

Here the T_1 and T_2 relaxation times of colicin E9 DNase S3C^{IA1} were measured under 3 conditions in 3 different environments. Colicin E9 DNase S3C^{IA1} samples were produced in water, in a 50% glycerol solution and a 50% glycerol-d₈ in D₂O. Ice samples were produced by slowly freezing the sample in the cryostat from 298 K to 50 K at 2 K/min. Glass samples were produced by flash freezing the sample in liquid nitrogen and being transferred to a cold cryostat at 50 K. Reannealed samples were produced by flash freezing the sample, and once inside the cryostat the temperature were raised at 5 K/min to just below the glass transition point (140 K [63]) and held for 20 min before being returned to 50 K at 2 K/min. Liquid nitrogen is a poor thermal conductor and the samples at 100 μ l are still relatively large, even with flash freezing any high density water area will see crystal lattice formation. By raising the sample to just below the glass transition (the maximum temperature that the glass can exist) a molten glass state is reached where the glass can freely move with almost liquid-like properties minimising any areas of high water concentration. The slow reduction in temperature ensures that the sample cools in a consistent manner so that no one area cools faster allowing lattice formation.

As expected, the longitudinal relaxation (T_1) (figure 6.11) vastly increases with the addition of glycerol. However, there is little difference between the deuteration level of the sample when the sample is slowly or flash frozen. In a protonated environment reannealing makes a small but unsubstantial change to the relaxation. However, in a fully deuterated environment the T_1 drops almost in half. We can only assume that the deuteration somehow affects the ability of hydrogen bonding in lattice formation.

Considering the transverse relaxation (T_2) (figure 6.12) we can again see that the addition of glycerol (a glass former) increases the relaxation time by ~ 8 fold regardless of the cooling method. Here we can also see that in a protonated environment there is little change to the relaxation time between cooling methods. However, in a deuterated environment a glass yields an increase in relaxation time over the ice, whilst the reannealed glass shows a significantly increased T_2 time of 1 μ s.

For samples where relaxation is very fast, this increase could prove vitally important. However, in this investigation the additional 2 hours required to reanneal each sample yields no gain. Most PELDOR distances under 8 nm can see sufficient oscillations in under 3 μ s, thus an increase in T_2 time from 5 μ s to 6 μ s simply increases the experiment time, where an additional 2 hours of sample averaging would be more beneficial.

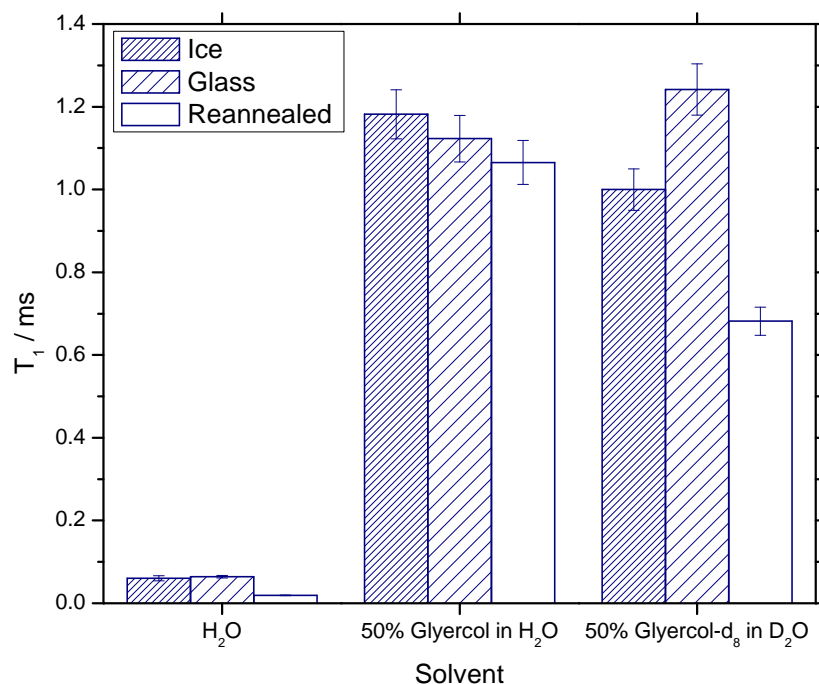


Figure 6.11 Effect of T_1 relaxation by different sample cooling methods of E9 DNase S3C. Error bars show 95 % confidence bands of the fitting. 50 K, 2 scans, 4 shots/point, 1024 points, 50000 μ s shot rep. time, 3455 G, 9.6903 GHz, 20 MHz video bandwidth, 60 dB video gain. 100 μ M protein.

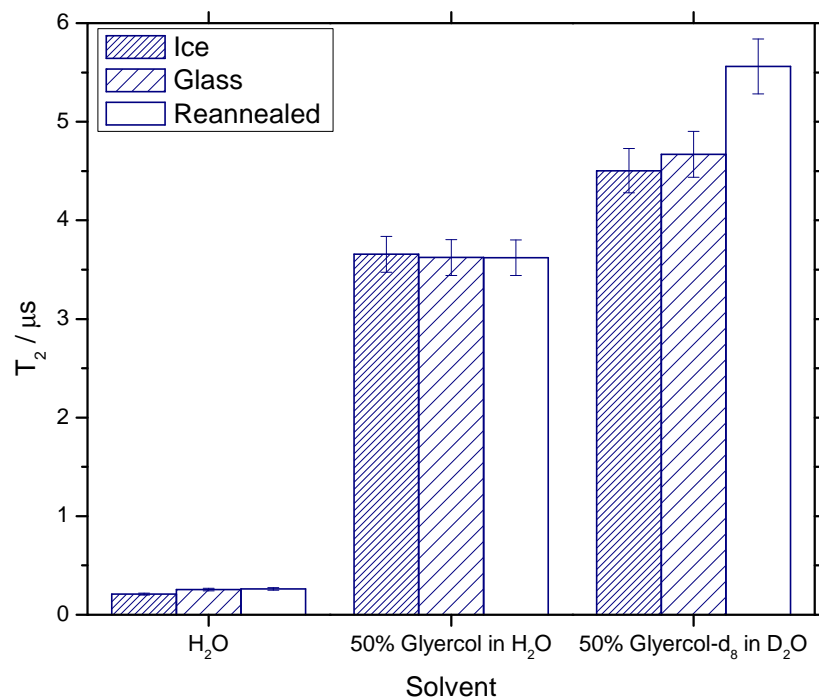


Figure 6.12 Effect of T_2 relaxation by different sample cooling methods of E9 DNase S3C. Error bars show 95 % confidence bands of the fitting. 50 K, 4 scans, 4 shots/point, 1024 points, 10000 μ s shot rep. time, 3455 G, 9.6903 GHz, 20 MHz video bandwidth, 51 dB video gain. 100 μ M protein.

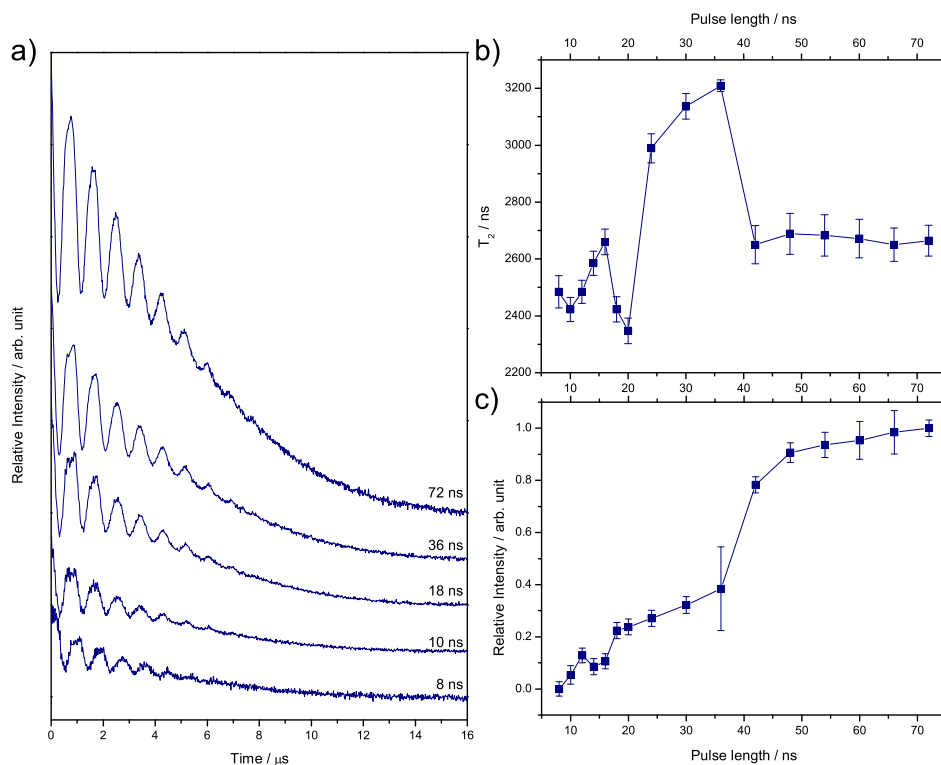


Figure 6.13 T_2 $\pi/2$ pulse length optimization of E9 DNase S3C. a) echo decay traces b) fitted T_2 times plotted against $\pi/2$ pulse length c) fitted echo amplitude plotted against $\pi/2$ pulse length. In b and c error bars show 95 % confidence bands of the fitting. 50 K, 4 scans, 4 shots/point, 1024 points, 10000 μs shot rep. time, 3454.5 G, 9.6846 GHz, 20 MHz video bandwidth, 48 dB video gain. 100 μM protein.

6.8. Pulse length optimization

With high levels of deuteration it is expected to have 2H ESEEM modulations with greater than 10 % modulation depth [94]. It was investigated whether a softening of pulses would create a decrease in the severity of the 2H ESEEM modulation depth [83]. For this, a standard Hahn echo sequence was used ($\pi/2 - \tau - \pi - \tau - \text{echo}$), with τ kept at an optimum 380 ns, whilst the $\pi/2$ length was varied. The π pulse was adjusted accordingly and the power optimised to give the largest echo.

Using the standard Bruker spectrometer configuration a 10 ns $\pi/2$ pulse was the shortest achievable (89.3 mW), at 8 ns there was not enough power to fully invert the spin population, seen in 6.13.a, whereby the echo modulations have a distorted frequency to all the pulse lengths. This results from uncertainty within the electronics and the ability to accurately measure such short times.

A 16 ns $\pi/2$ pulse (20 mW) gave a good T_2 relaxation time and experimental amplitude whilst maintaining a narrow Hahn echo of 140 ns. Moving to an 18 ns $\pi/2$ pulse there was a broadening of the echo to 160 ns, which corresponds to a dropping of the T_2 time and amplitude.

Pulses of 30 and 36 ns saw an increase in the signal amplitude which, when fitted, also gave a marked increase in the T_2 time. Unfortunately, the echo at this point increases to 200 ns in width. At a 48 ns $\pi/2$ pulse the echo width increases to 250 ns and at 66 ns $\pi/2$ pulses the Hahn increase to over 300 ns.

The increase in Hahn echo width is not a problem, for a simple collection of transverse relaxation time where the integral of the whole echo can be taken, however, for more complex

experiments this becomes a problem. As the pulse length increases more spins are excited but fewer of them are within the xy plane, instead there is a cloud of magnetisation about xy . For an experiment, such as PELDOR that has a level of orientation selectivity, a high degree of coherence in magnetisation is essential. Removing the orientation dependence can be achieved with strong samples by only integrating a small portion at the centre of the Hahn echo, rather than the entire echo encompassing all orientations and off-resonances.

Most interestingly, with all $\pi/2$ pulse lengths the first oscillation depth is 36.1 ± 1.1 % (1 standard deviation) of the total signal. The aim of this investigation was to see whether soft pulses would decrease the 2H ESEEM modulation effect, but upon this sample no effect can be seen. For this reason a $16\text{ ns } \pi/2$ pulse remained the best option, yielding good transverse relaxation times, signal amplitude and good polarisation homogeneity in spin magnetisation.

6.9. Summary and discussion

All data in this section, aside from the temperature assays, were collected at 50 K . Throughout the time period of these experiments the laboratory was installing closed circuit cryostats, with 15 K representing the lowest, reliable temperature with rapid cool down time this equipment could achieve at the time. At times when the closed circuit cryostats could not be used, such as to achieve temperatures below 15 K in the temperature assays a traditional flow cryostat was used.

In a PELDOR experiment the system must return to magnetic equilibrium before initiating the next pulse sequence for the next scan. The experiment repetition time is therefore typically set to five times greater than the longest relaxation time. Equally the dipolar coupling oscillations that are observed in a T_2 dependent relaxing echo. To capture the highest quality data and as many as dipolar oscillations as possible requires selecting a temperature to that maximise the T_2 relaxation time but minimise the much longer T_1 relaxation time. The best compromise then allows for the longest persistent echo with the highest scan rate, resulting in the most signal capture and thus, best possible data set.

In figures 6.2 and 6.4 (page 83) at 30 K the T_2 relaxation time is 94.2 % of the maximum, seen at 4 K . The T_1 relaxation time meanwhile has been much reduced to only 27.2 % of the maximum, seen at 4 K . At 20 K a small T_2 relaxation time increase of 3.1 % (relative to the maximum at 4 K) is seen, but also an increase in T_1 relaxation time of 40.9 %. Raising the temperature further to 40 or 50 K results in a drop in T_2 relaxation time (of 88.1 % and 86.1 % respectively, relative to the intensity at 4 K) with no real gain in T_1 relaxation time (25.6 % and 25.1 % respectively). Whilst these numbers reflect the situation of a pure relaxation experiment, the actual PELDOR experiments quickly showed it were far preferable to continue at 50 K as the echo duration with glycerol and other parameter optimisation showed little variation, however the experiment replication time could safely be dropped a further 10 ms .

The pH of the experiment showed little bearing on the experiment in terms of T_2 relaxation time, though a small maximum is seen around pH 7.5 . As the cytosolic pH of *E. coli* is 7.6 [10] and the optimum conditions for colicin E9 activity is pH 7.4 [91] an experimental pH 7.5 is most desirable for spectroscopic and biological relevance. Had the colicin E9 shown a strong pH dependence outside of this range then a compromise would have been needed to be made between protein structure and function with still having functional detection times.

6. Echo optimization

Individually, the deuteration and glycerolisation optimization experiments showed that increases in both gave better T_2 relaxation times and signal intensities. The deuteration assay showed an almost exponential increase in T_2 relaxation times using the standard fitting method up until 70 % deuteration (figure 6.6). However, on closer analysis it was found that the 70 % deuteration point was likely to be a quirk of environmental parameters and fitting. At 70 % deuteration of the solvent a large amount of modulations from both hyperfine coupling and ESEEM can be seen, allowing the sample to relax quickly via two methods. The relatively small amount of signal combined with inexact fitting methods artificially generated a high T_2 relaxation time. With data points above 70 % solvent deuteration any further increase in deuteration yields no significant increase in T_2 relaxation time, however, an increase in signal amplitude is seen. These observations made it clear that full deuteration of all samples was desirable.

Working with high glycerol concentrations is difficult as it requires that the protein be concentrated to very high levels nearing saturation before being added to the glycerol or require the difficult process of resuspending freeze-dried protein in glycerol. The high viscosity solutions associated with high glycerol concentrations also generate sample handling problems when requiring small volumes to be accurately measured and placed inside narrow bore sample tubes. The increase from a 50 % glycerol solution to 90 % glycerol solution yielded only a further increase of 0.1 μs in the T_2 relaxation time (a 5.3 % relative increase (figure 6.8)). With this in mind, the combined deuteration and glycerol experiments were conducted using 50 % glycerol, as very much a practical handling compromise over a best experimental result.

With individual assays of glycerol and deuteration complete, it was already known that the most desirable conditions were with high glycerol concentrations and deuteration. To investigate the effect of deuteration of the glycerol itself two experiments were conducted. In the first (figure 6.9) free in solution IA-PROXYL was placed in several combinations of solvent and glycerol deuteration. The addition of glycerol as before had marked increase on the effect of T_2 relaxation time, from 0.40 to 2.10 ms in protonated solvent and 0.40 to 3.32 ms in deuterated solvent. Deuteration of the glycerol in either solvent showed again a marked increase in the T_2 relaxation time. Most significantly the fully deuterated sample showed a 17.5 fold increase in T_2 relaxation time than the protein in water alone. Glycerol deuteration was further shown to be beneficial to signal amplitude. The increased viscosity, and thus molecular tumbling time as well as the reduced relaxation via solvent 1H hyperfine couplings, yield large experimental gains.

With the results of free in solution IA-PROXYL showing such drastic results, it was hoped that similar results could be achieved with the spin label attached to protein. In figure 6.10, the original IA-PROXYL experiment was extended slightly such that each solvent (protonated/deuterated) saw each level of glycerol deuteration. In protonated solvent environments the addition of glycerol increased the T_2 relaxation time but to a lesser extent than free IA-PROXYL. As the glycerol was deuterated the T_2 relaxation time increased, only slightly with glycerol- d_3 , but glycerol- d_8 showed a relaxation time almost exactly double that of glycerol. With deuterated solvent, the T_2 relaxation time almost matches that of the protonated solvent with glycerol- d_8 . When the deuterated solvent has glycerol added the T_2 relaxation time actually drops as the addition of protons and associated hyperfine coupling relaxation is more detrimental to relaxation than the increase in viscosity and associated molecular

tumbling time. As with IA-PROXYL, further deuteration of the glycerol in the deuterated solvent results in longer T_2 relaxation times but to a lesser extent than seen with IA-PROXYL free on solution. Overall, with glycerol and deuteration a 7.8 fold increase in T_2 relaxation time was seen with a near doubling of signal amplitude.

With infinite resources, it would be desirable that the proteins be grown in deuterated media, ensuring that the entire protein was completely deuterated, with samples using deuterated labels and being resuspended directly into 100% glycerol- d_8 for the experiments. Deuterated grow ups are prohibitively expensive for all of the required mutants. As a result, all samples were grown in normal, protonated, media and later buffer exchanged into D_2O . Deuterated spin labels in the quantity required for this investigation also proved prohibitively expensive. All experiments were conducted using the much easier to handle but similarly behaving 50 % glycerol- d_8 . This gave much of the benefit of deuteration and glycerolisation without extreme expense and difficulty in sample preparation.

7. Pulsed Investigation

With protein expressed, spin labelled and characterised, and all conditions optimised the PELDOR experiments could begin. This investigation includes 7 spin label sites across the E9 DNase domain and immunity protein Im9. This yields a binding cross of 42 possible distances. Mutations for every possible distance were not prudent, giving an experimental investigation of 18 PELDOR experiments.

Each PELDOR experiment was individually optimised to yield the best possible data. This meant that before each PELDOR run several experiments were conducted; a T_2 experiment to gather the maximum PELDOR observation window as well as the optimum τ_1 time (see chapter 1), a T_1 experiment to determine the minimum time between shots in an experiment, a field swept echo experiment to locate the spin resonance location and an ELDOR nutation experiment to determine both ELDOR pulse power and length.

In addition, PELDOR experimental background experiments were conducted in which the PELDOR experiment was repeated under identical conditions and parameters but the sample is replaced with a sample where only one of the binding partners is spin labelled and a further experiment where one of the binding partners is completely removed. By subtracting the experimental background of a spin label in the same position and environment with no PELDOR effect from the real experiment, the resulting trace can be attributed entirely as a result of a dipolar coupling.

This amounts to 18 PELDOR experiments each with; 4 PELDOR backgrounds, a T_2 experiment, a T_1 experiment, a field swept echo, an ELDOR nutation as well as experimental background subtracted PELDOR traces and PELDOR derived distance distributions. This amounts to a substantial amount of data, far more than would be appropriate to be show here in full. As a result the results of the T_1 and T_2 experiments are briefly summarised in §7.1, whilst full parameter setup experiment examples can be found in appendix D.

7.1. T_1 and T_2 measurements

Following optimization of echo parameters, each spin labelling site was investigated to ensure a level of consistency between the different positions. As each sample was run with near identical conditions, near equivalent solvent accessibility of label sites (with the expectations of colicin E9 D20C/E66C and E66C) and given that the protein size and weight should be of the same order, even between colicin and immunity protein, a large difference was not expected.

However, this experiment established the maximum acquisition time from the T_2 time for the PELDOR experiment and the minimum required shot repetition from the T_1 time. As seen from table 7.1 the maximum acquisition time for any sample was around 3 μs with

7. Pulsed Investigation

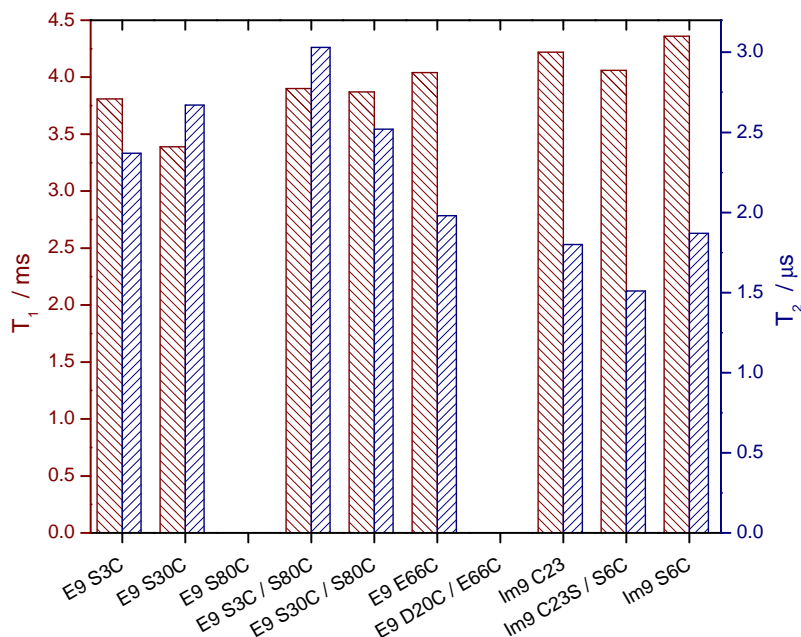


Figure 7.1 Relaxation times of each labelled protein in the E9 DNase - Im9 binding complex under optimised conditions from chapter 6. All proteins shown here use the IA-PROXYL label except E9 DNase S30C which uses MTSL. All experiments: 50 K, 4 scans, 4 shots/point, 1024 points, T_2 experiments used 6.0 dB attenuation, 20 MHz video bandwidth, 45 dB video gain and 10000 μ s shot rep. time. T_1 experiments used 9.0 dB attenuation, 20 MHz video bandwidth, 51 dB video gain and 50000 μ s shot rep. time. Frequency and magnetic field were optimised on a per experiment basis typically around 9.70 GHz and 345.70 mT respectively. Protein concentration 100 μ M in D_2O and 50 % glycerol- d_8 .

most sitting between 1.8 and 2.6 μ s. Whilst the T_1 times sit at approximately 4 ms, so as to ensure that the system is completely relaxed the shot repetition time was set at two and a half times this value (10 ms).

Protein, labelling position	T_1 / ms	T_2 / μ s
E9 DNase S3C	3.81	2.37
E9 DNase S30C	3.39	2.67
E9 DNase S3C / S80C	3.90	3.03
E9 DNase S30C / S80C	3.87	2.52
E9 DNase E66C	4.04	1.98
Im9 C23	4.22	1.80
Im9 C23S / S6C	4.06	1.51
Im9 S6C	4.36	1.87

Table 7.1. The T_1 and T_2 relaxation times of each labelled protein within the E9 DNase - Im9 binding complex under optimised conditions, at 50 K.

7.2. Distance measurement

Each PELDOR experiment sample was made according to the conclusions of previous chapter. Namely, samples were buffer exchanged into 100 % D_2O and further washed with D_2O to ensure that any labile protons on the protein had been exchanged with deuterons and combined with glycerol- d_8 , so that the final solution was 50 % glycerol.

Samples requiring the mixing of two proteins were mixed such that the protein stoichiometry remained 1:1 to avoid any free protein distorting the PELDOR trace. Any difference in 100

concentration was adjusted with the addition of further D₂O. A short mixing time of a few minutes was allowed before addition of the glycerol. The addition of glycerol yielded an 80 μ l total sample volume, which required agitation to stop formation of a biphasic system. Protein concentration was kept as high as possible to maintain signal but on occasion in unbuffered D₂O protein precipitation was unavoidable and at much lower concentration than expected for the sample protein in normal, protonated, buffered solution.

Each PELDOR experiment figure presented here shows an approximate 20 hour data collection and is presented with 4 subfigures. The first subfigure shows a cartoon model of the colicin E9 DNase - Im9 complex from PDB 2K5X. IA-PROXYL spin labels have been added to each labelling site (with the aid of MMM, with the calculated most populated rotamer) being displayed in ball and stick representation. Each label has been coloured uniquely and the near-solid line representing the PELDOR vector is coloured to match the first label in each pair. For example, as the colicin E9 DNase S3C^{IA1} label is black and by convention is the first label listed any distance involving this label is coloured black. This subfigure is included to aid the reader's perception of each experiment in relation to the others.

The second subfigure shows the raw data of the PELDOR experiment, including the brief rise before time zero. Whilst most experiments are kept with similar parameters it is worth noting the difference that the acquisition time for some experiments does vary.

The third subfigure shows the raw data PELDOR trace subtracted from the background and renormalized in navy. For particularly noisy spectra the raw trace passed through a Fourier transform noise filter to remove high frequency noise associated with very short distances, which would be below detection by PELDOR. When this technique was used it has been noted in the main text. The red line shows the fit to the experimental data from DeerAnalysis.

Finally, the last subfigure shows the distance distribution from DeerAnalysis after a Tikhonov regularisation L-curve analysis. In navy is the experimental data, plotted against the distance prediction, made from MMM rotamer population analysis, based upon the NMR minimised average complex (PDB: 2K5X) in red. In cases where the experimental and predicted distance distribution vary substantially, the experimental data is loaded into MMM and a rotamer analysis is performed to see if varying the rotamer populations could better fit the predicted distribution to the experimental. Where this is the case, the adjusted distribution is presented in maroon.

In addition to the figures shown in the body text here for each experiment a L-curve analysis plot of the Tikhonov regularisation is shown in appendix D.2 on page 317. Whilst the theory behind Tikhonov regularisation is not the focus of this investigation the resulting L-curve shows the quality of the regularisation and thus reflects the quality and confidence of the final distance distribution.

Additionally, appendix D.2 shows a superimposing of the 4 pulse PELDOR experimental data and background trace from which the experiment is subtracted. With this additional figure it can help visualise the quality of the experiment by highlighting how strong the dipolar oscillations vary from a simple decay.

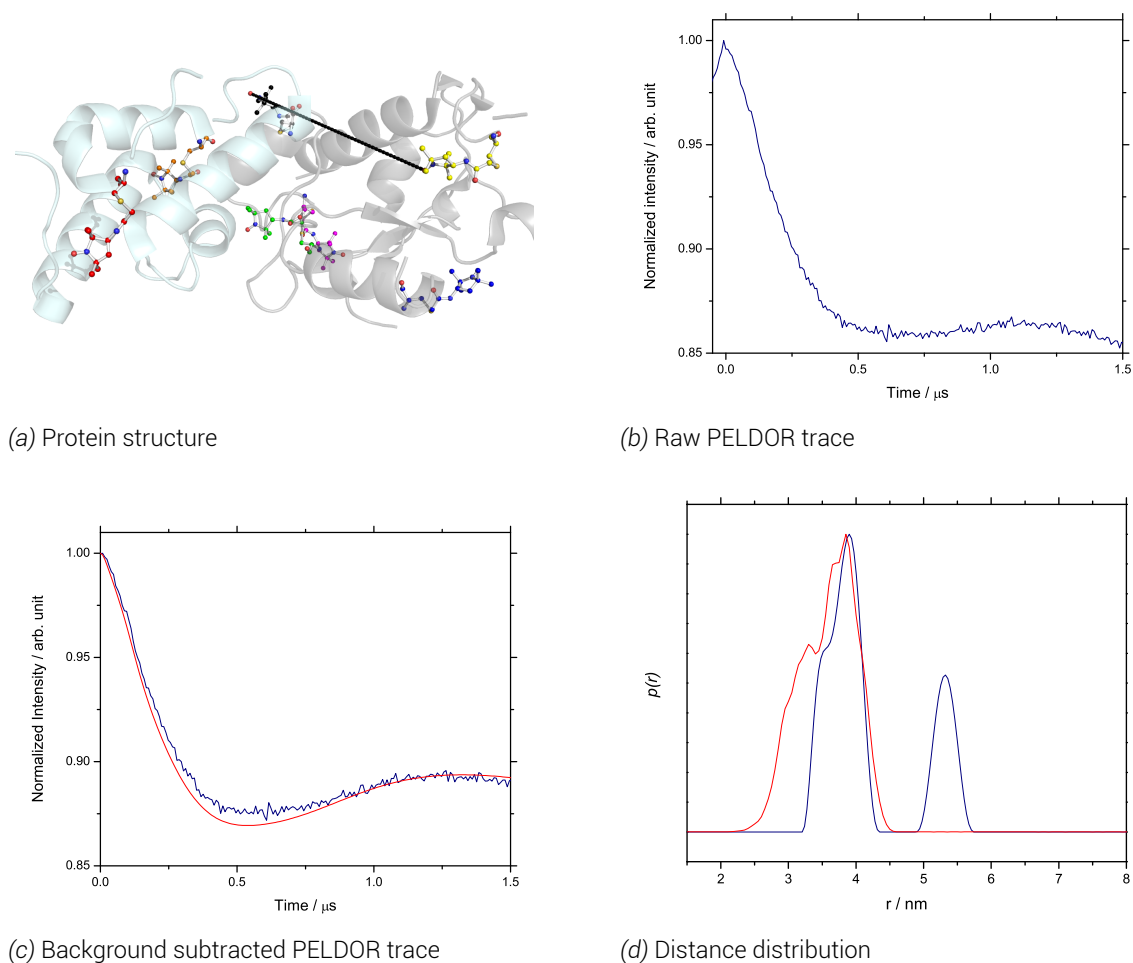
7.2.1. Colicin E9 DNase S3C^{I.A1}/S80C^{I.A1}

Figure 7.2 PELDOR experiment results for colicin E9 DNase S3C^{I.A1}/S80C^{I.A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. S80C position in the middle, S3C position at the right b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 276 points, 6.0 dB attenuation, 20 MHz video bandwidth, 57 dB video gain and 10000 μ s shot rep. time, ~9.70 GHz, 345.70 mT. $\tau_1 = 380$ ns, $\tau_2 = 2200$ ns, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

The colicin E9 DNase S3C^{I^{A1}}/S80C^{I^{A1}} sample clearly shows a mix of 2 distinct distances in the distance distribution. The peak at 3.91 nm shows good agreement with the MMM prediction in position but is much less broad. Perhaps suggesting that one or both of the sites is more restricted than predicted, thus not allowing the full motion of the rotamer cloud predicted by MMM. The second peak at 5.31 nm however is not predicted by MMM but cannot necessarily be discounted due to its high probability. Considering both peaks together could suggest that one or both of these positions must be more restricted than expected and jumps between 2 main conformations, resulting in two narrow distribution peaks rather than a single broad peak as expected.

When the MMM rotamer distributions of these two labelling sites are visualised, as in figure 7.3, both of these sites can be seen to be largely mobile on the surface of the protein. The colicin E9 DNase S3C^{I^{A1}} is very mobile and shows an almost equal sampling across the entire rotamer cloud. The colicin E9 DNase S80C^{I^{A1}} is more restricted and shows a smaller sampling than the colicin E9 DNase S3C^{I^{A1}} position and appears to have a slight split into two populations at the top and bottom of figure 7.3.b. It is likely that the shoulder of the MMM distance distribution trace (red line in figure 7.2.d) appears as a result of this split. However, this small split would not account for the distinct difference of populations seen in figure 7.2.d. The colicin E9 DNase S80C position lies upon a short α helix between a flexible hinge and the main, flexible DNA binding region. It is therefore not unreasonable to expect some molecular motion of this helix that could result in a two state conformation interchange, particularly in the absence of the immunity protein to stabilise the DNA binding region. This could explain the difference in experimental and calculated distance distributions.

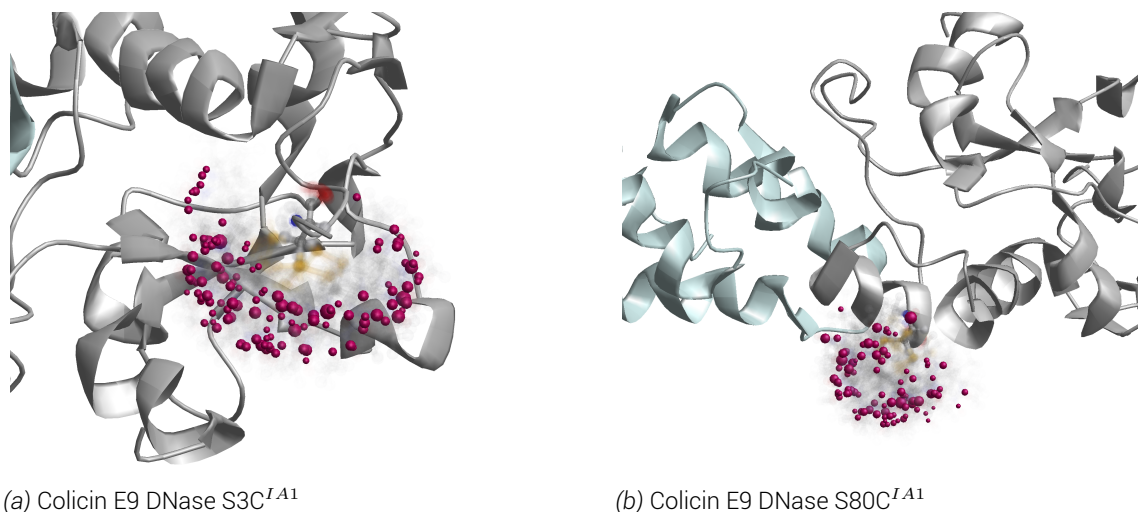


Figure 7.3 Rotamer comparison of colicin E9 DNase S3C^{I^{A1}} and S80C^{I^{A1}}. MMM computed rotamers are shown superimposed onto the cartoon structure of colicin E9 DNase/Im9 complex (PDB: 2K5X). Each individual rotamer's transparency relates to that rotamer's population, with solid rotamers being statistically more relevant. The magenta sphere represents the nitroxide radical position of each rotamer, with sphere size relating to rotamer population.

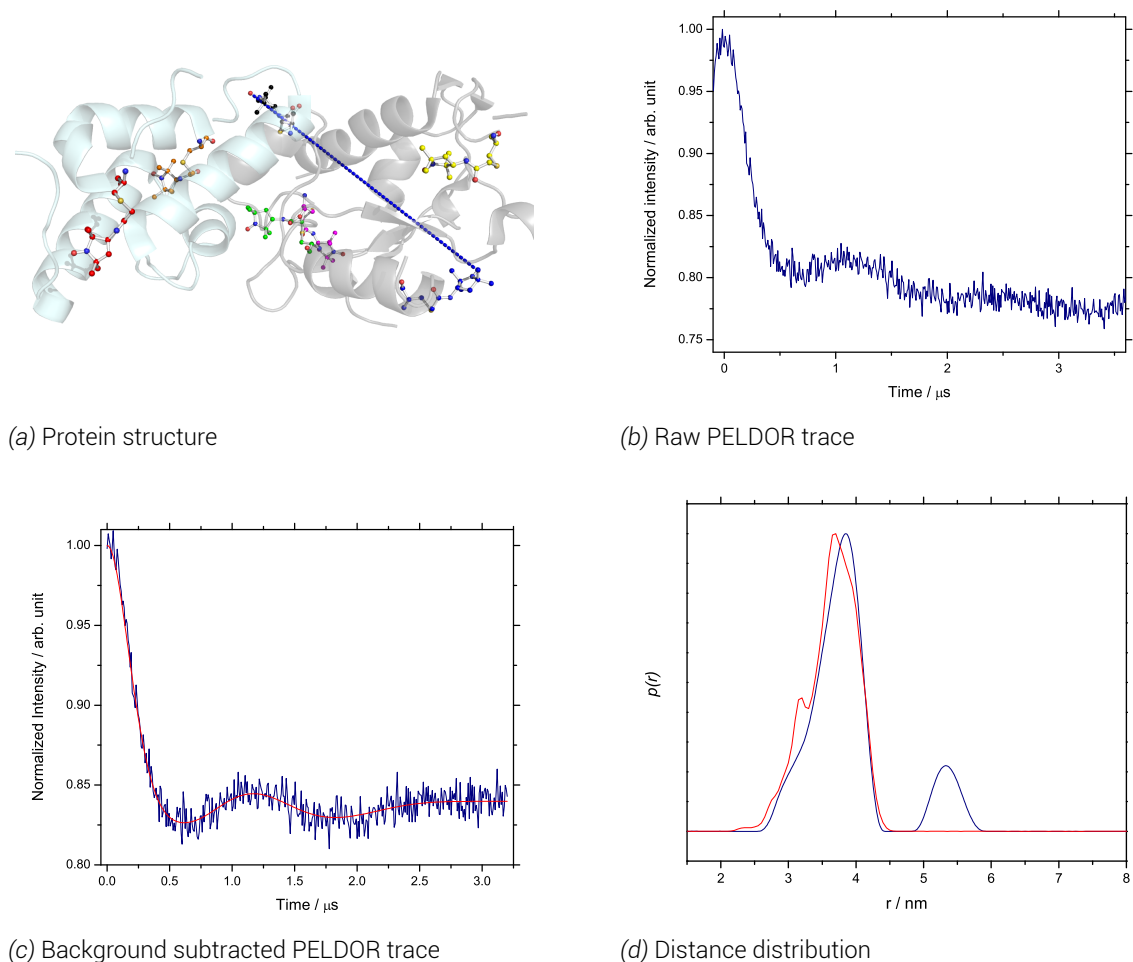
7.2.2. Colicin E9 DNase S30C^{I.A1}/S80C^{I.A1}

Figure 7.4 PELDOR experiment results for colicin E9 DNase S30C^{I.A1}/S80C^{I.A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. S80C position in the middle, S30C position at the bottom right b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 193 points, 10.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.69 GHz, 346.00 mT. $\tau_1 = 380$ ns, $\tau_2 = 4160$ ns, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

The colicin E9 DNase S30C^{IA1}/S80C^{IA1} sample again shows a mix of 2 distinct distances in the distance distribution. The peak at 3.86 *nm* shows good agreement with the prediction from MMM, suggesting that the rotamer population from MMM differs only slightly from the case *in vivo*.

As with the colicin E9 DNase S3C^{IA1}/S80C^{IA1} a significant second peak at 5.33 *nm* can be seen which is not predicted by MMM. Further analysis of this peak was conducted (see appendix D.2.2), to determine whether this peak results is simply a ghost peak, a truncation artefact from background subtraction or significant. By using DeerAnalysis' validation tool the fitting window was moved through 2.5 and 3.5 μ s at the same regularisation parameter. It was found that the oscillation after 2.0 μ s in the background subtracted PELDOR trace (figure 7.4.c) remains significant and fits the data well. This had little effect upon the shorter distance peak, varying only a little in the low distance tail. With the longer distance peak the validation process added significance to 5.40 *nm* as a significant distance.

The common element between this and the previous experiment is the colicin E9 DNase S3C site. When viewed in combination there is a suggestion that the colicin E9 DNase S3C site is more restricted than predicted and jumps between 2 conformations. One pointing into the colicin and giving the shorter distances and one pointing out and away from the protein giving rise to the longer distance. This could perhaps explain the slightly lower spin labelling efficiency and aqueous accessibility than that of the colicin E9 DNase S30C site, seen earlier in this investigation.

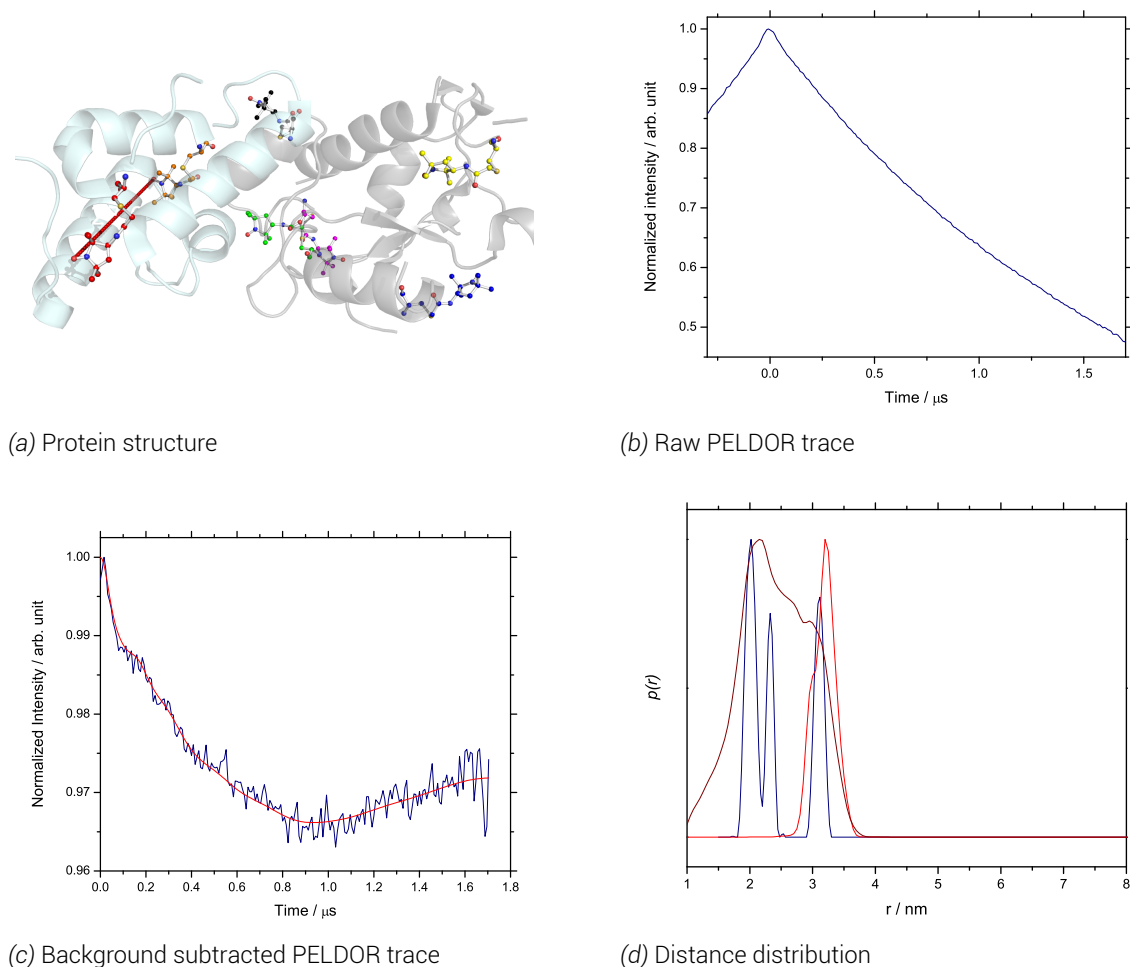
7.2.3. Im9 S6C^{I A1}/C23^{I A1}

Figure 7.5 PELDOR experiment results for Im9 S6C^{I A1}/C23^{I A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. C23 shown in orange, S6C shown in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 254 points, 8.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.69 GHz, 345.60 mT. $\tau_1 = 380$ ns, $\tau_2 = 1800$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

Unfortunately, this Im9 mutant was incredibly prone to precipitation and could only be used at concentrations around $25 \mu M$. The low (16 %) labelling efficiency, with both sites individually labelling poorly (S6C 14 %, C23 5 %) meant that the proportion of spins contributing to electron-electron dipolar oscillations were going to be small. Initially looking at the raw PELDOR trace, whilst being virtually noise free this experiment appears to be little more than a straight decay. However if the data between 1.0 and 1.5 μs (in figure 7.5.b) is extrapolated a slight rise can be seen. This is made more apparent in appendix D.2.3, where the PELDOR trace is shown with the Im9 C23^{IA1} relaxation run with the same experimental parameters.

An optimistic fit can be made with DeerAnalysis yielding a distance distribution with 3 short distances, though the L-curve (appendix D.2.3) is very shallow. MMM initially predicts a single, relatively narrow peak reflecting the slightly restricted labelling sites (red). This predicted peak matches well with the longest experimental distance at 3.2 *nm*, after loading the resulting DeerAnalysis distance distribution into MMM the rotamer population can be adjusted whereby the population of each rotamer is adjusted to see if the prediction and rotamers can match the experiment data. This experimentally adjusted prediction is given the maroon line distribution, which aligns not too dissimilarly from the experimental result. Nevertheless, this is a somewhat optimistic result and still yields an overall distance distribution of over 2.0 *nm*.

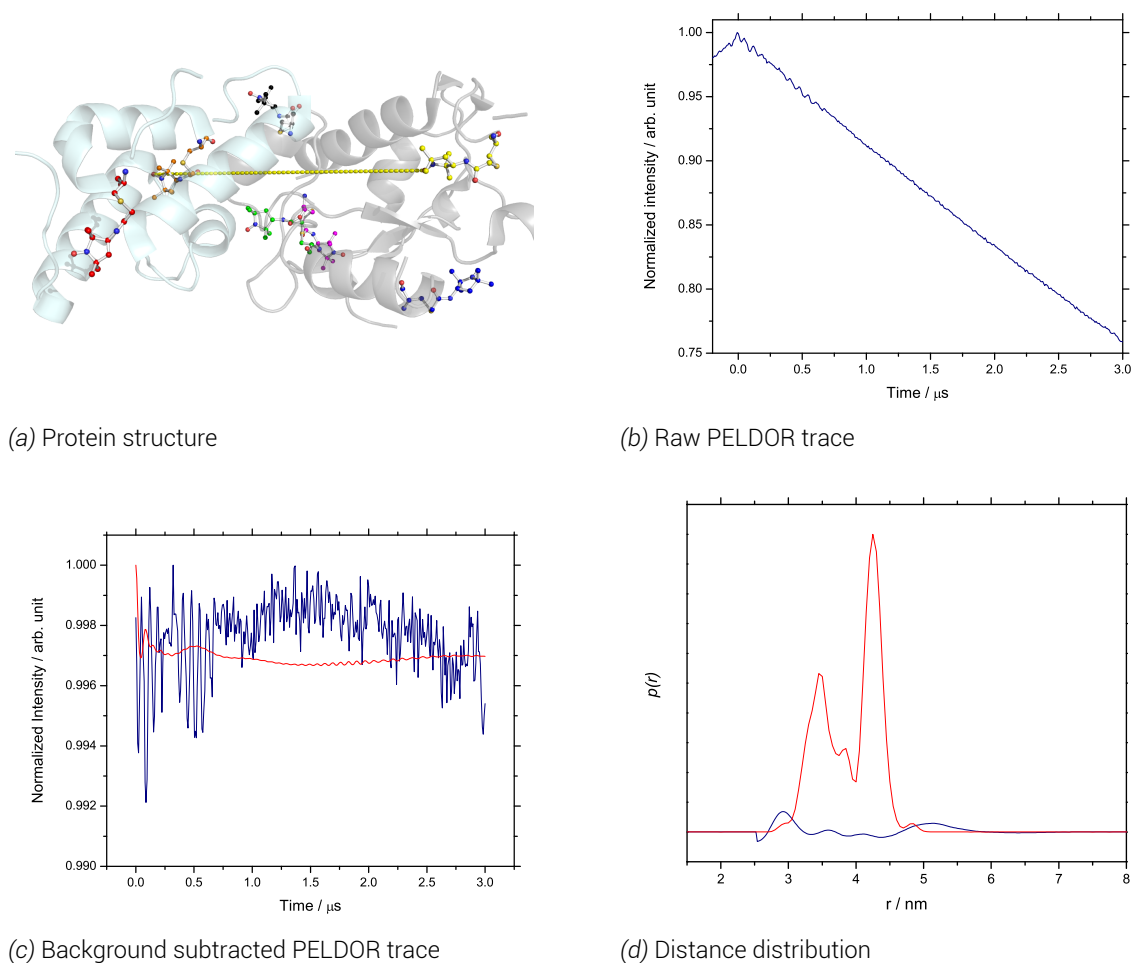
7.2.4. Colicin E9 DNase S3C^{IA1} - Im9 C23^{IA1}

Figure 7.6 PELDOR experiment results for colicin E9 DNase S3C^{IA1} - Im9 C23^{IA1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, Im9 C23 in orange b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 414 points, 5.5 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.69 GHz, 345.60 mT. $\tau_1 = 380$ ns, $\tau_2 = 3440$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

This experiment proved troublesome and was repeated no less than 4 times. In the first instance the effective label concentration was far too low to achieve a good SNR echo. The second experiment increased the protein concentration virtually to the point of precipitation. After still no PELDOR being seen the individual proteins were taken and protein concentration accurately confirmed with a nanodrop spectrophotometry and spin label attachment confirmed with room temperature X-band cw-EPR followed by a repeated spin labelling procedure repeated.

Figure 7.6 shows the result of the fourth PELDOR attempt with maximum possible protein concentration and both proteins having been spin labelled twice. Figure 7.6.b shows a clear, linear decay, which when subtracted from the background, in figure 7.6.c, gives a straight line distorted by noise. If the PELDOR experiment trace is plotted with the single colicin E9 DNase S3C^{I41} PELDOR decay (as in appendix D.2.4) it can be seen that there are no dipolar oscillations present in this sample.

This experiment clearly showing no dipolar contributions when analysed with DeerAnalysis, unsurprisingly gives non-sense results. The background subtracted trace shows less than 0.2 % intensity - little more than noise (figure 7.6.c). The Tikhonov L-curve (appendix D.2.5) and distance distribution (figure 7.6.d) give, expected, terrible data. The L-curve is not the expected shape and the distance distribution includes below zero probabilities of certain distances.

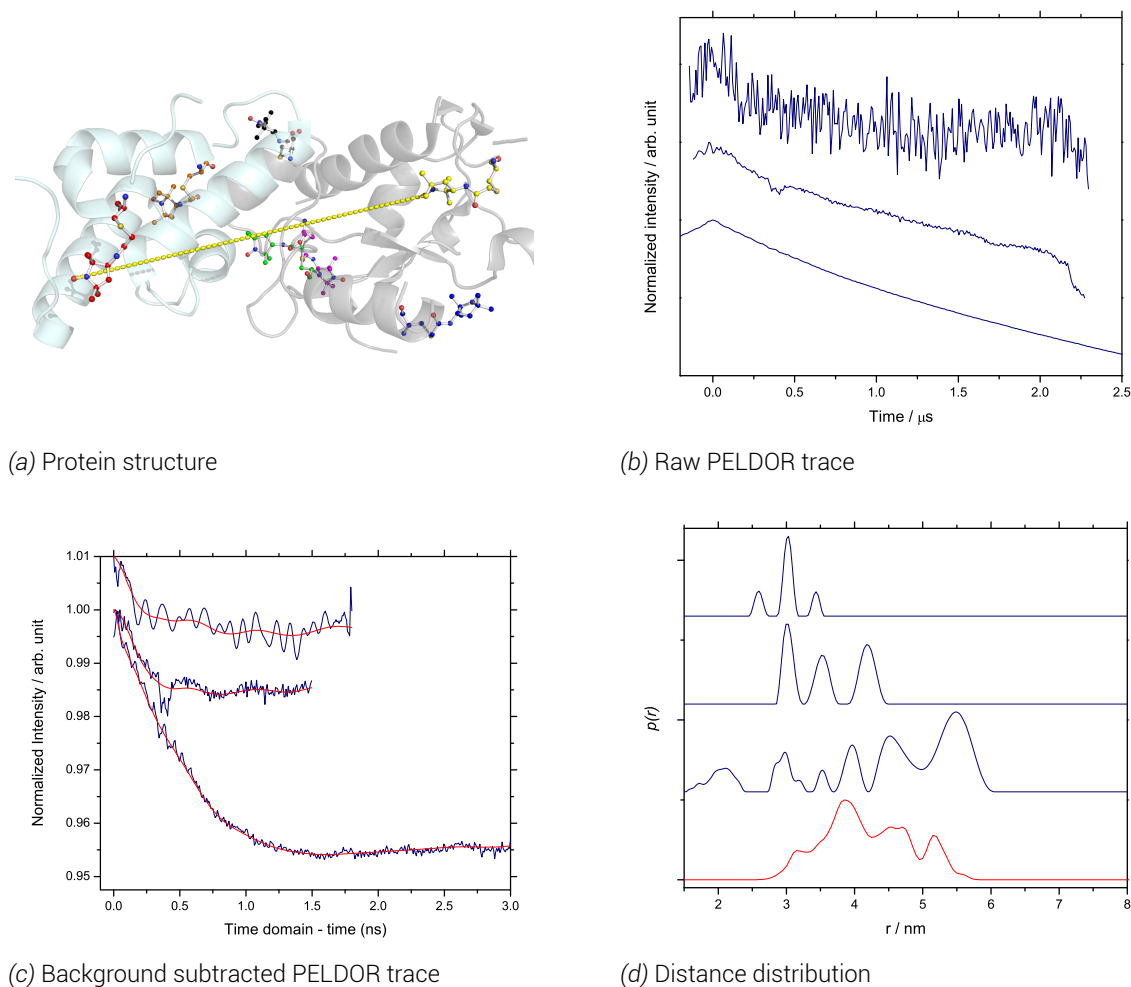
7.2.5. Colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23S

Figure 7.7 PELDOR experiment results for colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23S. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, Im9 S6C in red b) shows the raw 4 pulse PELDOR data for 3 different experiments (described in the text). c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red for each of the 3 experiments. d) Shows the DeerAnalysis Tikhonov derived distance distributions in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 20 MHz video bandwidth and 10000 μs shot rep. time, ~9.69 GHz, ELDOR pulse step 8 ns, no nuclear modulation. Data points, magnetic field, attenuation, video gain and delay times were optimised on a per experiment basis suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

This experiment again due to the very low concentration and labelling efficiency of colicin E9 DNase S3C^{I^{A1}} proved troublesome. However, unlike colicin E9 DNase S3C^{I^{A1}} - Im9 C23^{I^{A1}} some trace of information can be gleaned from each experiment. For this experiment the extent of difficulty has been highlighted by showing the initial experiment (top spectrum in each panel), after protein concentration to precipitation point (middle spectrum in each panel) and after protein concentration and a second attempt at spin labelling with a 100 fold excess of label (bottom spectrum in each panel).

Comparing the bottom spectrum of figure 7.7.a with a background decay of colicin E9 DNase S3C^{I^{A1}} (appendix D.2.5) shows only a slight raise in signal intensity in the first microsecond. After background subtraction (figure 7.7.c) fitting shows some oscillations, of which some are clear amongst at least one trace.

The MMM prediction for colicin E9 DNase S3C^{I^{A1}} - Im9 S6C^{I^{A1}}/C23S is shown in red in figure 7.7.d. Many of the peaks predicted by MMM appear in at least one of the experimental distributions. Key to notice is that peaks at 3.0 and 3.5 *nm* appear in all 3 experimental distributions, whilst a peak around 4.2 *nm* appears in both low noise spectra. Perhaps, it is to be expected to have such an array of peaks in the experimental distance distribution when the predicted distribution clearly is very broad and is a complex built up of many small populations. Consulting the Tikhonov L-curve of the final spectrum (appendix D.2.5), a discontinuity is evident but both lengths of the curve are far from the desired straightness or steepness. It is possible that in each experiment, conditions were varied sufficiently that different rotamers were isolated and contributed to different peaks; however with such weak signals these peaks could be fitting artifacts from such a weak signal.

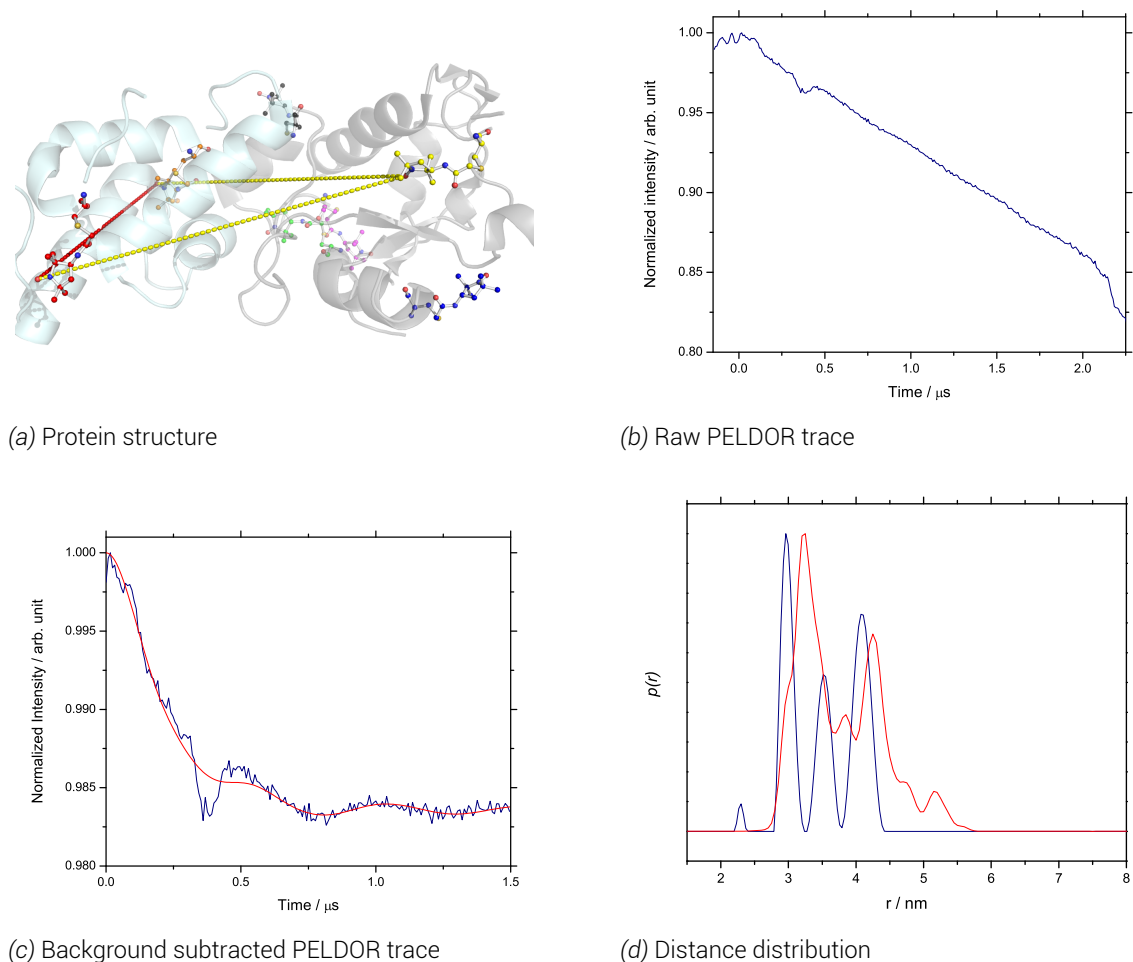
7.2.6. Colicin E9 DNase S3C^{I A1} - Im9 S6C^{I A1}/C23^{I A1}

Figure 7.8 PELDOR experiment results for colicin E9 DNase S3C^{I A1} - Im9 S6C^{I A1}/C23^{I A1}. PELDOR experiment results for colicin E9 DNase S3C^{I A1} - Im9 C23^{I A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, Im9 C23 in orange, S6C in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 602 points, 3.5 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μs shot rep. time, ~9.74 GHz, 347.30 mT. $\tau_1 = 380$ ns, $\tau_2 = 2580$ ns, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

This 3 distance experiment is predicted by MMM to have distances of 2.70, 3.48 and 3.83 *nm* so any broadness in rotamer population was expected to broaden distance distribution peaks and hide data.

The raw trace, shown in 7.8.b is peculiar in that it appears an almost linear decay with a slight wobble, and dramatic dip at 0.4 μ s. The sudden drop off in signal at 2.0 μ s corresponds to a mirrored increase in signal in the imaginary channel (not shown), suggesting a loss of phase in the echo toward the end of the experiment. To compensate for this the end point of the experiment was moved up to 1.5 μ s so that analysis was only conducted on reliable data. Background subtraction was conducted with colicin E9 DNase S3C^{IA1} complexed with unlabelled Im9, when the background experiment was conducted with identical the dramatic dip at 0.4 μ s was not seen (appendix D.2.6) and was considered as significant after repetition of the PELDOR experiment in the background subtracted data (figure 7.8.c). The dramatic dip at 0.4 μ s remains, but has been lessened and once again we see a very poor modulation depth associated with low labelling efficiencies, meaning that few of the molecules in the sample are contributing to the PELDOR signal.

The experimental distance distribution, with multiple distance DeerAnalysis ghost suppression applied, gives 3 clear peaks at 3.0, 3.5 and 4.2 *nm*, with a low probability and distance ghost peak at 2.3 *nm*. The Tikhonov L-curve (appendix D.2.6) shows an obvious discontinuity, though the curve is not strong, likely reflecting the weak oscillation depths. The resulting distances are a little longer than the distances predicted by MMM highest populated rotamer distances but are a little shorter than the MMM rotamer ensemble average shown in red in figure 7.8.d. Overall, despite the very weak oscillation depth and very unusual, unexplained dip the resulting distances have a high level of confidence in this data set.

However, after further analysis whereby the fitting window is moved with a width of a deuterium ESEEM modulation the distance distribution (as shown in appendix D.2.6) collapses down to a single, sharp distance just below 3.0 *nm* - where the black line is the average distance probability with red and blue representing upper and lower fitting bounds (2 standard deviations). As a result, it can be concluded from the very weak oscillation depth and Tikhonov validation test that this experiment is showing little more than deuterium ESEEM oscillations and should be disregarded from further investigation. The additional peaks seen in figure 7.8.d are evidently fitting artifacts from the choice of fitting window and resemblance to the predicted MMM distance distribution is coincidental.

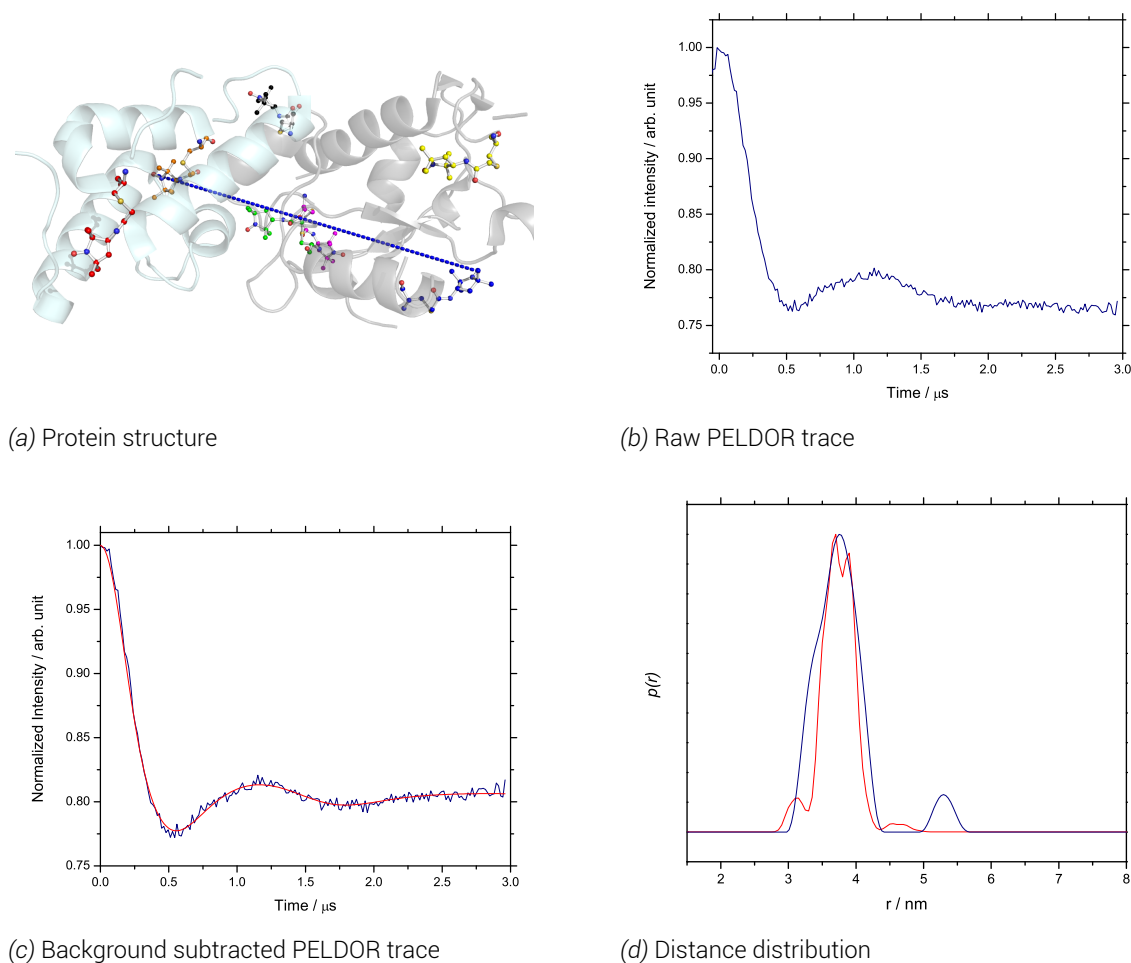
7.2.7. Colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}

Figure 7.9 PELDOR experiment results for colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S30C shown in blue, Im9 C23 in orange b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 253 points, 4.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μ s shot rep. time, ~9.66 GHz, 344.70 mT. $\tau_1 = 380$ ns, $\tau_2 = 3000$ ns, ELDOR pulse step 12 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

This experiment uses the scarce colicin E9 DNase S30C protein and is the only PELDOR experiment to use MTSL instead of IA-PROXYL. MTSL has a shorter linker between nitroxide radical and protein backbone, but at the expense of having slightly more flexibility in the linker. MTSL is still expected to have a narrower distance distribution than IA-PROXYL in the same positions.

Here despite the low concentrations of sample, a clean PELDOR trace is obtained, which when background subtracted against colicin E9 DNase S30C^{R1A} complexed with unlabelled Im9 (appendix D.2.7) yields a good fit (figure 7.9.c). The experimental distance distribution after a very clean Tikhonov regularisation (appendix D.2.7) shows an obvious main peak at 3.8 *nm* with a secondary ghost peak at 5.3 *nm*. Furthermore, this experiment shows an almost flawless agreement with the predicted distance distribution from MMM.

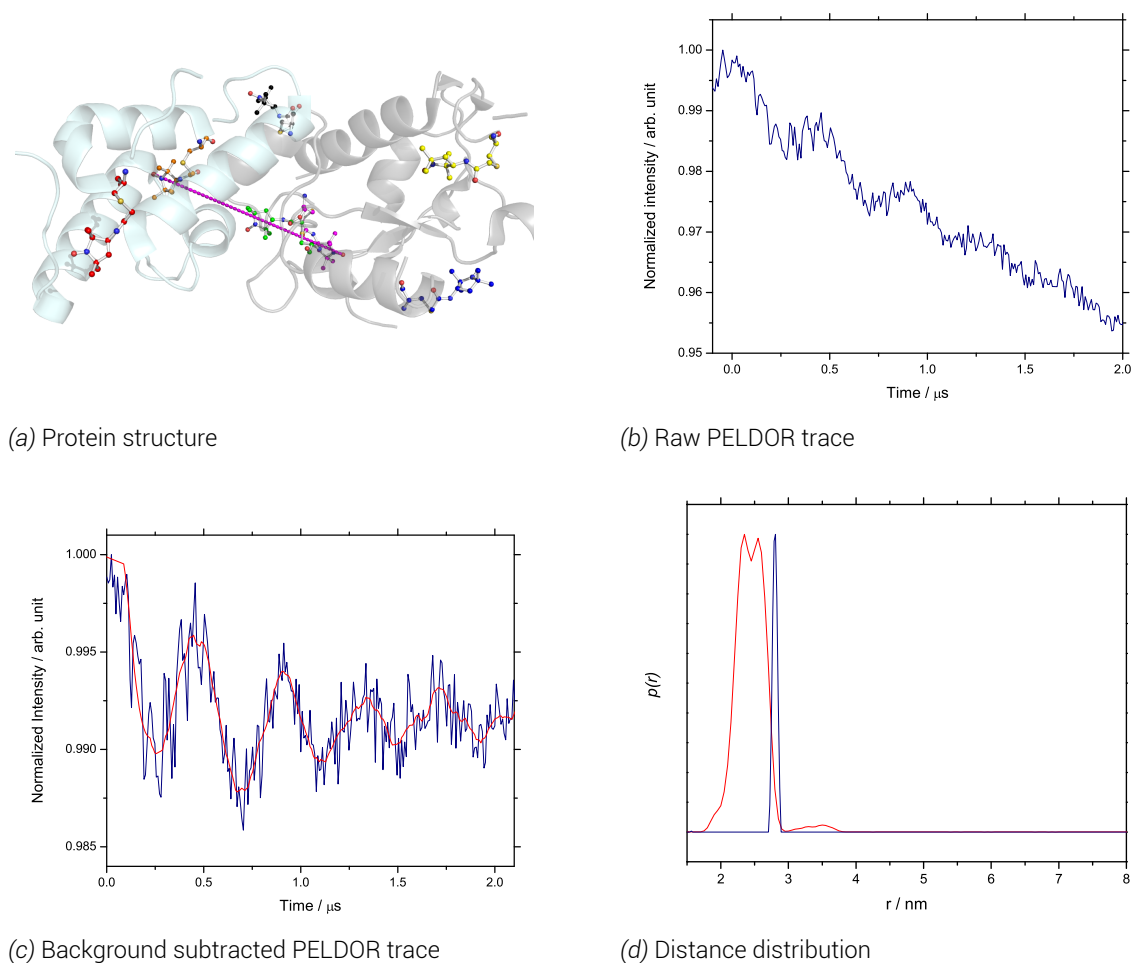
7.2.8. Colicin E9 DNase E66C^{I A1} - Im9 C23^{I A1}

Figure 7.10 PELDOR experiment results for colicin E9 DNase E66C^{I A1} - Im9 C23^{I A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase E66C shown in magenta, Im9 C23 in orange b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 352 points, 6.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μs shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ ns, $\tau_2 = 2600$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

This experiment is very curious, in that a clear oscillation can be seen in the raw spectrum shown in 7.10.b. These oscillations have a tiny modulation depth, but are incredibly regular yielding the very narrow distance distribution seen in figure 7.10.d.

In the case of colicin E9 DNase E66C^{I A1} there was not enough sample to run a background experiment of colicin E9 DNase E66C^{I A1} with unlabelled Im9 as the refocussed echo was so small that the experiment could not be accurately setup. As a result a 3 dimensional homogenous background was employed in DeerAnalysis to give a near straight line background (shown in appendix D.2.8).

The resulting Tikhonov L-curve and distance distribution from these very regular modulations gives a single, very sharp line at just below 3.0 *nm* - a distance that corresponds to the oscillation frequency of deuterium ESEEM effects. The only other possible explanation would be that both spin label positions are almost completely rigid and that despite the low labelling efficiencies of both proteins almost every labelled protein contributes to the dipolar oscillations. This scenario is highly unlikely statistically and would be in stark contrast to the earlier accessibility experiments where the Im9 C23^{I A1} position was found to have 60.2 % aqueous accessibility.

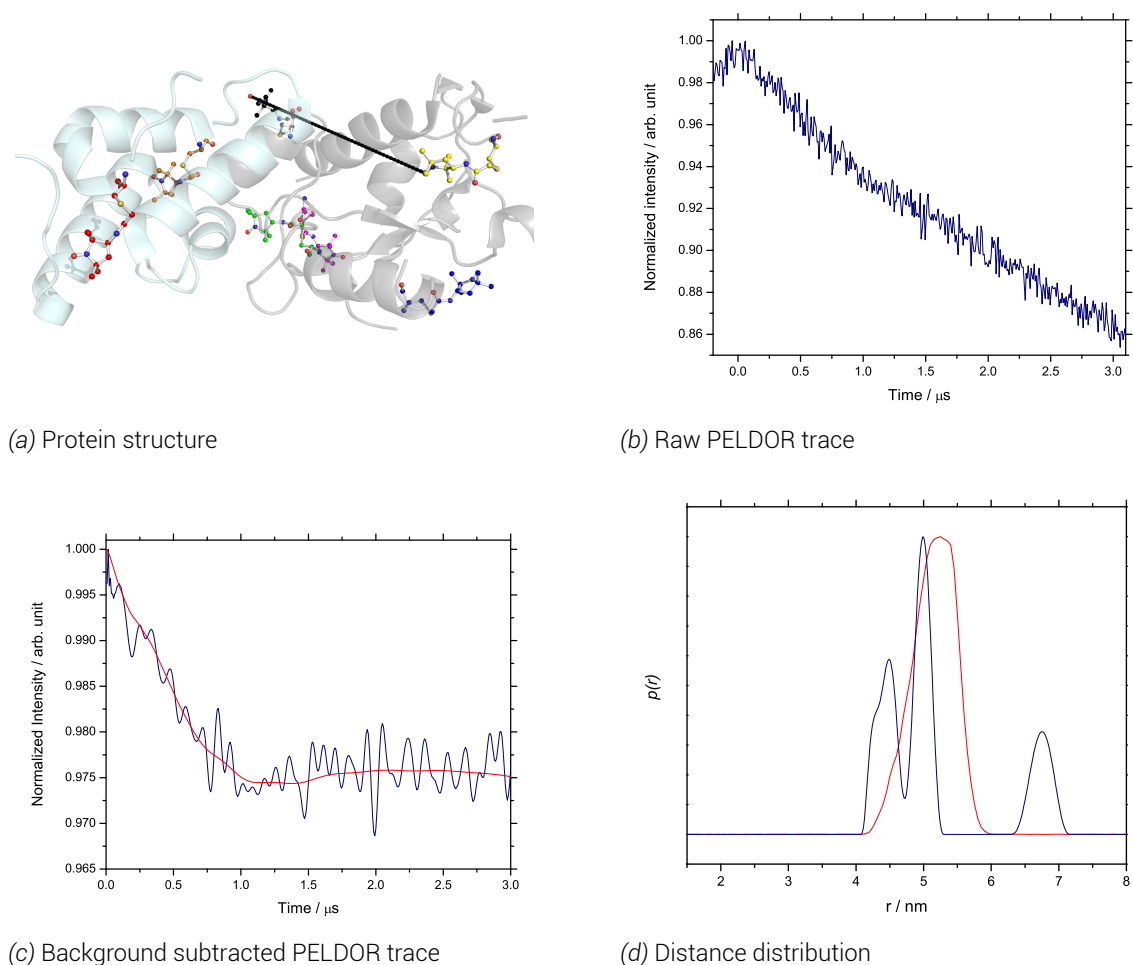
7.2.9. Colicin E9 DNase E66C^{I A1} - Im9 S6C^{I A1}/C23S

Figure 7.11 PELDOR experiment results for colicin E9 DNase E66C^{I A1} - Im9 S6C^{I A1} / C23S. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase E66C shown in magenta, Im9 S6C in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 500 points, 5.5 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μ s shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ ns, $\tau_2 = 4160$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

Unfortunately, during this experiment there were hardware issues overnight and so only the autosave of the first 500 scans could be used with confidence compared to the >2000 scans most of these other experiments have collected. Accounting for poor SNR in this experiment, oscillations can still be seen in the raw data.

A high frequency Fourier transformation filter to remove noise corresponding to distances less than 1.00 *nm* was applied. A background was constructed in DeerAnalysis using a 3 dimensional homogenous model as again the background experiment of colicin E9 DNase E66C^{IA1} with unlabelled Im9 yielded unusable results. The background was subtracted after the Fourier transformation filter to yield a still noisy, but a clear single oscillation, as shown in figure 7.11.c (and appendix D.2.9).

Tikhonov regularisation of this single, weak oscillation yielded a poor L-curve but one still with a discontinuity. MMM rotamer analysis predicted a single, broad peak, centred at 5.2 *nm*. However, the experimental distance distribution shows a clear splitting of the broad peak into two peaks at the 4.4 and 5.1 *nm* with an additional small probability peak at 6.8 *nm*. This longer range peak is likely attributed errors in the fitting of such weak, single oscillation experimental data. However, when the experimentally derived distance distribution is loaded into MMM and is analysed using the rotamer population adjustment tool, the two main, experimentally observed peaks fit easily into allowed rotamers (MMM experimentally adjusted rotamer distance distribution is omitted in figure 7.11.d for the sake of clarity).

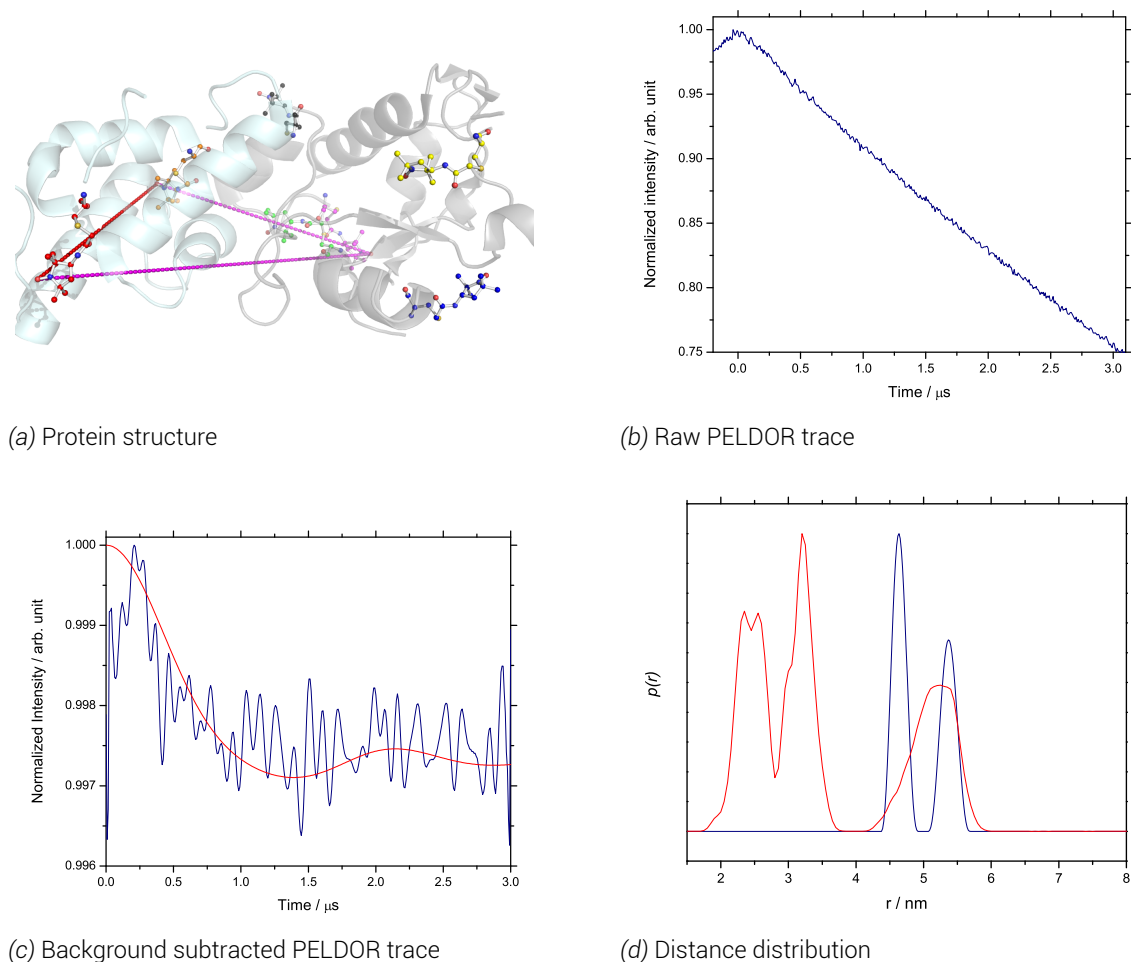
7.2.10. Colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}

Figure 7.12 PELDOR experiment results for colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1} / C23^{IA1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase E66C shown in magenta, Im9 C23 in orange, S6C in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 66 dB video gain and 10000 μ s shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

Again, here is a very weak PELDOR trace with dipolar oscillations of modulation depth less than 0.3 % of the initial signal. As a result the raw data can hardly be differentiated from a straight linear decay (highlighted in appendix D.2.10). Again a background was constructed in DeerAnalysis using a 3 dimensional homogenous model as the background experiment of colicin E9 DNase E66C^{IA1} with unlabelled Im9 yielded unusable results. Once the background is subtracted (figure 7.12.c) the modulation depth is very small indeed. After a high frequency noise filter had been applied to remove noise associated with distances less than 1.00 nm a noisy oscillation can be seen. With such a noisy spectrum, with such small modulation depth, it was important to ensure that the correct zero time was selected. To ensure as such the first 500 ns of data was taken, normalised and zero filled on either side to double the number of data points. The resulting data set was fitted to a Gaussian distribution to find the maxima, and this was used as the zero time. This method was found to be more reliable than simply using a moving point average smoothing function that, noise depending, could distort the starting position.

Tikhonov regularisation was conducted using ghost suppression for 3 spin label distances, but due to the very poor data quality yielded a near straight line L-curve (appendix D.2.10).

With 3 spin labels in this sample, 3 distances are expected, 2 short distances (between the E9 DNase E66C^{IA1} - Im9 C23^{IA1} and the intramolecular Im9 distance S6C^{IA1}-C23^{IA1}) and 1 longer distance (E9 DNase E66C^{IA1} - Im9 S6C^{IA1}). In the experimental distance distribution we see the longer, broad distance split into 2 peaks at 4.65 and 5.55 nm. However, the shorter distances relying upon Im9 C23^{IA1} position cannot be seen. The Im9 C23 position was shown in the earlier continuous investigation to label very poorly. This investigation is again not surprising when considered with figures 7.11 and 7.12 where the colicin E9 DNase E66C^{IA1} to Im9 C23^{IA1} shows no result but the experiment to Im9 S6C^{IA1} shows some suggestion of dipolar oscillations.

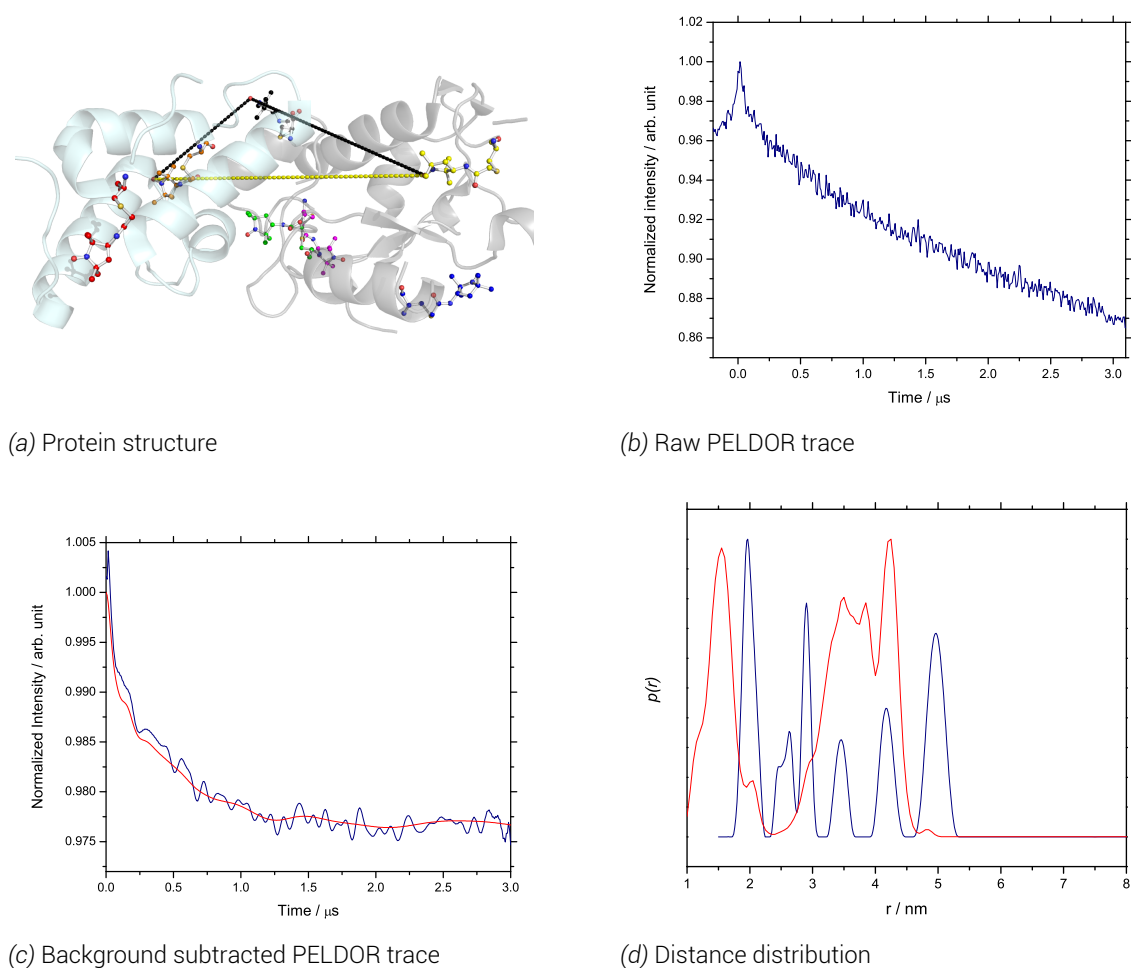
7.2.11. Colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}

Figure 7.13 PELDOR experiment results for colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, S80C in black, Im9 C23 in orange b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.68 GHz, 342.60 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

With an additional spin label in the system an additional distance is introduced. If each of these distances is of a suitably narrow distribution and the distances are suitably different then the distance distribution should resolve to only contain 3 peaks.

This sample is noisy and shows only very small oscillations, that only become apparent after using a high frequency Fourier transformation filter to remove noise corresponding to distances less than 1.00 *nm*. Only again, the modulation depth is only around 3 %, suggesting very few molecules in the signal are contributing to the dipolar oscillations.

Using, the rotamer cloud average from MMM this experiment should resolve to a single short distance (E9 DNase S80C^{IA1}- Im9 C23^{IA1} at 1.71 *nm*) and a likely convolution of two similar distances between E9 DNase S3C^{IA1}- S80C^{IA1} (3.19 *nm*) and E9 DNase S3C^{IA1}- Im9 C23^{IA1} (3.48 *nm*). Earlier in this investigation (§7.2.1) experimentally it was been shown that the E9 DNase S3C^{IA1}- S80C^{IA1} measurement has 2 distinct distances at 3.80 and 5.40 *nm*, of which peaks can be seen at 3.50 and 5.00 *nm* in figure 7.13.d.

Knowing that at least one of these positions exists in two states takes the total number of expected distances to at least 6. Given the weak nature of the signal it is difficult to say with any certainty whether the peaks seen here are true representations of the sample and the distances between the spin labels and which peaks are artifacts of fitting. The number of peaks is within what is expected, but the short E9 DNase S80C^{IA1}- Im9 C23^{IA1} distance appears to be too long and there are peaks covering the 2.00 to 3.00 *nm* range that would not be expected from MMM predictions.

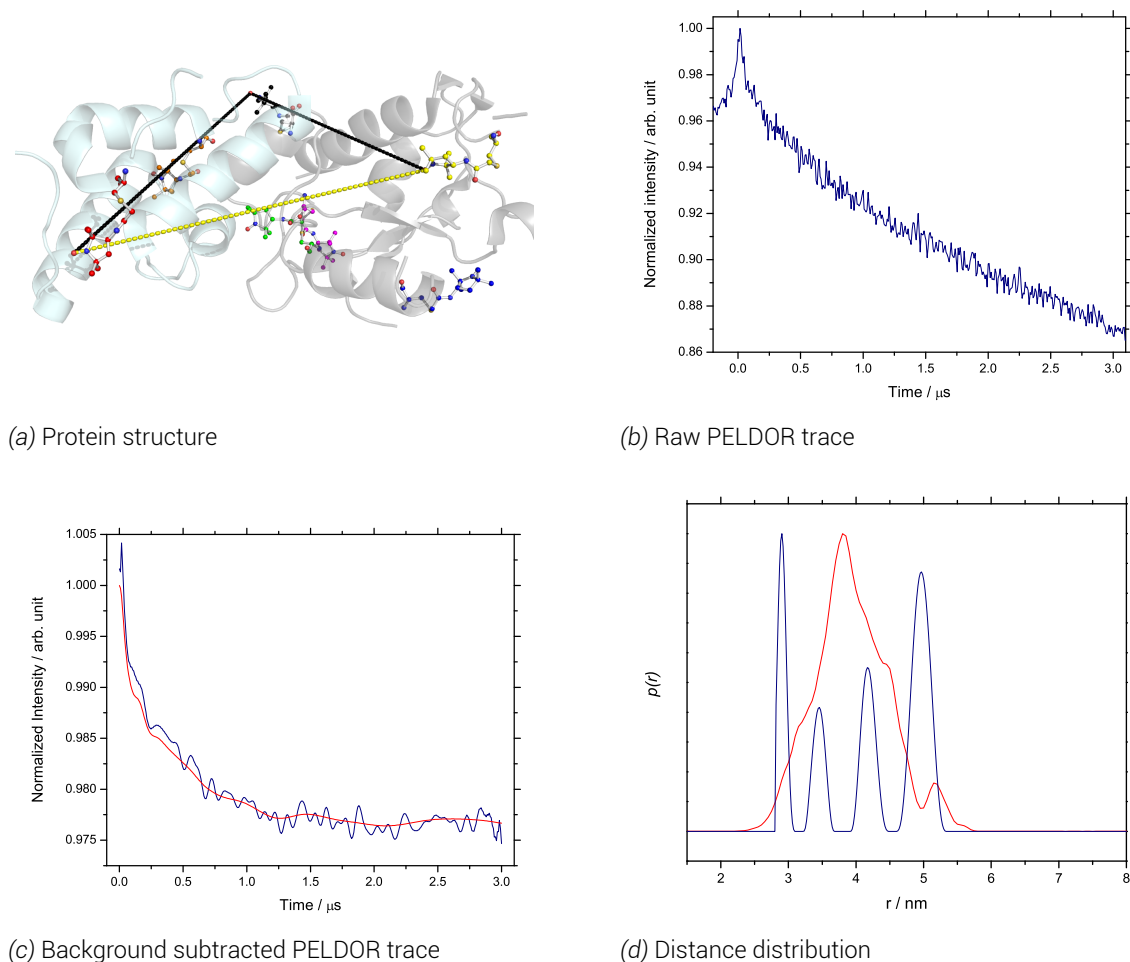
7.2.12. Colicin E9 DNase S3C^{I A1}/S80C^{I A1} - Im9 S6C^{I A1}/C23S

Figure 7.14 PELDOR experiment results for colicin E9 DNase S3C^{I A1}/S80C^{I A1} - Im9 S6C^{I A1}/C23S. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, S80C in black, Im9 S6C in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/-point, 428 points, 11.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.71 GHz, 346.30 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

Here again we see a poor PELDOR trace, but once the background is subtracted and noise filter applied a better trace can be seen than with Im9 C23^{I^{A1}} labelling position. Once again with 3 spin label sites, 3 distances would be expected, except that the E9 DNase S3C^{I^{A1}}-S80C^{I^{A1}} has been shown experimentally to have 2 distances at around 4 and 5 *nm*.

Using the highest populated MMM rotamer from each site we could expect 3 distances in this experiment at 3.19, 3.83 and 4.09 *nm* which align reasonably well to the experimental peaks at 2.90, 3.50 and 4.20 *nm*.

The overall MMM analysis with at least two of the spin label sites being very mobile shows a large rotamer cloud spread which collapses the predicted distance distribution to a single, broad peak spanning 2 *nm*.

Whilst this experiment shows a clear single oscillation (highlighted in appendix D.2.12), the experiment still yields a poor PELDOR trace. However, when a Tikhonov regularisation analysis is conducted with ghost suppression for a three spin system a reasonable L-curve is yielded (appendix D.2.12) with a clear discontinuity. The result is a distance distribution that matches the predicted distance distribution surprisingly well, one which would suggest that the distances are more resolved than predicted.

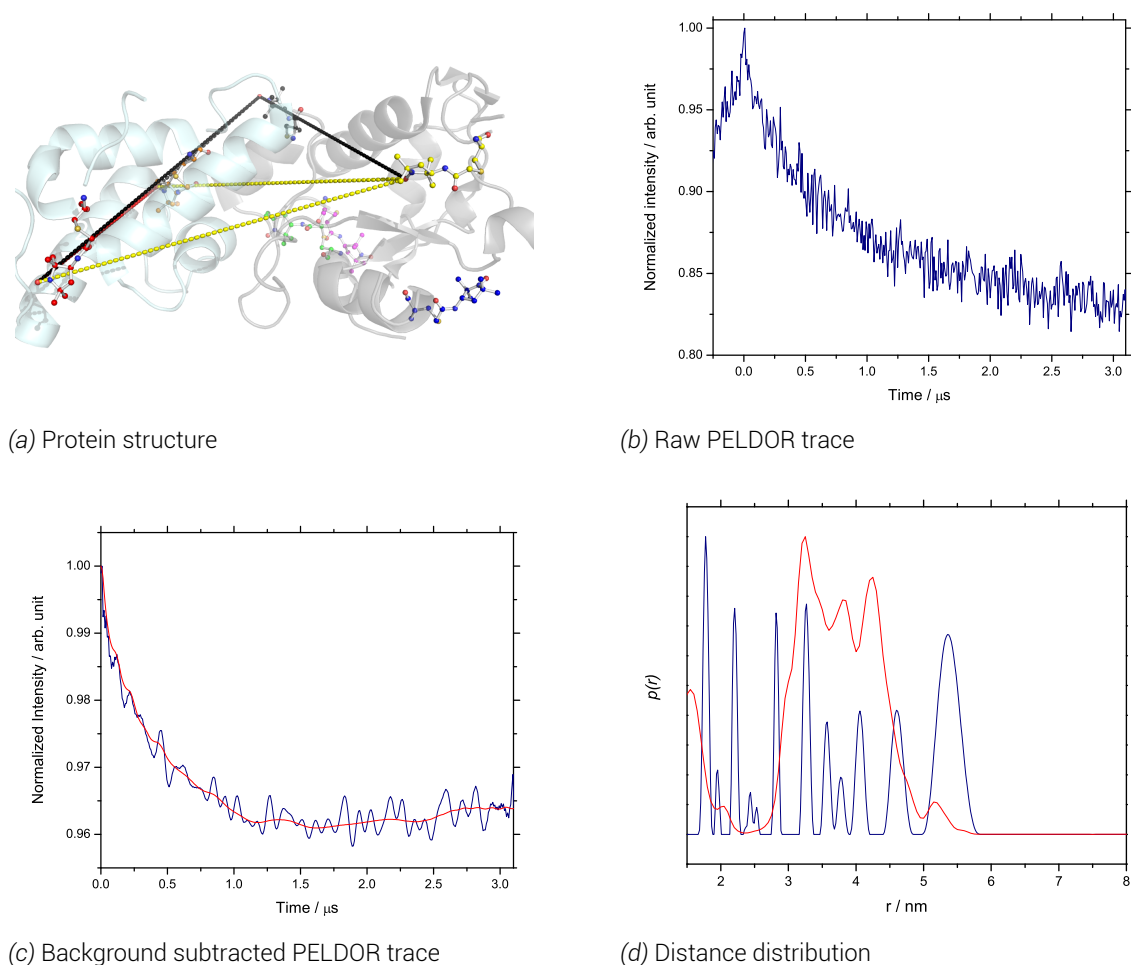
7.2.13. Colicin E9 DNase S3C^{I A1}/S80C^{I A1} - Im9 S6C^{I A1}/C23^{I A1}

Figure 7.15 PELDOR experiment results for colicin E9 DNase S3C^{I A1}/S80C^{I A1} - Im9 S6C^{I A1}/C23^{I A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S3C shown in yellow, S80C in black, Im9 C23 in orange, S6C in red b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 12.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.70 GHz, 342.60 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

This experiment shows a noisy PELDOR trace, but this is to be expected given the labelling efficiencies of all of these sites. A single oscillation is evident, which after a high frequency noise filter to remove frequencies associated with distances less than 1.0 *nm* and background subtraction using E9 DNase S3C^{I_{A1}} with unlabelled Im9 yields figure 7.15.c. Notice again that the renormalized signal intensity puts this single oscillation depth at around 3 % of the signal representing a very low number of spins contributing to the dipolar oscillation.

In this first experiment with 4 spin label sites, with resolved distances there would still be an expected minimum of 6 distances. With the splitting in population evidenced in §7.2.1 between at least two positions in the E9 DNase S3C^{I_{A1}}/S80C^{I_{A1}} this becomes extends the number of expected, individual distances to at least 13. In figure 7.15.d the distance distribution shows 12 distinct peaks; most distinctive of which being the clearly present long E9 DNase S3C^{I_{A1}}/S80C^{I_{A1}} at 5.40 *nm*.

Whether these peaks are artifacts from noisy data or genuine peaks is not possible to discern. The Tikhonov regularisation of such weak data results in an L-curve (appendix D.2.13) that shows no significant discontinuity. By moving to higher regularisation parameter values on the L-curve results in all of these peaks melding into one broad peak of many features and a total width of over 4 *nm*. MMM analysis of the highest populated rotamers of each spin label suggests individual distances between of 3.19, 3.48, 3.83 and 4.09 *nm*. Clearly these distances are too close to each to distinguish when individual peak broadening is considered, resulting in a large, 2 *nm* wide peak in the MMM prediction.

It is clear that 4 spin labels in a single sample is detrimental to analysis, and in this case would be impossible to analyse without first analysis of the 2 spin systems, rendering the 4 spin label experiment useless.

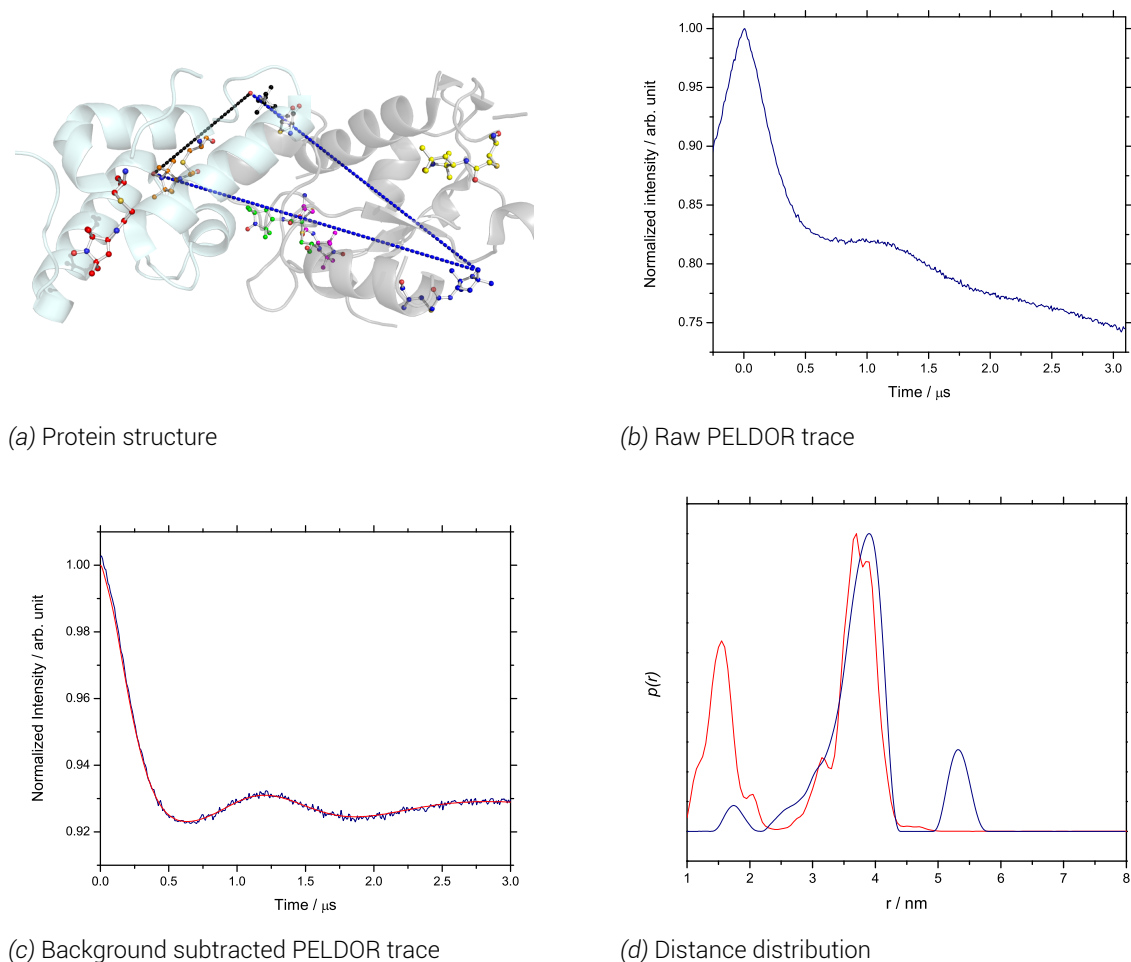
7.2.14. Colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}

Figure 7.16 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S30C shown in blue, S80C in black, Im9 C23 in orange. b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μ s shot rep. time, ~9.70 GHz, 342.50 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

With more than 2 spin label sites comes greater complication in the spin label distance analysis. Undoubtedly, due to the high labelling efficiency of the colicin E9 DNase S30C^{IA1}/S80C^{IA1} protein, this distance dominates this spectrum. However, with such a strong signal, even a small interaction from Im9 should differentiate it. Indeed, figure 7.16.d shows high similarity with the very strong signal seen from the E9 DNase S30C^{IA1}/S80C^{IA1} with some additional, but weaker signals.

Unsurprisingly, with such a strong signal analysis is not difficult. In the raw trace the first modulation shows a depth of 15 % of the total signal. After background subtraction (no high frequency noise filter was required), Tikhonov regularisation produces a characteristic L-curve (appendix D.2.14) with obvious discontinuity.

Again peaks at 3.9 and 5.4 *nm* can be seen from the colicin E9 DNase S30C^{IA1}/S80C^{IA1} interaction, but crucially a shoulder at 3.0 *nm* becomes more pronounced, as well as the appearance of the short E9 DNase S80C^{IA1} to Im9 C23^{IA1} distance, predicted to be only 1.70 *nm* by MMM. This peak however has a much lower intensity than predicted, reflecting the spin labelling efficiencies of the different sites.

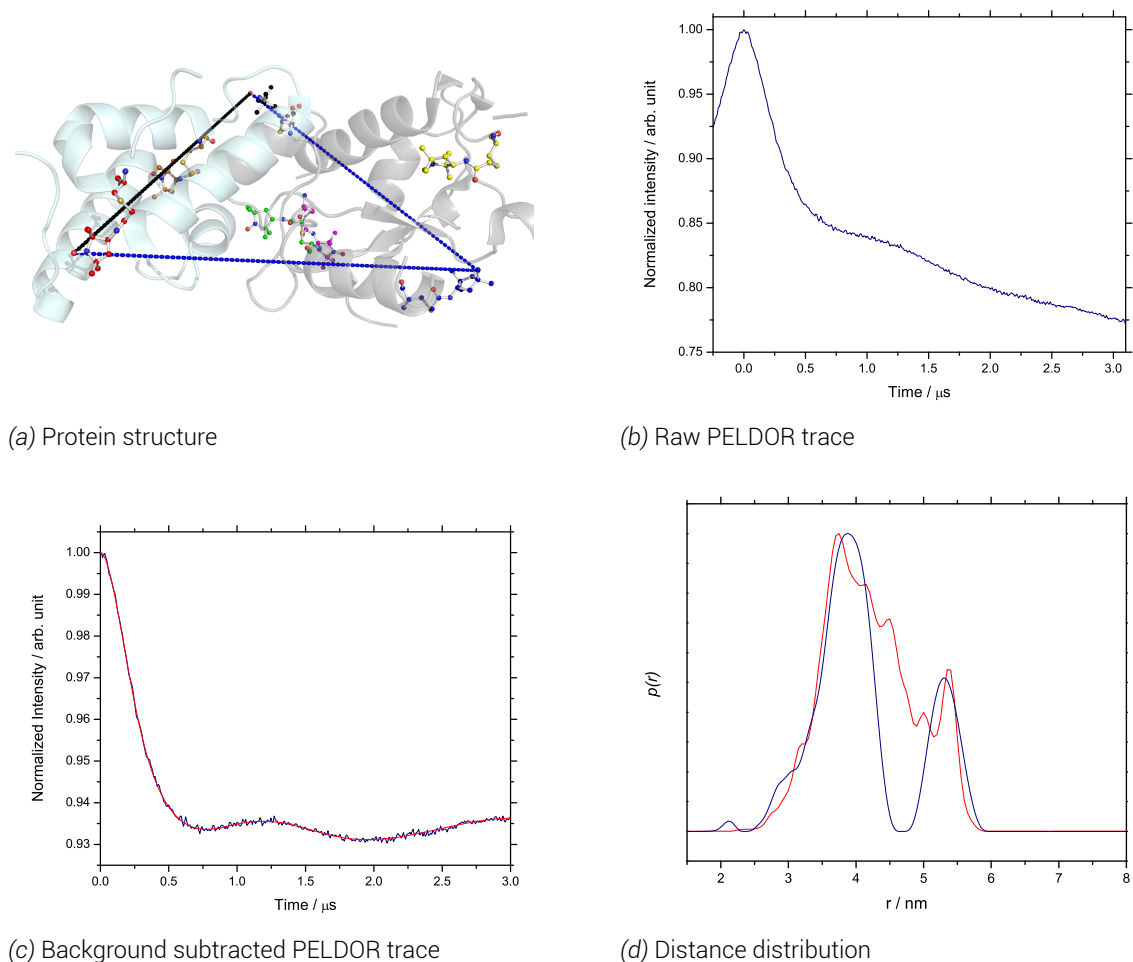
7.2.15. Colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S

Figure 7.17 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S30C shown in blue, S80C in black, Im9 S6C in red. b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/-point, 428 points, 8.0 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μ s shot rep. time, ~9.79 GHz, 345.50 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

From visual inspection of the cartoon representation of the protein complex in figure 7.17.a it would be expected that this experiment has 3 inter-spin distances of approximately the same distance. If it were not for the 5.20 *nm* rotamer spin distance of colicin E9 DNase S30C^{I_{A1}}/S80C^{I_{A1}} then all 3 distances could not easily be distinguished from one another. In earlier experiments it was easy to establish peaks at 3.9 and 5.4 *nm* from colicin E9 DNase S30C^{I_{A1}}/S80C^{I_{A1}}.

Once again, thanks to the strong labelling efficiency of colicin E9 DNase S30C^{I_{A1}}/S80C^{I_{A1}} a PELDOR trace was easy to optimise and collect. Whilst this distance still dominates the distance distribution, when compared with figure 7.4.d of just the colicin E9 DNase S30C^{I_{A1}}/S80C^{I_{A1}} shoulders to the main peak are evident. The signal strength also allows for easy background subtraction and Tikhonov regularisation (figure 7.17.a, 7.17.b and appendix D.2.14).

The longer distance previously associated with colicin E9 DNase S30C^{I_{A1}}/S80C^{I_{A1}}'s longer distance conformation, now appears in the predicted by MMM. Individually in this experiment, MMM would predict distances of 3.68, 4.78 and 4.83 *nm* when using only the highest populated rotamers, yet when the entire rotamer population is considered the prediction aligns very well with the experimental data.

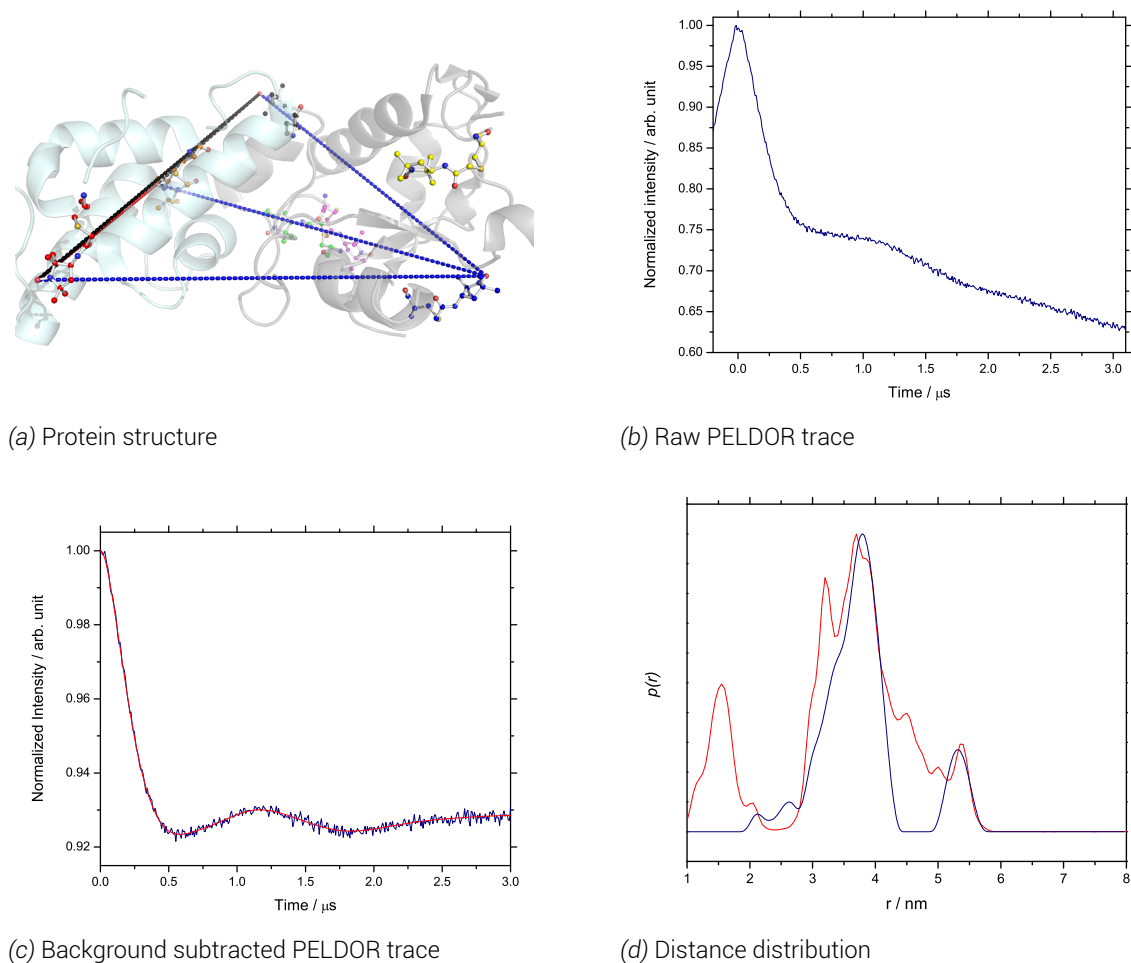
7.2.16. Colicin E9 DNase S30C^{I A1}/S80C^{I A1} - Im9 S6C^{I A1}/C23^{I A1}

Figure 7.18 PELDOR experiment results for colicin E9 DNase S30C^{I A1}/S80C^{I A1} to Im9 S6C^{I A1}/C23^{I A1}. a) shows the PELDOR experiment distance superimposed onto PDB structure 2K5X (colicin E9 DNase in grey, Im9 in cyan). Spin labels are highlighted in coloured ball and stick representation showing highest populated MMM rotamer. E9 DNase S30C shown in blue, S80C in black, Im9 C23 in orange, S6C in red. Distances from E9 DNase S30C shown in blue, E9 DNase S80C shown in black, intramolecular Im9 distance (not obvious, behind black lines) is in red. b) shows the raw 4 pulse PELDOR data. c) shows the PELDOR experiment after background subtraction in navy with the fitting made by DeerAnalysis is red. d) Shows the DeerAnalysis Tikhonov derived distance distribution in navy with the MMM predicted distance distribution in red. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 12.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.70 GHz, 346.30 mT. $\tau_1 = 380$ ns, $\tau_2 = 3200$ ns, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

With 4 spin label sites yields a theoretical 6 distances producing complex PELDOR data sets, pushing the limits of deconvolution. Unfortunately, due to the various spin labelling efficiencies at different sites it also means that some of the distances are drowned out by others. But for such a complicated experiment most of the experimental data corresponds well with the MMM prediction.

Like previous experiments using E9 DNase S30C^{IA1}/S80C^{IA1} the signal strength makes for easy optimisation of the experiment to collect practically noise free data. The high signal to noise ratio and modulation depth of nearly 25 % in the raw spectrum (8 % after background renormalization) makes Tikhonov regularisation easy with a strong L-curve (appendix D.2.16).

As with other experiments the strong signal from colicin E9 DNase S30C^{IA1}/S80C^{IA1} corresponding to distances at 3.68 and 5.20 *nm* can be seen. However, the short (1.71 *nm*) intramolecular distance between the poorly labelling Im9 sites S6C^{IA1} and C23^{IA1} has been lost, almost certainly due to the very low labelling efficiency of these labelling sites in comparison to the colicin E9 DNase S30C^{IA1}/S80C^{IA1} protein. The other distances have either adopted distances too similar or have been drowned out to be seen, this results from the low Im9 labelling but notice the main peak's shoulder at 3.20 *nm* is significantly larger than in the previous two experiments, as well as a very small peak appearing at 2.60 *nm*. Simply, there are too many distance of approximately similar ranges to deconvolute.

7.3. Orientation selection

The PELDOR experiment is designed to measure a vector between 2 points, which by definition, has a distance and direction. A spin label frozen in solution and bound to a protein will have molecular directionality relative to the external magnetic field. When considering 2 spin labels, each with anisotropy, in order to transform between one position and other requires both a transformation in space and orientation [45].

In analysis of the PELDOR spectrum it is possible to deconvolute both the distance between points and the angle of the vector, θ , from the external magnetic field, B_0 . By repeating the PELDOR experiment across the field swept echo spectrum, sampling all orientations and detecting at different g positions it is possible to build up a vector model from these angles.

Typically at X-band the g values from the nitroxide are not sufficiently resolved to accurately determine each of the required angles, though in colicin samples some amount of orientation selection can be seen. By moving to higher fields, such as Q or W-band the g tensors resolve and such angles can be deconvoluted more easily. Moving to higher fields also has the additional benefits of far greater sensitivity and requiring far smaller sample sizes (due to dielectric loss of samples in aqueous buffers).

Through funding from a COST short term scientific mission grant, time could be spent at Prof. R. Bittl's group at Freie Universität Berlin, Germany using their Q-band pulsed spectrometer. Due to the short time available on this machine, it was decided to investigate the colicin E9 DNase S30C^{R1A} to Im9 C23^{R1A} as the colicin E9 DNase S30C^{R1A} sample was very limited and could not provide any more experiments at X-band and had proved very successful at X-band.

Figure 7.19.a shows the field swept echo of the colicin E9 DNase S30C^{R1A}/Im9 C23^{R1A} complex at Q-band. Notice that at the higher magnetic field associated with Q-band that the g tensors have begun to separate.

PELDOR experiments were methodically taken across the field swept spectrum with the most important positions being sampled first. In figure 7.19.a the experiment pump positions have been highlighted. To ensure the best data acquisition initially the observation position was kept to the maximum signal intensity position within the spectrometer bandwidth to the pump position. Ideally, 6 PELDOR experiments should be conducted in total such that each of the three tensor orientations is recorded to every other.

In the corresponding PELDOR traces (figure 7.19.b) field dependence can be seen first by differences in modulation depth - this is the same sample simply with excitation occurring in different orientations. This means that there is a difference in populations between the different orientations corresponding to an orientation dependence. Further, whilst the oscillation periodicity does not vary significantly between black, green and blue traces (g_{xy} , g_y , g_x pump positions respectively) the magenta trace (g_z pump position), shows a very shallow modulation depth due to the small number of spins in this orientation but also the oscillation periodicity varies slightly. This suggests that the distance in the g_z frame is slightly longer than that in the g_x or g_y frames.

Indeed, in figure 7.19.d where the distance distributions are calculated using a Tikhonov regularisation approach that the g_{xy} - g_z distance distribution is split into two peaks. The

upper peak matches the other orientations well but a significant peak also appears at a shorter distance of 2.7 *nm*.

In figure 7.20 a comparison of the same sample run at X and Q-band can be seen. The X-band experiment (figure 7.20.a-c) shows observation at the low field peak with ELDOR pumping of the centre peak, marked with navy arrow. Total experiment time was 23 *h* to give a measured distance of 3.84 *nm* with a half height width of 0.5 *nm*. A secondary ghost peak is shown at 5.20 *nm*.

The Q-band experiment (figure 7.20.d-f) shows the extremes in PELDOR experiments with high field excitation (corresponding to g_z) (blue) and low field peak (corresponding to g_{xy}) (black). Both of these experiments shows a good SNR after only 90 *min* acquisition time, and show oscillations of similar periodicity but distinctly different modulation depths. In figure 7.20.f the distance distributions of the PELDOR traces show a slight differing in ghost peaks but the main peaks are almost identical at 3.75 *nm* for the high field spectrum with a half height width of 0.3 *nm*, whilst the low field peak shows a distance of 3.97 *nm* and a half height width of 0.7 *nm*.

Unfortunately, DeerAnalysis was used to calculate these distances in which zero orientation dependency is assumed. Whilst these distances can be assumed to be correct, DeerAnalysis is not the program to analysis the angular information contained within these spectrum. Additionally, to fully deconvolute the angular transformation would require 6 PELDOR experiments (g_x-g_y , g_x-g_z , g_y-g_z , g_x-g_x , g_y-g_y , g_z-g_z) and specially designed software. At Q-band in this case it proved difficult to resolve all of the g tensors due to a strong mixing of g_x and g_y into g_{xy} and despite the strength of the signal, measurements with g_z proved difficult due to low spin population, particularly in g_z-g_z where no PELDOR could be seen. Perhaps by further increasing the field further to W-band would provide sufficient resolution of the g tensors to easily measure each orientation. However, with no freely available analysis software data analysis would remain extremely difficult and time consuming.

7. Pulsed Investigation

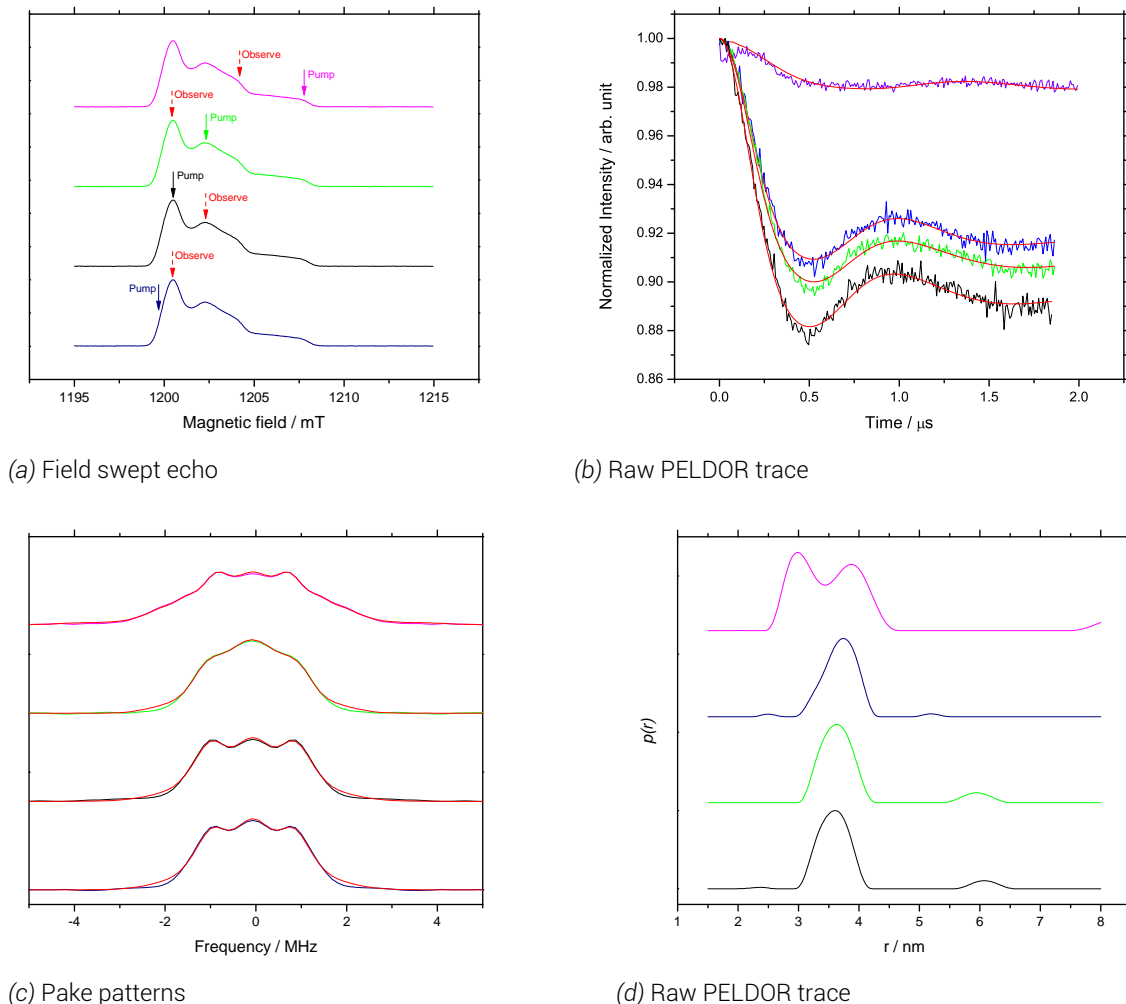


Figure 7.19 Q-band PELDOR experiment results for colicin E9 DNase S30C^{R1A} to Im9 C23^{R1A}. a) shows the field swept echo repeated 4 times, in 4 experiment colours, open each field sweep spectrum the pump and observe position for the PELDOR experiment is shown. The pump position is highlighted by a solid arrow in the experiment colour. The observe position is highlighted by a dashed arrow in red. b) shows the PELDOR experiment after a 3 dimensional homogenous model background subtraction. Experiment colours remain in the same colour, DeerAnalysis fits are shown in red. c) shows the PELDOR traces Fourier transformed into the frequency domain, known as a Pake pattern. d) Shows the DeerAnalysis Tikhonov derived distance distributions, maintaining experiment colours. All experiments 50 K, 0.0 dB attenuation, 20 MHz video bandwidth, 30 dB video gain and 1000 μ s shot rep. time, ~ 33.74 GHz. a) 4 shots/point, 10 scans, 512 points, 20 mT sweep width, 1000 μ s shot rep. time, 1202.40 mT. $\tau_1 = 180$ ns. b) 30 shots/point, 260 points, 4.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 2000 μ s shot rep. time. $\tau_1 = 380$ ns, $\tau_2 = 3000$ ns, ELDOR pulse step 60 ns, no nuclear modulation suppression. 4 hours collection per experiment, except magenta trace that required 36 hours. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

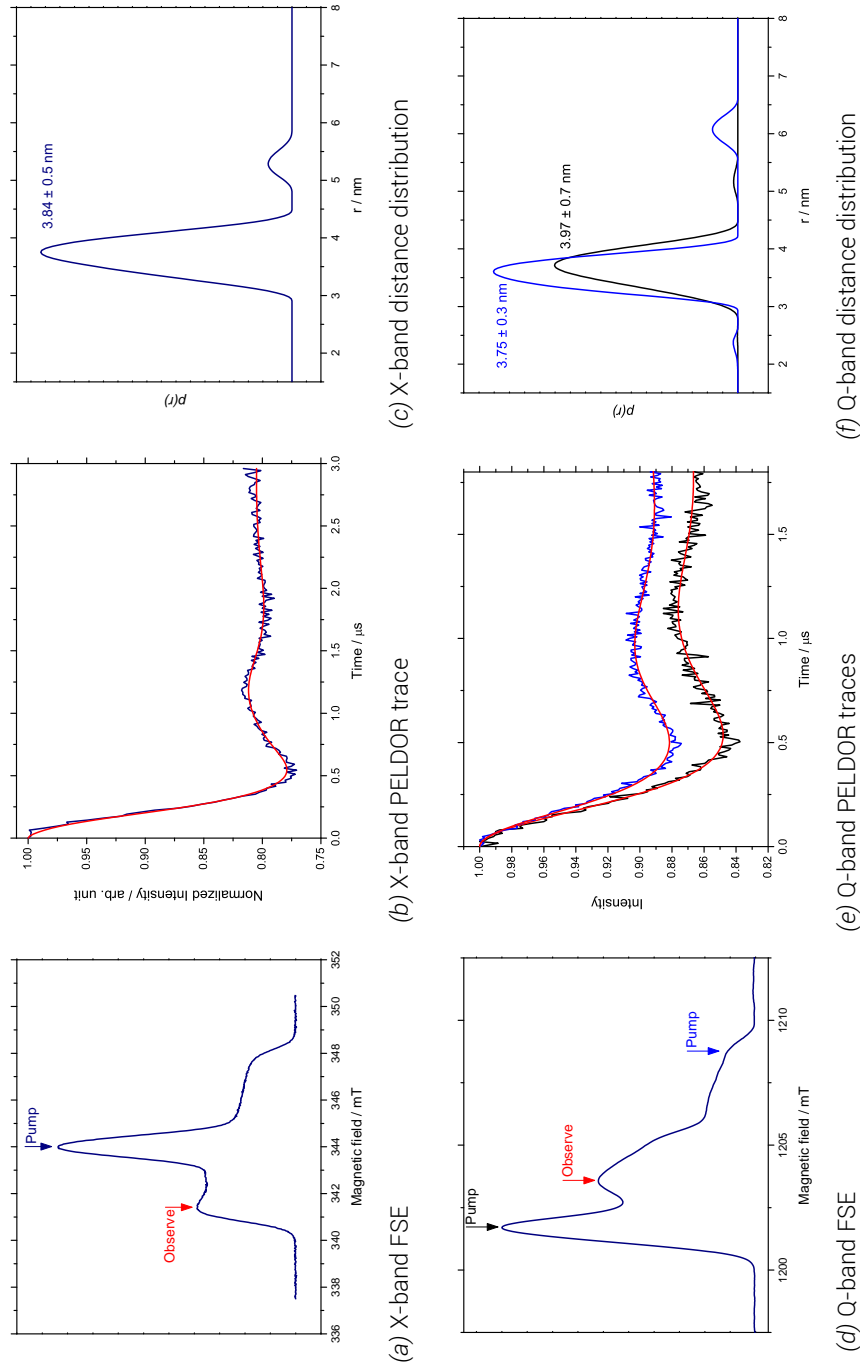


Figure 7.20 Comparison of X- and Q-band PELDOR experiment results for colicin E9 DNase S30C^{R1.4} to Im9 C23^{R1.4} showing the 2 extreme Q-band spectra. a) shows the X-band FSE with pumping at the blue arrow and detection upon the red arrow b) shows the resulting X-band PELDOR trace after background subtraction, with DeerAnalysis fit in red c) shows the X-band distance distribution from DeerAnalysis d) shows the Q-band FSE with pumping at the black and blue arrowed positions and observation at the red arrowed peak between them e) shows the resulting PELDOR traces after background subtraction, traces are colour coded to match the arrows of d). with DeerAnalysis fits in red f) shows the DeerAnalysis derived distance distributions with colours again matching the arrows in d). a) 50 K, 4 scans, 4 shots/point, 1024 points, 9.7047 GHz, 6.0 dB attenuation, 20 MHz video bandwidth, 45 dB video gain, 12 mT sweep width, 10000 μ s shot rep. time b) 50 K, 20 hours collection (\sim 2500 scans), 4 shots/point, 253 points, 4.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μ s shot rep. time, \sim 9.66 GHz, 344.70 mT. $\tau_1 = 380$ ns, $\tau_2 = 3000$ ns, ELDOR pulse step 12 ns, no nuclear modulation suppression. d) 50 K, 0.0 dB attenuation, 20 MHz video bandwidth, 30 dB video gain, 1000 μ s shot rep. time, \sim 33.74 GHz, 4 shots/point, 10 scans, 512 points, 20 mT sweep width, \sim 33.77 GHz, 1202.40 mT. $\tau_1 = 380$ ns, $\tau_2 = 3000$ ns, ELDOR pulse step 60 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-ds.

7.4. Summary

After 16 unique sample PELDOR experiments, numerous repeats, optimizations, 2 spin labels and 2 different microwave frequencies a database of distance constraints could be built using 5 of the original 7 designed labelling sites, graphically presented in figure 7.22. This database does not contain a fixed point distance, but, utilising a combination of peak width and multiple peak analysis creates an allowed region. A distance constraint with upper and lower limits (discussed later in §8.4).

Some experiments gave clear and easily interpretable data with obvious, concise distances. Others gave ambiguous results, where rotamer populations were giving rise to significant peaks in the distance distribution that were not anticipated from modelling of the spin labels.

In this investigation, a significant amount of protein was able to be produced allowing for a long time to be spent ensuring that the highest possible spin labelling efficiencies for each site was obtained. Despite, this effort the protein concentration before precipitation was little more than 200 μM making stock solutions of any concentration difficult. This, combined with the much lower than expected labelling efficiencies meant that many of the PELDOR experiments were always going to be challenging. With careful sample preparation and experimental setup it was possible to observe dipolar oscillations in some experiments.

Notable, the synthetic colicin E9 DNases E9 S3C^{IA1}/S80C^{IA1} and S30C^{IA1}/S80C^{IA1} with their much higher labelling efficiency gave clear dipolar oscillations which whilst should have yielded easy analyse the apparent mobility of the S80C^{IA1} site split both experiments into an expected, highly populated distance and a longer, less populated, but significant, longer distance. Unfortunately, as these mutants labelled approximately eight times better than the Im9 proteins the effect upon the distance distribution on addition of the Im9 was only slight. However, as the colicin E9 DNase in this case was providing so much signal that the Im9 could be seen, but only as shoulders to the intramolecular colicin E9 DNase distance, rather than the more useful to docking intermolecular distance. This made deconvolution of the exact distances much more difficult than had the labelling efficiencies of both proteins been approximately equal and provided equal intensity in the distance distributions.

In many experiments the combination of two proteins with low spin labelling efficiencies meant that the effective concentration of complex contributing to the sought after dipolar oscillations was very low and often sub-micromolar. In some instances this amounted to PELDOR traces that lifted only slightly from the background decay rather than giving a clear oscillation as was the case with experiments to the colicin E9 DNase S3C^{IA1}. In other instances, as with the colicin E9 DNase E66C^{IA1} to Im9 C23^{IA1}, the signal was so weak that all that could be detected was deuterium ESEEM modulations.

In the most extreme cases detection was almost impossible. The colicin E9 DNase D20C/E66C with spin label efficiency of 1.4 % meant that no echo could be detected to even begin the PELDOR experiment. In the case of the colicin E9 DNase S80C^{IA1}, the protein labelling was so low as to not be detectable with cw EPR above background noise of the instrument.

The labelling efficiencies were incredibly frustrating to work with significantly hampered the investigation. It was the original intention of this investigation to produce far more intermolecular distances to triangulate the position of the docking partners in the complex than were required. This would thereby allow an investigation to be made into the minimum num-

ber of distance restraints to come to an unique solution, and compare the result in biology to the theoretical basis that would require 6 distances, 1 for each motion of freedom.

The idea to engineer double mutations was also a mixed result. In most cases the mobility of the spin label gave rise to broad distance distributions. Given the size of the proteins in question, these broad peaks occur in a relatively similar distance range making deconvolution of each distance difficult had the structure of the complex not already been known. With less mobile spin label sites it was hoped that by simply running a PELDOR experiment of the doubly labelled protein with and without the complex partner would be sufficient to deconvolute the complex spectrum. To an extent this was shown and distances that could not be obtained via singly labelled mutants (such as colicin E9 DNase S30C and S80C) were obtained. However, recent work has shown that in systems with more than two spins show some multiple spin effects which may distort the distance distributions. The resulting spin interactions can sum and difference in the dipolar frequency contributions giving rise to additional peaks in the distance distribution which do not correspond to real distances. It has been shown that in some circumstances that these interspin effects can result in ghost peaks so prominent that they may be mistaken for real distance peaks or that the real peaks may move or disappear entirely [35, 95]. This ghost peak could perhaps explain why some of the later experiments, particularly those with 4 spin labels, such as the colicin E9 DNases E9 S3C^{I^{A1}}/S80C^{I^{A1}} to Im9 S6C^{I^{A1}} /C23^{I^{A1}}, showed such convoluted spectra with so many peaks, even despite so many distances expected to be obtained from this experiment.

Until these effects have been investigated further the doubly labelled mutant intermolecular distances need to be carefully scrutinised. As a result of this new information doubly labelled mutants would not be undertaken if this investigation were to start now.

Despite the failure of colicin E9 DNase S30C and S80C proteins to be expressed in any quantity or the colicin E9 DNase D20C/E66C to label, 11 distance constraints still remain to be taken forward to modelling.

To further aid the reader on the next page two summary figures are presented. In figure 7.21 the spin labelled colicin E9 mutants are plotted against the spin labelled Im9 proteins in a grid fashion showing the resulting distance distribution. The top row and first column are blank as a result of doubly labelled proteins being tested in isolation. Blank panels are a result of experiments that failed to yield results.

In figure 7.22, the highest probability peak from figure 7.21 for each distance is displayed upon a cartoon representation of the colicin E9 DNase / Im9 complex.

7. Pulsed Investigation

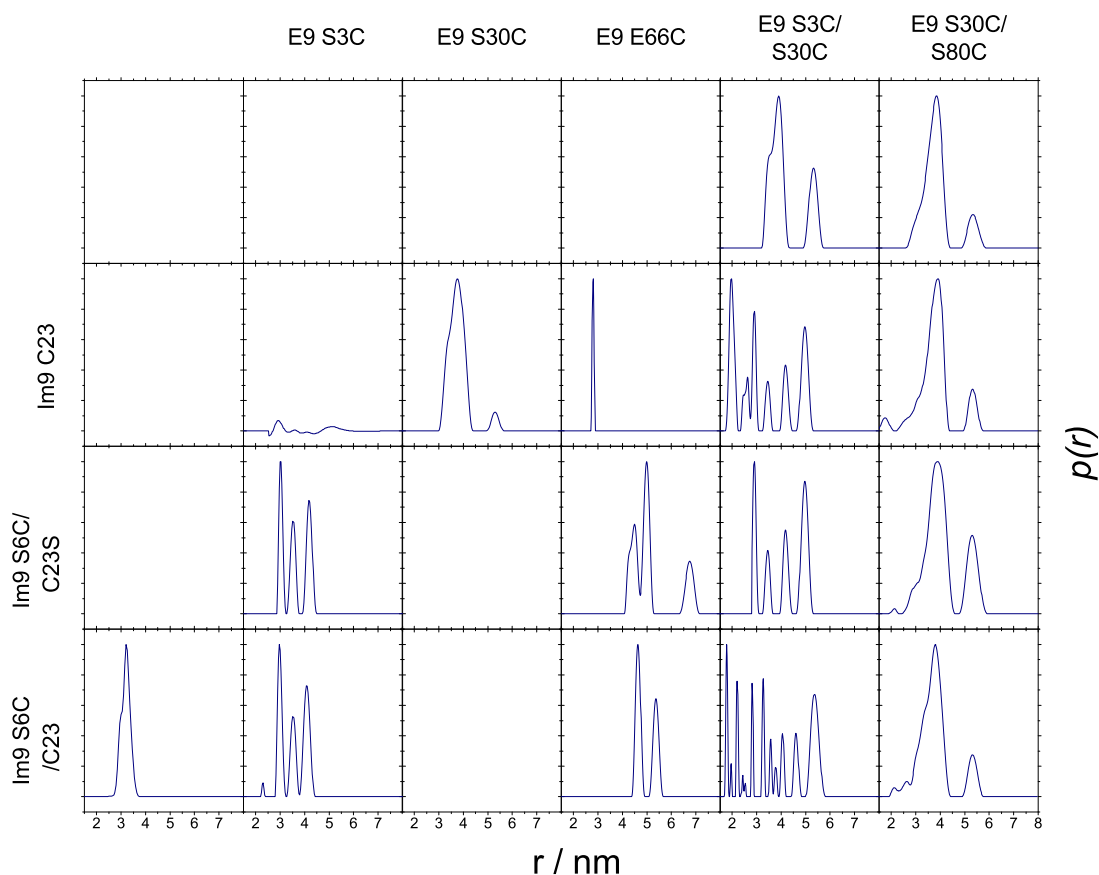


Figure 7.21 A binding cross summary of distance distributions from all PELDOR experiments. Presented horizontally are the spin labelled colicin E9 mutants plotted against the spin labelled Im9 proteins. The top row and first column are blank as a result of the protein being run in isolation.

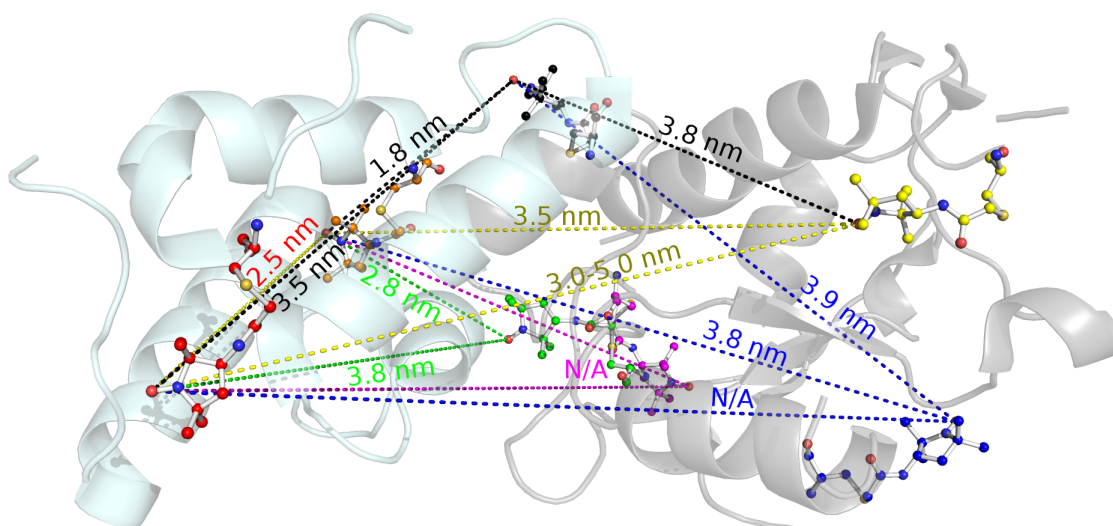


Figure 7.22 Graphical summary of PELDOR experiments with final distances superimposed onto the protein structure of PDB file, 2K5X. In grey, colicin E9 DNase, in light cyan, Im9. Dashed lines and labels are coloured by the origin of the distance. In ball and stick representation, the spin labels upon the colicin E9 DNase the blue is S3C, magenta is D20C, yellow is S30C, green is E66C black is S80C. Upon the Im9 red is S6C, orange is C23.

8. Protein - protein docking with HADDOCK

Reported structural binding sites for colicins in general, but specifically non-cognate binding partners, are very limited. Their transitory, dynamic nature and low K_d values has made NMR structural studies difficult and, as yet, crystallisation unsuccessful. Reflected by the low numbers of deposited complexes in the Protein Data Bank.

Traditionally the NMR approach to solving protein-protein complexes requires the lengthy measurement of nuclear Overhauser effects (NOEs) whereby spin coupling of electron orbitals in close proximity transfer nuclear spin polarisation from one nucleus to another by cross-relaxation. This typically requires atomic level assignment of side chains for sufficient accuracy requiring excellent protein purity and stability.

With the use of ^1H - ^{15}N heteronuclear single-quantum correlation (HSQC) NMR experiments changes in the chemical shift can be detected upon binding. Typically this is achieved by isotope labelling each protein in turn and titrating in the unlabelled binding partner and observing the change, or perturbation. Residues that are observed to have a chemical shift perturbation (CSP) have undergone a change in environment and can be used to identify regions of intermolecular interaction and thus can be used to map an interaction surface. In figure 8.1 the residues with a perturbation of greater than one standard deviation have been drawn showing the side chains in line format and each atom's Van der Waal radius drawn; the colicin E9 DNase shows a definite binding surface whilst the strongly α helical Im9 shows a spread of residues (as a conformational shift that moves one end of an α helix will alter all residues on that helix).

As computational power has become more available several *in silico* methods for the prediction of protein-protein complexes have been developed. Some operate through *ab initio* procedures requiring no experiment data, whilst others make use of various biochemical and

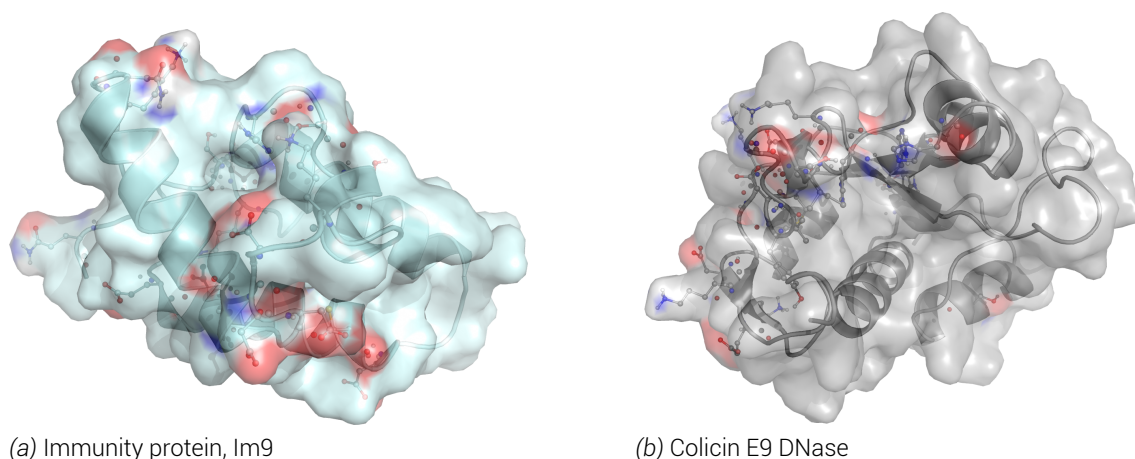


Figure 8.1 Residues with significant chemical shift perturbations mapped onto the NMR structure of the colicin E9 DNase / Im9 complex (PDB ID: 2K5X), side chains are coloured by atom (oxygen red, nitrogen blue)

biophysical techniques such as mutagenesis studies (including fluorescence and isothermal titration calorimetry), bioinformatics and, of course, NMR.

8.1. Docking principles

Docking methodologies assume that the lowest possible free energy state of a protein-protein system represents the structure of the complex. Therefore to find the lowest possible energy state a number of energy states need to be sampled and the free energy calculated. Early methods of docking used low resolutions (residue scale) [107] but as sampling methods have become more efficient and energy calculations can be evaluated more accurately the resolution of methods has been improved to the extent that matching of complementary shapes of partner molecules can be matched although the flexibility of such areas is often not adequately accounted for. As a result molecular dynamics and incremental simulated annealing techniques were later introduced [28, 26, 31, 14].

All docking procedures consist of at least two steps and usually three. First the molecules are randomly orientated, rotated and translated about each other. Using some form of scoring criteria, such as energies of electrostatic interaction, Van der Waal's (VdW) repulsion or other effect that is heavily weighted toward shape complementarity. A number of candidate structures are identified and are taken forward to the second round of optimisation whereby each structure is reranked according to the energy of association using more realistic energy functions, reflecting better native conditions. Finally the remaining structures are considered at the atomic level with rotation allowed on side chains and the protein backbone to further optimise the free energy.

The energy-score landscape, like the protein-folding landscape is full of local minima in which optimisation can easily get stuck. For this reason it is important to start with many random positions and to introduce experimental restraints as early as possible to improve the efficiency of low energy structure discovery. Literature reviews of *in silico*, NMR based, docking studies show that valuable information can be ascertained in this way, particularly useful in high molecular weight or dynamic systems [9, 92].

8.2. HADDOCK overview

For the docking calculations described henceforth the data driven program HADDOCK (High Ambiguity Data-driven biomolecular DOCKing) [26, 28] will be used. Whilst other programs are available, none provide the ability to use NMR data as docking constraints and include the simulated annealing permitting the rearrangement of both backbone and side chain atoms required for the optimisation of interface packing. The algorithms which underpin HADDOCK have been well tested and shown to perform well in CAPRI¹ evaluations [26]. In these evaluations various protein-protein complex prediction software packages are entered into a competition to evaluate the performance of various approaches upon a blind protein-protein experiment of, as yet, unpublished X-ray crystal or NMR structures. HADDOCK distinguishes

¹Critical Assessment of PRediction of Interactions - <http://www.ebi.ac.uk/msd-srv/capri/>

itself from most other *ab initio* complex docking packages as it uses an information driven docking approach, whereby the available information is used from the earliest possible point in time. This minimises the required computation time as opposed to generating a large sample space and then filtering those solutions that do not match the available information. As a result, HADDOCK has become the *de facto* standard in the NMR field with over 3200 registered users².

Since the original HADDOCK release in 2006, HADDOCK has been in constant development and has been subsequently expanded to include docking from proteins to and with DNA, RNA, select unnatural amino acids, cofactors (including hemes) and pharmaceutical based small ligands, to name but a few. This represents no small feat as the intrinsic flexibility of poly-nucleotide chains has often hampered the development of effective protein-nucleotide docking methods.

Importantly, HADDOCK is a high level program consisting of several programming languages, that uses many of its own scripts to generate a fast and usable interface for the complex docking. It originally evolved from several scripts in the Ambiguous Restraints for Iterative Assignment program (ARIA) (Institut Pasteur, FR), a program designed to automate the nuclear Overhauser effect assignments in NMR. The actual process of moving and rotating proteins in a virtual space as well as the important task of energetically scoring a structure is performed by calls to the Crystallography and NMR System (CNS) program CNSsolve (Yale University, US). CNS itself is an offshoot of X-PLOR (Yale University, US) as a result of X-PLOR development all-but stopping in the long-term stable release XPLOR-NIH. Of course X-PLOR itself was an offshoot of the much earlier Chemistry at Harvard Macromolecular Mechanics (CHARMM) package (Harvard University, US) which was originally released in 1983. As a result all four of these programs have a large amount of crossover in their programming languages and syntax.

HADDOCK consists of three stages similar to that described previously (summarised in figure 8.2). In the first stage the two or more molecules are positioned 25 Å apart in a virtual Cartesian coordinate with each dimension theoretically extending between -99,999 and 99,999 Å. In this virtual space each molecule is freely rotated about its centre of mass. Each random rotation step is followed by cycles of rigid body intermolecular energy minimisation where the molecules are moved together before allowing simultaneously rotation and translation to produce a complex. 1000 Monte-Carlo generated model structures are generated and are subsequently ranked according to a weighted sum of the model's Van der Waal's energies, electrostatic energies, buried surface area (hydrophobic energies) and ambiguous interaction restraints (AIRs).

The top 200 scoring models are taken forward into the second stage of docking where model in turn is subjected to three simulated annealing routines, whereby the model is heated in the virtual space and allowed to cool. With each annealing step it is assumed that the molecules will cool into the lowest possible energy state as they collapse down the energy landscape to the minimum, which likely represents the biological solution. Initially, in the first annealing step each docking object is considered as a rigid body. In the second annealing the side chains that are defined as being active residues are allowed some freedom of movement as well as the side chains of within two residues either side of the active residue. In this instance an active residue is one that is considered to have a large perturbation upon

²At August 2013, <http://haddock.science.uu.nl/enmr/services/HADDOCK/>

8. Protein - protein docking with HADDOCK

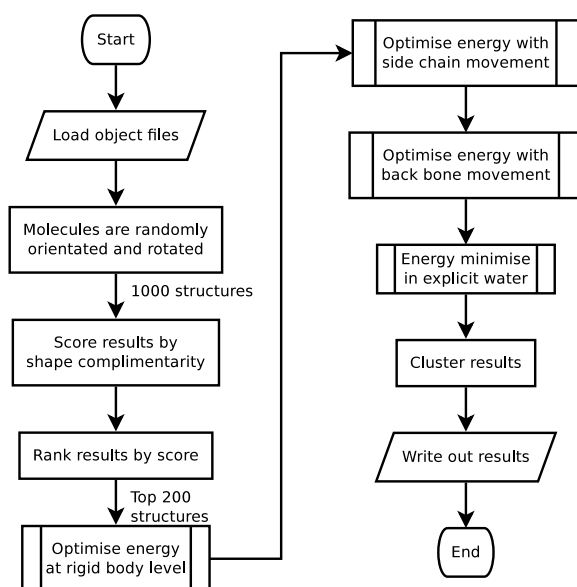


Figure 8.2 Flowchart representation of the HADDOCK modelling process, model restraints are applied at with each of the four energy optimisation steps

the creation of the complex. Typically, these are obtained from chemical shift perturbation measurements in NMR, but HADDOCK also allows for the random generation of restraints. In the third stage the entire protein backbone and side chain atoms are allowed freedom of movement - allowing for conformational rearrangement at the interface.

In the final stage of docking each model is energy minimised in explicit water or other solvent such as DMSO. For ease of analysis the final structures are typically clustered according to the pairwise backbone root mean squared deviation (RMSD). This means that each model is compared to every other model and the backbone atoms are aligned with the difference in space between them calculated. A cluster is defined by default as consisting of at least 4 structures with a backbone average RMSD of less than 7.5 Å from one another. In a good docking model with high confidence there will be few clusters, perhaps only one cluster containing most, if not all, of the final 200 models.

NMR chemical shifts have long been considered too unreliable and poorly understood to be used in structural information although with the advent of *ab initio* modelling they are now being used often as the only constraint in three dimensional structural determinations. However, this technique only measures a change in chemical environment and so a certain amount of attention needs to be paid such that long range conformational changes, remote from the binding surface typically through hinges or helices, do not have an unfair bias in the calculations.

8.2.1. Docking methodology

Both X-ray and NMR structures of the cognate colicin - immunity protein complexes have shown that neither molecule undergoes significant structural changes upon binding [7, 27] with backbone RMSDs between the free and bound colicin E9 DNase of 1.7 Å (X-ray only) and 1.6 Å for Im9. Furthermore there are minimal differences in inter-helix angles between free and bound forms suggests that the two proteins are interacting fundamentally as rigid bodies, and by extension it is assumed that the weaker binding of non-cognate partners be-

haves in the same way which allows us to justify our usage of CSPs as the driving restraints in docking.

After assignment of HSQC and triple resonance experiments a table of chemical shifts for each residue of each protein with and without the binding partner can be constructed. Residues that are considered as active in docking are defined as those with a calculated difference in chemical shifts outside of normal range associated with binding. Typically residues that are outside 1 standard deviation of the average chemical shift value are defined as active to binding.

These residues are refined such that only residues with greater than 50 % solvent accessibility are classified as *active* residues, whilst residues that are within 3 Å of active residues and have less than 50 % solvent accessibility are considered to be *passive* residues to binding. Together these active and passive residues make up the ambiguous interaction restraints (AIRs), where an AIR is defined as an intermolecular distance between any atom (m) of an active residue (i) of molecule A (m_{iA}) and any atom (n) of either active or passive residues (k) (N_{res} in total) of the partner molecule B (n_{kB}) and similarly for molecule B giving the effective distance calculated for each restraint as:

$$d_{iAB}^{eff} = \left[\sum_{m_{iA}=1}^{N_{atoms}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}}^{N_{atoms}} \frac{1}{d_{m_{iA}.n_{kB}}^6} \right]^{\frac{1}{6}} \quad (8.1)$$

Where N_{atoms} is all given atoms of a residue and N_{res} the total number of active and passive residues for a given molecule. Passive residues as such do not constrain the partner protein but they do allow for the fulfilment of active restraints. An upper limit of 3.0 Å is imposed by HADDOCK on the effective distance as a compromise between minimum hydrogen - hydrogen and heavy atom - heavy atom Van der Waal distances whilst the $\frac{1}{r^6}$ averaging guarantees that an AIR is fulfilled once any two atoms of the proteins are in contact.

8.2.2. Colicin protein structure

Any docking solution is only as good as the quality of the information put into the model. For this reason it is essential to consider all of the currently available information. Several different structures for colicin E9 DNase and Im9 have been published as work upon the system was conducted, and are accepted by the field as being of high quality.

A NMR structure for colicin E9 DNase alone does not exist, in solution there is too much intrinsic dynamic movement between at least two different states which broadens all peaks and makes assignment impossible. This dynamicity made crystallisation difficult and could only be achieved in 30% PEG 5000, 0.1 M sodium acetate, at pH 5.4 and at 285 K, conditions far removed from those found *in vivo*. Nevertheless, upon binding, the states of colicin E9 DNase collapse, which have allowed for successful NMR assignment. Fortunately the NMR structure of the complexed colicin E9 DNase agrees favourably with the crystal structures (see figure 8.3).

Of course any structural solution for either binding partner when put through the docking process should ideally generate the same docking solution. The quality of docked models was assessed by comparison to that of the NMR structure of the complex (PDB: 2K5X) by 1)

8. Protein - protein docking with HADDOCK

PDB (chain)	Type	Methodology	Experimental	Date of publication
2K5X (B)	Complex	NMR	Minimized average	July 2008
1FSJ	Free	X-ray	1.8 Å resolution	September 2000
1EMV (B)	Complex	X-ray	1.8 Å resolution	March 2000

(a) Colicin E9 DNase structures

PDB (chain)	Type	Methodology	Experimental	Date of publication
1IMP	Free	NMR	21 conformers	May 1996
1IMQ	Free	NMR	Minimized average of 1IMP	May 1996
2K5X (A)	Complex	NMR	Minimized average	July 2008
1E0H	Complex	NMR	20 conformers	March 2000
1EMV (A)	Complex	X-ray	1.8 Å resolution	March 2000

(b) Im9 structures

Table 8.1. Available PDB structures for the cognate pairing Colicin E9 DNase - Im9

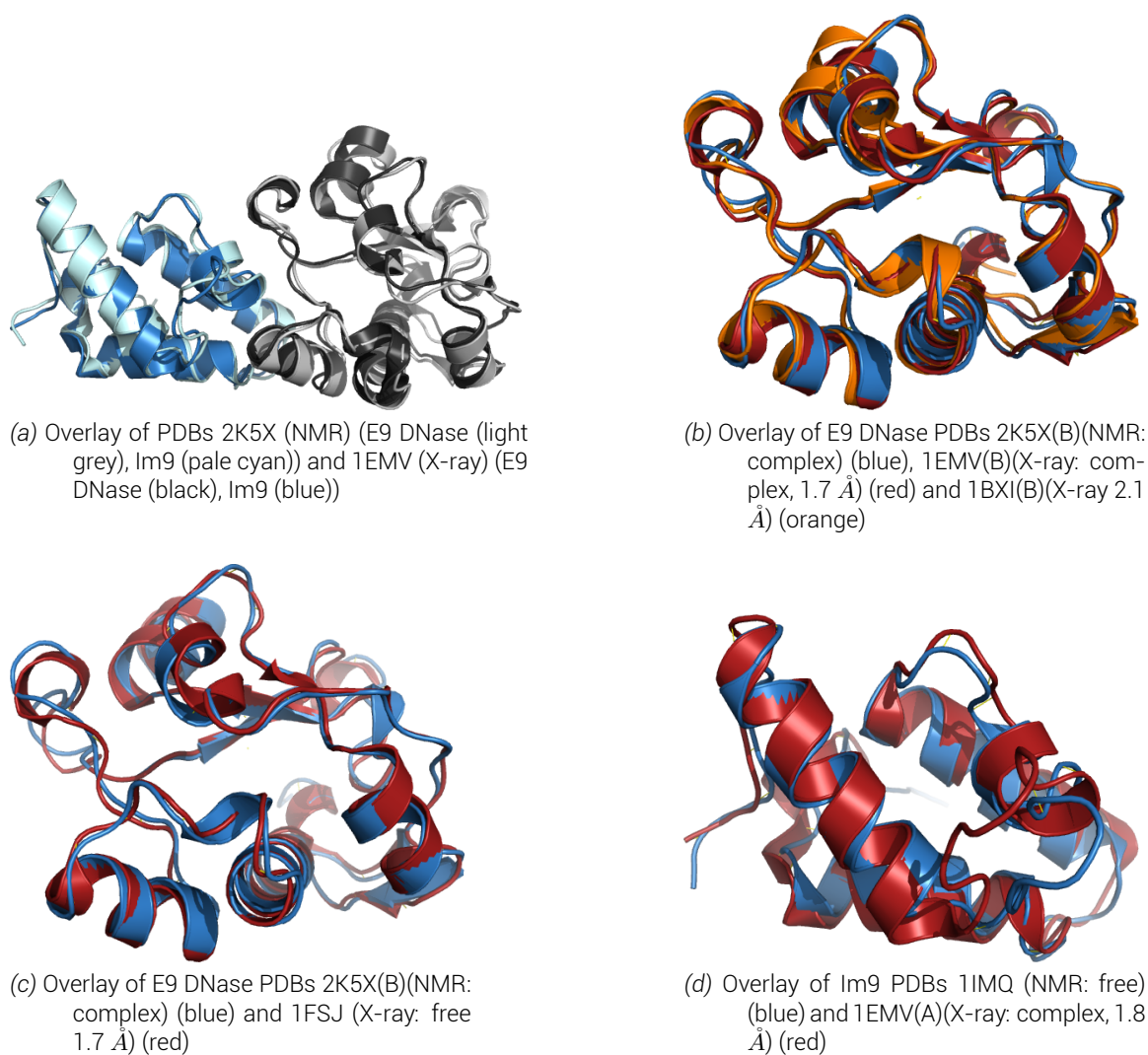


Figure 8.3 Visual overlays of various cognate partner structure files

the overall and interface RMSDs 2) backbone RMSDs of Im9 when the DNase co-ordinates are overlaid and 3) the fraction of native contacts.

8.2.3. HADDOCK's application to colicins

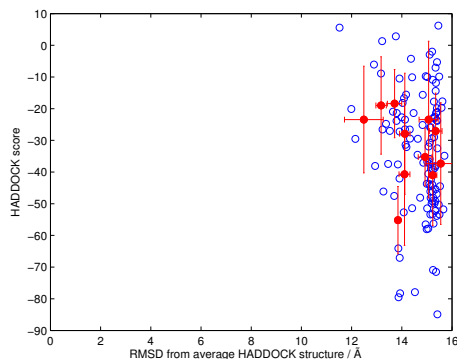
Performing a HADDOCK calculation using only the raw, unpublished CSPs from NMR experimentation (one standard deviation threshold) and Naccess to calculate the solvent accessibility, an interesting result occurs. In figure 8.4.a the HADDOCK score plot is seen, where the HADDOCK score (more negative is a better docking model) is plotted against the RMSD of each cluster from the average HADDOCK structure, where the average is calculated by the superimposing of each structure on to the backbone atoms of the flexible interface defined by the active residues calculated by CSP threshold and Naccess analysis. Here, almost all 200 final structures have an RMSD over 12 Å away from the average HADDOCK run solution. The reason for this does not become obvious until the protein complex solutions are plotted when it becomes apparent that there is an almost 50:50 split in models between Im9 sitting to the side of colicin E9 DNase (cyan in figure 8.4.b) and Im9 sitting to the back (blue in figure 8.4.b).

In figure 8.4.b the DNase activity loop has been highlighted with the black arrow, so it makes biological sense that the immunity protein directly sits in front of the active site in a competitive inhibition fashion. However, it is not beyond the realms of possibility that Im9 binding upon the back surface of E9 DNase, interact with the highlighted helix (red arrow), which could cause changes in dynamicity and binding functionality of the DNase activity loop and thus cause inhibition in a non-competitive fashion.

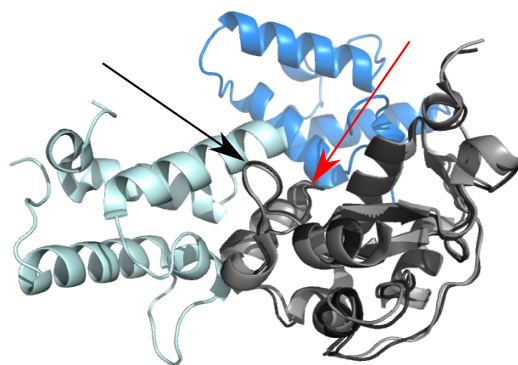
This model was never published or discussed in publications for a number of reasons. First, biologically considering the strength of the interaction between colicin and immunity protein it would be very unusual for this to be a non-competitive inhibition. Second, at the time of this NMR investigation the X-ray crystal structure had been in the public domain for over 5 years and had never been challenged by any other biochemical investigation. Third, using homology models of the full length colicin E9 it is likely that the other domains of the colicin E9 would prohibit the binding of the immunity protein at the back position, and that the docking process has noticed some weakly hydrophobic residues that would otherwise normally reside within the inner surface of the protein.

Colicin E9 is one of nine E-group colicins in a much larger family of bacteriocin proteins. Each with a unique immunity protein but due to common ancestry high homology, yielding some cross-protection. Colicin E9 can be inhibited by immunity proteins Im2, Im7 and Im8. Molecular structures for many immunity proteins have already been elucidated as well as several of the more common colicins. By combining the existing structures with only a few site-directed spin labelling protocols and PELDOR experiments the structure of many of these non-cognate complexes could be solved using EPR on a much faster and cheaper time scale than would be practical for NMR or X-ray crystallography.

8. Protein - protein docking with HADDOCK



(a) HADDOCK run results, HADDOCK score plotted against the RMSD of the cluster to the HADDOCK run average structure. Blue open circles are individual models, red filled circles are cluster averages.



(b) Overlay of E9 DNase PDBs 2K5X(B)(NMR: complex)(black/grey) with HADDOCK cluster 1 best model (cyan) and HADDOCK cluster 2 best model (blue). E9 DNase active site highlighted by the black arrow.

Figure 8.4 Results of HADDOCK run using PDB entry 2K5X structures and NMR CSP data from $^1\text{H} - ^{15}\text{N}$ HSQC and triple resonance experiments.

8.3. *in silico* spin labelling of proteins

EPR fundamentally uses paramagnetic centres as sites of investigation, as most native proteins have no paramagnetic centres (with the notable exception of some metalloproteins) the addition of paramagnetism is required, typically through the introduction of spin labels. The most commonly used in the study of proteins is that of nitroxide radical spin label compounds, of which the most commonly used are the spin labels S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonylthioate (MTSL) and 3-(3-(2-iodo-acetamido)-propyl-carbamoyl)-PROXYL (IA-PROXYL).

With the addition of spin labels to proteins the three dimensional structure is changed or rather more atoms are present. These changes must be taken into account when experimental data is to be interpreted. For calculations to accurately represent and be comparable with experimental data the *in silico* protein must also be spin labelled.

MTSL whilst giving paramagnetism to a structure does not place it at a fixed position even when attached to a rigid part of the protein. In fact, the majority of the time a fixed spin label is undesirable as it has the potential to distort the protein structure. The flexibility in MTSL is as a result of free rotation about the $\text{C}_\alpha\text{-C}_\beta$, $\text{C}_\beta\text{-S}_\gamma$, $\text{S}_\gamma\text{-S}_\delta$, $\text{S}_\delta\text{-C}_\epsilon$, $\text{C}_\epsilon\text{-C}_\zeta$, $\text{C}_\zeta\text{-C}_3$ bonds (visualised in figure 8.5). Using Ramachandran plot parameters and steric hindrances, or more correctly, torsion angle dynamics simulation, MTSL attached to a free cysteine has a rotamer library of 210 discrete conformations (at 175 K) with the nitroxide radical $\sim 10 \text{ \AA}$ from the C_α of the protein backbone.

IA-PROXYL has an additional atom in the flexible linker, making the label slightly longer with typical nitroxide radical $\sim 11 \text{ \AA}$ from the C_α . Free rotation is given by 5 bonds ($\text{C}_\alpha\text{-C}_\beta$, $\text{C}_\beta\text{-S}_\gamma$, $\text{S}_\gamma\text{-C}_\delta$, $\text{C}_\delta\text{-C}_\epsilon$ and $\text{N}_\zeta\text{-C}_3$ bonds) with only partial rotation across the $\text{C}_\epsilon\text{-N}_\zeta$ as a result of bond dissociation of the formal $\text{C}_\epsilon=\text{O}_\epsilon$ bond. This single bond restriction results in a rotamer library of 108 discrete conformations (at 175 or 298 K) by torsion angle dynamics simulation to a free cysteine.

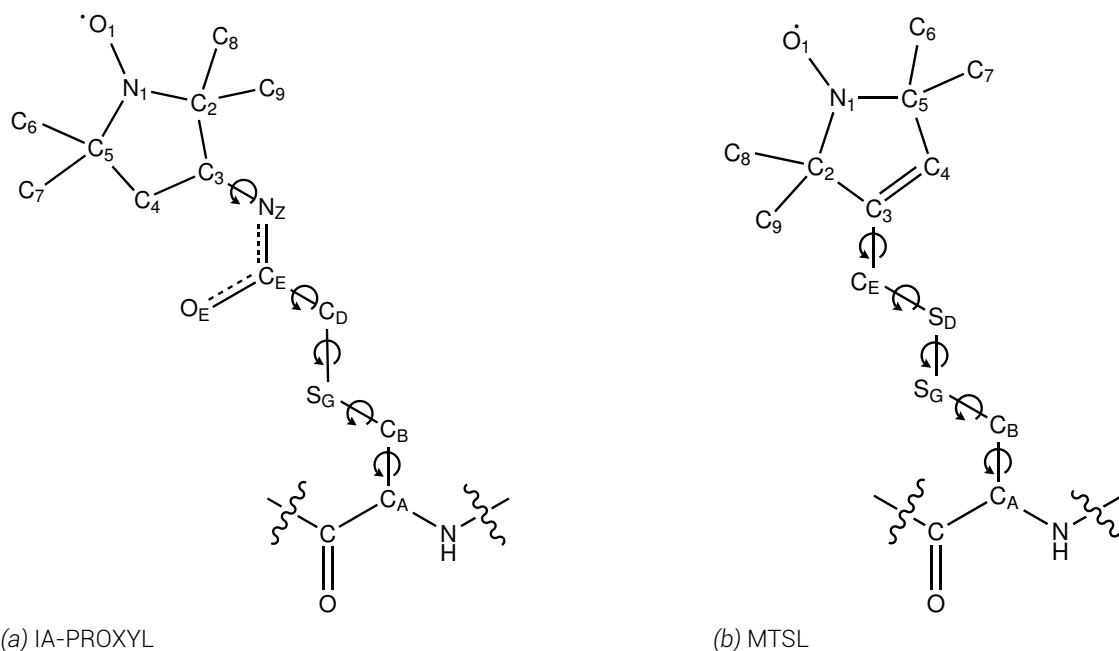


Figure 8.5 Structures of spin labels IA-PROXYL and MTSL attached to a cysteine residue with bond angle motion highlighted.

8.3.1. MMM

Obviously all 210 rotamers of MTSL will not be available to every SDSL site due to steric hindrance with the protein backbone or other side chains. To calculate which rotamers are thermodynamically acceptable the program MMM³ is used.

MMM takes a protein coordinate file (in protein data bank (PDB) format) and assumes that the protein is a rigid body and all atom positions absolute. A residue is selected and mutated to a cysteine, if required, assuming that the backbone atoms' co-ordinates do not change, which may not be the case biologically especially if the mutation involves bulky or charged amino acids. MMM then applies the rotamer library to the cysteine mutant and calculates the energy of each rotamer outputting the allowed rotamers into a pseudo-PDB format.

This assumes that the underlying model is correct, in that there is no difference between the EPR sample and model arising from crystal data versus solution or lipid environment, etc. and that the EPR distribution is free from experimental artefacts.

For the purposes of this investigation it is not sufficient to have separate rotamer libraries and protein co-ordinate files. Instead what is required is PDB files each with a single rotamer, an *in silico* snapshot in time.

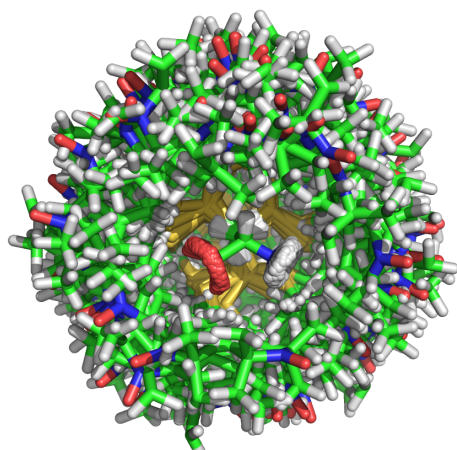
Initially, this was achieved through manually cutting rotamers from the pseudo-PDB file and inserting them into the original protein PDB file detailed elsewhere^{4, 5}. However, this method was extremely time consuming and an automated method was sort.

Through personal communication with Dr. Y. Polyhach (ETH Zurich, CH), a script was attempted which systematically scanned through the rotamer file, read each rotamer, opened the protein PDB file and overwrote the original residue, saved the new PDB file sequentially

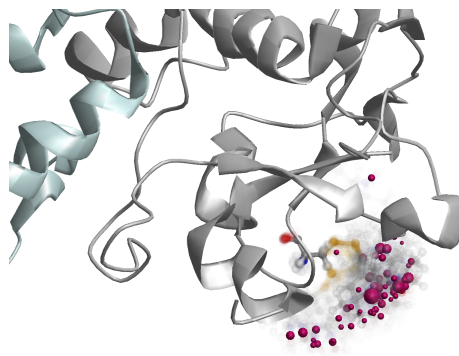
³<http://www.epr.ethz.ch/software/>

⁴<http://morganbye.net/research>

⁵http://wiki.uea-epr.eu/index.php?title=Adding_a_spin_label_to_a_PDB



(a) pseudo-PDB output from MMM



(b) Rotamers displayed within MMM

Figure 8.6 Graphical representation of colicin E9 S30C's 62 rotamer library calculated by MMM at 298K, showing a variation of 4.7 Å radical position. (a) the pseudo-PDB displayed in PyMOL (stick representation). (b) is the MMM display where the residue is shown in ball and stick format with opacity proportional to probability, the location of the nitroxide radical is shown by a purple sphere, with the size of the sphere proportional to probability.

before moving to the next rotamer. Instead Polyhach modified a series of his own old scripts to write out individual PDB files instead of writing the rotamer atom coordinates to the rotamers file⁶.

The problem with this approach was that the PDB file format had for a number of years been poorly enforced. As a result PDB files generated from one program are not guaranteed to work with all others. Many software approaches to clean PDB files have been implemented, whilst most developers resort to programming elaborate solutions with a large number of switches based upon the origin of the PDB file and include a number of safety features leading to large footprints to simply open a file⁷.

8.4. Conversion of distance distributions to distance restraints

The single most important part of modelling is the quality of the data that is used, therefore the conversion of the experimentally derived distance distributions into distance restraints is an important consideration. Unfortunately due to the constraints of the modelling software distance distributions can not be directly used, nor can multiple distance restraints exist between individual atoms.

HADDOCK uses a manipulation of CNS's unambiguous distance restraints input, initially designed for NMR NOE distance restraints. This requires that each restraint must be presented in CNS format⁸; naming the atoms involved in the distance restraint (which requires a protein chain and residue number for both). The distance constraint itself is presented as the distance (d), with additional parameters for dminus and dplus; the allowed distances from d.

⁶A total of 557 lines of code across 7 .m files

⁷The current PDB guidelines are a 191 page document available here: <http://www.wwpdb.org/docs.html>

MatLab's `pdbread` function is 1366 lines of code

⁸<http://cns.csb.yale.edu/v1.1/tutorial/formats/noe/text.html>

Using inspiration from both DeerAnalysis [47] and DEFit [84], a distance between two points can be thought of as a normal distribution when measured using a method, where the uncertainty causes the single distance to broaden into a distance measurement. As a result, even complex distance distribution peaks can be thought of as the sum of many superimposed Gaussian peaks, such that conversely a strong peak can be fitted using a Gaussian distribution with any shoulders being fitted with their own Gaussian distributions. If the shoulder peaks are close in proximity and intensity then the distance distribution will appear to be broadened.

Gaussian distributions have the form:

$$y = a \cdot \exp \left[-\frac{1}{2} \left(\frac{x - b}{c} \right)^2 \right] \quad (8.2)$$

Where a is the amplitude, b is the centre and c is the width (standard deviation). These common features have been shown in figure 8.7.a, additionally Gaussian distributions are discussed in terms of the full width at half height (FWHM), as well as the range at two standard deviations representing 95 % data coverage, these are also shown.

For strong signals, in a single conformation then the distance distribution obtained should be close to a pure Gaussian distribution. In this case, such as figure 8.7.b showing the distance distribution of E9 DNase S30C^{R1A} - Im9 C23^{R1A}, the distance distribution was fitted to a Gaussian model using a Levenberg-Marquardt non-linear least squares fitting algorithm. Here, a good fit to the main peak of interest can be seen, whilst the single standard deviation represents the data well. In this instance the d_{plus} and d_{minus} values are set symmetrically about d .

However, the Gaussian distribution here perhaps over-represents longer distances, whilst the shorter distance shoulder is under-represented. Here a multivariate normal or split normal distribution would yield a better fit to the data. In figure 8.7.c the symmetric Gaussian is shown in red, whilst the asymmetric Gaussian is shown in green. Here a clear increase in d_{minus} is seen as well as a decrease in d_{plus} however, the reality of the difference here is less than 0.1 Å which makes the increased computation impractical.

In the final scenario for this investigation there were scenarios where two peaks existed and shown to be significant such as in the colicin E9 DNase S3C^{IA1}/S80C^{IA1} experiment. Ideally an AND or OR statement would be introduced into the constraints file such that the constraint is for only the peaks and not the null space between them. As this is not possible, compromise was struck whereby the both the peaks in question were fitted using a symmetrical Gaussian method and each peak's integral was used to weight the overall result. For example in figure 8.7.d assume that the green peak has a quarter the integral of the red. In this instance, the distance would be weighted 75 % toward the red peak. d_{minus} was then set as the distance from d to the lower standard deviation of the red peak whilst d_{plus} was set to the upper standard deviation of the green peak.

In the instance of very broad or asymmetric peaks with non-Gaussian like distributions a similar single standard deviation approach would be used, in which 68.2 % of the peak area either side from the peak would be sampled. In such unsymmetrical, non-Gaussian systems the full width half height or standard standard deviation analysis would yield less meaningful boundaries as opposed to this approach which relies less upon peak shape.

8. Protein - protein docking with HADDOCK

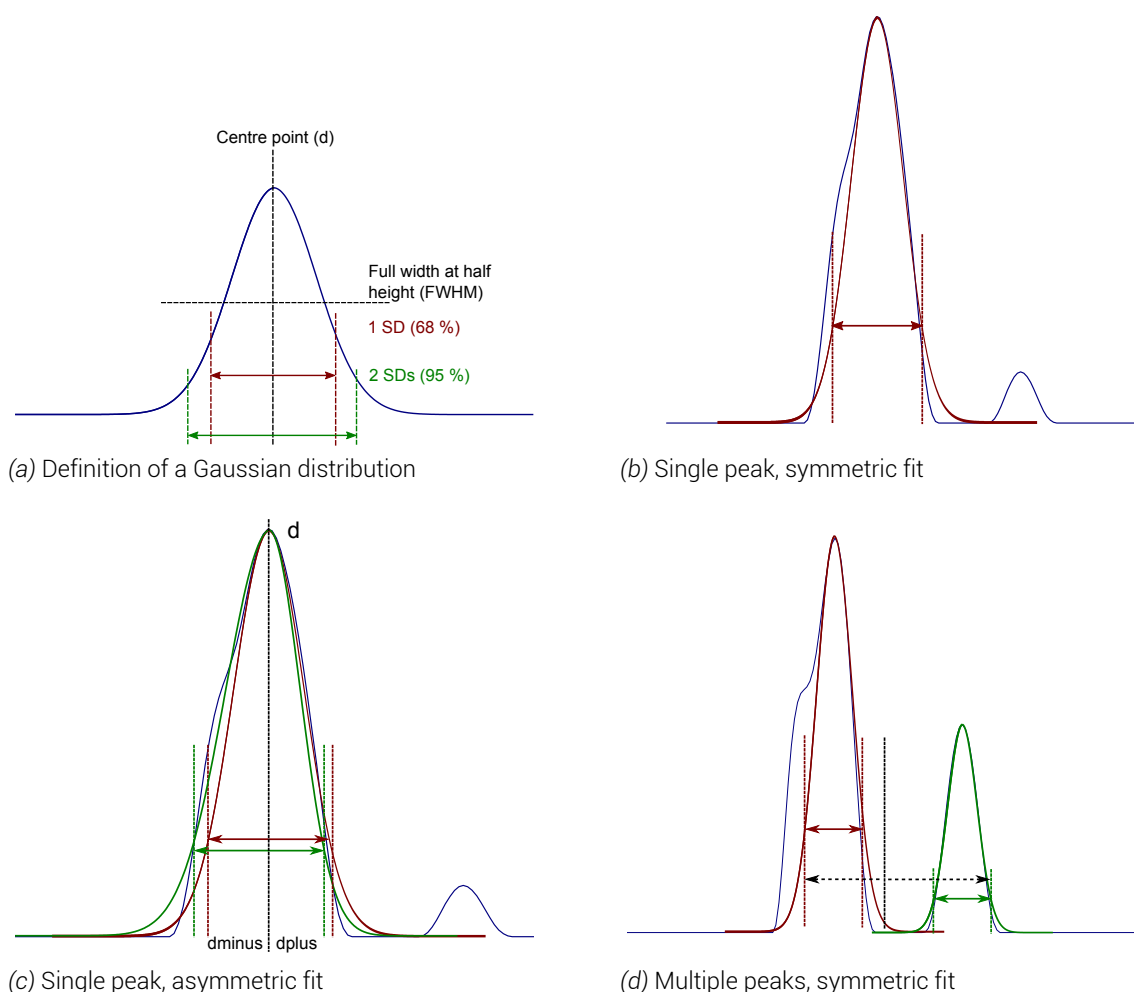


Figure 8.7 Graphical representation of the distance distribution fitting mechanism employed. a) defines the common terms in a Gaussian distribution, with the centre point (d), the full width at half height, and the first two standard deviations of data in red and green, representing 68 % and 95 % of the data. b) shows the symmetrical fitting of the single, largest peak to a Gaussian distribution in colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}, with the black lines highlighting the first standard deviation. c) shows the symmetrical fitting in red and asymmetric fit in green of colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}. d) shows the multiple peak fitting using symmetrical fitting upon colicin E9 DNase S3C^{IA1}/S80C^{IA1}, the first standard deviation for each peak is shown in the peak colour, whilst the overall d position is shown as a vertical black line with d_{plus} and d_{minus} highlighted as black, dashed arrows.

This broad distance range would cover most of the distance distribution with reasonable probability but also account for uncertainties arising in the dynamics in the spin label as well as in data analysis. Uncertainties can appear during analysis as broadened wings in the distance regularisation due to Tikhonov regularisation and low amplitude artifacts at the end of the distance distributions [47, 115]. This method could equally be used to multiple peak analysis with careful implementation. In this investigation, each distance distribution analysis was conducted manually using a Gaussian approximation, though an automated system using this final described method would be desirable and likely implemented later.

Using a combined approach of these described methods the PELDOR results of chapter 7 were analysed to yield table 8.2.

By contrast and for further investigation, MMM was used to generate a complete restraint list for every spin label site used in this investigation with the IA-PROXYL label from the NMR complex structure (PDB: 2K5X) shown in table 8.3. The distance distributions generated by MMM were analysed in exactly the same manner as the experimental distance distributions.

Protein, labelling position		Distance restraints / Å		
from	to	d	dminus	dplus
E9 DNase S3C ^{IA1}	E9 DNase S80C ^{IA1}	3.23	0.85	1.15
E9 DNase S30C ^{IA1}	E9 DNase S80C ^{IA1}	3.96	0.98	1.65
Im9 C23 ^{IA1}	Im9 S6C ^{IA1}	2.58	0.88	0.83
E9 DNase S3C ^{IA1}	Im9 S6C ^{IA1}	3.62	0.63	1.03
E9 DNase S30C ^{R1A}	Im9 C23 ^{R1A}	3.70	0.50	0.50
E9 DNase E66C ^{IA1}	Im9 S6C ^{IA1}	4.85	0.78	0.82
E9 DNase E66C ^{IA1}	Im9 C23 ^{IA1}	2.80	0.10	0.10

Table 8.2. Tabulated summary of the distance constraints generated as a result of PELDOR experimentation and processing.

Protein, labelling position		Distance restraints / Å		
from	to	d	dminus	dplus
E9 DNase S3C ^{IA1}	E9 DNase S80C ^{IA1}	2.46	0.55	0.45
E9 DNase S30C ^{IA1}	E9 DNase S80C ^{IA1}	3.68	0.75	0.50
Im9 C23 ^{IA1}	Im9 S6C ^{IA1}	3.21	0.33	0.20
E9 DNase S3C ^{IA1}	Im9 S6C ^{IA1}	3.89	0.87	1.41
	Im9 C23 ^{IA1}	4.10	0.30	0.92
E9 DNase D20C ^{IA1}	Im9 S6C ^{IA1}	5.37	0.93	0.20
	Im9 C23 ^{IA1}	3.75	0.30	0.30
E9 DNase S30C ^{IA1}	Im9 S6C ^{IA1}	4.28	0.89	0.42
	Im9 C23 ^{IA1}	1.55	0.25	0.60
E9 DNase E66C ^{IA1}	Im9 S6C ^{IA1}	5.27	0.87	0.18
	Im9 C23 ^{IA1}	2.48	0.22	0.39
E9 DNase S80C ^{IA1}	Im9 S6C ^{IA1}	5.25	0.65	0.30
	Im9 C23 ^{IA1}	2.43	0.28	0.32

Table 8.3. Tabulated summary of the distance constraints generated as a result of MMM rotamer analysis.

Even in the intramolecular distances, such as the colicin E9 DNase S3C^{IA1}/S80C^{IA1} a difference of 0.77 Å can be seen in the distance restraint with the constraint range varying by 1.00 Å between prediction and experimental result. Most of the intermolecular distance restraints vary by less than 0.50 Å between prediction and experimental data. However, the *in silico* spin labels, unhampered by protein concentration, labelling efficiencies or experimental convolution yield far more distinct distance restraints.

8.5. MISHAP

MISHAP is the open-source MMM *in silico* simulated spin label to HADDOCK program. It combines several MatLab scripts, functions, programs and user interfaces as well as a set of additional parameter files for HADDOCK.

MISHAP aims to take spin labelled proteins and the distance distributions between labels from *in silico* (MMM) calculations or experimental data (from pulsed electron-electron double resonance experiments which have been analysed with DeerAnalysis) and pass them to the biomolecular docking program HADDOCK as docking restraints.

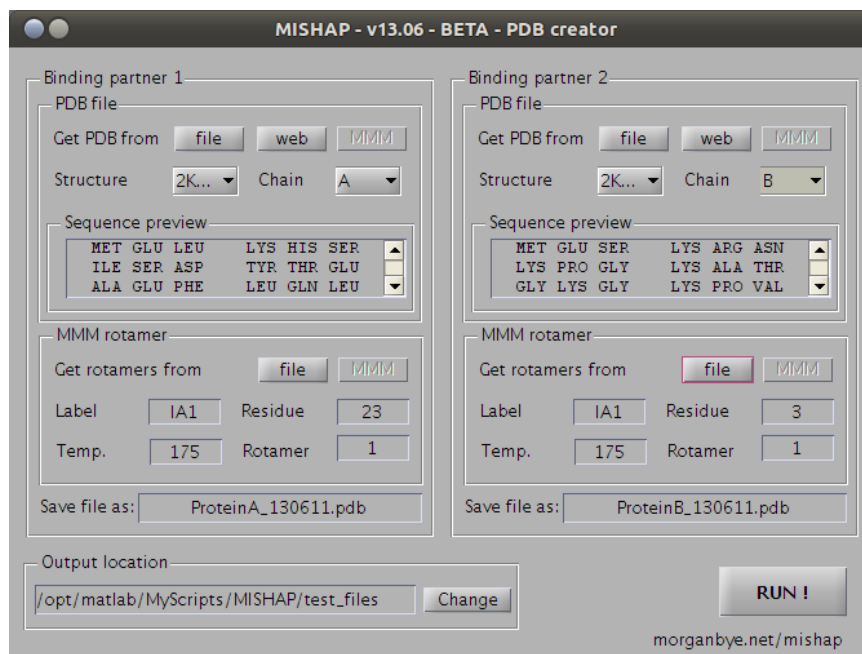


Figure 8.8 MISHAP PDB Creator user interface with 2K5X parameters loaded

8.5.1. Implementation of MMM rotamers into PDB format

With the PDB file format being so ill-defined, or rather, not strictly enforced, it is very unusual for a PDB file, even those obtained from pdb.org to be compatible with HADDOCK. Furthermore, HADDOCK expects PDB files in a particular way, for example a PDB file can only contain one amino acid chain, it cannot contain a segment identifier and must end with a correct END statement. These constraints rule out almost every published PDB file. Whilst HADDOCK provides some basic scripts to strip PDB files such as `pdb_blank_segid` the user is still left with a lot of manual work.

Furthermore, it is required not only to have PDB files compatible with HADDOCK but also to have PDB files including spin label attached residues. As a result MISHAP provides the ability to generate HADDOCK compatible PDB files of single chains of amino acids where one or more residues have been substituted for a nitroxide spin label attached to a cysteine residue.

At this time it is not known how successful HADDOCK is at optimising rotamer populations of amino acid side chains, particularly for unnatural amino acids. For this reason, MISHAP allows for the generation of a single MMM rotamer PDB file - which should be sufficient for HADDOCK. However, if more methodical work is required, functionality has been built in to generate rotamer PDB files for all MMM rotamers or for rotamers over a user defined threshold, such as only generating PDB files of MMM rotamers with greater than 5 % occupancy. This method utilises HADDOCK's ensemble structure input functionality, as would be commonly used by NMR users with an ensemble of structures rather than a minimized average structure.

This is implemented in MISHAP's PDB file creator graphical user interface, seen in figure 8.8 and further discussed in chapter 12 and appendix B.

8.5.2. Acquiescence of HADDOCK to spin labels

Since the original HADDOCK publication in 2003 by Dominguez *et al.* [28], HADDOCK has been extended to deal with a large variety of data and complexes including protein-DNA, protein-RNA, protein-oligosaccharides and protein-ligand complexes. Traditional docking methodologies are based upon a combination of energetics and shape complementarity. Instead, HADDOCK was initially designed differently from using a biochemical/biophysical experimental data approach using data resulting from NMR titration experiments, mutagenesis data or bioinformatic predictions and introducing it as ambiguous interaction restraints.

The use of spin labels as unnatural amino acids is not a problem with HADDOCK. HADDOCK is simply a complex top layer that drives docking whilst structural calculations are called by HADDOCK using HADDOCK-specific parameter files from the Crystallography and NMR System (CNS) program CNSsolve⁹. HADDOCK contains a series of files which contain all of the necessary parameters for amino acid residues, ligands, DNA, etc. These parameters largely are default parameters for CNS that have been extensively modified and added to for the purposes of HADDOCK.

8.5.2.1. Topology

In order to define a molecule or amino acid in CNSsolve requires the explicit definition of every atom type, atom size, atom charge, bond, bond energy, bond length, dihedral angle and improper dihedral angle in the molecule. Furthermore, in the case of amino acids the way in which the amino acid links to the $i - 1$ and $i + 1$ residues, as well as its behaviour as at the terminus of an amino acid chain requires defining.

topallhdg5.4.pep allows the labels to exist in a peptide, and defines how the peptide would bind to surrounding amino acid residues

topallhdg5.4.pep-noCter allows the labels to exist and bond to it's neighbours in a peptide where no C-terminus exists

topallhdg5.4.pep-noNter allows the labels to exist and bond to it's neighbours in a peptide where no N-terminus exists

topallhdg5.4.pep-noter allows the labels to exist and bond to it's neighbours in a peptide where neither a N nor C -termini exists

topallhdg5.4.pro defines the mass and charge of each atom, as well as each bond, angle, improper angle, dihedral angle and hydrogen bond acceptor/donor site

⁹CNSsolve, Yale University, USA, <http://cns-online.org>

8.5.3. Distance distributions and HADDOCK restraints

With the results of the PELDOR investigation of chapter 7, several distance distributions were generated between multiple sites across the colicin E9 DNase - Im9 complex.

These distance distributions require conversion into docking constraints that can be used by HADDOCK. To do so, manipulation of HADDOCK's/CNS's unambiguous distance restraints, initially designed for NMR NOE distance restraints are required. Restraints must be presented in a separate file, following CNS format¹⁰; naming the atoms involved in the distance restraint (which requires a protein chain and residue number for both). The distance itself is presented as the distance (d), with additional parameters for dminus and dplus (the extents either side of the distance, d)

For example, this line defines the distance between Im9 S6C^{IA1} and Im9 C23^{IA1} nitroxides as 3.21 Å with a lower limit of 2.88 Å and a higher limit of 3.41 Å.

```
assign (name "01" and resid "6" and segid "A") ...
      (name "01" and resid "23" and segid "A") ...
      3.21 0.33 0.20
```

MISHAP implements the generation of unambiguous distance restraints in a MatLab graphical user interface (shown in figure 8.9). The user loads a distance distribution from DeerAnalysis or MMM, the distribution is displayed, automatically scanned for peaks and a distance value presented. The user is free to change the automatically detected values and add more spin label sites (for 3 or 4 spin label experiments) before generating the constraint file. If the MISHAP constraint generator is used in conjunction with the MISHAP PDB generator then the atom address and label information will automatically completed for the user as well.

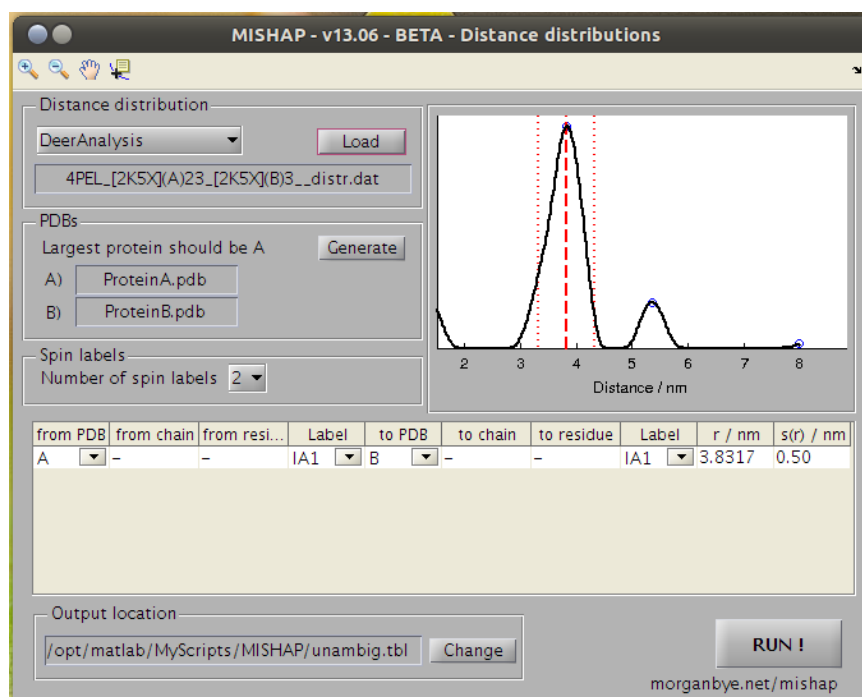


Figure 8.9 MISHAP distance constraint generator user interface

¹⁰<http://cns.csb.yale.edu/v1.1/tutorial/formats/noe/text.html>

8.6. HADDOCK runs

HADDOCK developers pride HADDOCK's ability to accept ambiguous data, small amounts of data or data from various different sources; including *in silico* methods. This means that HADDOCK can be run, and give results with little, if any, experimental data. As a starting position and example of HADDOCK's flexibility, the colicin E9 DNase and Im9 structures from the PDB entry, 2K5X were taken and run through HADDOCK using no restraints. It was expected that in the first stage of docking (solid body), when both molecules are randomly orientated and combined with no restraints, that there should equal sampling across the molecular surfaces, which after energy minimisation should yield little clustering of structures.

Next ambiguous interaction restraints derived from a typical NOE perturbation based NMR titration assay would be introduced to replicate the process used to derive the published complex structure. Having verified that the NMR complex is repeatable, spin labels were introduced and docking rerun with no further restraints to ensure that the small spin labels made no significant difference to the docking procedure. With the spin labelled proteins confirmed to be representative of the natural proteins then additional EPR derived restraints could be introduced.

Each HADDOCK run in this investigation was conducted using the same parameters such that no run should be unfairly biased. In addition, several parameters as a result of the small size of the docking molecules were adjusted such that a sufficient number of structures completed the docking calculations. Namely the molecular dynamics time step between energies calculation was reduced from 2.0 *fs* to 0.5 *fs* as well as reducing the annealing temperatures which the molecules are raised to before being allowed to cool. Without these adjustments the molecules would often move too far between molecular dynamics calculations allowing atoms moving toward each to clash. In the next round of calculations the massive potential energy from the repulsion of atoms is resolved by moving the atoms quickly far apart, often outside the calculation co-ordinate space (> 99,999 Å) causing the structure to fail.

In this section HADDOCK run results are displayed as both a statistical and graphical summary. In the statistical summary the structures that complete the docking process are superimposed onto the backbone atoms of the flexible interface before they are aligned and averaged to give an "average structure". The interface is defined by the atoms from the ambiguous interaction restraints (AIRs), though if AIRs are not available then the entire protein backbone is used. This average structure then serves as the basis of several calculations that compare each model of the ensemble in terms of both energies and spatial deviation.

Each structure has several bonded and non-bonded energies calculated, including the total complex energy, the bond energies, the bond angle energies, the dihedral angle energies, the Van der Waal's energies, the electrostatic energies, the buried surface area (and thus the desolvation energy). Finally, with consideration of hydrogen bonding, intermolecular hydrophobic contacts and violation analysis a HADDOCK score is generated.

In the visual statistical analysis the RMSD of each structure from the average structure is plotted against the structure's HADDOCK score. Each structure is shown as an open blue circle upon the axes. During analysis, as discussed in §8.2, structures of close RMSDs are grouped into clusters. In the statistical analysis the cluster averages are shown as closed

8. Protein - protein docking with HADDOCK

red circles with error bars showing the average for the best four structures in that cluster. A good docking run will show few, if not only a single cluster with small error bars. Poor docking runs with multiple solutions will show many clusters with all the structures showing a high RMSD from the average structure. Whilst a very bad docking run will show few structures complete annealing with almost no clustering as none of the results are close enough to be considered a cluster.

The graphical analysis shows the published NMR structure for the colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format. As with previous diagrams in this investigation the colicin E9 DNase remains in grey and Im9 remains in pale cyan, though the rotation has been altered slightly to allow full view of the binding interface. All HADDOCK runs maintain the same perceived orientation so that each run is visually directly comparable with all others.

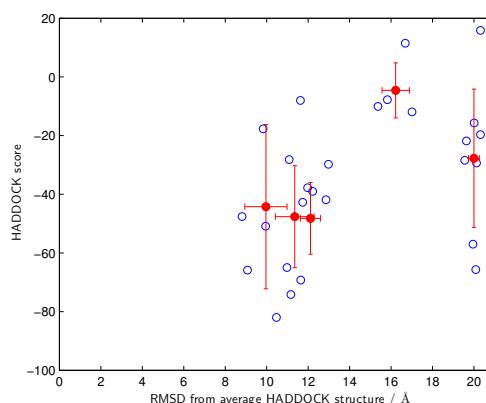
The HADDOCK structures are represented in line format, for the sake of clarity only the backbone atoms (N, C, C_{α}, O) are shown. In black, the top scoring structure's colicin E9 DNase is plotted after alignment to the reference structure. Other structures were considered however, the very low RMSDs upon the colicin E9 DNase only confuse the image and as a result are omitted from the visual representation. Each HADDOCK structure is then aligned to the top scoring structure. The Im9s from each HADDOCK structure is then shown in line format and coloured in a spectrum manner according to their scoring position with blue being the best scoring and red being the lowest scoring structure. By default, the top 4 structures from the top 4 scoring clusters are displayed. In later HADDOCK runs additional figures are presented where more structures from fewer clusters are displayed, as this becomes more appropriate given the cluster population. For some HADDOCK runs additional figures are included which remove the reference structure and reorientate the co-ordinate system to better show the Im9 positions, in these cases the cartoon model replaces the black line representation for the top scoring colicin E9 DNase.

When comparing the positions of proteins, such as comparing the position of a HADDOCK calculated structure to that of the published structure a two step analysis is performed. In the first the colicin E9 DNase is aligned using a root mean squared deviation minimisation method of C_{α} atoms, where matching atoms, v and w are superimposed such that

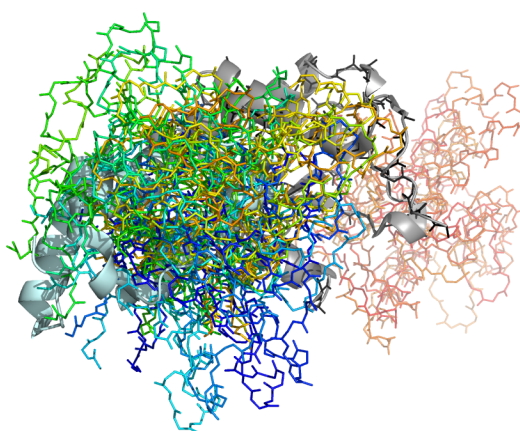
$$RMSD(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2 \right)} \quad (8.3)$$

Where x, y and z are the principle co-ordinate system. In the second step the distance between Im9 positions is analysed by taking the translational component of an ordinary Procrustes analysis for each C_{α} atom and averaging for the entire molecule. Using this method removes any bias from rotation and molecular orientation that would be present from using single representative atoms from the binding interface or a protein centre of mass pseudoatom based approach.

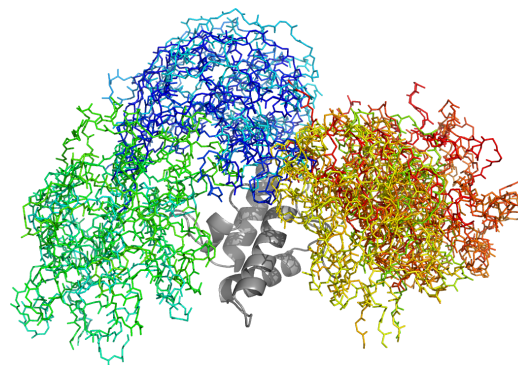
8.6.1. Without constraints



(a) HADDOCK statistical summary



(b) HADDOCK visual summary



(c) Reorientated visual summary

Figure 8.10 HADDOCK run results with no restraints, and only minimal, randomly generated AIRs. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing the cluster average of the top 4 scoring structures. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C α , O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey cartoon representation.

In figure 8.10.a we see a statistical summary of the HADDOCK run whereby each of the completing 200 structures taken forward from solid body analysis is plotted as a blue circle. Each structure's HADDOCK score (more negative is a better model) is plotted against the RMSD distance of that structure from the average structure of the 200 models.

In HADDOCK, by default, structures are pooled into a cluster if, four or more structures have an RMSD of less than 7.5 Å. In figure 8.10.a the top 4 scoring structures of any cluster are averaged and represented by filled red circles. A strong cluster should show many structures with similar RMSDs and HADDOCK score, thus small error bars. A perfect model of absolute certainty therefore should have only one cluster.

With no restraints in place the energy optimisation docking procedure proves difficult. On a protein with few solvent exposed charged residues and a small cross-sectional surface area in many models entropy drives the system such that the binding partners are infinitely separated. Figure 8.10.a shows far less than 200 blue circles as these structures are removed

8. Protein - protein docking with HADDOCK

from further calculation after failure.

Figure 8.10.b shows the visual summary of the HADDOCK run. In this instance the top scoring structure has had the colicin E9 DNase aligned to the NMR reference structure using a RMSD refinement process, after which all other HADDOCK models are aligned to the top scoring HADDOCK structure. As the colicin E9 DNase conformation varies little between structures, only the colicin E9 DNase of the top scoring structure is shown in black line format, whilst the reference structure is shown in cartoon format. The top 4 structures from the top 4 clusters have their Im9 structure shown in backbone line format. The Im9 structures have been colour coded such that blue is the best HADDOCK scoring structure and red is the fourth structure from the fourth cluster.

Visually from figure 8.10.b little can be seen of the position of the Im9. However, when the reference structure is removed and reorientated, as in figure 8.10.c, it can be seen that whilst there has been some clustering into 3 positions (blue, green and yellow/red) the spread is almost entirely spaced across the colicin E9 DNase's surface. Remember also that these representations only present the top 4 structures from the top 4 clusters, when more structures are considered the image becomes considerably more distorted, with the apparent clustering becoming less obvious.

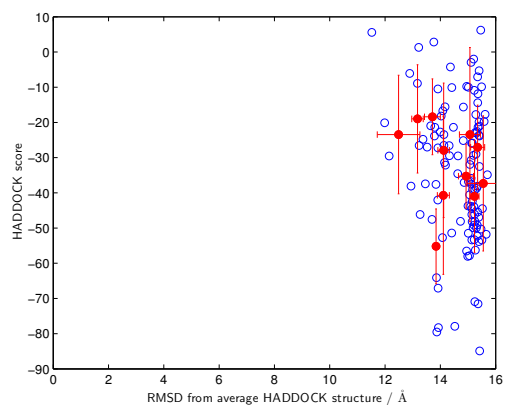
When both figures 8.10.a and 8.10.c are considered together then no docking solution is within 8 Å of the average HADDOCK structure. From the visualisation it can be imagined that the average position of the Im9 is to be almost directly superimposed on top of the colicin E9 DNase with these structures having an RMSD to the reference structure of 1.45 and 1.98 Å. However, there are a number of structures with a distance over 15 Å from the average structure representing approximately half the width of the protein.

8.6.2. With chemical shift perturbation constraints

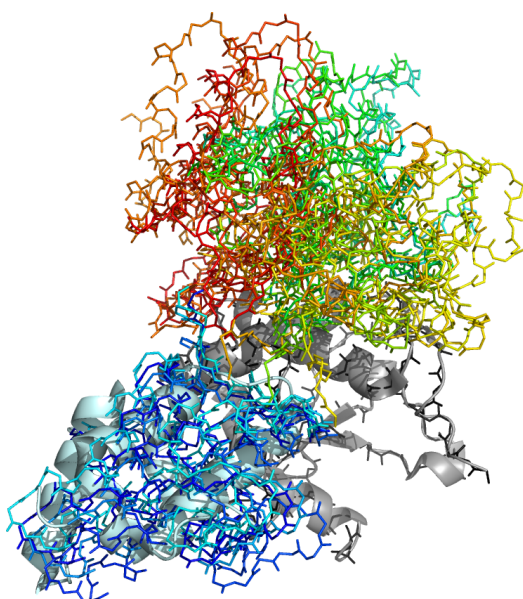
To investigate whether EPR distance restraints could be used to elucidate models that remain ambiguous from only NMR experiments requires an investigation into how CSP data influences the docking procedure.

In figure 8.11 we see a repetition of figure 8.4 now using experimentally derived ambiguous NMR restraints. In figure 8.11.a, more structures can be seen to complete the docking trials. In addition, a far greater clustering of structures can be seen whereby almost all 200 structures fall in a narrow 4 Å range from the average structure. However, almost all of these structures fall at least 12 Å from the average structure. Due in no small part to the splitting of the top scoring structures into 2 distinct positions, highlighted in figure 8.11.c as the blue/green and orange/yellow populations. As further, lower scoring, structures are introduced to the analysis the visual representation distorts as the populations show a broader interaction area.

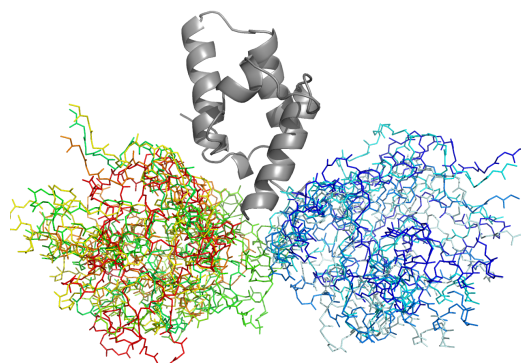
Key to this docking procedure are the AIRs, which not only keep a far greater number of structures within the confines of the virtual docking space, but also generate distinct clusters opposed to the random spreading of structures that was seen when these restraints were not present. In figure 8.11.a this causes the top scoring structures to have a greater RMSD from the average structure, when compared to figure 8.10.a but the top structure docking scores vary little.



(a) HADDOCK statistical summary



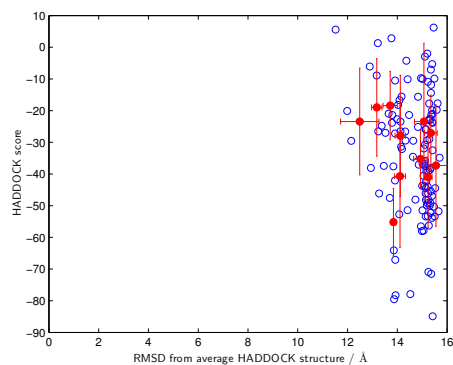
(b) HADDOCK visual summary



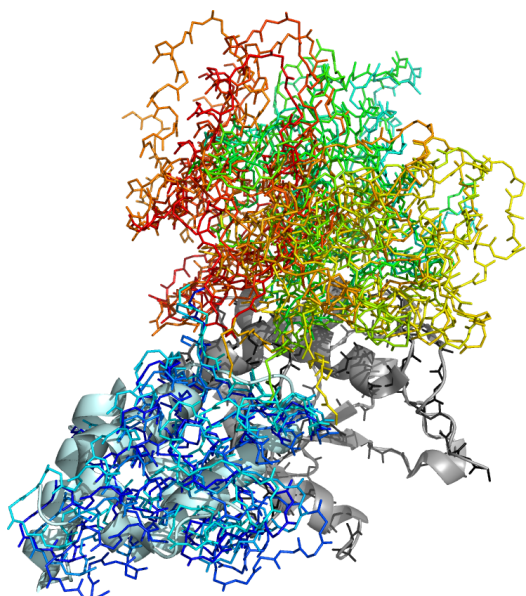
(c) Reorientated visual summary

Figure 8.11 HADDOCK run with NMR AIRs restraints. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing the top four scoring structures range. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey.

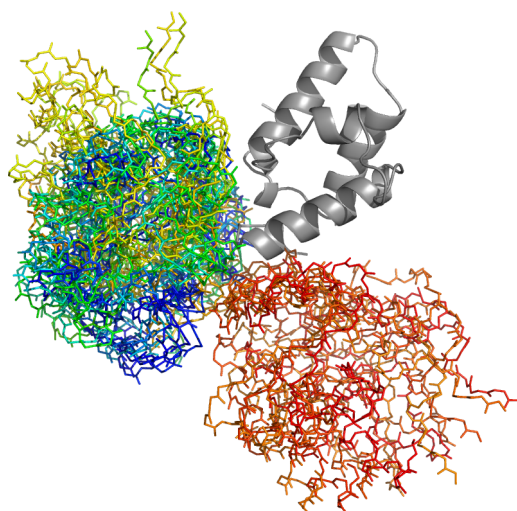
8. Protein - protein docking with HADDOCK



(a) HADDOCK statistical summary



(b) HADDOCK visual summary



(c) Reorientated visual summary

Figure 8.12 HADDOCK run with NMR AIRs restraints upon MISHAP generated proteins with all spin labels present. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing the cluster average of the top 4 scoring structures. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C α , O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey.

8.6.3. With CSP constraints and spin labels

Through the use of site directed spin labelling the physical properties of the protein are changed, especially in the case where the spin label locations are deliberately designed to be upon the surface or inside of the protein. Whilst the spin labels are attached by a flexible linker, they still have more bulk than a traditional cysteine or serine which they replace, this could influence docking. Whilst each site was chosen to be away from the docking interface in the published X-ray crystallography and NMR structures, it may still influence the docking procedure. To test whether the spin labels influenced docking a HADDOCK run was constructed using only the AIRs from NMR using MISHAP generated proteins with spin labels at every labelling position used in this investigation.

Figure 8.12 shows that the flexibility in the spin label linker means that the spin label makes little difference to the protein structure or the docking behaviour. Here the colicin E9 DNase with 4 spin labels and Im9 with 2 spin labels has given a near identical spread of structures and clusters as those shown in figure 8.11.

On detailed analysis without spin labels present, the top cluster of 50 structures (42.0 % of structures completing the run) have a backbone RMSD to one another of $2.22 \pm 1.74 \text{ \AA}$. With spin labels present the top cluster of 34 structures (36.9 % of structures completing the run) have a slightly increased RMSD $2.78 \pm 1.61 \text{ \AA}$.

Visualisation of the top structures in figure 8.12.b shows an identical representation as previously in figure 8.11.b. This confirms that the spin labels are away from the docking surface in the top scoring structures, and that the spin label presence does not unfairly bias one of the potential binding sites.

8.6.4. With intramolecular PELDOR restraints

Having established that the presence of spin labels across the docking partners made little difference to their docking behaviour the next step was to begin introducing distance restraints derived from the PELDOR experiments. The proteins in question are small, globular and delivered to HADDOCK as rigid bodies. As such, the introduction of intramolecular distance restraints to already restrained models should not yield any significant difference to the docking process. The restraint may ultimately allow less flexibility in the energy optimisation as domains and secondary structures will be restricted in their movement, but as the first step in HADDOCK assumes rigid bodies the overall docking should not vary greatly.

8.6.4.1. Colicin E9 DNase S3C^{IA1}/S80C^{IA1}

In figure 8.13, the unmodified Im9 protein from the NMR complex structure is docked with colicin E9 DNase with a single distance restraint between the terminal side chain oxygen atom upon the S3C^{IA1} and S80C^{IA1} residues.

The statistical summary (figure 8.13.a) here shows little variation from an unconstrained run with a great number of structures completing the docking procedure. However, all structures remain over an RMSD of 11 \AA from the average run structure. The number of clusters also

8. Protein - protein docking with HADDOCK

remains high with the error bars of each cluster covering a wide range of RMSD distances and HADDOCK scores.

When the top structures are visualised in figure 8.13.c a clear splitting in the clusters is seen again, with the top scoring structures (blue) sitting at the back of colicin E9 DNase in this orientation (with an RMSD of the Im9 72 Å from the reference structure). However a number of good scoring structures adopt the previously seen position on top of the colicin E9 DNase, which is also slightly closer to the reference position of the Im9 with a C_α average RMSD of 53 Å.

However, when figures 8.12.b and 8.13.b are directly compared in the same orientation to the reference structure it can be seen that the restraint in the colicin has a profound effect. The best scoring structures (blue) move to the opposite side of the colicin but also constrain the model such that the Im9s presented appear less distorted and almost approaching discrete conformations.

8.6.4.2. Colicin E9 DNase S30C^{IA1}/S80C^{IA1}

In the case of the docking with the colicin E9 DNase S30C^{IA1} to S80C^{IA1} distance restraint (figure 8.14) a shift in docking structures is seen. Figure 8.14.a shows a number of structures in one cluster is seen to move just below an RMSD of 12 Å from the average structure. 3 more small clusters with an RMSD from average in the range of 6 - 8 Å also appear, but these are not particularly well scoring.

By inspecting the clusters visually in figure 8.14.b, the spreading of structures across a greater RMSD range from the average structure is related to the spreading of clusters from the 2 distinct sites to an entire side of the colicin E9 DNase. This presents as a mix of situations between when no NMR AIRs and when no long range distance restraints are included. It could simply be a case that the single long range distance restraint is causing less flexibility in the colicin E9 DNase which causes unusual preferences upon introduction of Im9.

Notice however that the vast majority of structures lie in the 11 Å cluster and that the HADDOCK scores of this cluster are significantly better than the clusters closer or further away from the average structure. This cluster is represented in blue in figure 8.14.b and has a top scoring cluster of over -115. The fourth best scoring cluster (RMSD 8.5 Å, HADDOCK score -75 in figure 8.14.a) is shown in red in 8.14.b shows an ordinary HADDOCK score but a low RMSD. If the high RMSD structures are orientated on the opposite face as the red cluster it would artificially generate an average structure on top of colicin E9 DNase. Making the high scoring cluster at an RMSD of 11 Å the most relevant.

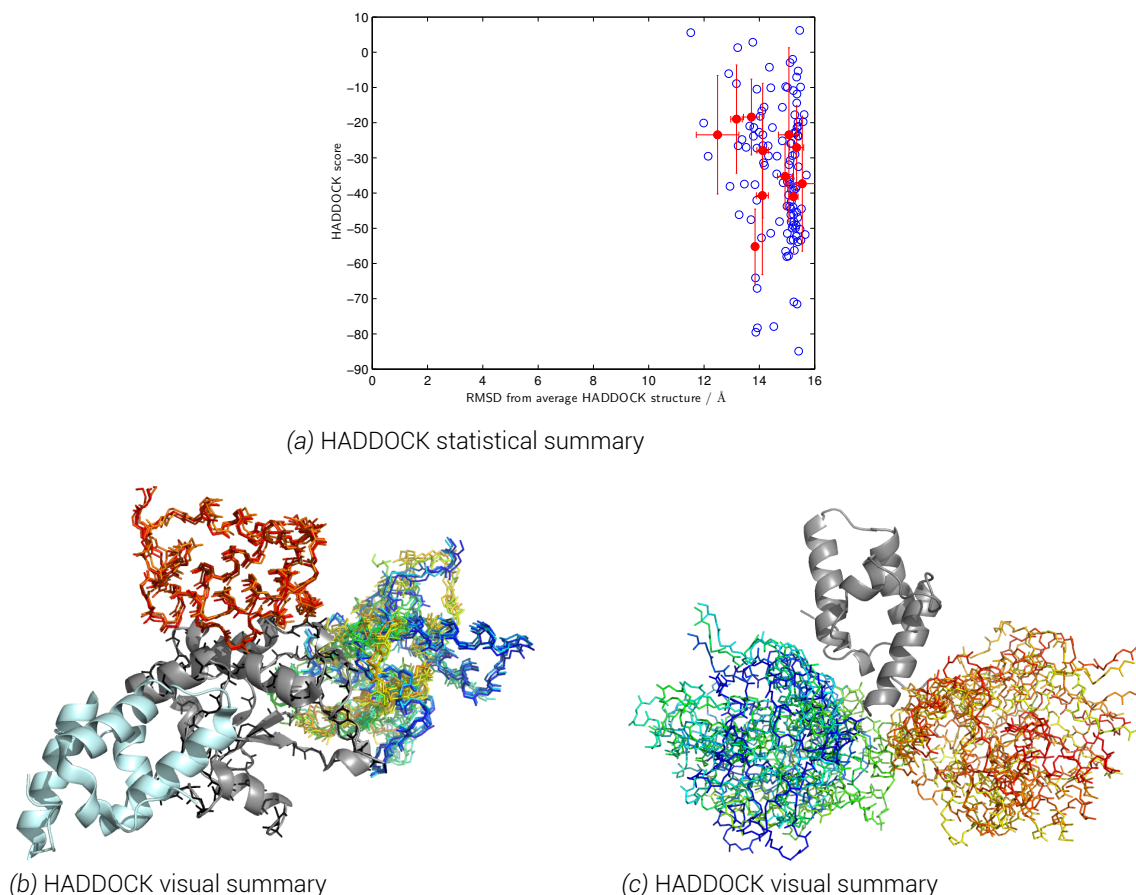


Figure 8.13 HADDOCK run with NMR AIRs and colicin E9 DNase S3C^{I A1}/S80C^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey.

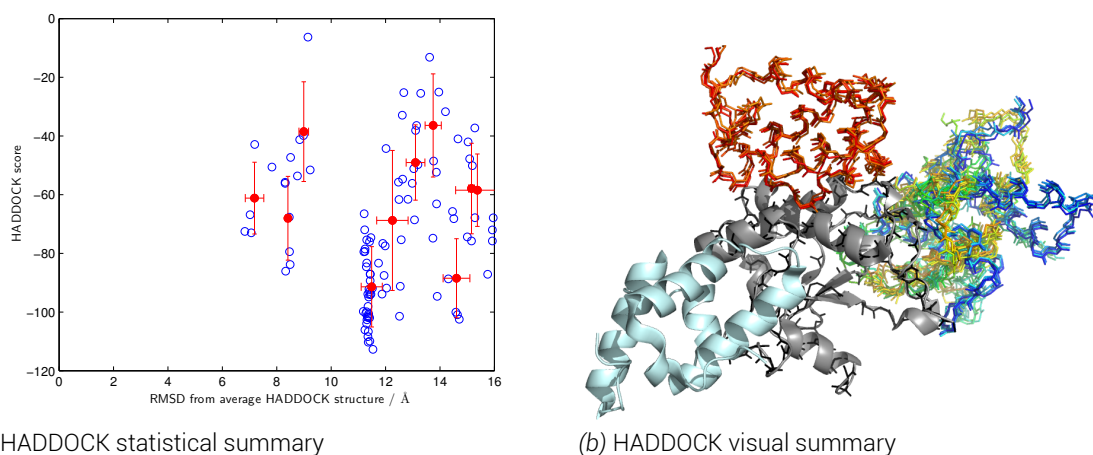


Figure 8.14 HADDOCK run with colicin E9 DNase S30C^{I A1}/S80C^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.

8.6.4.3. Im9 S6C^{IA1}/C23^{IA1}

Figure 8.15 shows the reverse situation of figures 8.13 and 8.14. Here the colicin E9 DNase remains without mutation or label whilst the Im9 has been restrained with an intramolecular distance between the terminal side chain oxygen of the S6C^{IA1} and C23^{IA1} residues.

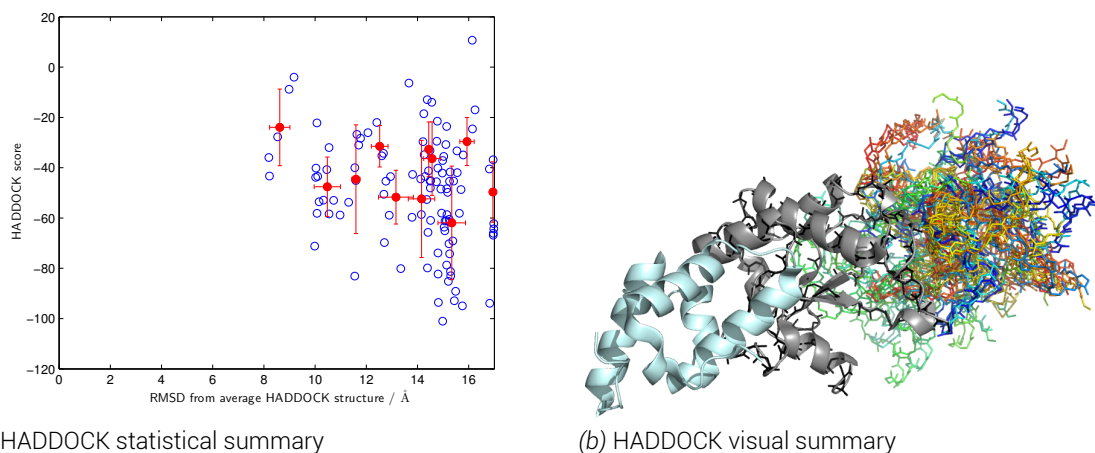
Figure 8.15.a shows a greater spread of structures than when there are no constraints. Noticeably, 3 cluster averages are below the 12 Å RMSD from average structure limit seen with no restraints. However, at the same time a significant number of structures are above the upper 16 Å RMSD from average structure limit seen with no restraints.

Figure 8.15.b confirms what is shown in figure 8.15.a, and the best scoring clusters all collect to the same side of the colicin E9 DNase, but a large distribution is still seen in these 16 structures. Translation of the top scoring clusters visualised here yields an average RMSD of the C_αs in the Im9 of approximately 75 Å.

8.6.5. With intermolecular PELDOR restraints from colicin E9 DNase S3C^{IA1}

Figure 8.16 shows the first HADDOCK run using both NMR AIRs and an experimentally derived, intermolecular spin label, PELDOR distance restraint. From visual inspection of the top scoring structures (figure 8.16.b) we see that from introducing a single EPR distance restraint that the structures fall into 3 prominent clusters. The yellow structure remains in a similar position at the back of the colicin E9 DNase (in this orientation) as has been seen previously in other HADDOCK runs.

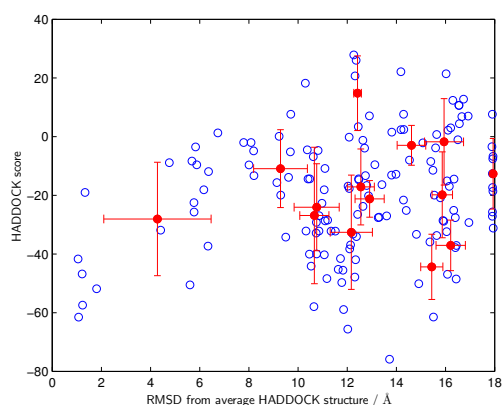
However, the visual representation here is misleading when considered with the statistical summary (figure 8.16.a). Here we can see that 6 structures deviate from the average structure by less than 2 Å, and the lowest cluster average is ~ 4 Å from the average structure despite variation. However, as the structures under an RMSD of 2 Å from the average structure do not form a cluster they can be assumed to coincidentally close to the average structure but not to each other. If the top 4 scoring structures all fell within the same cluster then only the blue/green distribution of figure 8.16.b is of interest whilst the others presented are by HADDOCK score and RMSD much further from the average. This is not to say that those results are the biological solution, it could be that the single structure at RMSD = 13.8 Å, HADDOCK score = -78, is the biological significant result and the other structures are simply calculation artifacts as a result of the energy folding landscape. When the low from average RMSD structures are calculated against the reference position of Im9 then it is found that they have large translation values varying between 65 and 80 Å.



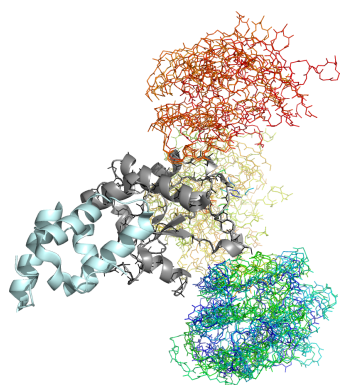
(a) HADDOCK statistical summary

(b) HADDOCK visual summary

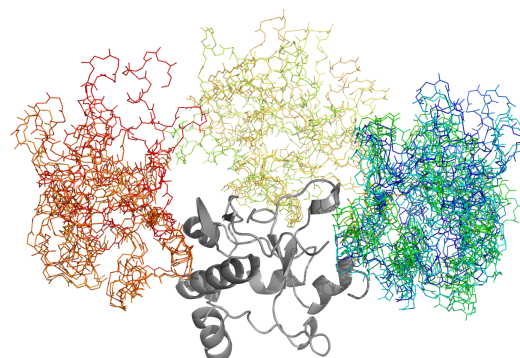
Figure 8.15 HADDOCK run with Im9 S6C^{I A1}/C23^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.



(a) HADDOCK statistical summary



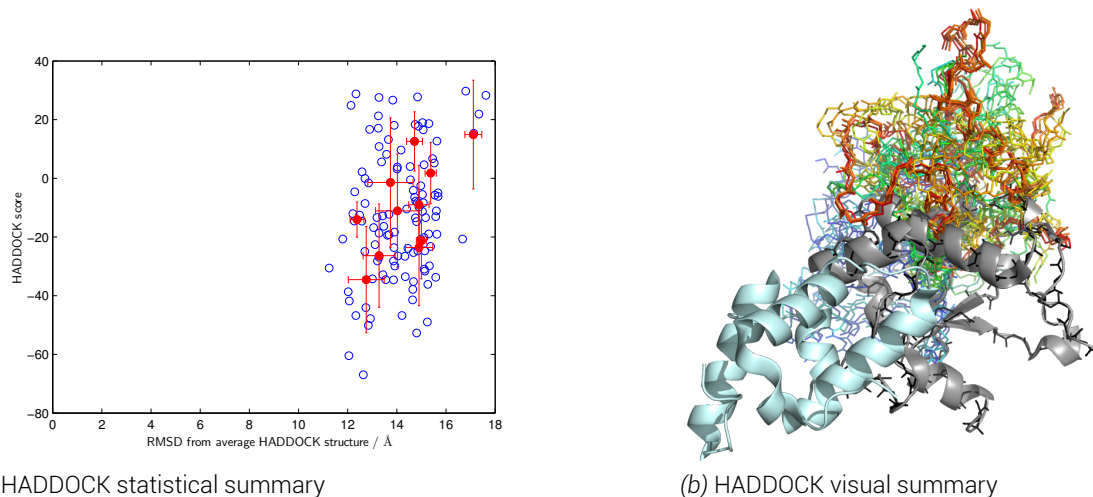
(b) HADDOCK visual summary



(c) HADDOCK visual summary

Figure 8.16 HADDOCK run with NMR AIRs and colicin E9 DNase S3C^{I A1}-Im9 S6C^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey.

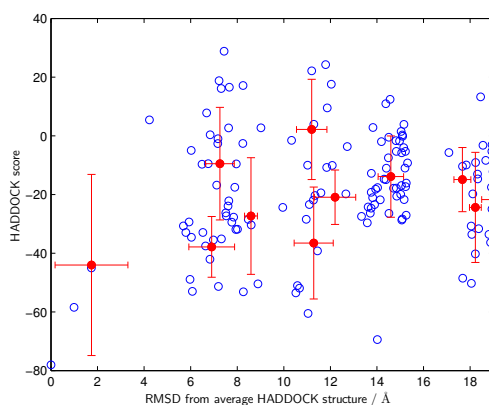
8. Protein - protein docking with HADDOCK



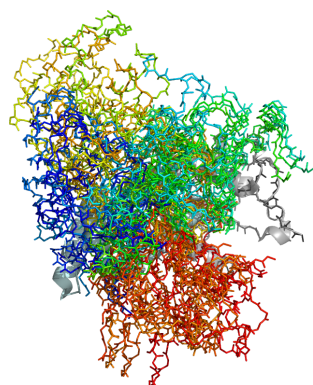
(a) HADDOCK statistical summary

(b) HADDOCK visual summary

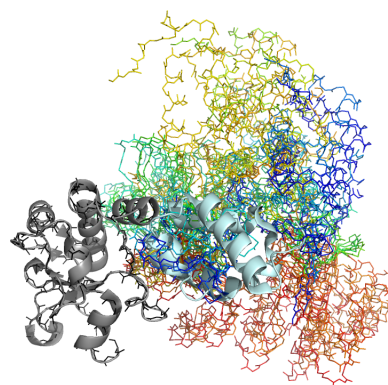
Figure 8.17 HADDOCK run with NMR AIRs and colicin E9 DNase E66C^{I A1}-Im9 C23^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.



(a) HADDOCK statistical summary



(b) HADDOCK visual summary



(c) Reorientated visual summary

Figure 8.18 HADDOCK run with colicin E9 DNase E66C^{I A1}-Im9 S6C^{I A1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b) maximising visualisation of all Im9 structures.

8.6.6. With intermolecular PELDOR restraints from colicin E9

DNase E66C^{IA1}

Figure 8.17 shows the results of a HADDOCK run with NMR AIRs and a PELDOR restraint between colicin E9 DNase E66C^{IA1} and Im9 C23^{IA1}. As the colicin E9 DNase E66C^{IA1} position is located almost centrally to the globular protein domain it was expected that this single distance would give the most ambiguous docking results in such a small, spherical protein as the docking would be driven almost exclusively by protein surface interactions.

In figure 8.17.a an almost identical distribution of structures to that of figure 8.11.a where docking was conducted with no distance restraints can be seen. On closer inspection of figure 8.17.b the top 4 clusters have shifted such they all reside on the same side of the protein, simply with a wide distribution. This could however simply be a result of the constraints as the experimental data could only resolve to 2.80 ± 0.50 nm. Again the top scoring structures would appear to show a relatively narrow distribution of structures, but as lower scoring structures are considered then the image distorts. The translation value though of the modelled Im9 positions to the reference position remains in the 65-80 Å range, showing that these structures are placing the Im9 on almost exactly the opposite side of the colicin E9 DNase.

The PELDOR distance restraint between colicin E9 DNase E66C^{IA1} and Im9 S6C^{IA1} whilst being longer could also be resolved to a narrower range of 4.65 ± 0.30 nm. This again splits the top 4 clusters such that no discernible position of the Im9 upon the colicin E9 DNase can be seen, as highlighted in figures 8.18.b and c. Crucially however, the Im9 now appears to be on the front side of the colicin (in the reference orientation shown in figure 8.18.b), close to the position of Im9 in X-ray crystallography and NMR structures.

Most interestingly this distance restraint causes a dramatic shift in structures and clusters in the statistical summary (figure 8.18.a). A large number of structures remain above 12 Å from the average structure but there seems to be definite blocks in the ranges 6-8 Å and 10-12 Å. Most significantly there is a single cluster under 2 Å, with 1 structure having not only the best HADDOCK score but a near zero RMSD from the average structure. The top scoring structure (blue) can even be seen in figure 8.18.c to have some backbone overlapping with helical structure of the NMR reference structure. The translation average of C_{α} s between modelled and reference Im9 positions remains quite high for these top scoring structures in the range of 10 Å, despite the structural overlap. The reason for this is simple, in the docking process for this docking run almost all of the structures resulted in the Im9 unfolding and rearrangement of the tertiary structure (as can be particularly evidenced in the yellow structure of figure 8.18.c). With the docking run repeated with lower annealing temperatures the same result was achieved as soon as the proteins were no longer treated as solid bodies and backbone freedom was allowed, perhaps suggesting a problem with the restraints.

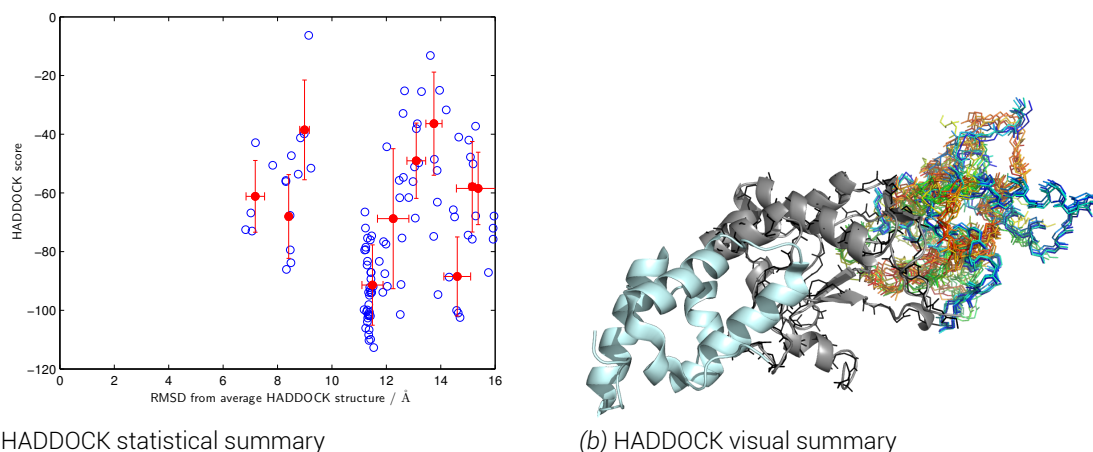
8.6.7. With intermolecular PELDOR restraints to Im9 C23^{IA1}

Figure 8.19 HADDOCK run with NMR AIRs and colicin E9 DNase S3C^{IA1}/S30C^{IA1}/E66C^{IA1}/S80C^{IA1}- Im9 C23^{IA1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.

In this docking run all experimentally derived intermolecular long range distance restraints to Im9 C23^{IA1} were combined (3 in total) with both colicin E9 DNase intramolecular distances. Docking models typically are at their best when there are multiple distance restraints across multiple sites on both docking partners. However, here it was investigated whether triangulation of distances to a single point on a docking partner would be sufficient to refine a model.

Most striking of this docking run is the visualisation in figure 8.19.b where the top 8 scoring structures (blue/green) almost exactly lie on top of one-another (RMSD < 2 Å to each other). Further clusters (yellow/red) show a distinct second orientation of the Im9 to the colicin E9 DNase but still very much in the same location upon the surface of the colicin E9 DNase.

As with previous runs, especially those using intramolecular distance restraints there is a strong cluster of many structures with good HADDOCK scores at an RMSD of ~11 Å from the average structure. A spread of clusters is seen at a slightly shorter and longer distance from the average. This suggests a distribution of sites where Im9 can dock and perhaps causes an unexpected average structure leading to artificially inflated RMSDs of the structures. What is clear, is that the single cluster not only scores the best structures within a very tight RMSD range of one another but (< 1.2 Å), this structure contains 40.4 % of the total structures for the run. This cluster, however remains upon the opposite side of colicin E9 DNase with a translation distance of the Im9 from modelled to reference of 68 Å.

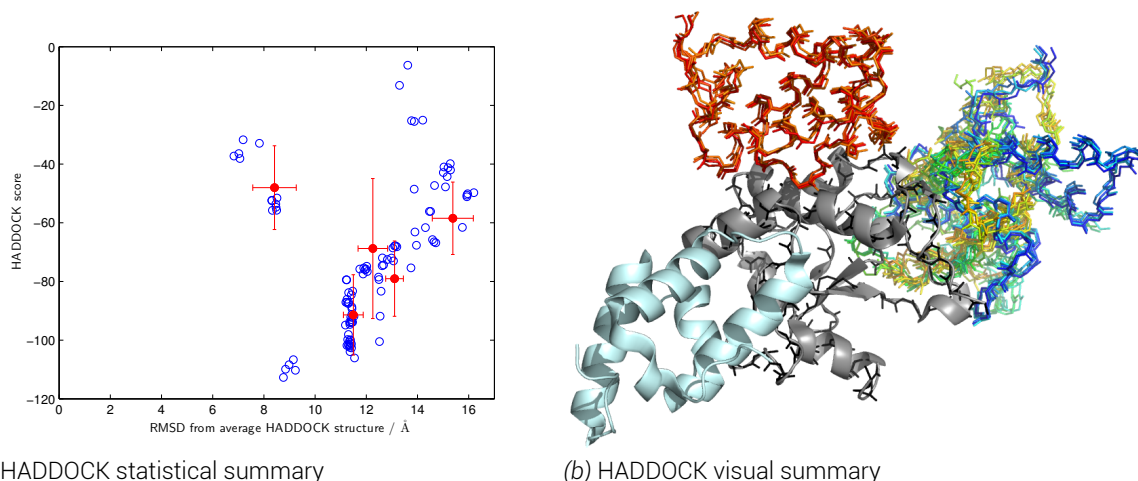
8.6.8. With intermolecular PELDOR restraints to Im9 S6C^{IA1}

Figure 8.20 HADDOCK run with NMR AIRs and colicin E9 DNase S3C^{IA1}/S30C^{IA1}/E66C^{IA1}/S80C^{IA1}- Im9 S6C^{IA1} PELDOR restraint. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.

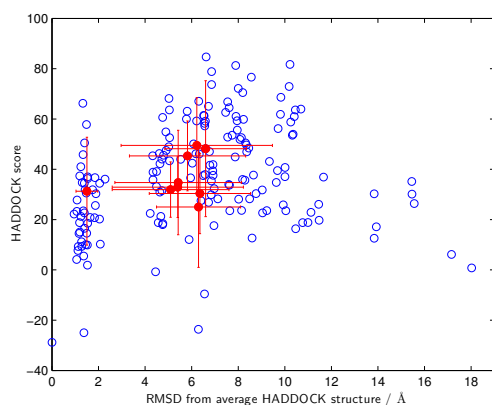
This docking run, similar to the previous docking calculation to Im9 C23^{IA1}, takes every intermolecular long range distance restraint from colicin E9 DNase to the Im9 S6C^{IA1} site (3 in total) with addition restraint within the colicin E9 DNase by 2 intramolecular distances.

As with figure 8.19, figure 8.20 shows a much more restricted system than in previous docking runs without long range distance restraints. In figure 8.20.b the top 3 clusters again are located upon a single side of the colicin E9 DNase (at the back in this reference orientation). The fourth cluster (orange/red) adopts the previously seen position on top of the colicin E9 DNase. Whilst the top clusters see a splitting of positions, the degree to which structures in a cluster of overlay is very high, and higher than was seen in the previous docking run. Im9 again assumes two orientations about the same sites on the surface of colicin E9 DNase, with the blue and yellow/green clusters 68 and 65 Å respectively from the reference position of Im9, whilst the lower scoring red/orange cluster is a slightly closer 53 Å.

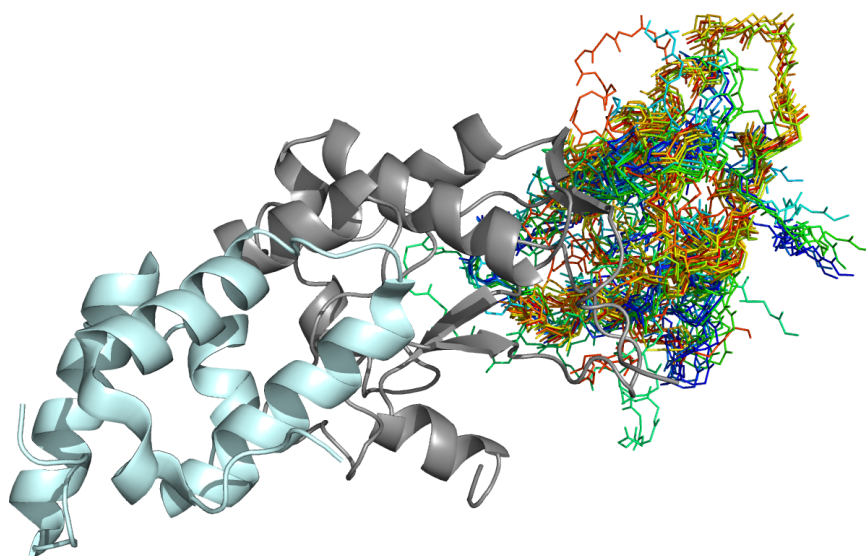
Whilst visually not as good as the Im9 C23^{IA1} site, the statistical summary of S6C^{IA1} shows even greater clustering of structures. There remains a few outlying structures which score poorly but the vast majority of structures fall into the 2 orientations of Im9 shown in figure 8.20.b.

The five best scoring structures at an RMSD from the average structure of approximately 9 Å, are not significant. Whilst appearing to at a similar HADDOCK score and RMSD from the average structure, when visualised, they have very different conformations to one another and cannot be considered a cluster and as a result need to be treated with suspicion as outliers.

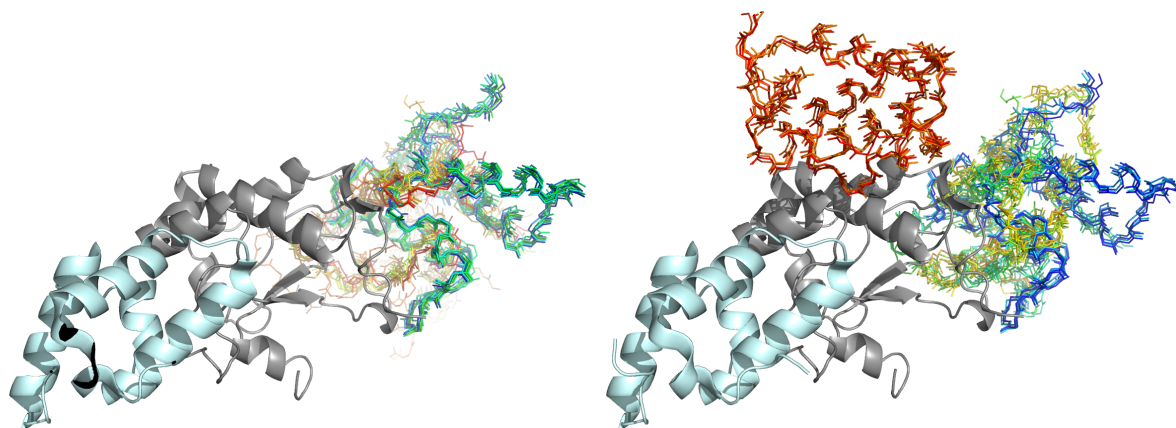
8. Protein - protein docking with HADDOCK



(a) HADDOCK statistical summary



(b) HADDOCK visual summary, 1 cluster, 16 structures



(c) HADDOCK visual summary, 2 clusters, 8 structures from each (d) HADDOCK visual summary, 4 clusters, 4 structures from each

Figure 8.21 HADDOCK run with AIRs and colicin E9 DNase S3C^{I A1}/S30C^{I A1}/E66C^{I A1}/S80C^{I A1}- Im9 S6C^{I A1}/C23^{I A1} PELDOR restraints. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing the average of the top four scoring structures in that cluster. b), c) and d) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Im9s in line format are coloured by HADDOCK score from blue to red. b) shows 16 structures from the top scoring cluster c) shows 8 structures from each of the top two scoring clusters d) shows 4 structures from each of 4 top scoring clusters.

8.6.9. With all PELDOR restraints and CSPs

Figure 8.21 shows the docking results when all four IA-PROXYL spin label sites on the colicin E9 DNase are used in combination with both sites on Im9. The run consists of 2 colicin E9 DNase intramolecular restraints, 1 Im9 intramolecular restraint and 4 intermolecular restraints. It was hoped that more intermolecular restraints could be used but the protein expression levels and labelling efficiency eliminated certain positions. Whilst doubly labelled proteins helped elucidate some restraints despite this, distances were often too similar to accurately restrain.

The statistical summary of the run shown in figure 8.21.a shows that there is one structure, with the highest score sitting almost exactly upon the average structure at 0 Å. With the addition of restraints the number of structures within cluster with a backbone RMSD of less than 2 Å from the average is very high. After this the clusters become somewhat cluttered with a large number of clusters and structures sitting in the RMSD from average structure range of 4-12 Å.

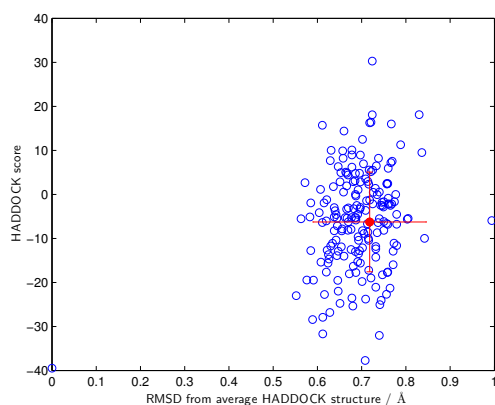
The first cluster contains 70 structures (35 % of the docking run), as a result figure 8.21.b shows the visual summary of only that cluster, showing the top scoring 16 structures. With the exception of the protein tails on the yellow structure, a near perfect overlay of the Im9 structures can be seen.

The second cluster contains a further 34 structures, and its cluster average HADDOCK score is slightly better than the first cluster. However, the cluster has a much large RMSD range from the average structure. Figure 8.21.c, shows the top 8 structures from each of the first 2 clusters, the first cluster remains tightly formed in black/red/orange. The second cluster in blue and green shows that whilst the absolute HADDOCK score is marginally better, the spread of structures narrows docking down to little more than an entire side of the colicin E9 DNase.

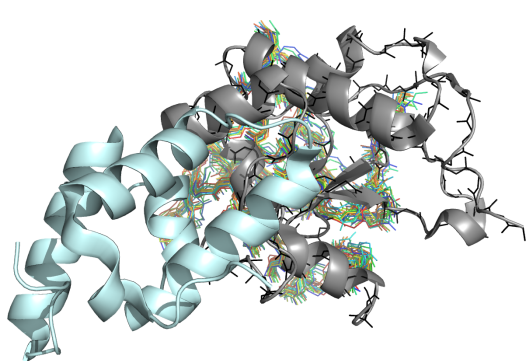
In figure 8.21.d we see a visual representation more in keeping with previous figures where the top 4 structures from each of the top 4 clusters are presented. Here cluster 3 has 20 structures (10 %) and cluster 4 has 12 structures (6 %). With the addition of 2 poorly scoring clusters with large RMSDs simply distorts the image. However at no point does the Im9 ever move to the other side of the colicin E9 DNase seen when the model was without PELDOR derived distance restraints.

Crucially, this docking run with experimental PELDOR restraints shows only structures upon one surface of the colicin E9 DNase and distinguishes between models from when only NMR data is used.

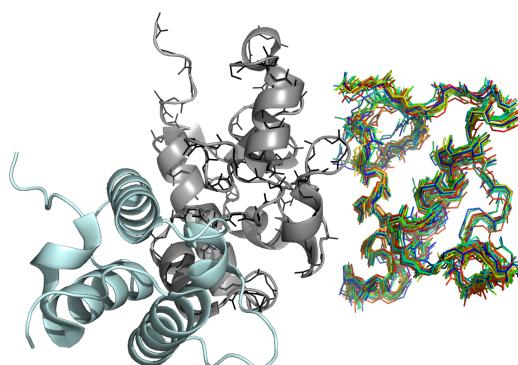
8. Protein - protein docking with HADDOCK



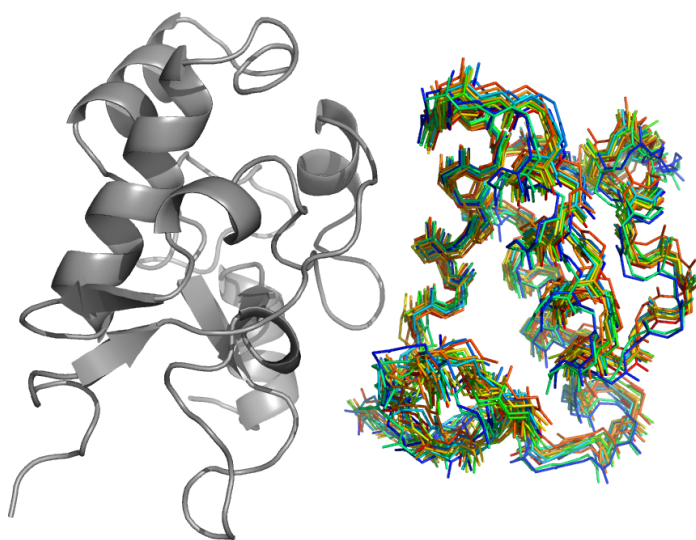
(a) HADDOCK statistical summary



(b) HADDOCK visual summary - reference orientation



(c) Reorientated visual summary



(d) HADDOCK visual summary - Im9 orientation optimised

Figure 8.22 HADDOCK run without AIRs, only colicin E9 DNase S3C^{I.A1}/S30C^{I.A1}/E66C^{I.A1}/S80C^{I.A1}- Im9 S6C^{I.A1}/C23^{I.A1} PELDOR restraints. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C α , O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red. c) reorientated version of panel b). d) reorientated version of b) maximising visualisation of all Im9 structures. The reference structure is removed and top scoring colicin E9 DNase structure is shown in grey.

8.6.10. With only PELDOR restraints

Using only 2 colicin E9 DNase intramolecular, 1 Im9 intramolecular and 4 intermolecular long range PELDOR derived distance restraints, could docking be successful? This was the question asked in the last experimentally obtained data driven docking run. From prior docking runs it was unclear as the number of long range restraints increased whether the remaining uncertainty was as a result of long range, unambiguous restraints or as a result of the ambiguous interaction restraints from the NMR experiment. A threshold of one standard deviation, in the NOE experiments means that 32 % of residues will be considered as important to binding, which in protein as small as Im9, 28 of the 86 residues will be active to binding before passive residues are considered.

In figure 8.22.a the statistical summary is presented, notice the difference in scale compared to previous figures. In this docking run all 200 structures have completed the docking procedure and have formed into a single cluster. The RMSD of every amino acid's C_{α} has been aligned to the top scoring structure and the RMSD averaged, with each structure having an average RMSD of less than 1 Å from the docking run average structure.

When this result is visually plotted, in figure 8.22.b the result is initially not clear, for this reason figure 8.22.c shows figure 8.22.b rotated approximately 90° clockwise. In these figures it can be seen that the top 16 scoring structures from this single cluster almost perfectly overlay, with only slight variation in the loop sections, whilst the central helices can still be seen even in this line representation of the backbone. Figure 8.22.d removes the reference structure and optimises the view upon the Im9 solutions to docking.

From this it is clear, in this case, that the ambiguity from using NMR derived ambiguous interaction restraints is hampering the collapsing of docking solutions to a single solution. Despite this the location of the Im9 is drastically different to that of the published X-ray crystallography and NMR structures and

By default, HADDOCK gives equal weighting to ambiguous interaction restraints as it does unambiguous, in this case the PELDOR derived distances. Clearly a stronger weighting to the unambiguous restraints would be beneficially. However, this assertion only holds true in a scenario where there are a sufficient number of distances in place to create a good model, with fewer distances then maintaining equal weighting would be prudent.

8.6.11. With a MMM derived restraint

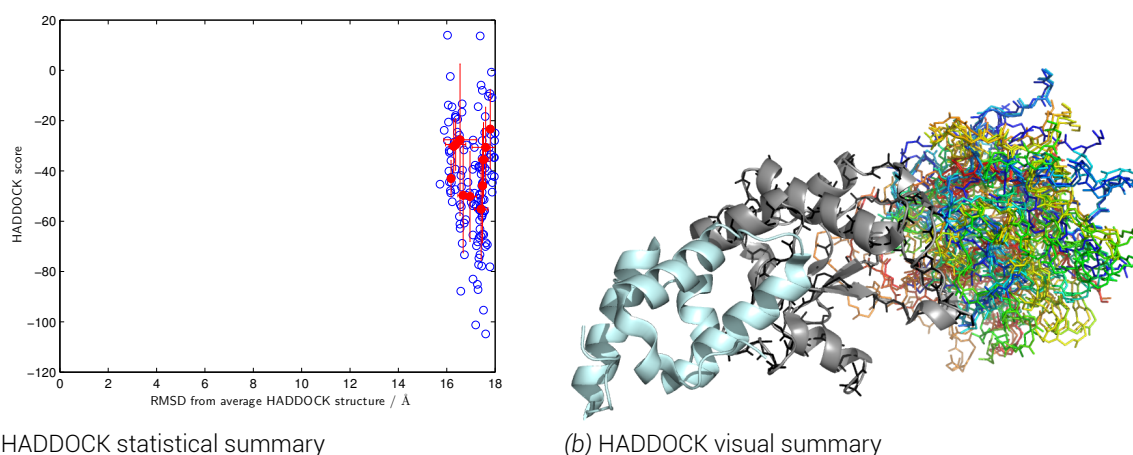


Figure 8.23 HADDOCK run with NMR AIRs and colicin E9 DNase S3C^{IA1}-Im9 C23^{IA1} MMM restraints. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing the top 4 structures average. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.

Figure 8.23 shows a HADDOCK run using NMR derived AIRs and an MMM predicted distance distribution restraint using the published colicin E9 DNase - Im9 complex (PDB: 2K5X) between positions colicin E9 DNase S3C^{IA1} and Im9 C23^{IA1}. The initial purpose of this run was to investigate whether *in silico* derived EPR distance restraints could be successfully integrated into a HADDOCK run and whether these restraints could yield useful data that could not otherwise be obtained experimentally.

Using MMM and the NMR complex structure (PDB: 2K5X) IA-PROXYL labels were introduced to the colicin E9 DNase S3C and Im9 C23^{IA1} positions at 175 K. The rotamer ensembles were then used to calculate a distance distribution in MMM before being converted to a single MMM derived, MISHAP generated distance restraint of $4.10^{+0.92}_{-0.30}$ Å.

From visual inspection of the top scoring structures (figure 8.23.b) it can be seen that from introduction of a single, long range, intermolecular distance restraint the top scoring structures have collapsed onto a single side of the colicin E9 DNase. This however, is on the opposite side of the colicin E9 DNase than the published. However, on inspection of the statistics presented (figure 8.23.a) whilst the spread of RMSDs from the average structure has reduced to ~ 2 Å the actual distances of all structures have increased from docking without restraint (as shown in figure 8.11) to between 16 and 18 Å. Figure 8.23.b shows a spread of Im9 positions amongst the top scoring structures, but it is unreasonable to think that the average structure could be over 16 Å away without a significant number of structures being upon the other side of the colicin E9 DNase. However these structures must obviously be of a much lower HADDOCK score to not be presented in the visual summary.

Most importantly, the capacity of MMM derived *in silico* distance restraints to be used in MISHAP and passed to HADDOCK has proved to yield meaningful results that were otherwise not possible to obtain experimentally.

8.6.12. With MMM derived restraints for all spin labels

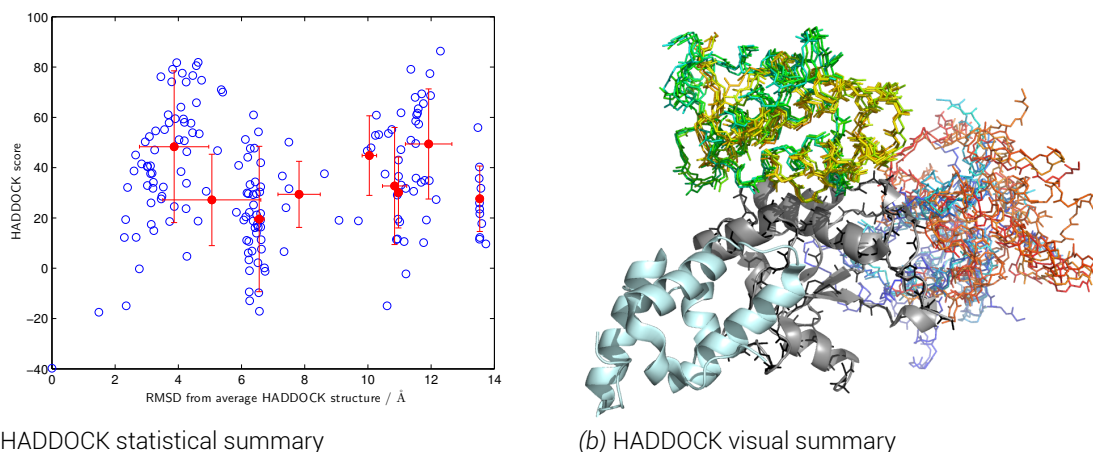


Figure 8.24 HADDOCK run with NMR AIRs and MMM derived distance restraints between every spin label. a) statistical summary, open blue circles show a HADDOCK solution structure, closed red circles show cluster averages with error bars showing one standard deviation. b) visual summary in fixed rotation showing the published colicin E9 DNase - Im9 complex (PDB: 2K5X) in cartoon format with colicin E9 DNase in grey, Im9 in pale cyan. Backbone atoms (N, C, C_α, O) for the top scoring colicin E9 DNase in black after structure alignment to the reference structure. Im9s in line format are coloured by HADDOCK score from blue to red.

Figure 8.24 shows a HADDOCK run using NMR derived AIRs as well as 3 intramolecular and 10 intermolecular long distance restraints from MMM calculation upon the NMR complex structure. Here IA-PROXYL labels were introduced to the 5 colicin E9 DNase sites and 2 Im9 sites in MMM at 175 K. The purpose of this docking run was to conclude whether with all these docking restraints in place whether there would still be uncertainty in the docking calculation and if the published structure would be achieved.

In figure 8.24.b it can be seen that despite having all of these restraints from the published structure the modelling fails to yield the top scoring structures in the same position. Instead the top scoring structures (blue) are favoured to be at the back of the colicin E9 DNase, whilst a notable amount are located in the previously seen position on top of the colicin E9 DNase.

Figure 8.24.a shows that a great deal of the structures complete the docking process with over half having a RMSD of less than 8 Å from the average structure. None of these structures score very highly and the cluster averages are shown to have both broad RMSD and HADDOCK score ranges.

The results here elude to two possibilities. First, the selection of labelling positions was not appropriate to give sufficient sampling of the protein structures and therefore enough restraint in the modelling. Alternatively, the ambiguous interaction restraints from NMR being given equal weighting to the unambiguous long range distance restraints creates artificially high levels of uncertainty. These uncertainties yield the possibility of poorly scoring structures in positions that would usually be prohibitive in the initial stages of the docking process. Additionally, it remains important to remember that colicin E9 DNase is only one domain of a multidomain, globular protein.

8.7. Summary and conclusions

In this chapter it has successfully been shown that using experimentally derived long range distance distributions between spin label sites can effectively restrict docking models. The extent of restriction varies greatly upon the position and environment of the spin labels in question.

The colicin E9 DNase E66C^{IA1} labelling position being almost central within the globular protein and having restricted movement yielded a sharper PELDOR derived distance distribution. With a narrow range upon the long distance restraint yielded less uncertainty within the docking models. However, in such a small, globular and largely spherical protein having a distance from close to the centre of mass does little to discriminate structures in the docking process. Of course, when using spin labels placed upon an internal protein surface the risk of distorting the protein secondary and tertiary structure is greatly increased, and without careful experiment design could yield many proteins inactive or unable to refold.

Using solvent exposed labelling positions, allows a higher probability of spin label - protein interaction and as a result makes for greater labelling efficiencies. Using solvent exposed residues also allows for the protein to be labelled without first having to unfold the protein chemically with the use of a denaturant. In this study the refolding of the colicin E9 DNase was carefully monitored using circular dichroism, but not all proteins would allow as such.

The largest cost of using solvent exposed labelling positions is that a far the spin label is typically much less sterically hindered. This removal of hinderance by the protein itself allows the spin label much more freedom of movement, which computationally can be calculated as allowing a larger rotamer population. The additional mobility ultimately is averaged in the PELDOR experiment as the experiment samples all molecules within the sample, resulting in broader PELDOR distance distributions. These broader distance distributions in turn result in less restriction during docking. However, by placing spin labels upon the protein surface the risk of distortion to the protein secondary and tertiary structure is minimised. Care must be taken such that the labels do not directly inhibit binding or that key residues are not mutated away rendering the protein insoluble.

In an idealised situation the protein in question could be roughly approximated as cubic and a restricted spin label placed upon the eight corners in separate protein mutants. In this situation it is easy to visualise the six degrees of freedom in which each cube can be rotated in x , y or z before translation onto each other. With suitable narrow distance distribution ranges a unique solution could be quickly and accurately found requiring only a single PELDOR target on the opposing docking partner yielding position and orientation. The same result could be achieved using less sites if the sites were suitably restricted and showed suitable orientation dependence (which may require higher frequency experiments). Replacing the cubes with theoretical spheres, with ideal conditions and distances, it should be possible to define an interface with four distances.

From the final experiments, particularly §8.6.9 and 8.6.10 where the docking procedure was conducted with all the PELDOR derived distance restraints with and without the NMR derived ambiguous interaction restraints it is apparent that the weighting of each data source needs to be carefully examined. When both NMR and EPR restraints are combined then the modelling process completes with many structures adopting a very close position on one

surface of the colicin E9 DNase but with several site orientation rotations of the Im9. These structures form many clusters but with low RMSD from each other and crucially there is a single cluster at less than a RMSD of 2 Å from the average structure and comprises 35 % of all solutions. However, with the removal of the NMR derived AIRs the docking process reduces to a single cluster comprising all structures with an RMSD of less than 1 Å.

The difference between the top scoring cluster of the docking run with AIRs and the single cluster of PELDOR only derived restraints is slight. When figures 8.21.b and 8.22.b are directly compared it can be seen in the visualisations that with the AIRs the Im9 sits mainly to the right and slightly behind the colicin E9 DNase but without the AIRs the Im9 position can hardly be seen as it is almost directly behind the colicin E9 DNase. The distance between the Im9 positions is only 9.8 Å when the top scoring structure from each run is translated in space. Here the docking run without CSPs is closer at 46.8 Å to the reference Im9 position than with AIRs (68.2 Å). Whilst both are in similar range from the reference position, these values speak little other than to inform what can be seen visually that the Im9 positions from modelling are significantly different than the reference structure.

Finally, in this investigation (§8.6.11), it was shown that it is perfectly possible to use MMM to *in silico* spin label proteins. This labelling procedure creates rotamer clouds which can then calculate an expected distance distributions for a PELDOR experiment, subject to variations in the protein and rotamer populations. Using the computing toolbox developed in this investigation these predicted restraints could be incorporated as long distance restraints and docking parameters in the same way as experimentally derived restraints.

Careful consideration of computed parameters should always be used as the simulation will only ever be as good as the information that is fed into it. In this instance, the proteins are considered to be rigid bodies and each rotamer is assumed to be largely equally populated. Obviously, proteins are not fixed in space and routinely undergo molecular breathing, conformational exchange as well as structural rearrangement upon docking.

In the final docking run (§8.6.12) it was hoped that with the use of idealised long distance restraints from MMM, including several of which could not be obtained experimentally, would restrain the model sufficiently to obtain the reference complex from which the distance restraints were calculated. The result of this docking procedure generated a rather ambiguous picture which echoed many of the previous docking runs. Statistically none of the final structures scored very well, with nine clusters - each with a varied spread of scores and RMSDs. On visual inspection of the better scoring clusters a clear split of populations is seen between the Im9 sitting on the right and slightly behind and on top of the colicin E9 DNase (in the reference orientation). These positions almost identically match those seen when the docking was conducted with experimentally derived distance restraints. This allows for a conclusion to be drawn that the curious nature of placing the Im9 in the right and top positions is as a result of the docking procedure and not the experimental data. In fact, the position of Im9 to the right of the colicin E9 DNase appeared even in the first HADDOCK run where no restraints of any sort were used, appearing in figure 8.10.b as the red/orange cluster. Clearly this surface of the colicin E9 DNase contains residues that are energetically favourable to be hidden from water. Im9 appearing on top of colicin E9 DNase also appeared early in the investigation when in figure 8.11.b, using only NMR derived AIRs, the Im9 position was clearly split between the reference position and on top of the colicin E9 DNase.

As a result then, it can be assumed that the top position of the Im9 in relation to the colicin

8. Protein - protein docking with HADDOCK

E9 DNase is as a result of the docking procedure's use of the AIRs. Whilst the right position is energetically favourable enough to occur in almost each docking run, especially in those runs where the long range distance restraints are not sufficient to discriminate and exclude it.

At this time, HADDOCK only provides the facility to turn on or off ambiguous or unambiguous restraints at each stage of docking. So that, for example, ambiguous results could be used during rigid body structure generation and then only unambiguous restraints are used during energy optimisation to discriminate models. This is not the same as having both restraints present but unambiguous restraints favoured. Nor is there a way to specify the tolerance of AIRs, so that AIRs could be abused if a favourable unambiguous restraint is nearly fulfilled. However, in each docking run a user definable fraction of AIRs are randomly deleted (20 % by default) to assist in ensuring random sampling. If this fraction of randomly deleted restraints was increased, with the docking process repeated several times from different initial seed values and results combined in some way then this could perhaps provide an inelegant solution.

Using only NMR derived AIRs of the final 200 structures calculated by HADDOCK there was a near 50% split in structures between 2 sites upon the surface of the colicin E9 DNase. That in which the Im9 directly inhibits the colicin E9 DNase activity by binding directly to the DNase activity region. Alternatively, that in which the Im9 approaches from behind the DNase activity region but could cause some conformational rearrangement and inhibit the DNase activity flexible region from binding DNA. It was hoped that with the introduction of increasing number of long distance restraints that one population would be increasingly favoured over the other. Instead, as a result of the poor spin labelling, poor PELDOR traces were obtained, which translated into poor and broad distance distributions. Rather than discriminating between the two Im9 binding positions, this modelling would suggest an entirely different site.

It is key to remember that in this docking we consider both binding partners to be rigid, solid bodies. In reality, the colicin E9 DNase domain in solution is shows enough intrinsically disordered character and dynamicity to stop the elucidation of its three dimensional structure by NMR. A X-ray crystallography structure of the colicin E9 DNase was obtained in 2000 (1FSJ) using conditions far from that found *in vivo*, whilst it does align well with complexed colicin E9 DNase structures, recent studies would suggest that this level of structure is unlikely in solution [43] with the structure arising as a result of crystal lattice stabilisation. The crystal structure (1EMV) is that of cocrystalised colicin E9 DNase / Im9 complex, whilst the NMR structure (2K5X) was collected and assigned in complex. The binding event then, with such a strong dissociation constant, must coincide with a massive conformation change in the colicin E9 DNase. It is reasonable then to assume that the colicin E9 DNase structure is induced by and forms around the Im9.

9. Discussion and conclusions

9.1. Protein production

When the investigation began, the vast majority of DNA work was thought to be already available after years of extensive investigation of the colicin E9 DNase system by NMR and X-ray crystallography groups at UEA. Unfortunately, much of the DNA stocks had been badly maintained and contaminated, resulting in several frame shifts which made a practical solution of fixing the genes with primers and PCR work impractical. Wherever possible DNA was sought from site active colicin research groups at the Universities of Nottingham, Oxford and York such that as many genes as possible in the investigation would be derived from naturally occurring gene. These groups, despite their generosity, did not have the desired genes or themselves only had copies of the contaminated genes from UEA.

Rather than attempting a time consuming and costly site-directed mutagenesis approach beyond the focus of this investigation and available facilities a synthetic gene approach was chosen. Here a desired protein has the amino acid sequence reverse translated into a simple DNA sequence, with certain optimisations made for species, codon usage and G/C content. The DNA is then produced in an oligonucleotide synthesis with appropriate ends to ensure the correct reading frame and insertion into a vector. The synthetic genes arrived delivered in a suitable cloning vector within 2 weeks of order at a financial cost comparable with DNA primers, cloning vectors and restriction enzymes. As the genes were being synthesised fully the codons could be optimised for maximum protein expression in an *E. coli* expression strain. In the case of the colicin E9 DNase S3C/S80C and S30C/S80C mutants, the resulting protein expression yielded excellent results, yet the whilst the naturally derived colicin E9 DNase S30C and S80C mutants barely expressed any significant amount of protein.

The constructed genes were designed such that they either the protein themselves or the co-expression partner contained a poly-histidine tail. Purification using an immobilised metal ion chromatography method proved a successful way of separating the protein of interest from the cell lysate, with only very minor contamination. A polishing purification step was completed using size exclusion chromatography to ensure a high quality final product. During size exclusion chromatography in the case where cysteines were highly solvent exposed, the necessity of using a gentle reducing agent to stop the formation of protein-protein homo-dimers was apparent.

Each purified protein was subjected to several tests to confirm both its identity and ensure its suitability for further investigation. Spectrophotometry confirmed protein concentration by measuring the absorbance of light at 280 *nm*. Circular dichroism confirmed the proteins' major secondary structure features as well as confirming that the protein had folded correctly, especially important later in the spin labelling of some mutants. Mass spectrometry confirmed the protein mass in the gas phase. Fluorescence enhancement assays confirmed

9. Discussion and conclusions

that the colicin E9 DNase was indeed binding to Im9 in both naturally and synthetically derived mutants of both proteins. Finally enzymatic assays were performed to confirm the kill ability to the colicin E9 DNase molecules, as well as the ability of the Im9 mutants to inhibit the action of the DNase activity.

As a result of many months hard work, 2 mutants of naturally derived Im9 were produced in high yields (C23, S6C/C23S), 1 synthetic mutant of Im9 (S6C) was produced in moderate yield, 3 naturally derived colicin E9 DNase mutants (S3C, E66C, D20C/E66C) were produced in high yields, 2 synthetic mutants of colicin E9 DNase mutants (S3C/S80C and S30C/S80C) were produced in high yields and 2 synthetic mutants of colicin E9 DNase mutants (S30C and S80C) were produced in poor yields.

9.2. cw-EPR

The process of spin labelling proteins was to use a slow and methodical approach to try to find the best labelling yields. It was known that any mutant moving forward to PELDOR experiments would require high labelling efficiencies as 2 proteins at 90 % labelling efficiencies yields 81 % (0.9^2) of protein complexes contributing to the PELDOR signal. Whereas a complex of proteins with 10 % labelling would yield 1 % (0.1^2) of those complexes contributing to the signal.

The spin labelling investigation tested multiple pHs, in aerobic/anaerobic environments, at multiple labelling concentrations, for various times, at different temperatures, with and without a reducing agent, with 2 different spin labels (MTSL and IA-PROXYL) and using three different separation methods (PD-10 columns, Amicon stirred cells and spin concentrators (must use Hydrosart membrane, not polyethersulphone as colicin E9 inversibly binds to it)). Despite all of this, labelling efficiencies were still lower than would have been desired. For most mutants the labelling efficiencies of greater than 15 % were achieved, when before optimisation a 1 % labelling efficiency was typical.

The largest problem that was encountered in spin labelling was the proteins themselves. The proteins used in this investigation are relatively small, globular proteins and due to their tight folds it was deliberately avoided to place spin label sites on the inside of the protein - with the exception of the colicin E9 DNase D20C and E66C positions that are inside a small pocket. It was also expected that by placing spin labelling targets in solvent exposed positions would allow access of free spin label to react with the exposed cysteine side chains. However, in such small, globular proteins exposed cysteines allowed the spontaneous formation of homo-dimers. The only way to remove the disulphide linked homo-dimers was with the use of a reducing agent, but this reducing agent would indiscriminately reduce the cysteine - spin label bond that was being attempted to be formed by the MTSL. Alternative strategies such as very low protein concentrations thereby minimising protein-protein interactions, anaerobic environments to minimise oxidation of formed disulphide bonds, pH shifts to move the reaction equilibrium, reaction time and reaction temperature did little to increase the labelling efficiency with MTSL. IA-PROXYL with its iodo leaving group however, reacts with the exposed cysteine sulphur atom in a non-reversible manner and greatly increased the labelling efficiency. The IA-PROXYL spin label also merited special attention as the labelling efficiency was independent of oxygen and yielded 90 % reaction completion

in one hour on the laboratory bench. It could also be used safely in a ten molar excess to the protein to encourage the labelling reaction without unspecific binding to the protein and washed away quickly as it passed through spin concentrator membranes.

Accessibility / power saturation experiments showed that the expected colicin E9 DNase S3C and S30C and Im9 C23 labelling sites which are highly solvent exposed in both X-ray crystallography and NMR structures were greater than 60 % solvent exposed. The expected internalised site, such as that of colicin E9 DNase E66C position showed less than 3 % aqueous accessibility. These accessibility measurements worked well to confirm that the spin labelling positions matched well with how they were expected to behave from previous structural studies, indicating that the protein production and spin labelling procedures were successful.

When the power saturation experiments were repeated in the presence of the complex binding partner the accessibility values did not change by any significant amount and in doing so proved that each labelling site was indeed away from the binding interface. Equally, low temperature line broadening experiments were attempted but showed no discernible broadening of the centre peak, indicating that the labels were at least 16 Å apart in the complex, agreeing with the NMR and X-ray crystallography structures. Had the colicin E9 DNase D20C/E66C double mutant expressed and labelled successfully it would have been a good candidate for seeing line broadening effects and in doing so allowed this investigation to incorporate two different EPR derived distance restraints. Continuous wave line broadening for distances under 15-20 Å and PELDOR for distances in the 15-80 Å regime.

9.3. Pulsed EPR

Time was spent in optimising every conceivable parameter to ensure the greatest possible transverse relaxation time required to keep spin coherence and maintain the longest possible observation window for PELDOR oscillations. Strong spin coherence also yields large signals which maximises the signal to noise ratio and reduces the required experiment time to achieve a good quality of data.

Simultaneously, a compromise was sought to maintain the shortest possible spin-lattice relaxation time. Such that the shot-repetition, the time in which the system was wait until it is fully relaxed and magnetically at equilibrium before it can be pulsed again, is kept low thereby minimising experiment length. In any investigation this optimisation is prudent. However, in this investigation with so many PELDOR experiments to be conducted an optimisation of a few hours per sample would quickly multiply. Additionally, from the cw-EPR investigation, the protein concentrations and labelling efficiencies were known to be very low and little above the detectable limit of the spectrometer. In order to obtain data of any significance it was known from the outset that optimisation here would be crucial to avoid huge experiment times from the required amounts of signal averaging.

Using a two pulse ESEEM pulse sequence to report the transverse relaxation time and an inversion recovery sequence to report the spin-lattice relaxation time, temperature was the first parameter to be optimised. Using a liquid helium flow cryostat samples from 4 K were tested until no further echo could be detected.

9. Discussion and conclusions

The transverse relaxation time showed a maximum at 4 *K*, with signal still being recorded after 1 μ s. As the temperature increased, so did the relaxation time in a linear fashion. At 30 *K* the transverse relaxation time remains at 94 % of that seen at 4 *K*. Rising further to 50 *K* sees the relaxation time drop to 86 % of that at 4 *K*, this is still a strong signal but the observable window had reduced by 100 ns. At 90 *K* a phase transition between ice states caused a large absorption of energy which generated an anomalous data point upon fitting.

Across this same range the spin-lattice relaxation sees a near exponential decay in length. At 4 *K* the spin-lattice relaxation stands at nearly 5.5 ms, at 30 *K* this has reduced to less than a third of what it was at 4 *K* to 1.4 ms. At 50 *K* the spin-lattice has reduced to a quarter of that seen at 4 *K*. No further data points could be recorded above 90 *K* as the signal became so weak.

As a result of these experiments a compromise was struck at 50 *K* between having a strong signal with a reasonable transverse relaxation time and having a reasonably short spin-lattice relaxation time. Had time been an infinite resource or the signal from the sample so strong as to require only a few scans then 4 *K* would have been used. Working at temperatures above 50 *K* would have been impractical as the spin coherence time would start to become so short as to not give a sufficient window in which to observe dipolar oscillations in the PELDOR experiment.

The pH of the protein was shown to have little bearing on the experiment in terms of the transverse relaxation time, though a small maximum is seen around pH 7.5. As the cytosolic pH of *E. coli* is 7.6 and the optimum conditions for colicin E9 activity is pH 7.4 an experimental pH of pH 7.5 was considered an acceptable compromise between biological relevances and spectroscopic desirability.

The deuteration assay showed an almost exponential increase in transverse relaxation times with deuterium composition of the buffer. Whilst some quirks were found here due to a complex mix of different relaxation methods, hyperfine interactions of protons and deuterium ES-EEM, ultimately the greater the level of deuteration the better. As a result all protein samples were concentrated in deuterated buffers such that all labile protons were exchanged. This simple exchange of buffers increased the transverse relaxation time by 300 ns from 500 ns.

One of the greatest increases found to the transverse relaxation time was found in the use of a glass forming agent. Glycerol interrupts the hydrogen bonding network of water and, when frozen quickly, stops the formation of crystalline ice - instead forming a glass. Physically this glass behaves more like a very slow moving liquid than a solid and crucially removes a large proportion of the crystal lattice and with it one of the major spin diffusion pathways. With a reduced crystal lattice, the energy and spin coherence lost by spin-lattice relaxation is greatly reduced. The glycerol also increases the viscosity of the solution, which at a given temperature greatly reduces the molecular tumbling time (correlation time). Even the small addition of 10 % glycerol increased the transverse relaxation time from 2.77 to 3.71 μ s. With further additions of glycerol the transverse relaxation time increased further but at a much reduced rate. It was decided that in future samples that 50 % glycerol would be used. Working with high glycerol concentrations is difficult as it requires that the protein be concentrated to very high levels nearing saturation before being added to the glycerol or require the difficult process of resuspending freeze-dried protein in glycerol. The high viscosity solutions associated with high glycerol concentrations also generate sample han-

dling problems when requiring small volumes to be accurately measured and placed inside narrow bore sample tubes.

The use of glycerol meant that the whilst the sample would now no longer form a large crystal lattice structure, it was now no longer completely deuterated. The effect of glycerol deuteration in both protonated and deuterated solvents was tested for completeness.

In protonated solvent environments the addition of glycerol increased the transverse relaxation time, with increased deuteration levels of the glycerol, with the relaxation time doubling from approximately $3 \mu s$ between no glycerol and 50 % glycerol- d_8 . Most impressive in this investigation was the dramatic effect of a fully deuterated sample of deuterated solvent and 50 % glycerol- d_8 which increased the transverse relaxation time from approximately $3 \mu s$ to $22 \mu s$ (a $7.8\times$ increase). It was concluded that the additional cost associated with deuterium oxide and glycerol- d_8 was worth the increased gain in signal. However, the main contribution to relaxation was now deuterium ESEEM and not proton hyperfine. The deuterium ESEEM modulations are approximately $500 ns$ in period and have a huge range in signal intensity which require careful consideration in setting up the PELDOR experiment to ensure that the observation is made upon a ESEEM modulation maximum.

It was found that a small increase could also be gained by taking the flash frozen sample and raising it in temperature to just below the glass transition before a slow drop in temperature. This resulted in a better glass being produced within the test tube with even less crystal characteristic, but the additional time required in warming and slowly cooling the sample again was decided against as the time could be better spent on signal averaging.

Finally a small investigation into pulse length was conducted to see whether short, high power hard pulses would yield a better signal than longer, soft pulses of greater length. It was found that longer pulses yielded longer transverse relaxation times as a result of a much greater signal intensity. However, this increased signal came with at a cost as the spin echo increased in spectral width, in the case of a $72 ns \pi/2$ pulse the echo increased to over $300 ns$. Intergrating over the entire echo yielded a large signal, but the spins show weak coherence to each other and would prove difficult when introducing more pulses as in the four pulse PELDOR sequence. Particularly, in the PELDOR experiment only a small part of the echo is integrated over to ensure minimal spin orientation dependence. Integrating over only the central $12 ns$ of the echo showed that using hard echoes and the resulting narrow echo of around $150 ns$ gave greater signal.

Equally moving to very short pulses meant that the spectrometer was unable to provide sufficient power to move the entire spin population. As a result with only some spins being correctly moved into the xy plane, with most still somewhere between the z axis and xy plane after the $\pi/2$ pulse, the detected echo was very weak and even distorted the deuterium ESEEM modulations seen in the decay. A compromise was made by using $16 ns \pi/2$ pulses as this gave an additional $300 ns$ to the transverse relaxation time over a $10 ns \pi/2$ pulse (the shortest pulse without distortion) whilst still giving a sharp echo of full width $140 ns$.

With all of this optimisation the colicin E9 DNase S3C^{IA1} transverse relaxation time was increased from $0.45 \mu s$ in water at $50 K$, to $2.37 \mu s$ in D_2O / glycerol- d_8 at $50 K$, a 527% increase. At the same time, the maximum spin-lattice relaxation time was $5.51 ms$ (recorded at $4 K$), was reduced to $3.81 ms$ by increasing the temperature and addition of a glassing agent, a decrease of 31% .

9. Discussion and conclusions

The increased transverse relaxation time allowed for a suitable dipolar observation window in the PELDOR experiment. The optimisations also allowed the PELDOR experiments to be reduced to a reasonable time frame of approximately 24 hours per sample, with some samples requiring 48 hours of signal averaging.

Ultimately, many of the PELDOR experiments are of very poor quality and low signal to noise. Unfortunately a number of less than ideal factors combined to cause issue with the experiment. First the low protein concentrations required to stop protein precipitation meant that no concentrated protein solutions could be made from which the EPR samples are usually made. Without these concentrated samples after mixing of two protein samples and addition of glycerol the protein concentration was further reduced. Any attempt at protein concentration, after purification, spin labelling or buffer exchange into deuterated solvents resulted in huge losses of protein as the protein precipitated against the membrane. A change of membrane material to Hydrosart helped, but concentrations were forced to be kept low and any concentration had to be conducted at a low spinning speed.

Second, the often poor spin labelling efficiencies meant that whilst the protein concentration may be approximately $100\ \mu\text{M}$ the amount of protein molecules that were contributing to the EPR signal were significantly less. In the case of the Im9 samples the C23 positioned labelled with only 5 % efficiency, so even with $100\ \mu\text{M}$ protein the effective sample concentration seen by the spectrometer was $5\ \mu\text{M}$. This problem is amplified when the PELDOR experiment will only show dipolar oscillations from complexes where both proteins are labelled. In the case of the colicin E9 DNase S3C^{IA1} (33.6 %) with Im9 C23^{IA1} (5.3 %) only 1.8 % of the sample will be contributing to the dipolar oscillations whilst the remaining 98.2 % of the sample is unused or worse contributes to the background only hiding the desired dipolar oscillations.

As a result of these factors long experiment times with huge signal averaging were required often before anything could be seen and even then the dipolar oscillations were often only just detectable from the background decay. Through careful analysis and background subtraction of the data, most experiments could yield some information. Far fewer experimental intermolecular distance restraints were obtained than had been hoped for, and fewer still were of the desired quality.

The quality of any modelling is a direct result of the parameters initially given. With such poor quality of data and often ambiguous distance distributions spanning almost $2\ \text{nm}$, the suitability of the experimental data was of constant concern. In proteins of less than $6\ \text{nm}$ in diameter, a distance restraint of range $2\ \text{nm}$ seemed hardly informative. Nevertheless, the only way to test such these constraints was to conduct the docking regardless.

9.4. Docking methodology

HADDOCK has one of the largest user bases of any biomolecular docking programs in the world and has constantly performed well in blind test evaluations [26]. Its design was originally for systems where high ambiguity is common such as in NMR experimentation, but it was also built with inherent flexibility to accept data from various sources both experimental and computational.

In order for HADDOCK to perform docking with EPR restraints, two approaches were required. First, in the PELDOR experiments conducted in this investigation the dipolar coupling measured is a vector between paramagnetic electron spins. These paramagnetic sources were introduced through the use of spin labels which do not occur naturally in the protein. Instead, the spins are located upon the nitroxide of a spin label that has been covalently attached to a cysteine residue.

Here two approaches could be used. Using a detailed rotamer analysis of each of the spin label sites an attempt could be made to convert the vector between electron spins on the N-O bond of the nitroxide into a distance between C_{α} s on the protein backbone. This long range backbone distance could then be introduced into HADDOCK in much the same way as any other distance restraint. The disadvantage of this method is that careful analysis of the rotamer ensemble requires a lot of processing power and detailed information of the three dimensional environment about the spin label site to accurately calculate the C_{α} - C_{α} distance. This calculation without huge amounts of supercomputer time would also likely be limited to considering the area within a narrow area around the spin label and consider all atoms except the spin label in that space as fixed points in space. Also there would be no way of telling whether the C_{α} - C_{α} distance would remain valid as HADDOCK allows backbone flexibility in later docking steps.

In the second approach, MMM is used to generate a rotamer cloud at the spin label site using knowledge of the entire protein (still as a rigid body), with some calculation as to which rotamers will likely be the most populated. The resulting MMM rotamer file could be taken with the most populated rotamers being merged with the protein coordinate file to generate a protein coordinate files where the spin label present in one of the many rotamer positions. These merged files are exported in a compatible HADDOCK format, and using HADDOCK's inbuilt ensemble functionality each rotamer could be loaded as a model of an ensemble. However, HADDOCK would require additional parameters such that new spin-labelled cysteine residues would be accepted as allowed amino acid residues. These parameters would require complete definition of every atom, bond length, bond angle, bond rotation and charge for each spin labelled residue. With HADDOCK now able to recognise and work with spin labelled residues, the PELDOR derived vector could be introduced as a distance between atoms upon the spin label. Combination of the initial ensemble of rotamers for the rigid body docking with nitroxide sidechain characteristics as well as inter-nitroxide distances would allow HADDOCK to optimise the distance restraint at every stage of docking. All that would remain is to convert the distance distributions from EPR experiments into distance restraints compatible with HADDOCK.

It was decided that the second approach would be computationally less demanding as well as give more representative docking. The first approach requires simulation of complex parameters to give a distance unrelated to that seen experimentally. These calculations would then be fed into a second calculation where the rigidity and isolation of atoms in the first calculations would be removed and flexibility allowed to the whole system in the second calculation process. In the second approach, a set of likely protein structures are given - which are nothing more than slightly edited original protein coordinate files, as well as a distance restraint between atoms reflecting an experimental observation. The docking calculation then proceeds allowing rotamer rotation, backbone flexibility and conformational rearrangement whilst maintaining the experimentally observed distance restraint.

9. Discussion and conclusions

Graphical user interfaces were generated for both the spin labelled protein generation and the conversion of PELDOR derived distance distributions such that future compatibility with users with little experience with coding could use this methodology.

With HADDOCK suitably prepared for spin labelled proteins and PELDOR derived distance restraints docking could be conducted. The investigation showed the various resulting docking solutions when various degrees of restraint were applied. From absolutely no restraints, to ambiguous restraints from NMR titrations, to fully restrained models using multiple PELDOR derived distances.

With increasing restraints, it was shown that the sample space of each docking run reduced to an extent where the final 200 structures after energy minimisation collapsed to a single solution with an RMSD of less than 1 Å. The final docking model, using only EPR restraints, showed a different docking site upon the surface of the colicin E9 DNase than in the published NMR and X-ray crystallography structures. The site of Im9 binding was still very close to the DNase activity region of the colicin E9 DNase but distinctly different. With the inclusion of NMR AIRs and EPR restraints the previous site was still by far preferential, in terms of total energy and docking score, but the published docking site could be seen in the structures of other clusters.

Despite the large experimentally derived dissociation constant of the colicin E9 DNase / Im9 complex ($9.3 \times 10^{-17} M$), the docking procedure here reports time and time again a better scoring site upon the opposite surface of the colicin E9 DNase than published NMR and X-ray crystallography structures. In fact this site came out scoring well in the very first HADDOCK run when no restraints were used at all. This is not to say that the docking process here is wrong, as with all modelling, the resulting model is only as good as the data that is passed into it. In this case the starting material was a rigid, folded structure of the complexed colicin E9 DNase that had been separated from Im9 computationally *in vacuo*.

Colicin E9 DNase is a protein that has been shown to be intrinsically disordered in solution [43]. This disorder is required for the colicin E9 to cross through the small OmpF pores in the target cell's cell membrane, a pore so small that aromatic amino acids are almost too large to pass through [22]. It is naive to think that an intrinsically disordered protein would spontaneously fold and form a rigid tertiary structure, before complexing with a binding partner, especially given the cytotoxic nature of the protein in question. It is more likely then that a small region of a few amino acids, likely within the flexible region of residues that contain the key DNase activity residues of the H-N-H motif, first bind the Im9 with the remaining colicin E9 DNase structure forming about the contact site. The associated protein folding and drop in enthalpy likely creates an artificially high dissociation constant. This is confirmed to some extent in the literature where despite the avidity of the interaction the binding of colicin E9 DNase has only a marginal effect on Im9 [40]. If it is the case that the colicin E9 DNase folds around Im9 then the fact that when NMR AIRs are maintained in the docking calculations the published structure is still found is impressive.

Of course, in this investigation we are also considering the DNase domain of colicin E9 to be representative of the domain within the full length protein. When the X-ray crystallography structure of the colicin E9 homology model, colicin E3 bound to Im2 is considered (shown in figure 9.1) the immunity protein can clearly be seen it site inside a pocket of the colicin E3. The translocation domain (black) appears to largely be unaffected by the presense of the immunity protein and remains globular. The cytotoxic domain (grey) has a long flexible

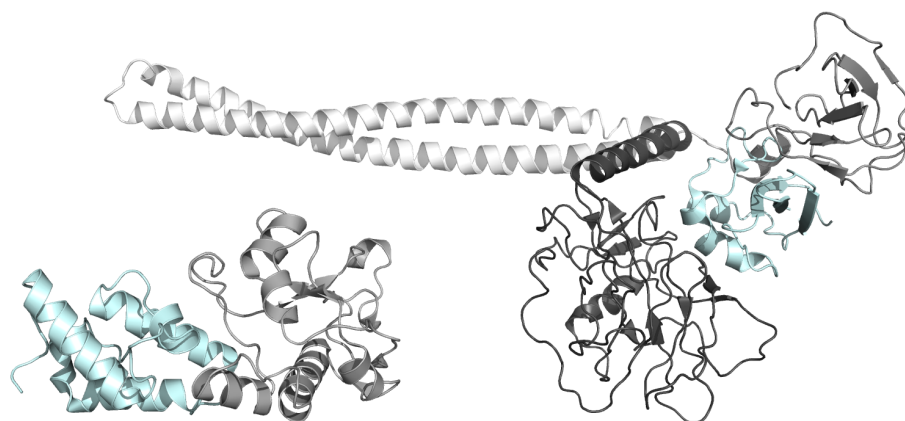


Figure 9.1 Comparison of colicin E9 DNase to the homology model colicin E3. On the left is the NMR complex structure of colicin E9 DNase (in grey) bound to the cognate immunity protein Im9 (in cyan), shown in the same orientation as earlier chapters (PDB 2K5X). On the right is the X-ray crystallography structure of the full length colicin E3, used as a homology model for the full length colicin E9, bound to the non-cognate immunity protein Im2 (PDB 1JCH). The colicin E3, cytotoxic (ribonuclease) (C) domain is shown in grey (matching the colicin E9 DNase), the receptor (R) domain is shown in white and the translocation (T) domain in black. Im2 is shown in cyan. Both proteins are scaled equally.

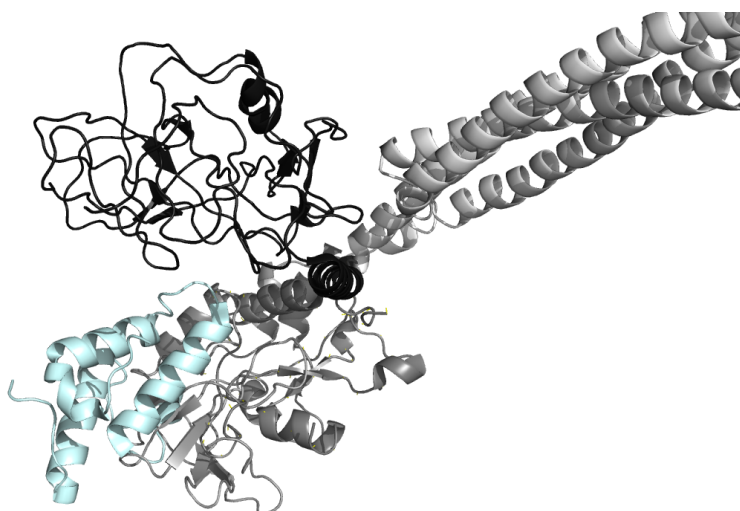


Figure 9.2 Colicin E9 DNase and Im9 complex structure from NMR (PDB 2K5X) has been superimposed onto the X-ray crystallography derived homology model colicin E3 structure (PDB 1JCH). Here the colicin E3 homodimer bound to the non-cognate immunity protein Im2, has had the colicin E3 ribonuclease domain and Im2 hidden. The colicin E9 DNase domain has been aligned using a RMSD minimisation routine to an average C_{α} distance of 0.56 Å. The NMR complex structure of colicin E9 DNase (in grey) bound to the cognate immunity protein Im9 (in cyan), is shown in approximately the same orientation as all previous docking figures. Colicin E3, bound to the non-cognate immunity protein Im2 has the receptor (R) domain shown in white and the translocation (T) domain in black.

9. Discussion and conclusions

linker to the receptor domain (white) and wraps around the immunity protein. Alignment between these models is difficult, and visually messy, due to the smaller size and sequence diversity in the domains of colicin E3 ribonuclease domain compared to the colicin E9 DNase and Im2 compared to Im9.

When the colicin E9 DNase domain is superimposed upon the colicin E3 structure, as shown in figure 9.2, the docking solutions become more obvious. In this figure the colicin E3 bound to Im2 has been used as a reference structure but the colicin E3 ribonuclease domain and Im2 have been removed. The colicin E9 DNase has been aligned using a RMSD minimisation procedure to place the colicin E9 DNase domain in the position of the colicin E3 ribonuclease domain. This model is not flawless, as there is now a small pocket above the Im9 which would likely cause the colicin E9 DNase domain and Im9 to move up and closer to the colicin E3 receptor domain (black). The figure is displayed in the same orientation as the reference structure used to display the protein-protein docking chapter results. In this orientation, it is clear that the docking results that place the Im9 to the right and slightly behind the colicin E9 DNase domain would not exist without steric clashes to the receptor domain of the colicin. Equally the docking results that place the Im9 on top of the colicin E9 DNase could not exist in the full length version without steric clashes to the translocation domain of the colicin. It is not unreasonable to think that these surfaces have residues which are designed to interact with the other colicin domains and maintain the colicin tertiary structure. It is likely then that these residues are sampled during the docking process and during energy calculations behave favourable when not exposed to water.

Somewhat peculiarly, maintaining NMR AIRs gave less certainty in the final model than an EPR constraint only docking calculation. From the HSQC experiment, a surprising number of residues showed significant perturbation resulting in an extensive AIRs list, far more than would be expected for proteins of this size. The decision to take residues beyond one standard deviation from the mean is an arbitrary one. By using one standard deviation results in 32 % of the molecule being considered as active to binding before passive residues are even considered. In the case of Im9 28 residues are defined as having greater than one standard deviation from average CSP, in such a small molecule a large number of these will be solvent accessible defining them as active residues to binding. Referring back to figure 8.1 (page 141) it can be seen that Im9 has active residues across most of the protein. Colicin E9 DNase shows active residues much more to one side of the protein. If the threshold was raised, to 2 standard deviations for example, only the key residues would be used in formation of the AIRs, making the ambiguous interaction restraints much less ambiguous. With less ambiguity the docking results may lose some of the lower scoring structures away from the main scoring clusters.

9.5. Outlook

This investigation has conclusively shown that EPR derived distance restraints can be used to restrain docking models and presents a practical user interface for the use of the nitroxide spin labels MTSL and IA-PROXYL with the docking program HADDOCK.

From a programming perspective MISHAP could easily be expanded to facilitate other nitroxide spin labels such as TOAC, TEMPOL and PYC derivatives, common paramagnetic

metal cofactors such as Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Cu^{2+} , paramagnetic structures such as $Fe - S$ clusters or even newly developed nucleotide labels for measurement of distances in DNA or RNA binding proteins. Greater integration with MMM has also been designed but as yet not implemented, such that users could send structures created in MMM directly to MISHAP for exporting to HADDOCK without having to separately save the PDB and rotamer files.

From an experimental view, having more spin labelling sites across the complex and having each site label efficiently would allow for a more accurate mapping of docking to be built. With increased labelling efficiency, would come PELDOR traces with far better signal to noise ratios with clear dipolar oscillations. These higher quality PELDOR traces would have far greater certainty and likely yield better resolved distance distributions and in turn better docking restraints. With more labelling positions it would allow a detailed investigation into the exact number of intermolecular distances and the ideal locations that are required to establish a well resolved docking model. Careful selection of the sites could have also allowed a comparison between solvent accessibility (and therefore label rigidity) and distance distribution broadening; in the current investigation only two environments, that of fully solvent exposed and solvent hidden, were available and with the labelling efficiencies as they were little could be distinguished.

Biologically, the project still has much scope for expansion. Collaborators at the University of Oxford have DNA constructs for full length colicin E9 and are routinely generating mutants for X-ray crystallography and mass spectrometry investigations. A small amount of protein was provided in the course of this investigation corresponding to full length equivalents of colicin E9 DNase S3C, S30C and S80C. Unfortunately, the samples provided proved to yield little more than cw-EPR spectra, as once again the protein concentration and labelling efficiencies inhibited any PELDOR experiments. However, a PELDOR investigation into the full length colicin E9 would yield useful information. Currently, it is accepted that the DNase domain is an accurate model of the full length colicin E9 in terms of interaction and binding to immunity proteins, however as no full length colicin E9 has yet been made publicly available, it cannot be certain.

As the proteins used in this investigation are small with only a single naturally occurring cysteine upon the Im9 the spin labelling could be greatly improved by moving to a cell-free protein expression system. In these systems the desired gene is placed into a reaction vessel of cell extract that contains the energy source, cofactors and organelles required to generate simple proteins. As each amino acid is placed individually into the reaction vessel the spin label could be added to free cysteine before being added. With careful preparation, it could be that all cysteines added to the reaction vessel are already spin labelled. This would ensure that all protein synthesised has a 100 % labelling efficiency. The limitations of this method are the initial cost of the required equipment and the relatively small batches of protein that are produced. However, the typical protein production yields are enough for an EPR sample and a typical cell-free expression system takes one day to produce the protein, requiring only a very simple one step purification process.

The initial hope for this investigation was to establish a protocol for docking proteins on an established system where the outcome was known by both X-ray crystallography and NMR. With an established and tested protocol then the methodology could be extended to non-cognate colicin-immunity protein interactions. Im2, Im7 and Im8 are known to bind

9. Discussion and conclusions

and inhibit colicin E9's biological activity, just with a much reduced dissociation constant. Likewise, Im9 can give cross-protection to other E-group colicins. By generating a few cysteine mutants for each E-group colicin and immunity protein the structures of tens of non-cognate complexes could be solved using only a few weeks of PELDOR experiments and existing NMR structures.

Part III.

Methods and materials

10. Molecular biology

10.1. General techniques

10.1.1. SDS-PAGE

A sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) is made of two parts; a running and a stacking gel. The running gel was made by mixing 15 % acrylamide (7 ml) with 10 % ammonium persulphate (AMPS) (700 μ l) and N,N,N',N'-tetramethylene diamine (TEMED) (70 μ l), which was mixed then poured between BioRad glass plates and allowed to polymerise and set. This gel was topped with ethanol to ensure a flat surface. The stacking gel was prepared by mixing 5 % acrylamide (5 ml) with 10 % AMPS (50 μ l) and TEMED (3 μ l). The ethanol on top of the running gel was removed once the gel had set and the stacking gel was then poured on top. A Teflon comb was added to the stacking gel and allowed to set. The BioRad glass plates and gel within were moved into the running tank and filled with running buffer (12.1 g Tris base, 17.9 g Tricine, 10 ml of 10 % SDS, ddH₂O to a total volume of 1 l). Upon removal of the Teflon comb from the gel wells were produced, approximately 15 μ l in size, giving the origin of the lanes. Protein samples (10 μ l) were combined with 2 \times SDS loading buffer (10 μ l) (100 mM Tris-HCl, pH 6.8, 20 % glycerol, 4 % SDS, 0.2 % bromophenol blue, 200 mM dithiothreitol) and the protein denatured by heating on a mantle (95 °C, 5 min). Samples (10 μ l) were then loaded into the wells of the stacking gel, ensuring that one well was left empty and later filled with low molecular weight markers (10 μ l) (Sigma-Aldrich, 66, 45, 36, 29, 24, 20, 14.2 kDa). Electrophoresis was run at 200 V for approximately 1 hour until the visible bands had progressed to the bottom of the gel. The gel was stopped and removed from the BioRad glass, before being stained in Coomassie Brilliant Blue (1 hour), before being placed in destainer (25% methanol, 10% glacial acetic acid). When sufficiently destained the gel was visualised using a standard PC scanner.

10.1.2. UV/Vis Spectroscopy

Optical densities of 1 ml protein samples were measured at 595 nm using a PerkinElmer Lambda Bio+ photometer. When used in conjunction with a LB broth blank, taken from the reading, concentration was calculated using the Beer Lambert Law:

$$\text{Optical density} = \epsilon c l \quad (10.1)$$

Where ϵ is the extinction coefficient, c is the concentration and l is the path length of the sample.

10.1.3. Dialysis

Dialysis is conducted to gradually lower the salt concentration of a solution, thereby desalting protein solutions. In Gdm solutions this results in refolding the protein back to native conformations. Dialysis was conducted by placing the protein solution elutant from the affinity chromatography inside Thermo Scientific SnakeSkin Pleated Dialysis Tubing (3500 *Da* molecular weight cut off) (catalogue number 68035) (1 *cm* tube for approximately every 3.5 *ml*), and sealing the ends before adding to dialysis buffer (5 *l*) (5 *mM* EDTA). The dialysis buffer is advantageous as the EDTA also acts as a metal chelator removing any excess metals that could shorten relaxation times in EPR experiments. The buffer solution was left with gentle stirring overnight to equilibrate; any sediment that formed overnight was discarded.

10.2. DNA

10.2.1. General DNA techniques

10.2.1.1. Agar plates

Agar plates were created by taking LB broth (400 *ml*, 25 *g/l* (10 *g/l* tryptone, 5 *g/l* micro granulated yeast extract, 10 *g/l* sodium chloride) and combining with agar (15 *g/l* (6 *g*)) before autoclaving. Whilst the solution was still warm (~ 50-60 °C) ampicillin (100 $\mu\text{l/ml}$) was added and the solution was poured onto a plate under sterile conditions and allowed to set. Once set the plates were stored upside down to minimize contamination.

10.2.1.2. Transformation of DNA plasmid into *E. coli*

On ice, 10 μl of *Escherichia coli* BL21 (DE3) cells were added to 1 μl of plasmid DNA. The solution was mixed in an eppendorf with a flick and left on ice for 30 *min*. Samples were then heatshocked (45 *seconds*, 42 °C) before being returned to ice (2 *min*). Samples were made up to 1 *ml* with LB broth (25 *g/l*) and incubated (1 *h*, 37 °C) before being plated (50 $\mu\text{l/plate}$). Plates were incubated overnight (37 °C) and a single, isolated colony selected to proceed to large scale expression.

10.2.1.3. Glycerol stocks

A single colony was picked from §10.2.1.2 and was added to LB broth (5 *ml*, 25 *g/l*) containing ampicillin (50 $\mu\text{l/ml}$) and incubated overnight (180 *rpm*, 37 °C). 1.5 *ml* of the suspension was centrifuged and the supernatant discarded. The cell pellet was resuspended in 50:50 LB media:glycerol to produce a working glycerol stock if later required.

10.2.1.4. DNA sequencing

A single colony was picked from §10.2.1.2 and was added to LB broth (5 *ml*, 25 *g/l*) containing ampicillin (50 $\mu\text{l/ml}$) and incubated overnight (180 *rpm*, 37 °C). 1.5 *ml* of the suspension was centrifuged and the supernatant discarded. The cell pellet was resuspended

in GTE buffer (100 μ l, (50 mM Glucose, 25 mM Tris-Cl, 10 mM EDTA, pH 8)), vortexing if necessary. Once resuspended lysis solution was added (200 μ l, (0.2 M NaOH, 1 % SDS)) and the tube inverted 6 - 8 times to ensure good mixing. Potassium acetate solution (150 μ l, 5 M, pH 4.8) was immediately added to neutralise the NaOH whilst the DNA precipitates. The solution was spun (10000 $\times g$, 1 min) and the supernatant transferred to a new tube. The nucleic acids were precipitated with isopropanol (0.5 ml, on ice, 10 mins) before being spun again (10000 $\times g$, 1 min). The supernatant was discarded and the pellet dissolved in TE buffer (0.4 ml, (10 mM Tris-Cl, 1 mM EDTA, pH 7.5)) to which RNase A solution (10 μ l, 20 mg/ml) was added and incubated (30 min, 37 °C) to remove any remaining RNA. Proteins were then extracted from the plasmid using PCIA (0.3 ml, (equal parts phenol/chloroform/isoamyl alcohol)), the solution was quickly vortexed and then spun (10000 $\times g$, 5 mins). The upper aqueous layer was removed from the organic PCIA layer at the bottom, carefully avoiding the white precipitated protein. The remaining solution was added to ammonium acetate solution (100 ml, 7.5 M) and absolute ethanol (1 ml) to precipitate the plasmid DNA (on ice, 10 min). The solution was spun again (10000 $\times g$, 5 mins) and the DNA pellet resuspended in pure water (to a total volume \sim 15 μ l). 5 μ l of DNA solution was taken and made up to 1 ml and the optical density recorded at 260 nm (measured as described in §10.1.2) giving a DNA concentration.

The samples were then sent for DNA sequencing at The Genome Analysis Centre, Norwich, UK¹. The results of which were analysed using a combination of FinchTV², T-COFFEE³, EXPASY⁴ and UniPROT⁵.

10.3. Plasmids

All protein samples were produced from plasmid DNA using similar growing and purification procedures (§10.4). Plasmids were kindly provided by N. Cull, University of East Anglia, UK or Dr. M. Vankemmelbeke, University of Nottingham, UK.

Naturally derived proteins were available from several sources, but not all were viable. For example Colicin E9 DNase S3C was available from plasmid pTG2 (pET21d based) (UEA) or pMV45 (Nottingham); the Nottingham plasmid being viable. Colicin E9 DNase S30C available from plasmid pTG5 (pET21d based) (UEA) or pMV46 (Nottingham); neither was viable. Colicin E9 DNase S80C from plasmid pTG6 (pET21d based) (UEA); not viable. Colicin Immunity Protein Im9 from plasmid pRJ345 (pTrc99a based) or pSE10 (both UEA); only pRJ345 viable.

In addition to the naturally derived mutants that were available through discussion with collaborators at the Universities of East Anglia, Nottingham and York more were designed and created using a standard procedure.

The desired protein sequence was reverse translated into a DNA sequence that was optimised for codon expression in *E. coli*. To the 5' end a NcoI restriction site was added (CC)

¹<http://www.tgac.bbsrc.ac.uk>

²<http://www.geospiza.com/finchtv/>

³<http://www.tcoffee.org>

⁴<http://expasy.org/tools/blast/>

⁵<http://www.uniprot.org/>

Protein, mutant	Origin	Plasmid name	Base plasmid	Cell line
E9 DNase S3C	Natural, Nottingham	pMV45	pET21d	BL21 (DE3)
E9 DNase S30C	Synthetic	pMB4	pEX-A	BL21 (DE3)
E9 DNase S80C	Synthetic	pMB5	pEX-A	BL21 (DE3)
E9 DNase S3C / S80C	Synthetic	pMB1	pET21d	BL21 (DE3)
E9 DNase S30C / S80C	Synthetic	pMB2	pET21d	BL21 (DE3)
E9 DNase E66C	Natural, UEA	pAK4	pTrc99a	BL21 (DE3)
E9 DNase D20C / E66C	Natural, UEA	pAK5	pTrc99a	BL21 (DE3)
Im9 C23	Natural, Nottingham	pRJ345	pTrc99a	BL21 (DE3)
Im9 C23S / S6C	Natural, Nottingham	pMV16	pTrc99a	BL21 (DE3)
Im9 S6C	Synthetic	pMB3	pET21d	BL21 (DE3)
Im2	Natural, UEA	pCD01	pTrc99a	JM109
Im7	Natural, UEA	pRJ347	pTrc99a	JM109
Im8	Natural, UEA	pKL340	pTrc99a	JM109

Table 10.1. Listings of all proteins, their derived plasmids and plasmid sources

and a XhoI (CTCGAG) to the 3' end. This sequence was then checked using the software package - GENEius⁶ - to avoid repeats or hairpin structures and to ensure that guanines and cytosines are equally distributed throughout the sequence; thus ensuring DNA stability and longevity.

The sequence was then passed to Eurofins MWG GmbH, Ebersberg, Germany who artificially synthesised the gene using overlapping oligonucleotides, employing ligase chain reaction (LCR) and polymerase chain reaction (PCR) technologies. The gene was then subcloned into a standard commercially available pET21-d vector (Merck KGaA, Darmstadt, Germany), before being mini-prepped to a suitable concentration. The gene was finally sequenced (in both directions) to ensure quality of gene and surrounding restriction map.

Once the plasmid was received from Eurofins it was transformed into viable *E. coli* BL21 (DE3) cells (as per §10.2.1.2) grown up and split, half forming a glycerol stock with the other half being miniprepmed to reform a plasmid stock.

The proteins (as a result of the expression vector) were all His₆-tagged, which allowed for a simple purification procedure. Raw, sonicated cell lysate could be passed over a nickel affinity column, which removed most of the impurities, with a secondary size exclusion column to give maximum purity.

10.3.1. Synthetically derived plasmids

10.3.1.1. Colicin E9 DNase S3C / S80C / H127A - pMB1

The double Colicin E9 DNase mutants could not be produced in isolation as they would kill the host cell, instead they are usually co-expressed with the immunity protein Im9. However, the start of the immunity protein Im9 gene overlaps the end of the colicin E9 gene in a different reading frame [19]. This proved very difficult to try to create a viable colicin E9 mutant whilst keeping immunity protein Im9 functional.

As a result of previous work [34] a further point mutation was introduced. The colicin E9 DNase H127A mutant has been shown to still bind DNA and the immunity protein Im9 but

⁶GENEius - for optimising synthetic genes [online], Available: <http://www.geneius.de/> (Accessed 3 June 2011)

shows no DNase action. With this mutation it is possible to overexpress the colicin E9 DNase in *E. coli* without killing the host cell. The His₆ tag and stop codons are included in the pET21-d vector and were ensured to be in the same reading frame as the insert.

Amino acid sequence

MEC KRN KPG KAT GKG KPV GDK WLD DAG KDS GAP IPD RIA DKL RDK EFK SFD DFR KAV WEE VSK DPE
LSK NLN PSN KSS VCK GYS PFT PKN QQV GGR KVY ELH HDK PIS QGG EVY DMD NIR VTT PKR AID IHR GK

DNA sequence (with restriction site tags)

cc ATG GAA TGC AAA CGC AAC AAA CCG GGC AAA GCT ACC GGG AAA GGC AAA CCA GTT GGC GAT AAA
TGG CTG GAT GAT GCC GGC AAA GAT TCA GGT GCA CCA ATT CCT GAT CGC ATT GCC GAC AAG TTA CGC
GAC AAA GAG TTC AAG TCG TTT GAC GAT TTC CGG AAA GCG GTT TGG GAA GAA GTG TCC AAA GAT CCC
GAA TTG AGC AAG AAT CTG AAT CCG AGC AAC AAA AGC TCT GTC TGT AAG GGG TAT TCG CCG TTT ACG
CCG AAG AAT CAG CAA GTA GGA GGT CGC AAA GTG TAT GAG CTG CAT CAC GAC AAA CCG ATC AGT CAG
GGT GGT GAA GTG TAC GAT ATG GAT AAC ATC CGT GTC ACT ACC CCT AAA CGT GCG ATT GAC ATC CAT
CGT GGC AAA ctcgag

10.3.1.2. Colicin E9 DNase S30C / S80C / H127A - pMB2

This mutant was also designed with the H127A mutation to stop DNase activity.

Amino acid sequence

MES KRN KPG KAT GKG KPV GDK WLD DAG KDC GAP IPD RIA DKL RDK EFK SFD DFR KAV WEE VSK DPE
LSK NLN PSN KSS VCK GYS PFT PKN QQV GGR KVY ELH HDK PIS QGG EVY DMD NIR VTT PKR AID IHR GK

DNA sequence (with restriction site tags)

cc ATG GAA TCC AAA CGG AAC AAA CCA GGC AAA GCA ACC GGT AAG GGC AAA CCT GTT GGG GAT AAA
TGG CTG GAT GAT GCG GGC AAA GAC TGT GGT GCT CCA ATT CCC GAT CGT ATT GCG GAC AAA TTG CGC
GAT AAA GAG TTC AAA TCG TTC GAT GAT TTT CGC AAA GCC GTT TGG GAA GAA GTC AGC AAA GAT CCG
GAG TTA AGC AAG AAC CTG AAT CCG TCA AAC AAG AGC TCT GTG TGC AAA GGC TAT TCG CCT TTT ACC
CCG AAG AAT CAG CAA GTG GGA GGT CGT AAG GTC TAT GAA CTG CAT CAC GAC AAA CCG ATT AGT CAG
GGT GGC GAA GTG TAC GAC ATG GAC AAT ATC CGC GTA ACG ACT CCG AAA CGT GCC ATC GAT ATC CAT
CGC GGG AAA ctcgag

10.3.1.3. Colicin Immunity protein Im9 S6C - pMB3

Mireille Vankemmelbeke provided DNA for Im9 S6C / C23S, a single cysteine mutant with a spin label site at a different location. In addition an immunity protein mutant with two spin label sites was designed for improved docking triangulation.

Amino acid sequence

MEL KHC ISD YTE AEF LQL VTT ICN ADT SSE EEL VKL VTH FEE MTE HPS GSD LIY YPK EGD DDS PSG IVN
TVK QWR AAN GKS GFK QG

DNA sequence (with restriction site tags)

cc ATG GAG TTG AAA CAC TGC ATT TCG GAC TAT ACC GAA GCG GAA TTT CTG CAG TTA GTG ACT ACG ATC
TGT AAT GCC GAT ACA TCC AGC GAA GAA GAG CTG GTT AAA CTG GTA ACC CAT TTC GAG GAA ATG ACC
GAA CAT CCG TCT GGA AGC GAT CTC ATC TAC TAT CCG AAA GAA GGT GAC GAT GAT TCG CCA TCA GGC
ATT GTC AAC ACG GTG AAA CAA TGG CGC GCT GCA AAC GGC AAA AGT GGG TTT AAG CAG GGT ctcgag

10.3.1.4. Colicin E9 DNase S30C / H127A - pMB4

Amino acid sequence

10. Molecular biology

MES KRN KPG KAT GKG KPV GDK WLD DAG KDC GAP IPD RIA DKL RDK EFK SFD DFR KAV WEE VSK DPE
LSK NLN PSN KSS VSK GYS PFT PKN QQV GGR KVV ELH HDK PIS QGG EVY DMD NIR VTT PKR AID IHR GK
HHHHHH

DNA sequence (with restriction site tags)

```
cc ATG GAG TCC AAA CGA AAC AAA CCA GGA AAA GCA ACG GGC AAA GGC AAA CCA GTT GGG GAT AAA
TGG CTG GAT GAC GCT GGG AAA GAC TGC GGC GCG CCC ATT CCG GAT CGT ATT GCG GAC AAG CTC CGC
GAT AAA GAA TTT AAA TCG TTC GAT GAT TTT CGT AAG GCC GTA TGG GAA GAA GTG TCG AAA GAT CCT
GAA TTA AGC AAG AAC CTG AAC CCG AGT AAT AAA AGT TCA GTG AGC AAA GGC TAT AGC CCT TTC ACC
CCG AAA AAT CAG CAG GTG GG CGG TCG GAA AGT CTA CGA ATT GCA CCA TGA CAA GCC GAT CTC TCA
AGG TGG TGA GGT CTA TGA CAT GGA TAA TAT TCG TGT TAC TAC CCC GAA GCG CGC CAT CGA TAT CCA
TCG CGG TAA A ctgag
```

10.3.1.5. Colicin E9 DNase S80C / H127A - pMB5

Amino acid sequence

MES KRN KPG KAT GKG KPV GDK WLD DAG KDS GAP IPD RIA DKL RDK EFK SFD DFR KAV WEE VSK DPE
LSK NLN PSN KSS VCK GYS PFT PKN QQV GGR KVV ELH HDK PIS QGG EVY DMD NIR VTT PKR AID IHR GK
HHHHHH

DNA sequence (with restriction site tags)

```
cc ATG GAG TCT AAA CGA AAC AAA CCA GGC AAG GCC ACC GGC AAA GGT AAG CCT GTC GGG GAT AAA
TGG TTA GAC GAC GCT GGT AAG GAC AGC GGA GCC CCC ATC CCA GAT CGC ATC GCG GAT AAG TTG CGC
GAC AAG GAA TTC AAA TCC TTC GAC GAT TTT CGG AAA GCG GTC TGG GAG GAA GTA AGC AAA GAT CCG
GAA CTC AGC AAA AAT CTG AAC CCG TCG AAC AAA TCG AGT GTT TGC AAA GGC TAC TCA CCT TTT ACT
CCG AAA AAT CAA CAG GTT GGT GGG CGC AAA GTG TAT GAA CTG CAT CAT GAT AAA CCG ATC AGT CAG
GGT GGC GAA GTG TAT GAT ATG GAT AAT ATT CGT GTG ACG ACC CCG AAA CGT GCA ATT GAT ATT CAC
CGT GGC AAA ctgag
```

10.4. Protein production

10.4.1. Colicin E9 DNase production

10.4.1.1. Large scale over-expression

E. Coli BL21 (DE3) cells were transformed with the constructed E9 plasmid (§10.2.1.2) and then grown overnight in LB broth (2x 5 ml, 25 g/l) with ampicillin (5 µl, 100 mg/ml) (37 °C, 180 rpm, 5 h). 400 µl was taken from the initial growth of 5 ml and added to 40 ml LB and grown overnight (37 °C, 180 rpm, 16 h) with the addition of ampicillin (40 µl, 100 mg/ml). 20 ml was taken from the overnight culture and added to LB broth (4x 500 ml) mixed with ampicillin (500 µl, 100 mg/ml) and incubated with shaking (37 °C, 180 rpm) until reaching an optical density (less LB) at 595 nm of approximately 0.6 (~ 4 h) (measured as described in §10.1.2), at which the culture is considered to be at the middle of the log growth phase. More ampicillin (500 µl, 100 mg/ml) was added along with isopropyl-β-D-thiogalactoside (IPTG) (5 ml, 100 mM) which activates the *lac* operon and induces gene expression of the desired E9 gene. The culture was further incubated with shaking (37 °C, 180 rpm, 4 h). 1 ml samples were taken from the culture at hourly intervals after induction with IPTG, the OD_{595nm} measured, then analysed by SDS-PAGE to monitor protein expression (§10.1.1). The remaining culture after 6 h was centrifuged (7000 ×g, 15 min), the supernatant was discarded and the pellets, containing cells were resuspended in Tris buffer (10 mM Tris-HCl, minimum required) and stored overnight at -4 °C.

10.4.1.2. Liberation of protein from cells

The resuspended pellets were thawed slowly on ice. The solution was sonicated on ice (2 runs of 500 second length with 50% power, half second bursts every second with a 5 minute on ice rest between runs) using a Homogenisers Status 200 sonicator. This process destroyed the cell walls releasing the protein into solution. The solution was centrifuged ($45,000 \times g$, 45 min, 4 °C) the supernatant containing soluble protein was removed and placed on ice, the pellet containing cell remnants was discarded. A sample was taken from the supernatant for SDS-PAGE analysis and comparison with various purification stages.

10.4.1.3. Immobilised metal ion affinity chromatography

GE Healthcare Life Sciences HisTrap HP columns with bed volume 5 ml were washed with 10 column volume equivalents (CV) (ie 50 ml) water, then activated with charge buffer (5 CV, 50 μM Ni₂SO₄), then rewashed with water (5 CV). The column was then run into the binding buffer (5 CV, 20 mM Tris, 500 mM NaCl, 50 mM imidazole). The protein was directly loaded onto the column in one movement. The column and bound protein were further washed with binding buffer (10 CV) to remove any none specific binding. Elution was conducted by titrating in elution buffer elution buffer to the column (4 CV, 20 mM Tris, 500 mM NaCl, 50 mM imidazole, 6 M GdmCl). The GdmCl was required to unfold the colicin E9 and unbind it from the Im9, as it is one of the few ways to separate the complex; the Im9 with its His₆ tag remained bound to the column. The column was finally stripped of Im9 using strip buffer (4 CV, 20 mM Tris, 500 mM NaCl, 100 mM EDTA) before being stored in 20% ethanol at 4 °C. Chromatography was conducted using a GE Healthcare Life Sciences Äkta Explorer FPLC system.

10.4.1.4. Dialysis

Dialysis was conducted overnight as per §10.1.3 to lower and remove Gdm from the solution, allowing the colicin to refold.

10.4.1.5. Lyophilisation

The protein solution was placed inside a large round bottom flask; this was then placed in a carbon dioxide (s) / ethanol (l) bath and flash frozen with rotation to ensure the maximum possible surface area. The top of the flask was then sealed with parafilm and electrical tape before being pierced with a surgical needle. The flask was then placed in a Thermo Sovant MicroModuly vacuum desiccator, brought to vacuum and left > 48 h.

The protein was then resuspended in gel filtration buffer (2.5 ml, 50 μM potassium phosphate (4.21 g/l KH₂PO₄, 3.32 g/l K₂HPO₂), potassium chloride (200 μM), pH 7.00).

10.4.1.6. Size exclusion chromatography

Size exclusion chromatography was conducted using a self poured 26/60 column consisting Amersham Pharmacia Biotech glassware, connectors and 350 ml of GE Health Life Sci-

ences Superdex 200 connected to a GE Health Life Sciences Äkta Explorer FPLC system. The column was equilibrated by running through gel filtration buffer (300 ml, (50 μ M potassium phosphate (4.21 g/l KH_2PO_4 , 3.32 g/l K_2HPO_2), potassium chloride (200 μ M), pH 7.00)). The concentrated protein (less than 2.5 ml) was injected onto the column and the progress of protein elution was monitored by measuring the UV absorption at 280 nm. Relevant fractions with UV absorption had samples (10 μ l) taken for SDS-PAGE analysis (§10.1.1). Relevant fractions, as determined by SDS-PAGE were pooled and taken forward.

10.4.1.7. Purification and concentration

The protein was twice dialysed to pH adjusted water to remove any salt (§10.1.3). The final volume (\sim 50 ml) was reduced to usable concentrated sample stock volume of approximately 3 ml by a combination of Amicon filtering and lyophilization (§10.4.1.5). Filtering was conducted using a Millipore Amicon Stirred Cell (model 8050), using a Y3K micro-filter (3 kDa molecular weight cut off). The filter was loaded into the cell and stirrer added, the volume to be concentrated was added. The cell was closed and sealed before being pressurised under argon (3 atm) with gentle stirring until final volume was \sim 5 ml. The flow through was discarded.

10.4.1.8. Protein concentration determination

Protein concentration was determined using a modified method as described in §10.1.2 (page 195).

$$\text{Optical density} = \epsilon c l$$

The molar extinction efficient, ϵ in equation 10.1 (repeated above) is unique to every macromolecule at differing wavelengths and is characteristic of concentration. It was found from extensive studies that at 280 nm the molar extinction coefficient comes almost exclusively from aromatic amino acids [36, 71]. Thus from knowing the sequence the extinction coefficient can be found according to:

$$\epsilon_{280\text{ nm}} M^{-1} \text{cm}^{-1} = x.Trp(5500) + y.Tyr(1490) + z.Cys(125) \quad (10.2)$$

Giving wild type colicin E9 DNase a molar extinction coefficient of 15470 $M^{-1} \text{cm}^{-1}$ (cytosolic), single mutant 15595 $M^{-1} \text{cm}^{-1}$ (cytosolic) and double mutant 15620 $M^{-1} \text{cm}^{-1}$ (cytosolic). Single cysteine Im9, 9970 $M^{-1} \text{cm}^{-1}$ (cytosolic) and double cysteine Im9, 10095 $M^{-1} \text{cm}^{-1}$ (cytosolic)

10.4.1.9. Protein storage

After the protein concentration was determined the protein solution was aliquoted into 1 ml amounts in 1.5 ml eppendorfs. These were flash frozen in liquid nitrogen and stored in a -80 °C freezer.

10.4.2. Colicin E9 production

Full length Colicin E9 mutants S451C, S478C and S528C (corresponding to E9 DNase domain S3C, S30C and S80C) were produced and purified by Dr. N. Housden (University of Oxford, UK) from his own gene construct.

10.4.3. Im9 production

10.4.3.1. Large scale over-expression

Escherichia Coli BL21 (DE3) cells were transformed with the desired Im9 plasmid (§10.2.1.2) and grown up as described for E9 (§10.4.1.1).

10.4.3.2. Liberation and purification

Liberation of Im9 was conducted by sonication, as described in §10.4.1.2. IMAC and size exclusion chromatography were conducted as described in §10.4.1.3 and 10.4.1.6, except that the IMAC did not require the GdmCl wash. Instead the protein, without unfolding, could be liberated as the imidazole concentration was increased.

10.4.3.3. Purification, concentration and storage

The protein was desalted by dialysis (§10.1.3) and concentrated by lyophilisation (§10.4.1.5). Protein concentration was determined using the method outlined in §10.4.1.8 concluding in Im9 having a molar extinction coefficient of $9970 \text{ M}^{-1}\text{cm}^{-1}$ (cytosolic). Again samples were flash frozen for long term storage, as described in §10.4.1.9.

10.5. Spin labelling

10.5.1. MTSL

Inside of a low oxygen glovebox ($\sim 1 \text{ ppm O}_2$) 13.5 mg of protein was dissolved into 2.0 ml of degassed potassium phosphate buffer (50 mM (4.21 g/l KH_2PO_4 , 3.32 g/l K_2HPO_2) pH 7.00). The protein solution was added to dithiothreitol (DTT) solution (1 mg in 2.5 ml), a reducing agent to ensure that no formation of disulphide bridges. The combined solution was passed down an GE Healthcare PD-10 column per manufactures instructions, to remove any excess DTT.

A five fold excess of S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulphonothioate (MTSL) (50 mM in acetonitrile) was added to the elutant of the PD-10 column (90.4 μl E9 mutants, 134.1 μl Im9). The solution was covered in aluminium foil (as the reaction is light inhibited) and left for 3 h at room temperature (298 K) with stirring every 30 min.

After the reaction time, the solution could be removed from the glovebox. The protein was desalted and excess label removed using 2 sequential PD-10 columns (equilibrated as before in potassium phosphate buffer).

The protein concentration was then evaluated using UV/Vis (as described in §10.1.2) using a 10 μ l sample being diluted to 1 ml in potassium phosphate.

10.5.2. IA-PROXYL

10.0 mg of protein was dissolved into chelexed and autoclaved metal free buffer (1 ml, 50 mM (4.21 g/l KH_2PO_4 , 3.32 g/l K_2HPO_4), 2 mM EDTA, pH 7.00), to which a 5 \times molar excess of IA-PROXYL was added (50 mM in DMSO).

Colicin E9 DNase D20C/E66C and E66C due to the labelling sites required the addition of 6 M GdmCl to the buffer to unfold the protein, and due to the inhibiting effect of GdmCl to labelling a 10 \times molar excess of IA-PROXYL was added.

Solutions were wrapped in foil, to stop any light inhibition and left overnight at 4 °C.

The samples were made up to 15 ml with further buffer and added to a VivaSpin 15R spin concentrator and spun (3000 \times g) until the volume reached 1 ml (~ 2 h). The samples were again made up to 15 ml and spun again until the volume reached 1 ml.

Samples were taken before spinning as well as after spinning from the retained volume and flow through. Each sample was evaluated by UV/Vis (as per §10.1.2) and cw-EPR (X-band), to ensure that free label was removed from the sample and to evaluate labelling efficiency to the protein.

11. EPR and other spectroscopies

11.1. cw-EPR

11.1.1. X-band

Unless otherwise noted all X-band experiments (~ 9.5 GHz) were conducted upon a Bruker ELEXSYS E500 or Bruker ELEXSYS E680 instrument.

11.1.1.1. Room temperature continuous wave

Experiments were conducted using a Bruker ELEXSYS E500 running Bruker software Xepr v2.6; with an X-band Bruker ER 041 MR microwave bridge, Bruker ER 4123D cw-Resonator (with the cavity being flushed with nitrogen gas, except in aerobic power saturation experiments), Bruker B-E251B9 magnet, Bruker ER 082 C magnet power supply, Marconi Instruments 2440 Microwave Counter. Experiments were conducted in an environmentally controlled room, where the temperature was maintained at 298 ± 1 K.

Samples were initially tested at ~ 9.7 GHz, 0.63 mW microwave power (25 dB power attenuation), 100 G sweep width (centered around the expected $g = 2$ position), 16 scans, 60 dB gain, 1 G modulation amplitude, 100 kHz modulation frequency, 20.48 ms conversion time, 10.24 ms correlation time. Any sample variation from these settings are noted as such in the main text.

11.1.1.2. Low temperature continuous wave

Experiments were conducted using a Bruker ELEXSYS E500 as described in §11.1.1.1 however the resonator was replaced with a Bruker ER 4122 SHQ (Super High-Q) resonator in combination with a Bruker ER 4112HV helium temperature control system.

Samples were initially tested at ~ 9.4 GHz, 1.997 mW microwave power (20 dB power attenuation), 100 Gauss sweep width (centered around the expected $g = 2$ position), 32 scans, 60 dB gain, 1 G modulation amplitude, 100 kHz modulation frequency, 20.48 ms conversion time, 10.24 ms correlation time. Any sample variation from these settings are noted as such in the main text.

Temperatures of 80 K and above were obtained through liquid nitrogen cooling whilst temperatures below 80 K were obtained through liquid helium cooling using an Oxford Instruments 503S ITC and VC41 gas flow controller.

11.1.2. S- and Q-band

S and Q-band experiments were conducted using a Bruker ELEXSYS E560 running Bruker software Xepr v2.6; comprising a Bruker ER 023 M signal channel controller, Bruker ER 032 M field controller, Bruker ER 048 R microwave controller, Bruker magnet B-E251B9, Bruker ER 083 Z magnet power supply and Hewlett-Packard 5352B frequency counter.

11.1.2.1. S-band

S-band measurements were performed using a Bruker ER 061 SR microwave bridge and a Bruker ER 6103 S-MS5 resonator, initial room temperature experiments were conducted inside a Bruker ER 4118CF cryostat.

Temperatures below room temperature were obtained using an American Research Systems closed cycle displax system comprising an ARS-4HW compressor, ARS-DE204SF-FMX-19EC cryostat and LakeShore LS331S Intelligent Temperature Controller. The closed cycle displax system is extensively evaluated in Appendix A.

Samples were initially tested at ~ 4.0 GHz, 0.63 mW microwave power (25 dB power attenuation), 100 Gauss sweep width (centered around the expected $g = 2$ position), 32 scans, 60 dB receiver gain, 1 G modulation amplitude, 100 kHz modulation frequency, 81.92 ms time constant. Low temperature experiments were conducted at 80 K. Any sample variation from these settings are noted as such in the main text.

11.1.2.2. Q-band

Q-band measurements were performed with a Bruker ER 053 QRD microwave bridge and a Bruker ER 5106 QTW resonator.

Temperature was controlled using using an American Research Systems closed cycle displax system comprising an ARS-4HW compressor, ARS-DE204SF-FMX-19EC cryostat and LakeShore LS331S Intelligent Temperature Controller. The closed cycle displax system is extensively evaluated in Appendix A.

Samples were initially tested at ~ 34.0 GHz, 0.63 mW microwave power (35 dB power attenuation), 450 Gauss sweep width (centered around the expected $g = 2$ position), 20 scans, 60 dB receiver gain, 1 G modulation amplitude, 100 kHz modulation frequency, 20.48 ms time constant. Low temperature experiments were conducted at 80 K. Any sample variation from these settings are noted as such in the main text.

11.1.3. W-band

Experiments were conducted using a Bruker ELEXSYS E680 running Bruker software Xepr v2.6; with an X-band Bruker XPB and Bruker W-band EC 01 (ESP 900-1041) microwave bridge, a Varian 5T-EPR hybrid magnet fitted with a Bruker W-band resonator and cryostat.

Samples were initially tested at ~ 94.0 GHz, 3.57 μ W microwave power (31 dB power attenuation), 280 Gauss sweep width (centered around the expected $g = 2$ position), 32 scans, 60

dB receiver gain, 1 *G* modulation amplitude, 100 *kHz* modulation frequency, 1.28 *ms* time constant, 40.96 *ms* conversion time. Any sample variation from these settings are noted as such in the main text.

Temperatures below room temperature were obtained through liquid helium cooling using an Oxford Instruments 503S ITC and VC41 gas flow controller.

11.2. Pulsed EPR

11.2.1. Sample preparation

All pulsed EPR experiments used samples of total volume 80 μl , and unless otherwise stated final protein concentrations of 100 μM . Where this involved mixing multiple components a stock protein solution of 200 or 400 μM were used.

All samples were extensively dialysed against high purity (Milli-Q) water (pH 7.5) and 5 *mM* EDTA to remove all salt and metal from the samples. Samples were then concentrated using VivaSpin 15R spin concentrators until the sample absorbance at 280 *nm* matched the desired concentration.

All samples were prepared in 200 μl centrifuge tubes, as per the following instructions, before being transferred using TPX tubing to Wilmad 707-SQ-250M 4 *mm*, thin wall precision quartz EPR tubes.

11.2.1.1. Protonated samples

At the final spin concentration stage, the sample was reduced to 0.5 *ml* with spinning, made up to 15 *ml* with high purity water (pH 7.5) and reduced again. This was repeated 3 times to remove any EDTA before final sample preparation.

11.2.1.2. Deuterated samples

At the final spin concentration stage, the sample was reduced to 0.5 *ml* with spinning, made up to 15 *ml* with high purity D_2O (>99.99 %)(pH 7.5) and reduced again. This was repeated 3 times to remove any EDTA and to ensure deuteration of the solvent and all labile protons of the protein in the final sample.

11.2.1.3. pH shifted samples

Samples were prepared as per §11.2.1.1, but solutions were made back up with buffered solution at the desired pH. 25 *mM* Tris (tris(hydroxymethyl) methylamine) was used for pHs 8.0 and above, 25 *mM* HEPES (4-2-hydroxyethyl -1-piperazineethanesulfonic acid) for pHs 7.0 and 7.5 and 25 *mM* MES (2-(N-morpholino) ethanesulfonic acid) for pH 6.5 and below.

11.2.1.4. Samples with glycerol

Samples with glycerol were prepared from pure, unbuffered, high protein concentration solutions. Typically, 40 μl of a 200 μM protein solution was added to a 200 μl centrifuge tube to which 40 μl was added, to make a 50 % glycerol solution with final protein concentration of 100 μM .

In the case of §6.5, a 400 μM stock protein solution was used (at the boundary of protein precipitation) and in the case of the 90 % glycerol solution was used directly to give a final 40 μM protein concentration. For lesser glycerol concentrations the protein solution was first diluted with high purity water before addition of glycerol to give the same final protein concentration.

11.2.1.5. Sample mixing

Samples once prepared required mixing to stop separation into aqueous and organic layers. Samples were sealed in 200 μl centrifuge tubes and flicked by hand several times until distinct layers were no longer clear. Samples were further mixed with a benchtop vortex mixer (300 *rpm*, 30 *s*), before being left at room temperature (10 *min*) to allow diffusion of protein throughout the solution.

11.2.1.6. Sample freezing

Typically, samples were desired to be in the glass phase. Glass samples were produced by flash freezing the sample, inside the EPR tube by immersion in liquid nitrogen.

For §6.7, ice samples were produced by placing the sample inside a sample tube inside of a cryostat at 298 *K* and then slowly cooling, inside the cryostat at 2 *K/min* to the experimental conditions. Care needed to be ensure a high quality sample tube was used to avoid the sample smashing the sample tube.

§6.7 reannealed samples were produced by flash freezing the sample in liquid nitrogen as above before being transferred to a cold cryostat at 50 *K*. Once inside the cryostat the temperature were raised at 5 *K/min* to just below the glass transition point (140 *K* [63]) and held for 20 *min* before being returned to 50 *K* at 2 *K/min*.

11.2.2. General experimental setup

Experiments were conducted using a Bruker ELEXSYS E580/680 hybrid running Bruker software Xepr v2.6; with an X-band Bruker XPB microwave bridge, Bruker ER 4118X-MD5 resonator inside a Bruker ER 4118CF cryostat, Bruker ER 073 magnet, Bruker ER 083 CS magnet power supply, Applied System Engineering 117X (1 kW) TWT amplifier. Experiments were conducted in an environmentally controlled room, where the temperature was maintained at 298 ± 1 *K*.

11.2.3. T_1

T_1 experiments were conducted at 50 K, ~ 9.7 GHz, with the magnetic field optimised for maximum signal around $g = 2$, typically 344 mT. 2 scans, 1024 points, 4 shots/point, time increment 12000 ns, 20 MHz video bandwidth, 50000 μ s shot repetition time (approximately $4\times$ the T_2 time), 4 dB attenuation, 57 dB video gain. These settings were optimised on a sample basis, particularly attenuation and video gain, any significant variation from these settings being noted in the main text.

11.2.4. T_2

T_2 experiments were conducted at 50 K, ~ 9.7 GHz, with the magnetic field optimised for maximum signal around $g = 2$, typically 344 mT. 4 scans, 16 shots/point, 1024 points, time increment 4 ns, 10000 μ s shot repetition time, 8.0 dB attenuation, 45 dB video gain. These settings were optimised on a sample basis, particularly attenuation and video gain, any significant variation from these settings being noted in the main text.

11.2.5. FSE

FSEs were conducted at 50 K, ~ 9.7 GHz, 187.5 nW microwave power (60 dB power attenuation), 100 G sweep width centred about $g = 2$, 4 scans, 2 shots/point, 512 points, time increment 4 ns, 10000 μ s shot repetition time, 8.0 dB attenuation, 45 dB video gain. These settings were optimised on a sample basis but matched the T_2 experiment.

11.2.6. ELDOR nutation

ELDOR nutation experiments were conducted at 50 K, ~ 9.7 GHz, with the magnetic field optimised for maximum signal around $g = 2$, typically 344 mT. 4 scans, 2 shots/point, 256 points, time increment 2 ns, 10000 μ s shot repetition time, 8.0 dB attenuation, 45 dB video gain, ELDOR power optimised to give maximum inverse, typically 11 dB.

11.2.7. PELDOR

PELDOR experiments were conducted at 50 K, ~ 9.7 GHz, with the magnetic field optimised for maximum signal around $g = 2$, typically 344 mT. >2000 scans, 4 shots/point, points (optimised to experiment ~ 400), time increment 8 ns, shot repetition time - determined from T_1 experiment (typically 10000 μ s), attenuation and video gain - determined from T_1 experiment (typically 8.0 dB and 45 dB respectively), ELDOR power - optimised to give maximum inverse (typically 11 dB).

Typically, the observer sequence was set to the low field peak of the FSE, whilst the ELDOR channel was set to excite the centre peak. Observing with less signal intensity but stronger ELDOR modulations was found to be preferable than observing with more signal intensity and less ELDOR modulation.

11. EPR and other spectroscopies

PELDOR parameters were set according to prior FSE, T_2 , T_1 , ELDOR power and ELDOR nutation experiments. These parameters were passed to a custom 4 pulse ELDOR program in PulseSPEL which allows for independent control of each pulse. Phase cycling of the first π pulse was employed to aid in noise suppression. Nuclear modulation suppression was not employed for deuterated samples as the deuteron modulation time is near to the ELDOR modulation time of a 3.0 *nm* distance.

11.3. Other spectroscopy

11.3.1. NMR

Standard 'BioPack' ^1H - ^{15}N HSQC and triple resonance (CBCACONH, HNCACB, CCONH and HCCONH) NMR experiments were conducted at ^1H frequencies of 500 and 600 MHz on Varian Inova spectrometers by Dr. R. Boetzel (UEA). Binding experiments were conducted using doubly labelled $^{15}\text{N}/^{13}\text{C}$ protein samples prepared by A. Reilly (UEA) at pH 6.2. Final ^1H - ^{15}N HSQC experiments were acquired at a molar ratio of immunity protein to DNase of 1.54:1.

11.3.1.1. Chemical shift perturbations

The assignments of the HSQC and triple resonance experiments were tabulated into chemical shifts for each residue of each protein with and without the binding partner. Residues with a difference outside 1 standard deviation of the average chemical shift were deemed to be significant.

The structure of each protein was then run through the program Naccess¹. This program runs a virtual sphere over the surface of the structure and returns the solvent accessibility of each residue. Any residues that had less than 50% solvent accessibility were considered not relevant and disregarded.

The remaining residues were defined as *active residues* to docking whilst residues with <50% solvent accessibility within 3 Å to active residues were considered to be *passive residues* in docking. Together these active and passive residues make up the ambiguous interaction restraints (AIRs).

11.3.2. Ultra-violet / visible spectroscopy

Experiments were conducted using an Hitachi U2900 Spectrophotometer running Hitachi UV solutions software (build: 461).

3 *ml* of the sample was loaded into a Hellma Precision Cell of Quartz Suprasil (10 *mm* light path) and tested at 298 *K*, medium response, 1.5 *nm* slit width, over a range of 800 - 220 *nm*, at a scan rate of 200 *nm/min*.

¹Hubbard,S.J. & Thornton, J.M. (1993), 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.

On each day of testing blanks of an empty cavity, empty cuvette and buffer were taken, which were post-experimentally taken from the signal.

11.3.3. Circular dichroism

Experiments were conducted using a Jasco J-810 Spectropolarimeter running Jasco CD software (v1.24.00)

3 *ml* of the sample was loaded into a Hellma Precision Cell of Quartz Suprasil (10 *mm* light path) and tested at 298 *K*, low sensitivity (1000 *mdeg*), data pitch 0.5, 1 *s* response, 2 *nm* bandwidth, continuous scanning mode, over a range of 500 - 220 *nm*, at a scan rate of 100 *nm/min*.

On each day of testing blanks of an empty cavity, empty cuvette and buffer were taken, which were post-experiment taken from the signal.

11.3.4. Fluorometry

Experiments were conducted using a Perkin Elmer LS 55 Fluorescence Spectrometer running FL WinLab software with thermostatted cell holder accessory.

3 *ml* of the sample was loaded into a Hellma fluorescence cuvette of Suprasil quartz (path-length 10×10 *mm*) and tested at 298 *K*, continuous excitation mode with excitation at 295 *nm*, detection between 300 and 500 *nm* with a scan rate of 100 *nm/min*.

On each day of testing, blanks of an empty cavity, empty cuvette and buffer were taken, which were post-experimentally subtracted from the signal.

12. Computational

12.1. EPR toolbox

The EPR toolbox began as a simple collection of scripts written to streamline repetitive tasks such as loading files into the MatLab workspace, autozeroing, normalization and the like. As time has gone on the EPR toolbox has grown in complexity and functionality. At the time of writing (August 2013) the toolbox's key highlights include over 45000 lines of code (plus extensive commenting), 21 public releases, over 30 months of development and over 1100 unique downloads from 41 countries.

12.1.1. cwViewer

cwViewer is a self contained program, included as part of the EPR toolbox, for the fast visualisation and basic manipulation of cw-EPR, although over time basic functionality for the pulsed EPR techniques of FSEs and PELDOR traces has been included.

cwViewer allows the user to quickly load a single file or folder of files into the workspace and instantly visualise them. 3-dimensional Bruker BES3T files (.DTA/.DSC/.YGF) are supported and can be plotted on top of one another or staggered in x and/or y axes.

The user can perform a number of basic manipulations upon the dataset including autozeroing, normalization (between 1 and -1), smoothing (by Savitzky-Golay, digital moving window averaging or FFT methods) and roughening (single point within a range).

The abscissa can be toggled (in an appropriate dataset) between magnetic field and g value through either an approximation method or exactly, when a known standard (such as DPPH) has been loaded.

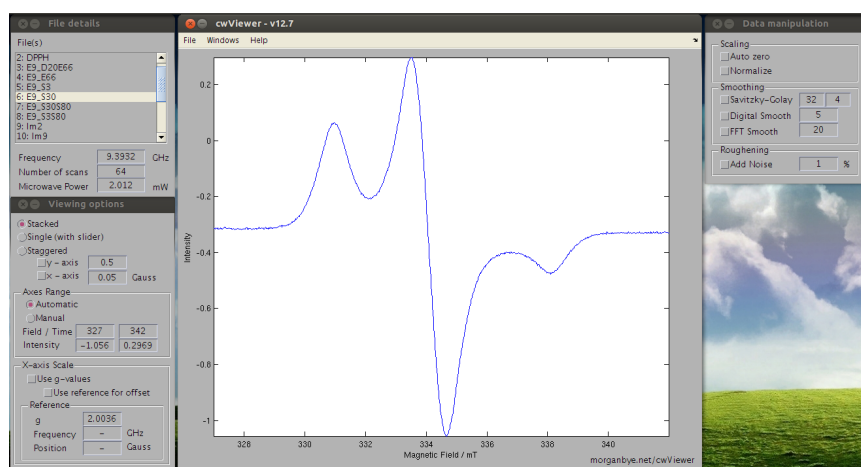


Figure 12.1 Screenshot taken from cwViewer, showing the room temperature X-band spectrum of colicin E9 DNase S30C^{R1A}.

12. Computational

The dataset can then be exported in a variety of common file formats such as bitmap images (.png), vector images (.eps), comma separated value number tables, with magnetic field and g values (.csv) or to the MatLab workspace, for use with other MatLab programs.

In later versions, an option was given that allowed the user to pass the dataset after manipulation directly to EasySpin for simulation through a simple graphical user interface (in that it hides most of the complexity of EasySpin and limits the parameters available).

12.1.2. EasyRefiner

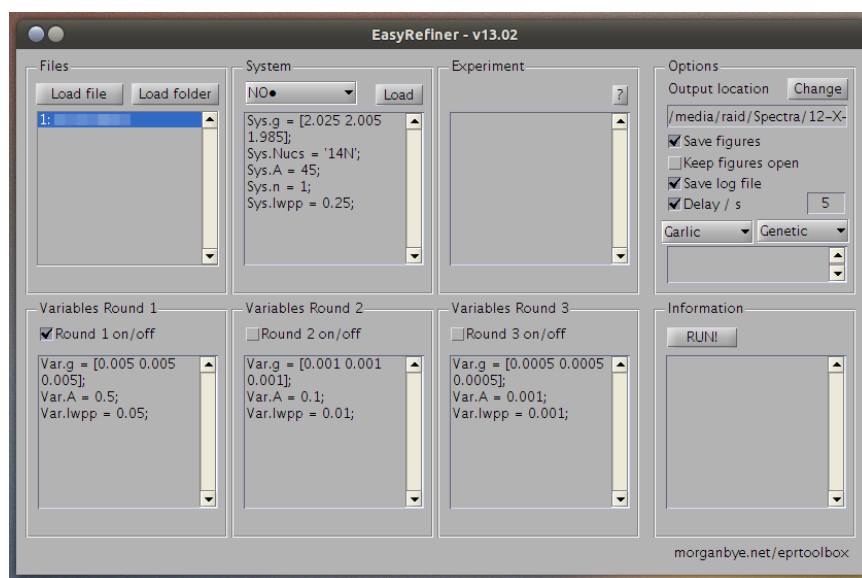


Figure 12.2 Screenshot taken from EasyRefiner, showing a file with default nitroxide system parameters loaded

In this investigation over 30 cw-EPR spectra were collected across different temperatures and multiple frequencies. Each sample was broadly similar to every other sample at that temperature and frequency, as all were nitroxide spin labels. Rather than attempt to manually fit each spectrum with EasySpin and reach the desired accuracy an automation facility was built.

EasyRefiner, is a MatLab written graphical user interface that allows the loading of files and folders to a queue list. The queuing functionality meant that fitting could be conducted upon multiple spectra overnight.

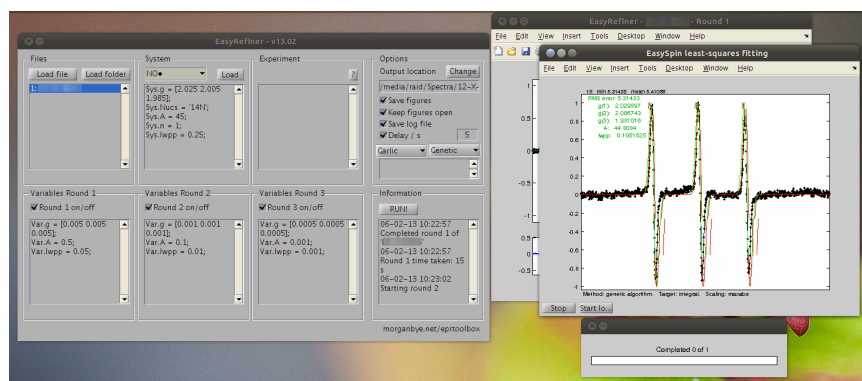


Figure 12.3 Screenshot taken from EasyRefiner, showing fitting of a nitroxide system

The spin system is stated in the normal EasySpin syntax in the system box, or the user can load defaults for nitroxides, Fremy's salt or DPPH. The experiment is defined in normal

EasySpin syntax. EasyRefiner then allows up to 3 rounds of variables fitting, the idea is that the first round of fitting gets variables that are roughly correct, which are passed to the next rounds of fitting with finer variables. This allows the finding of very accurate g values in a short computation time. Finally in the options menu there are options to save each EasySpin figure, the EasyRefiner log file and introduce a delay between fits (to prevent CPU overheating in long runs) as well as the EasySpin fitting options.

12.1.3. PowerSat

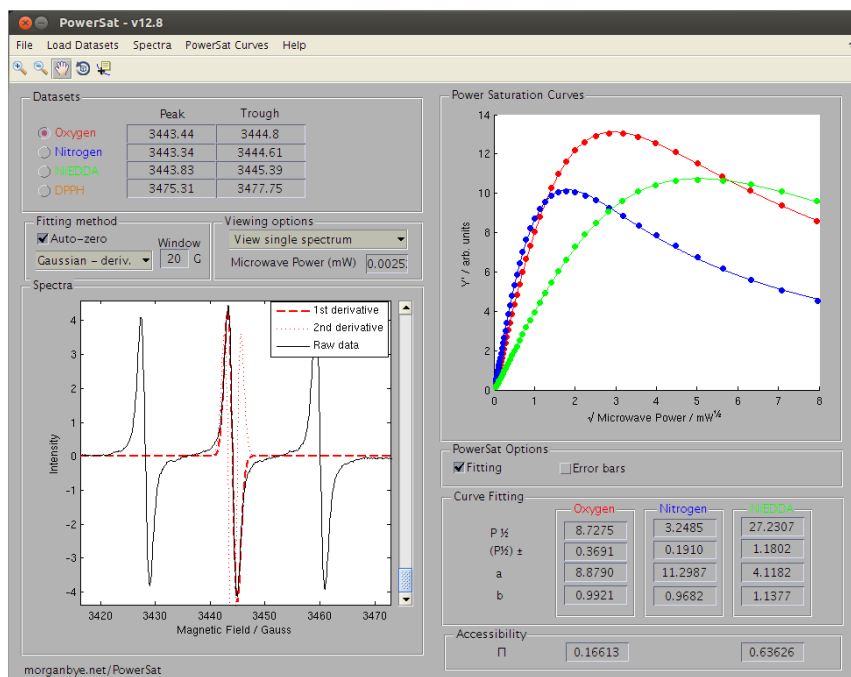


Figure 12.4 Screenshot taken from PowerSat, showing a Gaussian derivative fitting method

PowerSat is a self contained program written to quickly and efficiently evaluate power saturation data. PowerSat is comprised of over 1500 lines of MATLAB code and has been released under an open-source license to the public as part of the EPR toolbox and has gained over 75 unique downloads (though its inclusion in the EPR toolbox prohibits exact user numbers, but the PowerSat webpage has been accessed over 2000 times¹).

PowerSat loads raw Bruker EPR spectra, as either a folder of files in Bruker BES3T (.DTA/.DSC) or Bruker EMX (.spc/.par) formats or as a single 3-dimensional Bruker BES3T file (.DTA/.DSC/.YGF). A combination of these file formats can be used for each of the power saturation experiments; oxygen (lipophilic), nitrogen (control), NiEDDA (hydrophilic) and DPPH (standard).

When files are loaded the program automatically finds the centre line peak and trough using a search function for each power (though this can be specified by the user manually). The centre line is described using either a simple point selection or by fitting. The fitting can be completed quickly by splitting the peak and trough and fitting each using a Gaussian or Lorentzian function. A more correct analysis can be made at the expense of computing time by considering the peak and trough together and fitting first derivative Gaussian or Lorentzian functions. The coefficients of the first derivative fit are then used to solve the

¹as of 1st August 2013

12. Computational

second derivative equation for zero to give the turning points (the peak and trough positions in the first derivative spectrum).

The peak-to-trough height is then plotted against the square root of microwave power for the spectrum to give the power saturation curve data points. Curve fitting of the data to:

$$y(x) = a * x * (1 + (((2^{(1/b)} - 1) * (x^2))/P))^{(-b)} \quad (12.1)$$

yields the $P_{1/2}$ (power at half of saturation) for that data set. With datasets and $P_{1/2}$ s for the standard and control, the spin label accessibility parameter, Π for lipophilic and hydrophilic environments is calculated using the method outlined by Altenbach *et al.* [1].

12.2. NACCESS

Naccess is a program written in Fortran 77 code by Simon Hubbard² which calculates the atomic accessible area of a macromolecule by rolling a probe over the Van der Waal's surface. The path of the probe's progress is traced out by its centre to give the accessible surface. Typically the probe is set to 1.4 Å, the same radius as water, allowing for the macromolecule surface to be considered as the solvent accessible surface.

12.2.1. Define spin labels

Naccess by default has Van der Waal's parameters for all natural amino acids, stored within the programs `vdw.radii` file, this however does not include the unneutral combined Cys residue with MTSL (R1A) or IA-PROXYL (IA1).

To line 160 of `vdw.radii` the following was added (displayed in 2 columns for space)

```
160 RESIDUE ATOM R1A 18          179
161 ATOM C 1.87 0              180 RESIDUE ATOM IA1 20
162 ATOM CA 1.87 0            181 ATOM C 1.87 0
163 ATOM N 1.65 0            182 ATOM CA 1.87 0
164 ATOM CB 1.76 0           183 ATOM N 1.65 0
165 ATOM SG 1.85 0           184 ATOM CB 1.87 0
166 ATOM SD 1.85 0           185 ATOM SG 1.85 0
167 ATOM CE 1.87 0           186 ATOM CD 1.87 0
168 ATOM C3 1.70 0           187 ATOM CE 1.87 0
169 ATOM C2 1.70 0           188 ATOM OE 1.40 1
170 ATOM C9 1.70 0           189 ATOM NZ 1.65 0
171 ATOM C8 1.70 0           190 ATOM C3 1.87 0
172 ATOM C4 1.70 0           191 ATOM C2 1.87 0
173 ATOM C5 1.70 0           192 ATOM C9 1.87 0
174 ATOM C7 1.70 0           193 ATOM C8 1.87 0
175 ATOM C6 1.70 0           194 ATOM C4 1.87 0
176 ATOM N1 1.65 1           195 ATOM C5 1.87 0
177 ATOM O1 1.40 1           196 ATOM C7 1.87 0
178 ATOM O 1.40 1            197 ATOM C6 1.87 0
```

²of formerly UCL, UK, currently University of Manchester, UK
<http://www.bioinf.manchester.ac.uk/naccess/>

```

198 ATOM N1 1.65 1
199 ATOM O1 1.40 1
200 ATOM O 1.40 1

```

Line 160: Residue identifier, instructions that the following are atom assignments, 3 letter code for residue, number of atoms to expect.

Lines 161-178: Identifier, unique atom identifier within residue (corresponding to the third column of a pdb file), Van der Waal radius in Angstroms corresponding to the sixth column of a pdb file), indication of whether atom is charged (0 = false, 1 = true).

The Van der Waal radii are taken from Bondi's compilation [8].

12.3. MMM

MMM is an open-source program written initially by Prof. G. Jeschke and Dr. Y. Polyhach at ETH Zurich, CH. Perhaps its most useful function is to take PDB files and *in silico* attaches a spin label. These spin labels can then be evaluated by a simulated PELDOR experiment.

12.3.1. Selenomethionine

MMM by default was designed for proteins that were to be examined by EPR. Many PDB files are from X-ray crystallography experiments which occasionally require heavy atoms to start assignment; this is often achieved with the use of selenomethionine (MSE).

The original release of MMM (2009) did not have the Van der Waal terms for selenium included, thus the following addition was proposed to `get_simpleCH27forLJ.m` file, line 60

```

60 % Selenium , 34
61 % we use sulphur values here , assuming that the protein for spin-label
62 % experiments is made with native amino acids
63 vdW(34)=1.80; % VdW radius
64 p_q(34)=0; p_Rmin2(34)=2.000; p_eps(34)=-0.450; % S "Sulfide Sulfur"
65 % p_q(16)=0; p_Rmin2(16)=1.975; p_eps(16)=-0.380; % SM "Disulfide Sulfur"%

```

In correspondence with Jeschke (February 2010) it was decided that whilst the actual VdW distance for selenium is 1.9 Å, the value of 1.8 Å corresponding to sulphur would be used. This is under the assumption that no magnetic resonance experiment would use selenomethionine over methionine.

This addition appeared in MMM 2009 Patch 1 as

```

60 % Selenium , 34 (taken EQUAL to Sulfur!)
61 vdW(34)=1.80; % VdW radius
62 p_q(34)=0; p_Rmin2(34)=p_Rmin2(16); p_eps(34)=p_eps(16);

```

This modification also required the `residues.mat` file to be updated with the addition of the MSE residue to the `residue_defs.hettags` array.

12.3.2. PDB file write function

MMM has the functionality to write out PDB files. Unfortunately, the PDB file format is old, having undergone many revisions and is used by many programs that do not explicitly enforce all PDB standards.

MMM versions 2009-2011.2 wrote out a PDB file a single line at a time using the file write command `fprintf`, with no real control over line length instead using the new line character `“\n”`. Some programs within the scope of this work, read PDB files on the older assumption that every line is exactly 80 characters. Thus to make MMM PDB files more widely compatible with these programs required a major rewrite of the MMM `wr_PDB.m` file. Each section of the PDB file is now written using a subfunction to an output string with `sprintf` which pads strings under 80 characters with spaces before a new line character.

The final `wr_PDB.m` file stands at 883 lines (too long to reasonable be shown here) and was implemented in the next MMM release (2013).

Part IV.
Appendices

A. Useful references

Physical Constants

<i>Name</i>	<i>Symbol</i>	<i>Value</i>
Avogadro's number	N_A	$6.02214 \times 10^{23} \text{ mol}^{-1}$
Bohr Magneton	μ_B	$9.27401 \times 10^{-28} \text{ J.G}^{-1}$
Boltzmann constant	k	$1.38065 \times 10^{-23} \text{ J.K}^{-1}$
Electron charge	e	$-1.60218 \times 10^{-19} \text{ C}$
Electron mass	m_e	$5.486 \times 10^{-4} \text{ amu}$
Free electron g-value	g_e	2.002319304
Gas constant	R	$8.31447 \text{ J.mol}^{-1}.\text{K}^{-1}$
Planck constant	h	$6.62607 \times 10^{-34} \text{ J.s}^1$
Speed of light	c	$2.99792 \times 10^8 \text{ m.s}^{-1}$

EPR frequencies

<i>EPR band</i>	<i>Frequency / GHz</i>	<i>Wavelength / mm</i>
L	1	300
S	3	100
X	10	30
K	24	12.5
Q	35	8.5
W	95	3.2
J	263	1.0

B. Closed cycle low temperature EPR

Following the development and design work of the Hinderberger group (MPIP Mainz, DE) in 2008 the EPR group at the University of East Anglia took delivery of 2 closed cycle cooling cryostats in December 2011.

The Hinderberger group's system is used primarily in polymer research where experiments are long and sample changes infrequent. In this instance, sample changing can occur by turning the machine off, warming to room temperature and deconstruction of the cryostat. At UEA the primary research is upon biological samples where experiment times are much shorter and sample changes more frequent.

The primary benefit of moving to a closed cycle system is that the requirement of liquid cryogens being passed over the sample is replaced by a cold environment being maintained around the sample using only continuously recycled gaseous helium. Whilst the delivered system pertains to a commercially available system it required adaption to the necessities of EPR spectroscopy.

The following material is the documentation generated during the installation, calibration and testing of the units. The version here is the 5th iteration and includes technical drawings of all additionally required hardware and an "Operator's guide" detailing resonator installation, equipment turning on, standard operation, sample changing and safe turn off procedures.

Closed cycle low temperature EPR - displex unit installation, calibration and operation

v13.08

Morgan Haydn Bye

This document was created during study for the degree of Doctor of Philosophy whilst at the University of East Anglia, UK.

11th August 2013



morganbye.net

...always accepting a worthy challenge



This document has been supplied on condition that anyone whom consults it is understood to recognise that its copyright rests with the author and that no quotation from this work, nor any information therefrom, may be published without the author's prior written consent.

Contents

1	Installation	1
1.1	Compressor	1
1.2	LakeShore intelligent temperature controller	1
1.3	Cryostat	2
1.3.1	Fitting the cryostat inside the magnet	2
1.3.2	Cryostat height	3
1.3.3	Vacuum shroud connections	3
1.3.4	Other connections	4
1.4	Hall probe	5
1.5	Cryostat to DN40 adapter	6
1.6	Gas line set up	7
1.7	Sample temperature sensing	8
1.8	Vibration dampening	9
1.9	Technical drawings	12
2	Calibration	17
2.1	Hall probe	17
2.2	Pressure testing	17
2.2.1	Empty cavity testing	18
2.2.2	Leak testing	18
2.2.3	Gas desorption	18
2.3	Pressure optimisation	19
2.4	Final calibration	20
2.4.1	Temperature offset	20
2.4.2	Temperature sensors and pressure	21
2.4.3	Cooling time	22
3	Operation	23
3.1	Installing a resonator	23
3.1.1	S-band	23
3.1.2	X-band	25
3.1.3	Q-band	25
3.2	Turning the dispex unit on	25
3.3	Changing samples	27
3.4	Turning the dispex unit off	28
4	Troubleshooting	29
5	Postface	33

List of Figures

1.1	A schematic diagram showing the connections on the cryostat	2
1.2	A schematic of the cryostat situated inside the magnet	3
1.3	Photograph of the installed cryostat with all connections made	4
1.4	Photographs of the Hall probe holder in place	5
1.5	Schematic of cryostat in place, with cryostat header and resonator	6
1.6	Schematic for the gas line set up into the cryostat sample space	7
1.7	Cernox photographs	8
1.8	Noise from the dispex system shown in Xepr at Q-band	9
1.10	Signal averaging affect on noise from the dispex system shown on a weak nitroxide labelled protein at S-band, 80 <i>K</i>	10
1.9	Construction and installation of vibration brace clamp	11
1.11	Technical drawing of the cryostat plinth	12
1.12	Technical drawing of the vacuum shroud adapter so that the cryostat can be attached to a vacuum pump inside the magnet	13
1.13	Technical drawing of the dispex header for conversion of the cryostat to a DN40 connection.	14
1.14	Technical drawing of the Hall probe holder	15
2.1	Temperature offset between that set on the LakeShore and that measured at the sample	21
2.2	The temperature from different sensors and pressure against the LakeShore set temperature	21
2.3	Cooling profile of a simulated sample from liquid nitrogen into a cryostat running at base temperature, shown from point of isolation of the system from the helium over-pressure immediately after sample insertion	22
3.1	Assembly of the waveguide to a Q-band resonator	23
3.2	Insertion of a Bruker Q-band resonator (Bruker ER 5106 QTW)	24
3.4	The Advanced Research Systems ARS-2HW compressor, taken from arscryo.com	26
3.3	LakeShore front panel, taken from lakeshore.com	26

Foreword

This manual is primarily concerned with the displex unit available from ARS Cryogenics (arscryo.com). As such the installation section will only be valid for the specified displex unit. However, the calibration and operation sections should be relevant to any displex units including those available from SHI Cryogenics (shicryogenics.com), Cryogenic (cryogenic.co.uk) or Bruker (bruker-biospin.com).

Each system manufacturer will have taken different approaches to overcome certain experimental problems. For example, Bruker systems operate with a much larger compressor with the cryostat space at an overpressure of helium ensuring that any leaks in the setup result in helium leaving and not air entering which minimises any icing of the cryostat.

Finally, this documentation is the product of approximately 9 months of work. That is 9 months from point of delivery until continuous operation by non-development members of the laboratory. A lot of this time was spent waiting, waiting for parts to turn up, waiting for trades to install infrastructure and waiting for parts to be manufactured. However, at time of writing the displex units in our laboratory have been operational for almost 2 years with the only associated cost having been 2 gas cylinders of gaseous helium.

Not all points discussed in this material is significant to the operation of the displex units and to most ordinary users they can move straight to chapter 3. However, a lot of what is discussed in chapters 1 and 2 were unknown until they were tested.

1 Installation

Upon delivery, the displax system from ARS Cryogenics arrives as one crate (containing the compressor) and one pallet of boxes containing everything else.

The system comprises:

- ◇ One compressor
- ◇ One LakeShore temperature controller
- ◇ One cryostat unit
- ◇ Two helium hoses
- ◇ One LakeShore to cryostat lead (2 pin banana plug to 9 pin DIN)
- ◇ One compressor to cryostat lead

1.1 Compressor

The compressor comes with a transformer and is quoted as suitable for single or three phase electrical connections. In the UK single phase connections are rated to only 13 *A*, whereas the compressor draws 25 - 30 *A*. For this reason the compressors were wired into three phase sockets using only one of the phases and placed onto a 30 *A* riser.

The compressor cooling water requirements are <2.5 *l/min* at <24 °C, which our building supply more than covers. A spur was taken from the magnet cooling water and inline taps added for ease of isolation.

The compression brass fittings for the supplied water cooling tubing are very tight, and require a heat gun to make the tubing pliable, as well as PTFE tape to ensure a water tight seal. The tubing can then be attached to the taps and compressor using adjustable spanners.

1.2 LakeShore intelligent temperature controller

The LakeShore should be completely unpacked and placed in an accessible location close to the spectrometer.

The LakeShore requires a standard IEC C13/C14 (kettle) lead for power and can be plugged into a normal mains socket.

The lead from the cryostat splits into 3 connections 2 × 6 pin connectors. The lead marked "Control Sensor" goes to Input A on the LakeShore whilst the lead marked "Free Length Sensor" goes to Input B

1 Installation

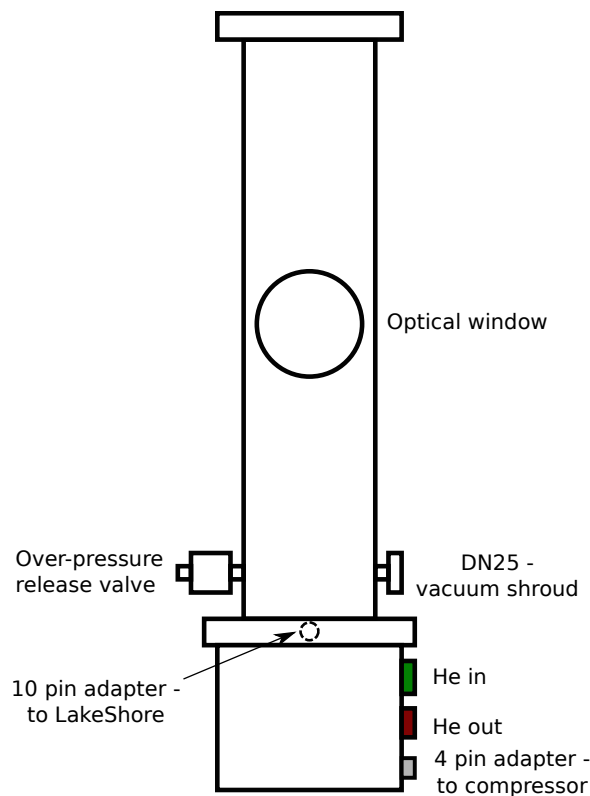


Figure 1.1 A schematic diagram showing the connections on the cryostat

The other lead is a 2 pin banana plug type connection which goes to the "Heater Output" of the LakeShore. IT IS VERY IMPORTANT YOU GET THIS CONNECTION THE RIGHT WAY AROUND, yet it is not mentioned in any of the instructional material. The banana plug with the white plastic lead goes to the "High" whilst the blue lead goes to "Low". Getting this wrong will see one of your heater coils receive 10 times the current it is designed for.

1.3 Cryostat

Whilst the displax system is sold as a complete system, several adjustments are required to the cryostat to make the system operation.

1.3.1 Fitting the cryostat inside the magnet

Remove the rapid scan plates from the magnet. Remove the Hall probe from its casing within the left rapid scan plate and put somewhere safe.

Get a set of calipers and accurately measure the diameter of the cryostat at its mid-point (at the optical window if you have one), and compare this to the width between magnet shoes. You may be able to **convince** the cryostat between the poles of the magnet, depending on the combination of magnet and cryostat (and at which end of engineering tolerances they lie).

In our case we had to get some fine grade sandpaper and slowly take back the shoes of the magnet by 0.5 mm. We corresponded with Bruker first and were assured that this "should be fine".

With Bruker supplied units, they have been designed to be used with rapid scan plates in place and should not have this problem.

1.3.2 Cryostat height

The cryostat comes as a fixed size. Unfortunately, the size of the cryostat does not coincide with the sample height of the Bruker magnets (the older brown BE25 or newer grey/white ER073). For this reason it was necessary for the cryostat to be raised off of the ground to optimise the position such that with the addition of the resonator that the sample would sit correctly between the poles.

To this end polymer plinths were constructed (see figure 1.11).

Note that, our pulsed magnet itself is on 100 mm girders so that the dual-doored console can fit under the microwave bridge shelf. For this reason a second, wider plinth with a larger recess was constructed to securely hold the first in place.

1.3.3 Vacuum shroud connections

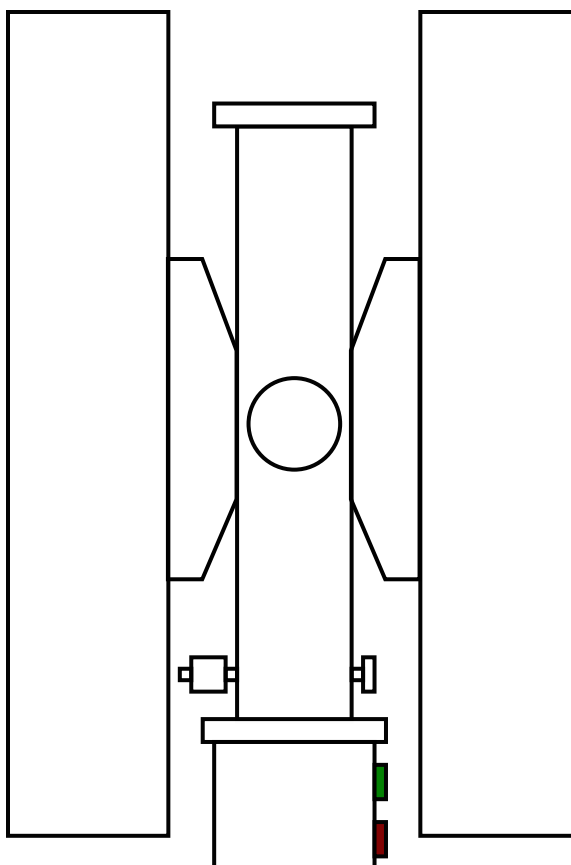


Figure 1.2 A schematic of the cryostat situated inside the magnet

With the cryostat in place between the magnet poles, it quickly becomes apparent that the vacuum connection (DN25) at the side of the cryostat (see figures 1.1, 1.2) points straight into the magnet such that whilst an O-ring and clamp could be placed upon the connection, no tubing could fit as well.

1 Installation

For this reason, an adapter is required to turn the connection on the cryostat away from magnet shoe, allowing for the vacuum tubing to come in from the front or rear of the magnet. A technical drawing for such an adapter is enclosed as figure 1.12 on page 13.

Connect the right angle end of the adapter to the cryostat using a DN25 O-ring and clamp, whilst the other end should be connected to a turbo-vacuum pump via DN25 vacuum tubing and an isolation valve.

Rough pump on the vacuum shroud for 30-60 *min* to take the pressure to $\sim 10^{-3}$ *mbar*. Once down to pressure the turbo pump can be turned on, taking the vacuum shroud down to high vacuum. The turbo pump should be left on overnight before the displax unit is turned to completely degas the system.

1.3.4 Other connections

Once the cryostat is in place all the connections can be made (see figure 1.3). First make the electrical connections to the LakeShore and compressor.

With all other connections made, take the protective screw caps off of the helium transfer lines. WEARING EYE PROTECTION attach the helium transfer lines to the compressor using spanners, and ensure a firm connection. Repeat the connection of the helium lines to the cryostat. This will take approximately 10-15 *min* to attach the transfer lines, as once within the magnet you can only use $\sim 1/6$ turns on the cryostat.

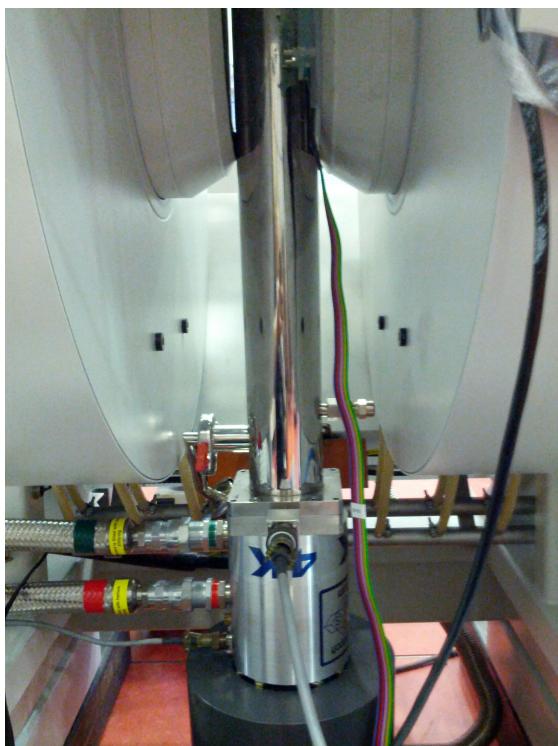


Figure 1.3 Photograph of the installed cryostat with all connections made

1.4 Hall probe

The Hall probe that usually sits in the rapid scan plates needs to be removed from the plate and put back between the poles of the magnet, as the displacer unit will not fit between the magnet poles with them in place.

The power supply to the magnet is controlled via a feedback mechanism from the Hall probe and without it in place the power supply will keep ramping up before it fail-safes.

The Hall probe needs to accurately reflect the magnetic field experienced by the sample. For this reason it needs to be as close to the sample location as is physically possible.

For initial testing it is sufficient to align the Hall probe at the correct height with the sample space (approximately the centre of the magnet shoe). Ensure that the Hall probe is flat against the magnet shoe and the correct way around (it is directional) and Duct tape the Hall probe in place.

This is not a long term solution, and certainly not suitable for magnetic field accuracy. For this reason a Hall probe holder needs to be created, before the spectrometer is calibrated using a known standard (such as DPPH), preferably using a teslameter if possible.

Technical details of the created Hall probe holder are shown in figure 1.14, created using 3 mm perspex sheets and brass connections. The base screws onto the magnetic pole of the magnet with Teflon screws, using the same holes as the rapid scan plates. The Hall probe is inserted into the gap, then the top plate (with padded foam tape) is screwed in place on top holding the Hall probe safely and firmly in place.

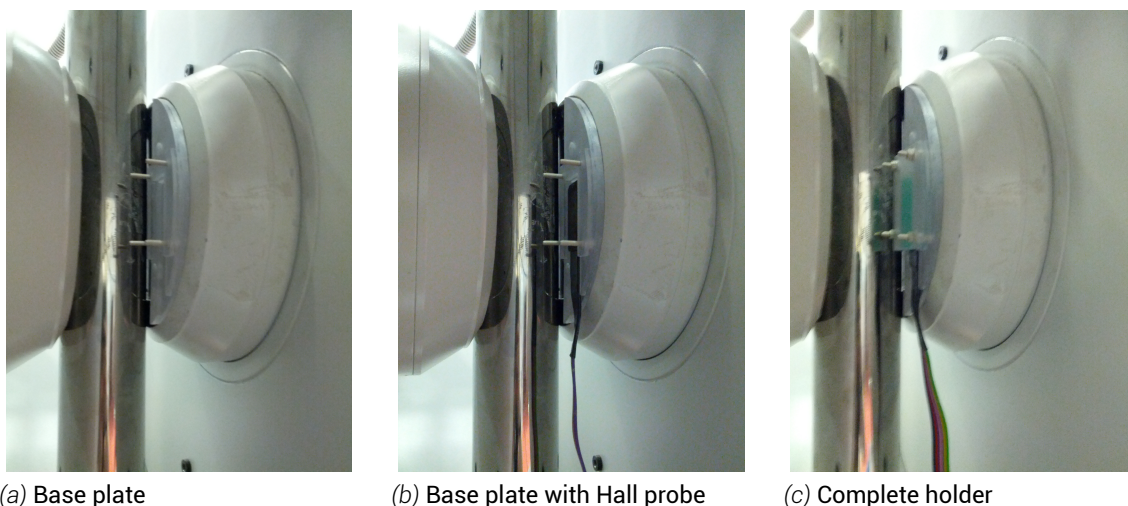


Figure 1.4 Photographs of the Hall probe holder in place

1 Installation

1.5 Cryostat to DN40 adapter

Like most flow cryostats the displax cryostat comes blank with a flat top surface with 6 screw holes for flexible attachments (at 60° to each other of outer diameter 64.5 mm). Bruker resonators use a standard DN40 connection and flow cryostats from Bruker come with an adapter for this purpose.

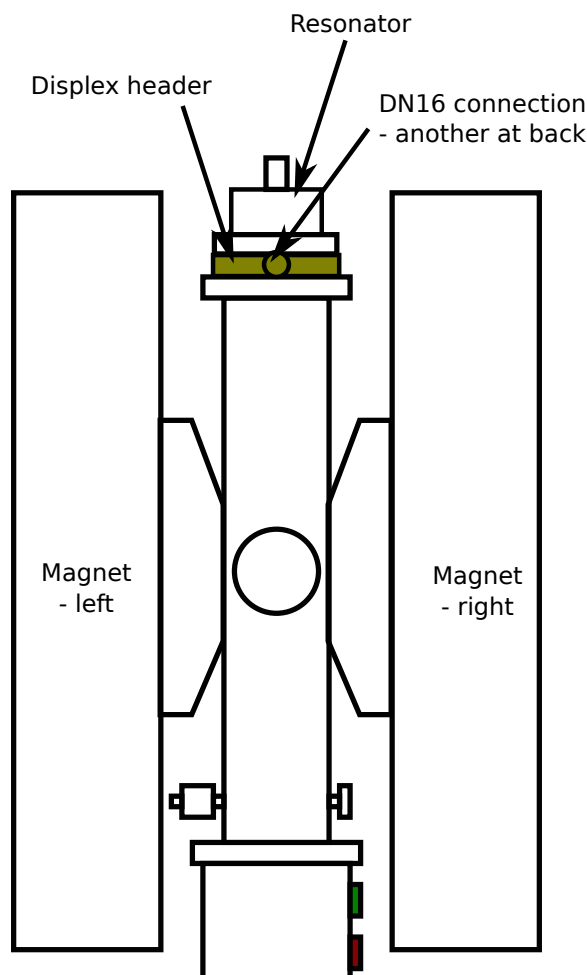


Figure 1.5 Schematic of cryostat in place, with cryostat header and resonator

However, we need access to fill and evacuate the sample space with resonator in place. For this reason a cryostat header was designed and manufactured of near identical design to the standard Bruker cryostat header but with 2 DN16 connections, one for gas control and one for an overpressure valve. Exact specifications can be seen in the technical drawing in figure 1.13 on page 14. The cryostat was manufactured out of brass to minimize interference with the magnetic field.

A 50 mm ID, 3 mm width O-ring is required between the cryostat and header, this can be greased with high vacuum grease to ensure a good seal. The header is then sealed in place with 6 × 5 mm washers and screws.

Between the header and resonator a standard DN40 O-ring and placement ring can be used in conjunction with a DN40 compression clamp.

1.6 Gas line set up

The sample space requires to be filled with helium gas and held at an underpressure. For this reason a simple system of vacuum parts was assembled such that the sample space could be controlled and isolated.

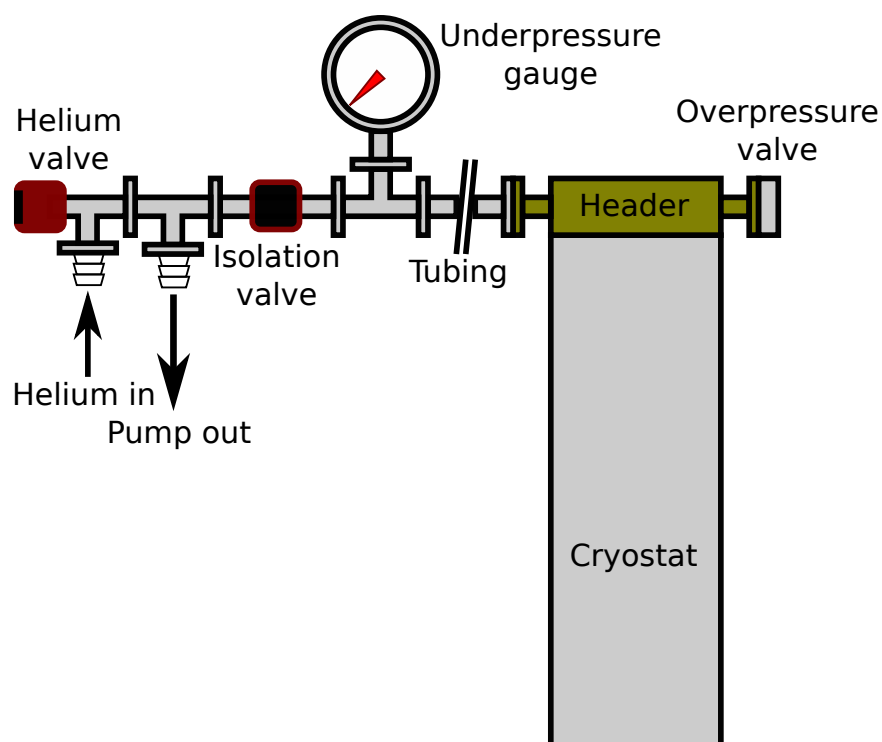


Figure 1.6 Schematic for the gas line set up into the cryostat sample space

The system shown in figure 1.6 allows for the helium to be turned on and flood the system. An overpressure valve (rated to 1150 *mbar*) stops any damage to the cryostat. The helium flow can then be shut off and pump turned on to evacuate the sample space whilst monitoring on the underpressure gauge (10-1000 *mbar*), the Oerlikon Leybold TM101 is recommended by the author. The isolation valve then allows the sample space to be held at this pressure even when the pump is turned off, allowing judgement as to whether there are any leaks in the system.

Depending upon your system, you may find it easier to put isolation valves inline off of the T-piece such that the pump and helium source can be isolated from one another. This gives the benefit of not having to turn the pump off during flushing of the cryostat.

1 Installation

1.7 Sample temperature sensing

To test the temperature at the sample it is necessary to construct a temperature sensor that can sit in the same position as the sample would and accurately report the temperature.

For this reason, a LakeShore Cernox CX-AA was purchased. A Cernox was chosen over other cryogenic temperature sensors due to its high temperature sensitivity and low response to magnetic fields. The relatively small size allowed the Cernox to be placed inside of a standard low temperature X-band EPR sample tube.

The sample tube was placed into a sample stick and wires run through, the challenge then lay in making a connection through the sample stick end whilst maintaining the vacuum. For this 10 pin cryogenic adapters (similar to that used by Oxford instruments) were used with a hood being soldered onto the sample stick and a male connection screwed down onto an O-ring inside the hood. The female connection was then run to a multimeter to read the resistance, which could later be transformed into a temperature.



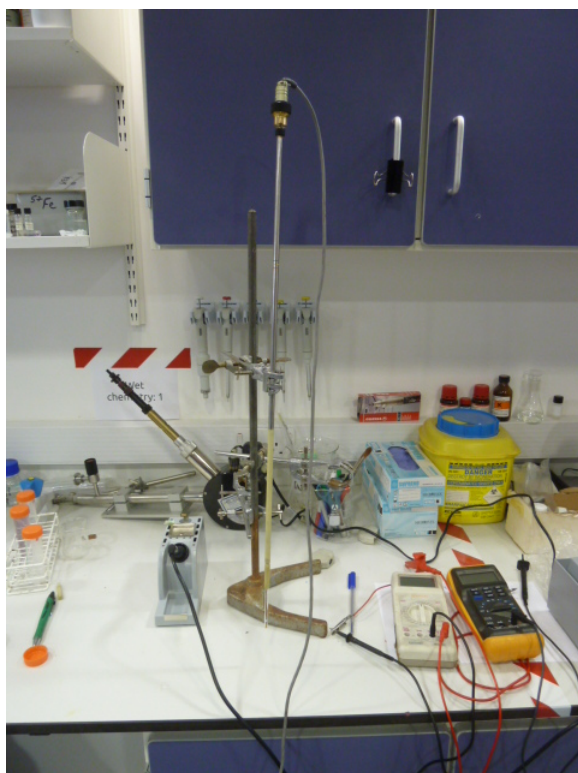
(a) Connector hood



(b) Male connector



(c) Female connector

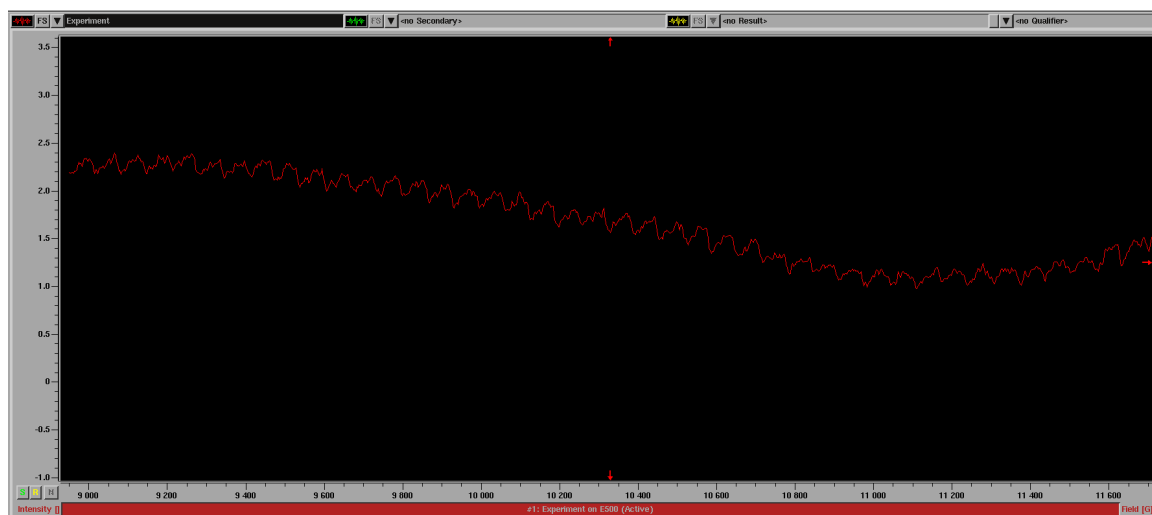


(d) Complete temperature sensor on the bench

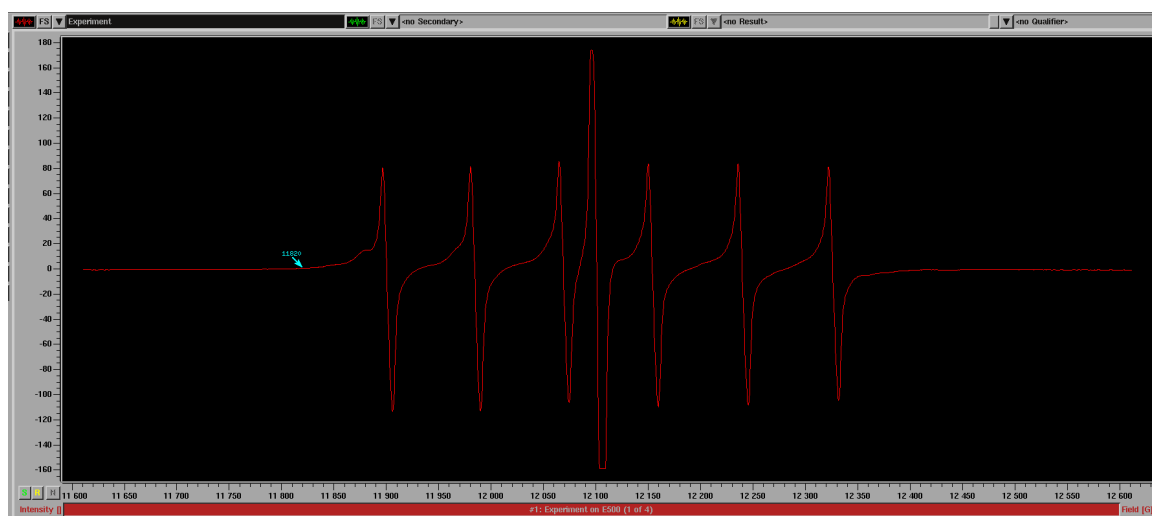
Figure 1.7 Cernox photographs

1.8 Vibration dampening

Once the cryostat is in place with all necessary connections made for operation it is important to brace the cryostat in place so that vibrations are minimised during operation. All vibrations from the valves on the cold head will be able to be seen on a cw experiment (see figure 1.8), for strong signals this should not be a problem but may cause problems for very weak signals; compare the intensities of figure 1.8a to 1.8b.



(a) The baseline taken from a non-signal part of the spectrum



(b) The full spectrum of the Mn/DPPH-KCl standard

Figure 1.8 Noise from the dispex system shown in Xepr at Q-band

It is important to isolate the microwave bridge as much as is possible from the cryostat, if the cryostat is in direct contact with the bridge platform it will be impossible for you to lock your signal during tuning.

With no vibration protection there is a risk that the cryostat falls over, this should have already been protected against through use of the heavy resin plinth. The most effective solution however was to place both microwave bridge and cryostat on a cut-down yoga mat. For further stabilisation (and thus reduce noise) it is recommended to surround the base of the cryostat with sand bags.

Finally X-band waveguide stabiliser arms can be used to wedge the top of the cryostat to the top of the magnet or the microwave bridge shelf cut out in the short term. A more effective long term solution, is to remove the cryostat from an Oxford flow cryostat system,

1 Installation

which leaves a horse-shoe shaped platform which fits between the magnet poles. A brace can then be constructed using padded tape and installed at the top of the cryostat, detailed in figure 1.9.

Finally, it is important to remember that signal averaging is most beneficial here. For pulsed experiments the timescale of the experiment is much shorter than the mechanical action and so will not be a problem, but for cw a single scan will contain many pump actions. The statistical chance that the spectrometer scan rate is exactly matched to the mechanical action is remote, thus signal averaging will average away any remaining vibrational noise. This can clearly be seen in figure 1.10 where an $\sim 10 \mu M$ nitroxide spin labelled protein is being run at S-band, 80 K. This is both a weak sample and a low resolution frequency and represents the worst possible situation, at X- and Q-band these vibrations average away almost instantly.

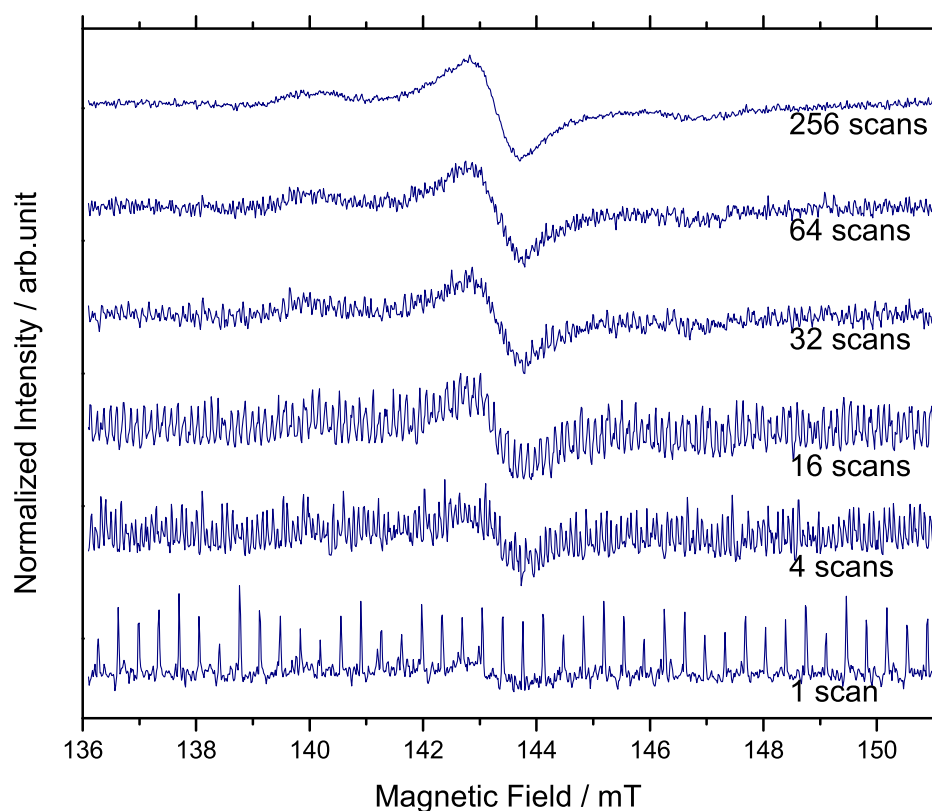
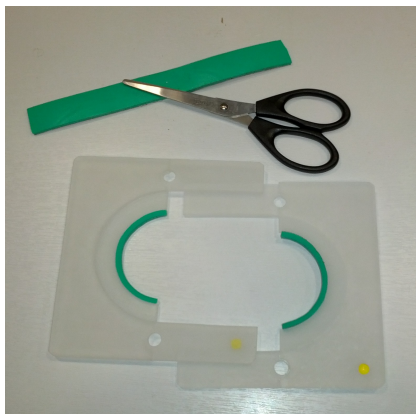
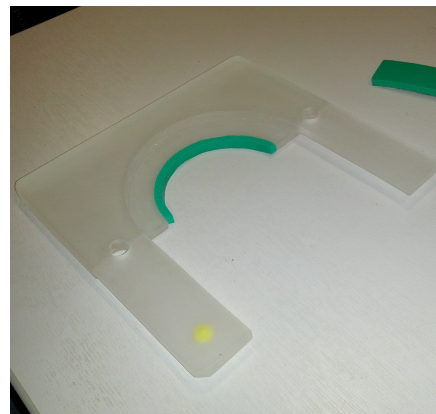


Figure 1.10 Signal averaging affect on noise from the dispex system shown on a weak nitroxide labelled protein at S-band, 80 K

1.8 Vibration dampening



(a) Cut down tape to correct width and fit inside brace piece



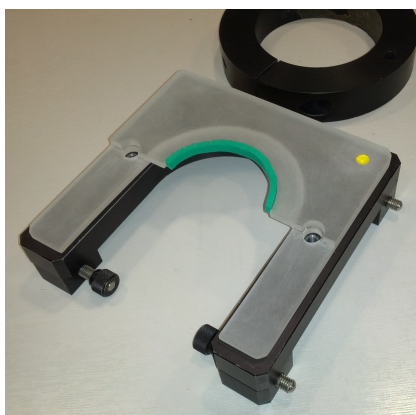
(b) One half of brace shown from side



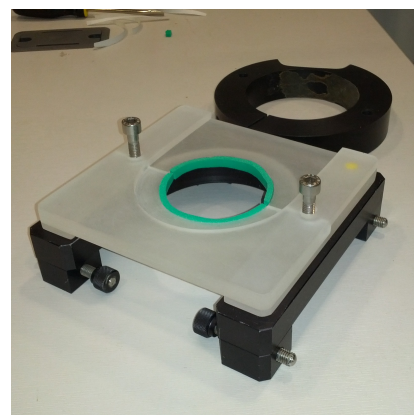
(c) Oxford flow cryostat with cryostat removed



(d) Remove top screws and top ring



(e) Attach bottom brace piece to cryostat holder



(f) Completed brace unit



(g) Brace in place - side view



(h) Brace in place - top view

Figure 1.9 Construction and installation of vibration brace clamp

1 Installation

1.9 Technical drawings

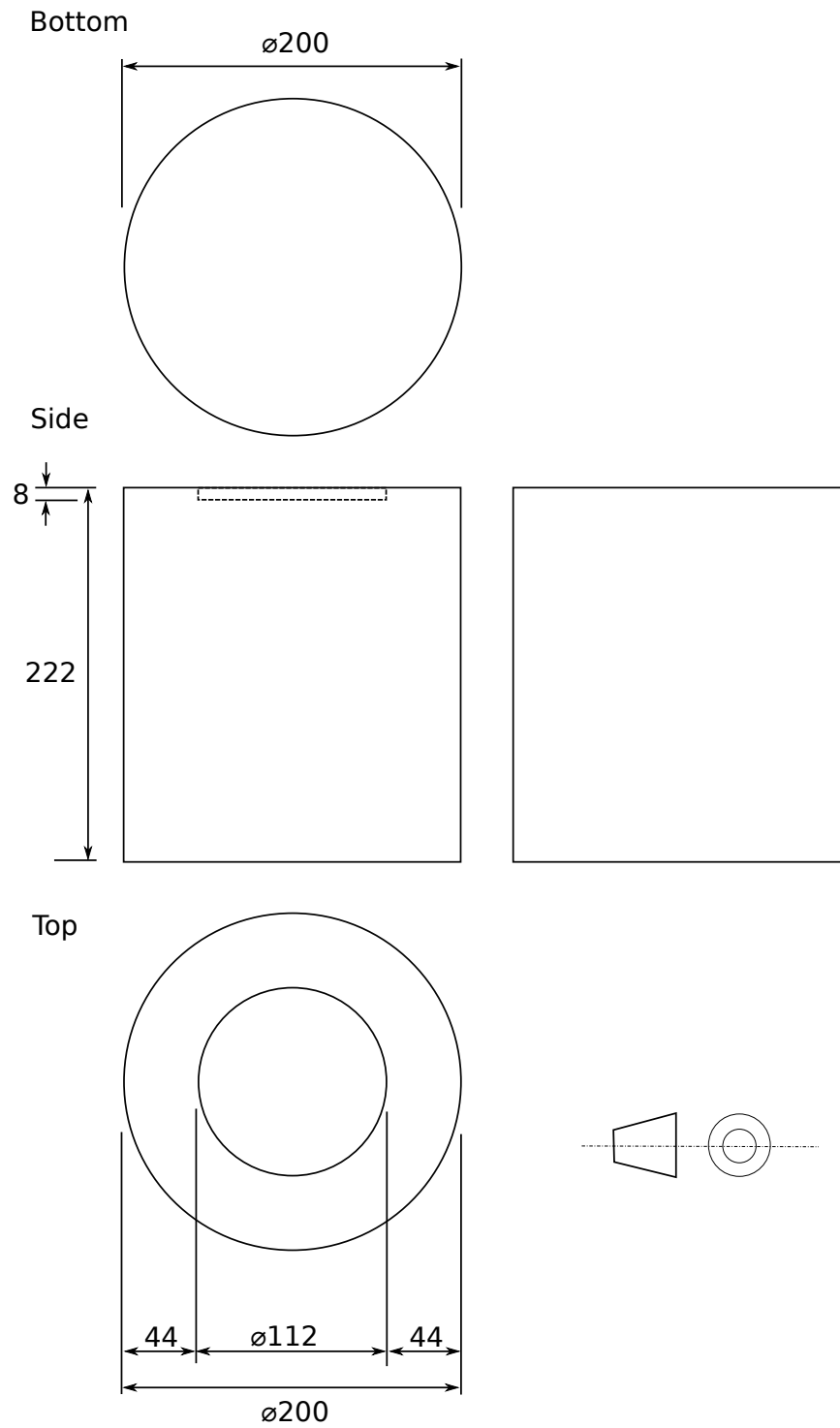


Figure 1.11 Technical drawing of the cryostat plinth

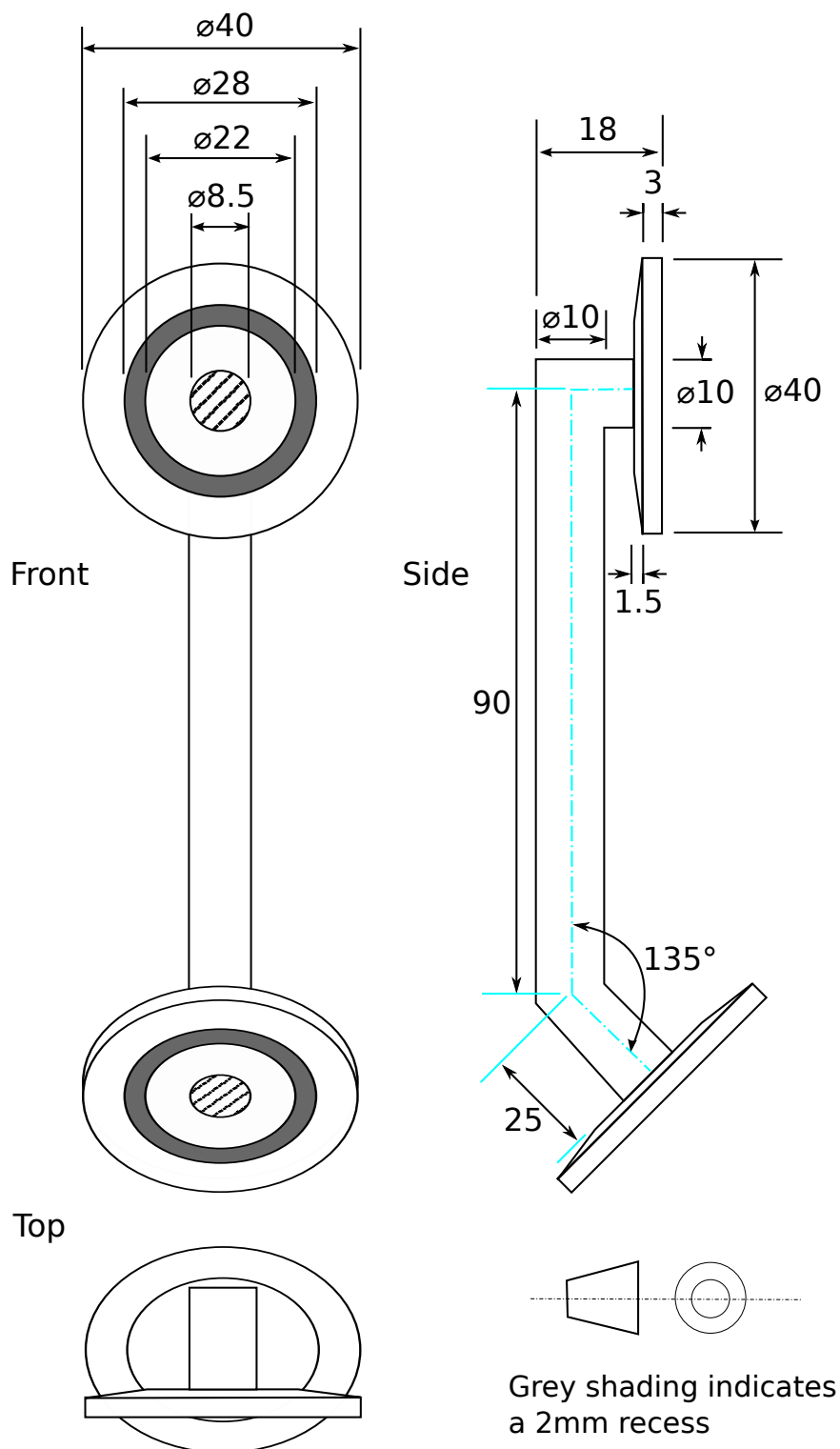


Figure 1.12 Technical drawing of the vacuum shroud adapter so that the cryostat can be attached to a vacuum pump inside the magnet

1 Installation

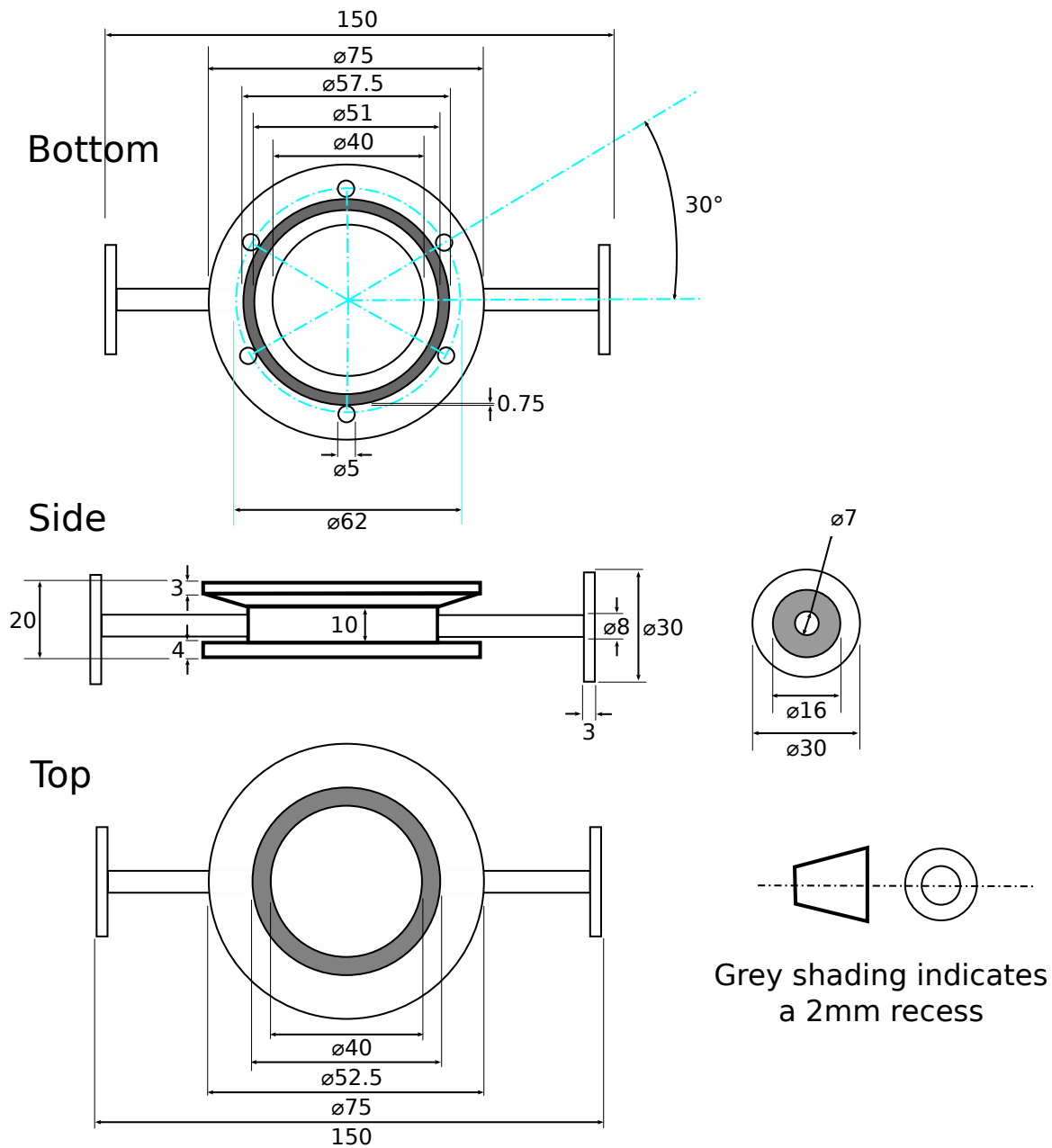


Figure 1.13 Technical drawing of the duplex header for conversion of the cryostat to a DN40 connection.

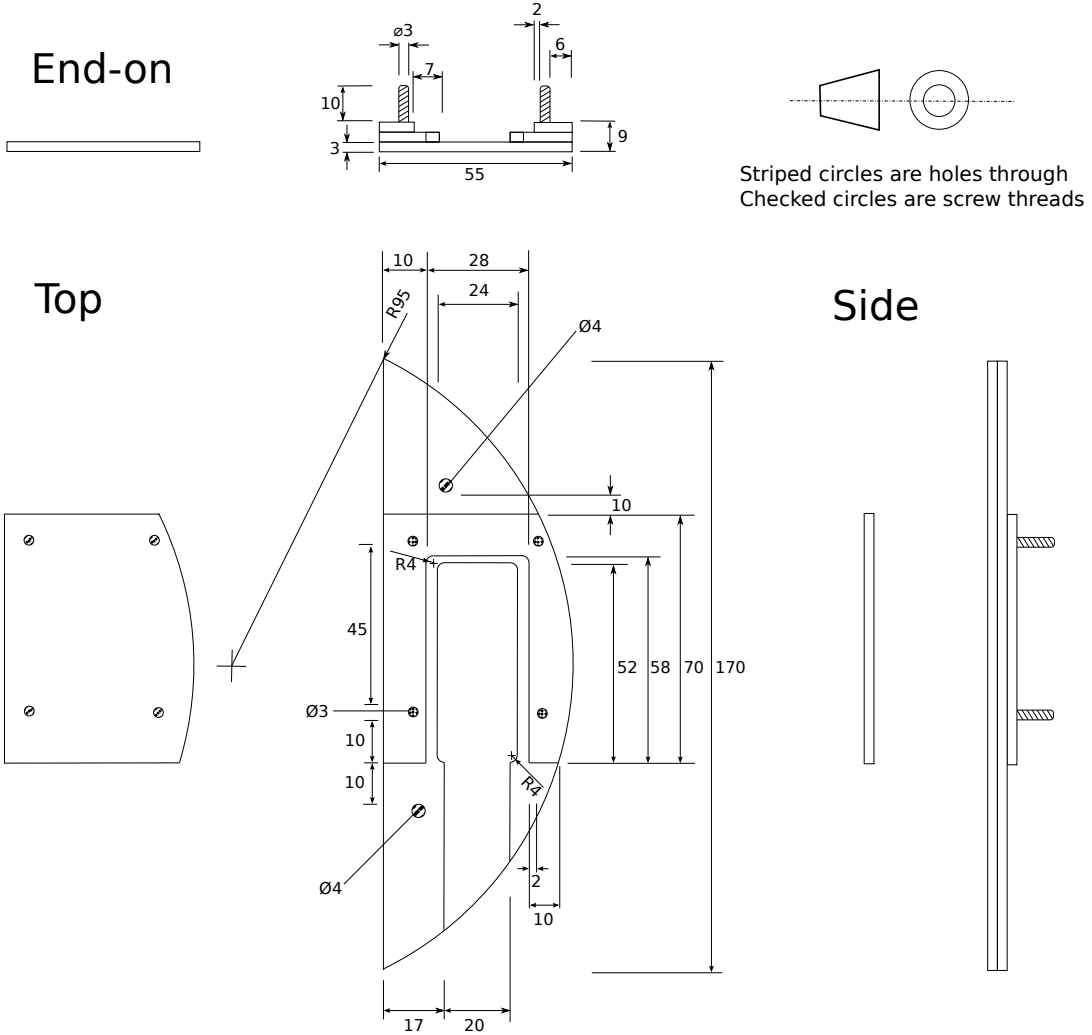


Figure 1.14 Technical drawing of the Hall probe holder

2 Calibration

2.1 Hall probe

Since the Hall probe has been relocated from within the rapid scan plates to a slightly different position upon the magnet shoe (and thus within the magnetic field), it requires to be calibrated again.

Ideally with the additional use of a Gaussmeter / Teslameter to accurately measure the magnet field, insert an EPR standard into the spectrometer and run a simple cw experiment to precisely find the field location of the standard centre line.

Using the spectrometers frequency counter and g-value of the standard calculate the expected peak position.

The relationship of magnetic field and radiation is given as

$$h\nu = \mu_B \cdot g \cdot B_0 \quad (2.1)$$

Where h is Planck's constant ($6.626 \times 10^{-34} \text{ m}^2 \cdot \text{kg} \cdot \text{s}^{-1}$), ν the frequency of radiation, μ_B the Bohr magneton ($9.274 \times 10^{-24} \text{ J} \cdot \text{T}^{-1}$), g the g value of the standard and B_0 the magnetic field. Rearranging equation 2.1 gives:

$$\text{Expected magnetic field} = \frac{6.626 \times 10^{-34} \times \text{Microwave frequency}}{g_{\text{standard}} \times 9.274 \times 10^{-24}} \quad (2.2)$$

Calculate the field offset by taking the experimental magnetic field from the expected magnetic field. In Xepr, go to **Acquisition** menu and select **Spectrometer Configuration**. In the new window click on the **Field** tab and insert the calculated value into the **Field Offset** field.

2.2 Pressure testing

ARS in their documentation state that the displax units should be operated at a 50 *mbar* helium environment. Helium is required as at the very low temperatures (< 20 K) expected in the normal operating conditions normal laboratory gases such as nitrogen, oxygen, carbon dioxide, compressed air, argon would all condense and freeze forming a solid layer over the cold plate surface and minimise heat transfer.

Helium pressure needs to be carefully regulated within the displax cryostat as the helium is the primary thermal energy transfer agent between the sample and cold plate. Too low a

2 Calibration

pressure and heat transfer will take several hours to cool the sample. Too high a pressure will allow for thermal transfer energy from through the resonator to the cold plate and raise the base temperature.

2.2.1 Empty cavity testing

Upon initial investigation it was indeed found that the displax units cool the cold plate at an optimum efficiency of 50 *mbar*. Higher pressures of helium allow for a much increased thermal energy transfer from the blanking plate on top of the cryostat; these convection currents mean that the cold plate never reached temperatures below 20 *K*. Pressures below 50 *mbar* whilst being achievable are difficult to accurately control using standard lab equipment and required the use of a turbo pump. Additionally the further dropping of pressure provided little additional drop in base temperature due to the very little amount of gaseous helium left to transfer thermal energy.

2.2.2 Leak testing

Fundamentally, trying to test the displax cryostat with a resonator in place at low pressures causes problems from leaking. Operating at the recommended 50 *mbar* means that the inside of the cryostat is at $1/20$ th of atmospheric pressure and represents a large pressure gradient.

S- and X- band resonators with their semi-flexible microwave connections showed initial leaking of > 50 *mbar/min* from 50 *mbar*, with a reduction to ~ 10 *mbar/min* at ~ 400 *mbar*. Q-band resonators with solid waveguides could only achieve 50 *mbar* after extensive pumping and leaked at a rate of over > 200 *mbar/min* with stabilisation at 1000 *mbar*, even with careful placement and replacement of mylar sheaths.

Clearly, leaks needed to be minimised. The following steps were required:

- ◇ Replacement and vacuum greasing of the O-ring between cryostat and cryostat header
- ◇ Replacement and vacuum greasing of the O-ring between cryostat header and resonator
- ◇ Replacement and vacuum greasing of the O-ring at the entry port to the resonator
- ◇ Slight greasing of entry port, as the entry port blanking piece is slightly thinner than a sample stick

2.2.3 Gas desorption

Despite all the measures taken for leak testing, it remained very difficult to establish a stable under pressure within the cryostat. In the testing with S- and X- band resonators the pressure stabilised around 400 *mbar* after being pumped down to 50 *mbar*. We needed to test whether 400 *mbar* represented a pressure gradient low enough between atmosphere and cryostat not to leak further or whether there was no leak and that the 350 *mbar* increase in pressure was from gas desorption of the cryostat walls.

In the first test under air the cryostat was pumped down to 50 *mbar* on a rotary vein pump, and the pressure traced for a minute. As expected it raised ~ 10 *mbar/min*, after the minute the cryostat was pumped back down to 50 *mbar* and monitored. This time the leak rate reduced to ~ 8 *mbar/min*. After a further 4 pump downs, the pressure remained steady at 50 *mbar* for the minute. Leaving the cryostat overnight, the pressure crept up to 200 *mbar*.

The second test was to leave the cryostat overnight attached to a rotary pump. After 15 hours the cryostat was isolated from the pump and the cryostat pressure returned to 50 *mbar* with gaseous helium. Over the next 24 hours, the cryostat showed no increase in pressure. The same effect could be achieved using a turbo-pump in a much shorter time.

Further testing with helium pressurization shows that helium does not adsorb to the walls of the cryostat. For this reason, when the cryostat is not in use it should be stored under atmospheric helium to minimise air adsorption to the walls of the cryostat.

If the cryostat is unused for any extended period of time then it should be pumped on with a rotary pump for at least 12 hours (overnight) before use.

2.3 Pressure optimisation

Charles's and the ideal gas laws tell us that as the temperature of a closed system changes the volume will also change or where the volume cannot change the pressure of the container will. Thus, it follows that a pressure set for the cryostat at room temperature will drastically reduce as the cryostat is cooled according to the rearrangement of the ideal gas law

$$P = \frac{nRT}{V} \quad (2.3)$$

Where, P is the pressure of the gas, V is the volume of the gas, n is the number of moles of gas, T is the temperature of the gas and R is the ideal gas constant (8.31 $J.K^{-1}.mol^{-1}$ (3 sf)).

Taking the volume of the cryostat space to be $\sim 4.89 \times 10^{-4} m^3$ (350 *mm* height \times 42 *mm* diameter). A room temperature pressure of 50 *mbar* becomes 1.68 *mbar* at 10 *K* assuming an ideal gas. For this reason the gas pressure inside the cryostat needs to be carefully monitored once the cold plate is down to temperature, so that the experimental pressure reflects the desired.

The further addition of any helium, will be warm helium and so cannot be accurately judged when adding. It is therefore recommended that the cryostat is cooled at a deliberately higher pressure, allowed to cool to temperature before being pumped upon. The pumping will result in a slight change in temperature but the result is negligible.

2.4 Final calibration

Following optimisation of parameters a final calibration experiment was conducted in which the set temperature of the LakeShore, reflecting the cold plate, was compared to the temperature experienced at the sample. Temperatures were recorded using a pre-calibrated Cernox semi-conductor from LakeShore. Details of the Cernox assembly construction can be found in section 1.7.

Cryogenic temperatures ($< 80\text{ K}$) were allowed to equilibriate for 300 s before any measurements were taken, at which point Cernox resistance and pressure were monitored for 30 s to ensure that they were both stable. For above cryogenic temperatures the Cernox resistance and pressure were not stable after the 300 s and so were given 1200 s before confirming stability and taking a measurement. These results hold true for the Bruker pulsed ER 4118X-MD5 resonator.

Set / K	LakeShore			Cernox		Pressure Gauge
	Sensor A / K	Sensor B / K	Heater / %	Resistance / Ω	Temperature / K	Pressure / $mbar$
8.0	7.94	8.18	3	2870.0	9.12	126.9
9.0	8.95	9.18	30	2523.0	10.14	138.2
10.0	9.99	10.24	45	2253.0	11.14	148.9
11.0	10.98	11.22	60	2042.0	12.13	158.4
12.0	11.98	12.24	75	1872.0	13.06	167.9
13.0	12.99	13.26	85	1718.0	14.10	176.9
14.0	14.00	14.28	88	1604.0	14.99	186.2
15.0	14.98	15.26	94	1492.0	15.98	194.7
17.5	17.48	17.73	30	1278.0	18.45	214.5
20.0	19.99	20.18	31	1125.0	19.78	232.6
25.0	25.01	25.11	37	895.0	25.63	265.6
30.0	30.00	30.16	39	741.6	30.71	296.8
35.0	35.00	35.20	41	633.0	35.35	324.1
40.0	40.00	40.22	42	551.3	40.41	348.8
45.0	45.00	45.22	44	488.0	45.20	371.8
50.0	50.03	50.02	45	444.0	49.60	393.7
55.0	55.00	55.18	45	398.4	54.94	422.5
60.0	60.00	60.14	46	365.1	59.54	449.7
70.0	70.02	69.97	48	320.7	67.34	486.3
75.0	75.00	74.97	48	292.9	73.32	511.0
80.0	80.00	79.00	49	274.1	77.94	529.5
85.0	85.00	84.73	50	264.0	80.88	544.1
90.0	90.01	89.70	50	245.0	86.89	563.0
100.0	99.87	99.16	53	227.0	93.31	588.7
150.0	150.00	148.08	63	149.7	138.54	703.0
293.0	292.94	293.37	0	72.4	278.00	1062.0

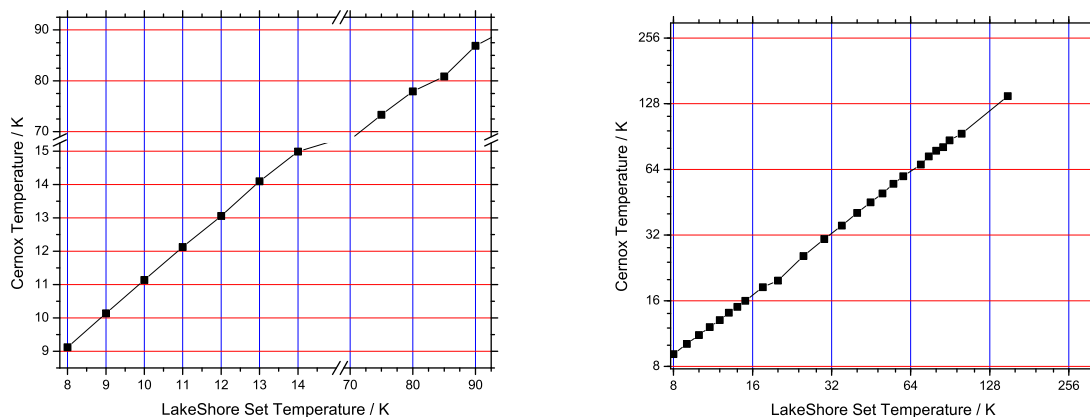
Table 2.1 Final calibration data table for the pulsed X-band resonator. The LakeShore columns show the set temperature, readings from the 2 sensors inside the displax unit and the extent of heater current at each temperature (temperatures in italics show the "Medium" heater range (8 - 15 K), normal type indicates the "High" heater range. The Cernox columns show the experimentally measured resistance across the semi-conductor as well as the transformed temperature.

2.4.1 Temperature offset

As previously explained, the temperature at the sample position will never completely match that of the LakeShore sensors within the cold plate of the cryostat due to a slight

temperature gradient across the cryostat and the fact that the sample is reliant upon the passive cooling of cold helium removing the thermal energy.

The LakeShore had temperatures set such that areas of particular experimental interest (<15 K, ~50 K and ~80 K) were sampled more in the hope of giving a better local fit.



(a) Key low temperature ranges (<15 K and ~80 K), with linear scale (b) Wide temperature range, with \lg_2 scale

Figure 2.1 Temperature offset between that set on the LakeShore and that measured at the sample

Fitting of the data yields $f(x) = 0.9242x + 2.5535$ ($R^2 > 0.999$) showing that across the cryogenic range the relationship between the LakeShore set and sample temperature is linear and almost 1:1 with an offset of 2.5 K.

2.4.2 Temperature sensors and pressure

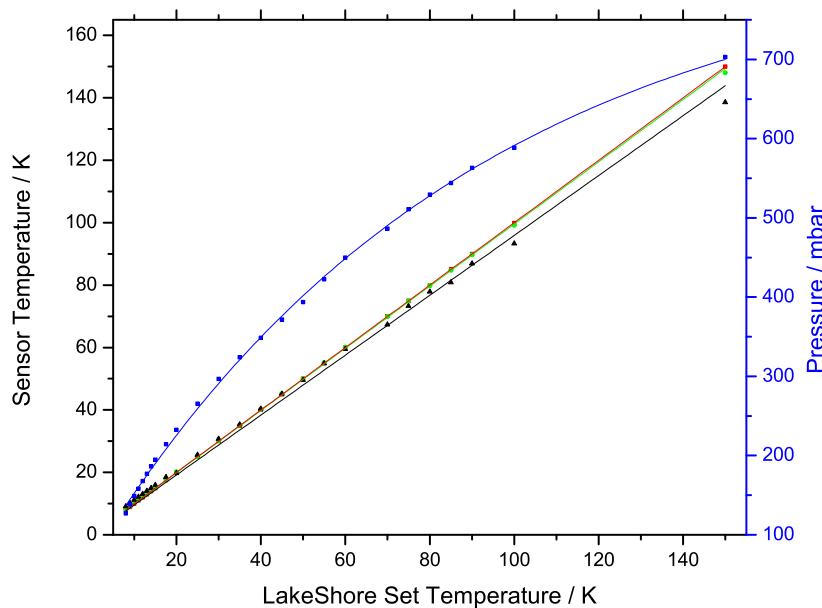


Figure 2.2 The temperature from different sensors and pressure against the LakeShore set temperature. In black the Cernox temperature, in red LakeShore Sensor A, in green LakeShore Sensor B, in blue the pressure. The temperatures are fitted with a straight line, whilst the pressure is fitted with a first order exponential.

Fitting of the data yields Sensor A $f(x) = 1.0000x - 0.0128$, Sensor B $f(x) = 0.9889x + 0.4971$, Cernox $f(x) = 0.9242x + 2.5535$ (4 dp) and pressure $f(x) = 70.96 + 387.57 \exp^{-x/89.96}$ (2

2 Calibration

dp) ($R^2 > 0.995$ for all). This figure highlights the experimental range of certainty, as temperatures rise above the cryogenic a greater discrepancy between the LakeShore sensors within the cold plate and the sample is seen. This could be as a result of the increased pressure of helium within the cryostat allowing for more transfer of thermal energy from outside the cryostat to the sample.

The pressure recordings clearly show a curve that fits an exponential very well, and not the linear relationship expected for an ideal gas.

2.4.3 Cooling time

The final experiment was to simulate a sample insertion or sample change. Samples will be held in liquid nitrogen before being put into the cryostat to pre-cool the sample to 77 K, this ensures that the sample is flash frozen before entry and minimises the heat load put onto the cryostat upon sample insertion. It is important that any sample has reached thermal equilibrium before any experiment is conducted as physical magnetic properties are heavily influenced by temperature.

For this experiment the cryostat was set to base temperature, representing the largest temperature difference possible for the set up.

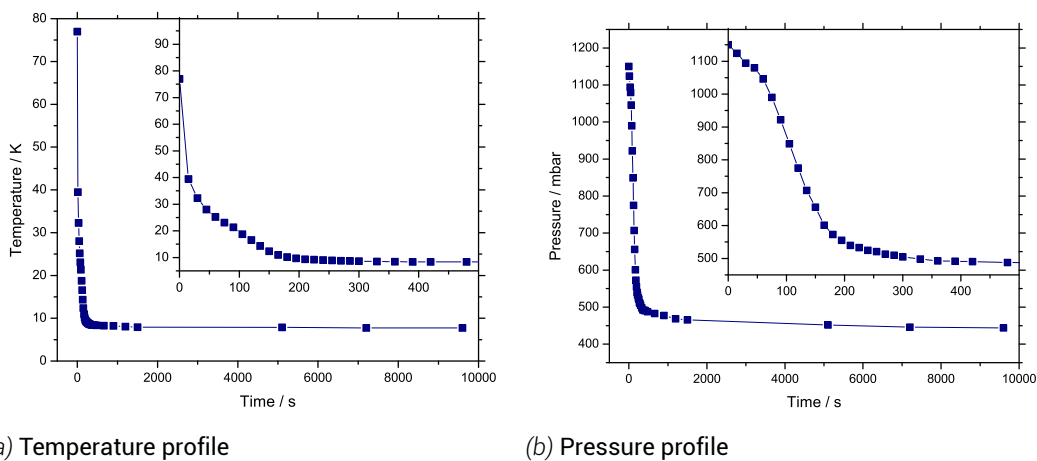


Figure 2.3 Cooling profile of a simulated sample from liquid nitrogen into a cryostat running at base temperature, shown from point of isolation of the system from the helium over-pressure immediately after sample insertion

The figures show that the temperature of the system drops following an exponential decay fashion whilst the pressure shows more of a reverse sigmoid profile. Suggesting that the thermal energy in the sample is sufficient to keep the helium pressure high, and as the energy in the sample is removed the helium pressure begins to drop rapidly. Both temperature and pressure being to equilibrate after 300 s (5 minutes), with no further changes after 400 s.

3 Operation

Warning: dispex units should never be in operation filled with anything other than high purity helium

3.1 Installing a resonator

With the dispex unit in place, prepare the resonator. S- and X-band resonators can use semi-flexible waveguides and are not a problem. For Q-band resonators assemble the waveguide before installation

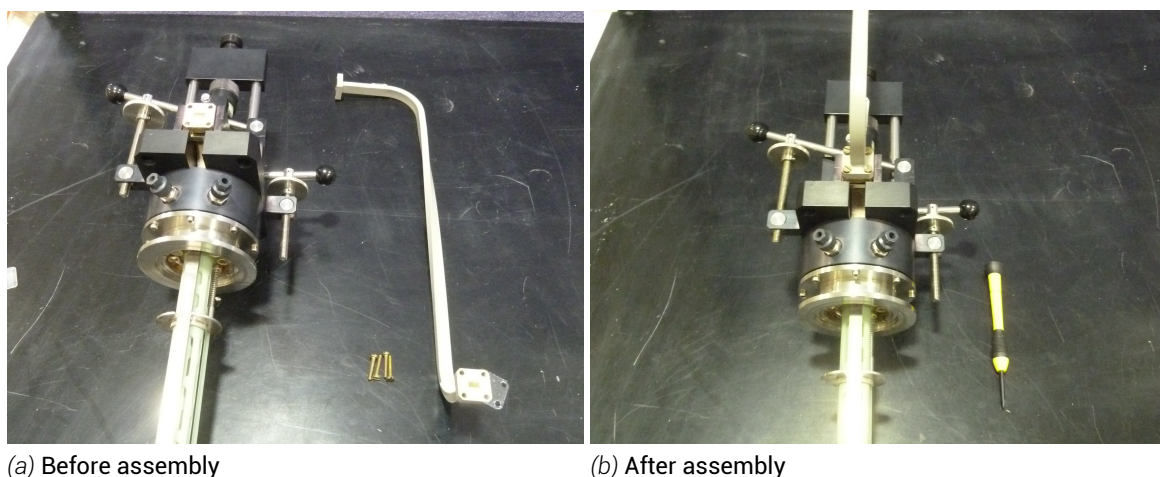


Figure 3.1 Assembly of the waveguide to a Q-band resonator

Bruker resonators are not (despite appearances) circular, instead being elliptical in one direction; with windowed dispex systems (for use with optics/lasers) it is important to rotate the resonator so that the black block (containing the modulation coils) on the bottom so they can pass the window lip, whilst some resonators will be unable to be used in their standard position.

3.1.1 S-band

The S-band resonator is the easiest of resonators to install as the resonator is the smallest and requires no rotation and can be simply slide straight into position.

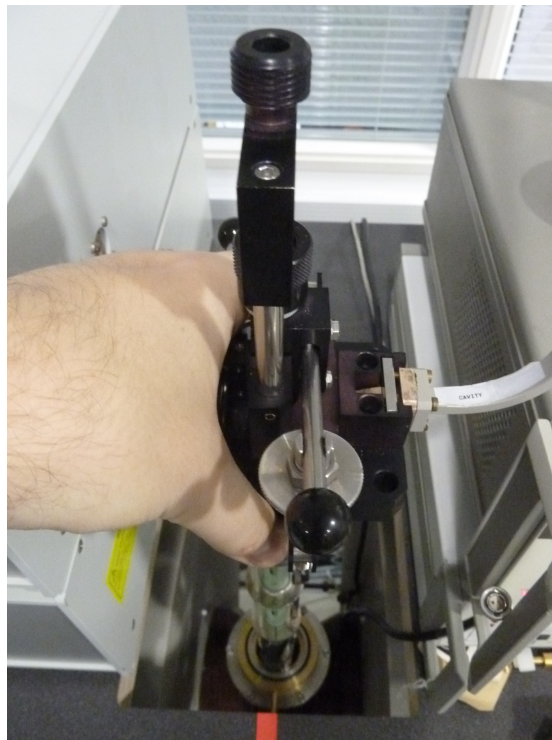
Attach the clamp and tighten, so that the O-ring is compressed. Attach the semi-flexible waveguide. Finally attach the modulation cable to the resonator.

On the console's Microwave Bridge Controller set the pins to (pins 1-4 down, down, up, up).

3 Operation



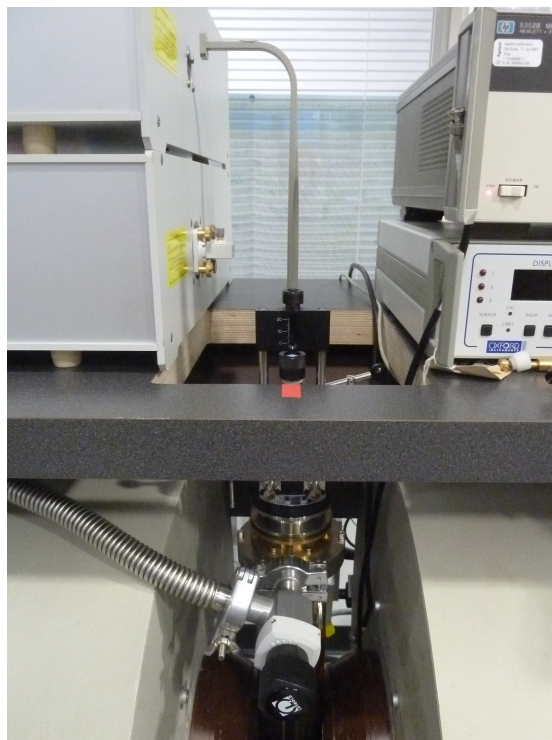
(a) The bottom of the resonator



(b) Insertion at 90° clockwise



(c) Turning back 45° at the window



(d) Turned completely back at full insertion

Figure 3.2 Insertion of a Bruker Q-band resonator (Bruker ER 5106 QTW)

3.1.2 X-band

The X-band resonator needs to be inserted at 90° anticlockwise until past the window, at this point the resonator needs to be rotated a further 90° anticlockwise such that it is rotated 180° from the normal flow cryostat position (with modulation and temperature connections at the back).

Attach the clamp and tighten, so that the O-ring is compressed. Attach the semi-flexible waveguide. Finally attach the modulation cable to the resonator (remember that the resonator is now rotated 180°).

On the console that the Microwave Bridge Controller has the pins set to (pins 1-4 down, down, up, up).

3.1.3 Q-band

The Q-band resonator is the trickiest of the resonators to install (and also the most expensive) and is shown pictorially in figure 3.2. It is recommended to attach the waveguide to the resonator before insertion into the displer unit, as once in place access to the back of resonator is very difficult.

The resonator is inserted at the 90° clockwise position. At the window, turn the resonator anticlockwise 45° and lower to the bottom. At the bottom the resonator can be rotated a further 45° anticlockwise so as to return to its normal, front facing position.

Attach the clamp and tighten, so that the O-ring is compressed. Attach the waveguide to the bridge. Finally attach the modulation cable to the resonator.

On the console that the Microwave Bridge Controller has the pins set to (pins 1-4 down, up, up, down)

3.2 Turning the displer unit on

Ensure that all O-ring connections are tight and air tight

Pressurize the helium bottle valves to approx. 0.5 *bar*

Open the helium bottle valves to a gentle flow, ensure that there is an overpressure on the flow gauge

Open the displer isolation valve for 2-3 *s* filling the cavity space

Shut off the helium at the needle valve, and turn on the vacuum pump

Pump down the system to under 50 *mbar*

Turn off the pump and open the needle valve and allow to fill to 1000 *mbar*

Repeat, pumping and flushing of system 5 times to ensure high purity helium inside sample space

Mostly close the fine control on the flow gauge, turn on vacuum pump and pump down sample space to exactly 50 *mbar*

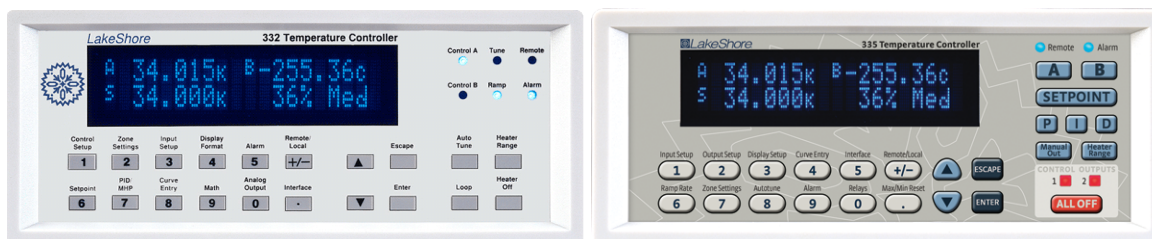
3 Operation



Figure 3.4 The Advanced Research Systems ARS-2HW compressor, taken from arscopy.com

Close the sample space isolation valve

Turn on the intelligent temperature controller (LakeShore)



(a) The discontinued 332 front panel

(b) The new 335 front panel

Figure 3.3 LakeShore front panel, taken from lakeshore.com

Press **6**, "Set point" on model 332 or **SETPOINT** on the 335 model

Set temperature using the number pad and press **Enter**

Press **Heater Range**, using the arrow buttons select **High** and press **Enter**

The front panel should now read that A and B are room temperature ($\sim 298\text{ K}$), S the temperature entered and in the bottom right "0% High"

Move to the compressor

Ensure the **GAS SUPPLY PRESSURE** gauge is at $\sim 15\text{ bar}$. Do not continue if it is significantly lower than this

Ensure that the cooling water is flowing (open taps if necessary)

WEARING EYE PROTECTION: Press the green power button located at the top left of the compressor unit

The unit will very rarely start on first press. Typically the compressor requires up to 5 attempts to start. If it still will not start check the cooling water supply, the unit has a built in safety feature that measures the water pressure and temperature. The water must be below $35\text{ }^\circ\text{C}$ and free from air.

Upon successful powering on of the compressor the characteristic valved system noise will be audible and the **GAS SUPPLY PRESSURE** gauge will increase to ~ 20 bar.

Check the readings on the LakeShore to ensure that the heat stages are cooling

For a well constructed system during cooling at 200 K, isolate the vacuum shroud from the turbo-pump.

At approximately 200 K any remaining gaseous particles will condense to the cold side of the shroud and create a near perfect vacuum. At this point even a turbo-pump will not have enough draw to stop molecules being pulled back through the pump into the vacuum shroud. This can potentially damage the pump. Whilst 200 K does not need to be strictly adhered too it certainly needs to be done before 100 K (boiling points of common air components are: CO_2 195 K, O_2 90 K, N_2 77 K). For systems with even slight leaks in the vacuum shroud then the pump should be left open.

Wait for the displax unit to get to temperature by monitoring the LakeShore. 80 K \sim 90 min, 10 K \sim 2 h.

3.3 Changing samples

Ensure the spectrometer is in a state suitable for changing of samples (tune mode).

Open the helium bottle valves so that the system is under pressure, open the needle valve so that there is an helium overpressure.

Open the sample cavity isolation valve, allow pressure to raise from an underpressure to an overpressure, with the overpressure audibly releasing the overpressure release valve.

If unfamiliar with the system it is better to extensively overpressure the system to the extent that the overpressure valve on the can audibly be heard than risk getting moisture in the system. This will freeze your next sample in place and require warming of the system back to room temperature to remove.

Loosen the sample O-ring on top of the resonator (around the sample stick), and remove the previous sample or resonator blanking piece.

Slide the sample stick into the correct position within the resonator.

Close the sample O-ring on the resonator.

Close the sample space isolation valve and monitor the pressure inside the cryostat as it closely reflects the sample temperature.

Allow approximately 3 - 5 min for the sample to equilibriate to temperature before conducting any experiments.

In the process of changing samples the LakeShore will read a jump in approximately 10-30 K (depending upon the speed of switching). This is normal and due to the increased amount of (warm) helium in the sample space. As soon as the sample space has been isolated the LakeShore will stabilise again by adjusting the heater current.

NOTE: As samples are maintained in an inert helium atmosphere, the samples should never boil off liquid nitrogen on removal as would occur with traditional flow cryostats.

3 Operation

3.4 Turning the displax unit off

Remove the sample and replace with the resonator cap as per §3.3.

Close the sample space isolation valve. It is not necessary to pump down the cavity, but doing so will keep the displax unit cold for longer. Doing this ensures that no moisture enters the cavity whilst it is still cold, else upon warming this moisture would condense and sit on top of the copper heat stage and corrode it.

Go to the compressor and turn it off at the (green) main power switch. This should be accompanied by a hiss as the pressure drops.

For a slow warm up the Lakeshore can simply be turned off.

For a quick warm up (for a change of resonator for example) then the best way is to use the LakeShore's ramping facility.

On the LakeShore 332:

1. Change the setpoint (press **6**) to desired temperature (ie 80 K, 290 K). The ramp will only go to the set point.
2. Press **1** "Control Setup"
3. Press **Enter** (5 times) until the display reads "Select for Loop 1, Setpoint Ramp Off"
4. Press the **Up arrow** so that the "Setpoint Ramp" is "On", press **Enter**
5. Enter the ramp rate in Kelvin/minute on the number pad and press **Enter**. It is not recommended to set this above 10 K/min as rapid heating could be dangerous to the heater coil and cause rapid expansion in the displax unit.
6. To cancel the ramping press **1** "Control Setup" and immediately press **6** "Setpoint", this pauses the ramping. To completely turn it off press **1** "Control Setup", press **Enter** (5 times) until the display reads "Select for Loop 1, Setpoint Ramp On", and turn it off.

On the LakeShore 335:

1. Change the setpoint (press SETPOINT) to desired temperature (ie 80 K, 290 K)
2. Press **6** "RampRate"
3. Enter the ramp rate in Kelvin/minute on the number pad and press **Enter**. It is not recommended to set this above 10 K/min as rapid heating could be dangerous to the heater coil and cause rapid expansion in the displax unit.
4. To cancel the ramp press **6** "RampRate" and set it to 0 K/min

Finally turn the LakeShore off, the power switch is at the back (top right when viewed from the front).

4 Troubleshooting

During the course of installation and running of the dispex units you will come across a couple of common problems. The following section hopes to address and solve some of these frequent problems.

Why can I not see any signal?

First of all check your sample, check it in a standard room temperature machine and see that there is a signal there. If there is no signal, is the modulation cable attached to the resonator.

If you are confident in your sample, then check your sample height. During the isolation and cool down, if your sample was not completely secure then it may have been sucked out of the holder and into the sample space. If adjusting your sample height shows no movement in your tuning dip then this is likely your problem.

Remove your sample, if the sample is still there then try insertion, otherwise you need to allow the sample space to warm to at least 80 K, remove the resonator and recover your sample. Make sure the dispex unit and resonator are completely dry before reinserting the resonator and cooling down again.

The magnet power supply keeps overloading

The magnet power supply is controlled via direct feed back from the Hall probe. Check that the Hall probe is still between the poles of the magnet and has not fallen off.

Samples keep freezing inside the resonator and cannot be removed

Ramp up the LakeShore to 80 K and quickly remove the resonator from the sample space as you turn off the heater on the LakeShore.

As the resonator is removed does it drip? Does the LakeShore heat stages hold at 80 K for a few minutes despite the system being open? If you look inside the sample space is there any liquid boiling?

If yes, this is liquid nitrogen.

The cause of this is the O-rings and seals to the dispex unit are not vacuum tight. Each time the sample space is pumped down the underpressure is sucking in air, which is then condensing and covering the base of the sample space. This effect is cumulative so often only your 3rd or 4th sample will freeze in place.

If this only happens after many sample changes then your helium flow is not high enough during sample changes, or the sample tubes have condensation over them when inserted into the resonator.

The dispex unit will not (/ is taking a long time) to get down to temperature

This is usually caused by the sample space containing some amount of nitrogen or air, as the dispex unit cools the gases will first condense and pool at the bottom of sample

4 Troubleshooting

space as cooling continues the pool will freeze. This is potentially damaging for copper heat stage and insulates the sample space from the cold.

Ramp up the LakeShore past 80 K, this should boil off any nitrogen, then pump-flush the sample space with dry helium before pumping down and cooling down again.

The vacuum pump will not start

The vacuum pump has a safety feature where it will not spin up if the pressure is too low as it may damage the pump.

Open the line to the helium cylinder and return it to atmospheric pressure (1000 mbar). Start the pump.

After opening the needle valve it takes a long time for the pressure to rise

Are the other valves on the helium bottle open? Is the helium bottle empty (as shown by zero pressure on the inlet pressure gauge)? Has the hosing between the bottle and system come lose?

The compressor will not start

First check that the compressor is plugged in and turned on

Is the cooling water on? Are the taps open? Is the water cold? Is the system pressurized?

Are the helium lines securely attached?

Finely check the compressor fuses and the number of operational hours. If the operational hours are ~ 10,000 hours then the compressor needs a service. Consult the manufacturers instructions.

After pumping down the system there is a high pitched whine

After flooding the sample space with helium occasionally the overpressure valve can get caught and stay slightly open. This can be verified by a lose in underpressure from the sample space.

Simply pull on the overpressure valve and release, hopefully it should relocate into its normal resting position.

The entire cryostat has visibly frosted up on the outside

This will occur when the vacuum shroud between the cavity space and outside world is not at high vacuum. This will also be seen on the LakeShore, as the additional heat load as heat enters the cavity space will cause the LakeShore struggle to maintain the temperature and may have crept up.

As we do not know the vacuum level, ensure that the turbo on the vacuum pump is off and open the vacuum shroud to the vacuum pump. Once the vacuum pump has reached a good vacuum again turn on the pump and turn the turbo back on, you can continue your experiment with the turbo pump open.

If a good vacuum cannot be achieved, immediately stop the experiment and check the fixtures between the vacuum shroud and pump. If these appear fine and but the problem persists try the tubing and connectors on another system with a blanking plate to simulate the cryostat. If they are fine, consult the manufacturer, it may be that there is a small leak and that the cryostat simple needs to be run with the turbo pump constantly.

The LakeShore seems to have no effect on temperature

Make sure that the LakeShore has a temperature set (the S value on the main display) and that the heater is turned on (the main display does not show "Htr Open"). If the LakeShore appears to otherwise be operational check that the connection to the displax unit is firmly attached and clipped into position.

The LakeShore frequently displays the error "Htr Open"

The LakeShore 332 has a protection circuit that monitors the heater current and voltage and shuts down the output when it detects a short or resistance that is less than 10 Ω or an open condition which is around 400 to 500 Ω .

First make sure that everything with the cryostat is fine.

- ◇ Check the pressure in the cavity space and ensure that is approximately correct or at least dropping. If it is rising there is a leak and the heat load is causing the open.
- ◇ Feel the outside of the cryostat. The vacuum shroud should maintain that the outside of the cryostat is at room temperature. It will be a little cold, but should be approximately the same as the magnet. If it is cold then the vacuum shroud is not at high vacuum. Follow instructions for "The entire cryostat has visibly frosted up on the outside".
- ◇ Check that the connections between LakeShore and cryostat are firmly in place.

If the equipment appears fine, have you recently set a new temperature that is far from the previous temperature? Having the heater on 100 % for more than a few seconds will safety it.

The LakeShore can be prone to spikes on movement. For example if the LakeShore has been on for a week with the heater at 29 % and the temperature is changed that a change to the heater to 30 % may cause a spike and safety the heater open.

This can be tested by slowly stepping up the temperature and seeing if the error comes as a result of crossing a particular temperature or heater current. Regretfully, thermal opens are very difficult to troubleshoot as they close up almost immediately after the heater current is removed.

With everything said above, it is possible that the protection circuit in the LakeShore is defective. You can test the circuit using a fixed 50 Ω / 50 *W* resistor in place of the heater. Set the setpoint above the control temperature and let it run until the output reaches 100 % and if it does not shut off within 30 *min*, the controller is OK. One word of caution is that the resistor will become very hot and you must use caution if you try this test.

5 Postface

Glossary

Base temperature the coldest possible temperature that the displer unit stabilises at

Acknowledgements

The following people were instrumental during the set up and calibration of the displer units at the Henry Wellcome Unit for Biological EPR at the University of East Anglia, UK

Morgan Bye - unit location, plumbing, wiring, calibration, system testing, documentation

Justin Bradley - initial set up and testing of units on arrival, cernox sensor construction, design of displer header

Tim Lane - UEA technical staff - addition of cooling water splitting and taps, construction of displer header

Fraser MacMillan - head of Henry Wellcome Unit for Biological EPR

Jeff Maynard - LakeShore Cryotronics service manager - for discussions on LakeShore safety circuits

Images

Images from the follow sites are used under a fair use policy and acknowledged in the text:

lakeshore.com

arscryo.com

Further information

Further information is available about displer units is available from the following sources

Morgan Bye - author of this document

morgan.bye@uea.ac.uk

Fraser MacMillan - head of the research unit

fraser.macmillan@uea.ac.uk

Bruker - EPR manufacturer

bruker-biospin.com

5 Postface

American Research Systems - displax unit manufacturer

arscryo.com

SHI Cryogenics Group - alternative displax unit manufacturer

shicryogenics.com

LakeShore - temperature controller products

lakeshore.com

Oxford Instruments - temperature controller products

cryospares.com

C. MISHAP manual

MISHAP has been designed and built to be as intuitive as possible for a novice user. It is intended that MISHAP be distributed with exemplary files and an user manual to facilitate understanding of the software as well as provide some explanation as to core modelling concepts and assumptions made in the modelling process.

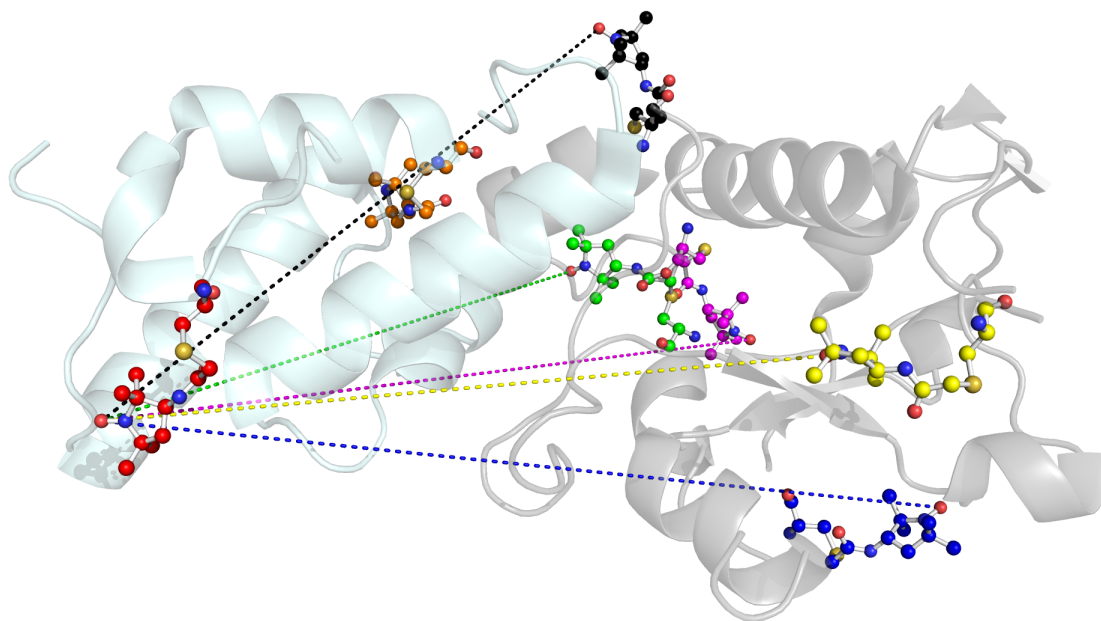
The following material is the intended MISHAP user manual to be distributed with MISHAP.

The MISHAP manual is split into 3 key sections:

1. A general introduction to MISHAP, including what MISHAP does and the problem to which it address. In addition, a getting started guide showing the installation and configuration of MISHAP and HADDOCK is included.
2. The main documentation. A discussion of main graphical user interfaces used in MISHAP (and MMM), including how each operates and common mistakes. Finally a description and discussion of HADDOCK parameter files and how MISHAP adapts them for the use of EPR.
3. A full worked example of using MISHAP in a docking run.

MISHAP

v13.08



Morgan Haydn Bye

This document was created during study for the degree of Doctor of Philosophy whilst at the University of East Anglia, UK.

17th August 2013

morganbye.net

...always accepting a worthy challenge

UEA University of
East Anglia
Henry Wellcome Unit
for Biological EPR

This document has been supplied on condition that anyone whom consults it is understood to recognise that its copyright rests with the author and that no quotation from this work, nor any information therefrom, may be published without the author's prior written consent.

"An expert is a man who has made all the mistakes which can be made, in a narrow field."

- Niels Bohr

Contents

1	General	1
1.1	Overview	1
1.1.1	What is MISHAP?	1
1.1.2	Functionality	1
1.1.3	Interface	1
1.1.4	Help	2
1.2	Getting started	2
1.2.1	Getting MISHAP	2
1.2.2	Installing MISHAP	2
1.3	Change history	3
2	User interface	5
2.1	MMM	5
2.1.1	Save rotamers from MMM	5
2.2	MISHAP Installer	7
2.3	MISHAP distance distribution window	11
2.3.1	Loading a distance distribution	11
2.3.2	Multiple spin systems	12
2.3.3	Protein information	12
2.3.4	Output location	12
2.3.5	RUN!	12
2.4	MISHAP PDB creator window	15
2.4.1	Load a PDB file	15
2.4.2	Load a MMM rotamer file	16
2.4.3	Output location	17
2.4.4	RUN!	17
3	HADDOCK configuration	21
3.1	Spin label definitions	21
3.2	Protocols	22
3.3	Topology parameters (toppar)	23
3.4	Run settings	23
3.4.1	CNS address	24
3.4.2	Topology selection	24
3.4.3	Parallelisation	24
3.4.4	Initial seed number	24
3.4.5	MD time step	25
3.4.6	Annealing temperature	25
3.4.7	Protocols	26
3.4.8	Structure generation	26
3.5	Run analysis	26
4	An example docking experiment	29
4.1	Example files	29
4.2	MISHAP	29
4.3	HADDOCK run generation	32
4.4	HADDOCK run analysis	33
5	Other tools	35
5.1	Naccess	35
5.2	HADDOCK processing tools	35
5.2.1	HADDOCK cluster processing	35
5.2.2	HADDOCK clusters to PyMOL	36
6	Postface	37
6.1	License	37
6.2	Third-party software	40
6.2.1	Software within MISHAP	40

Contents

6.2.2	MMM	40
6.2.3	HADDOCK	40
6.3	Acknowledgements	40
6.4	Hardware and computation time	41
6.5	References	41

List of Figures

2.1	Loaded protein model in MMM	5
2.2	Select a residue in the command line	6
2.3	Message board update showing the select residue(s)	6
2.4	MMM menu showing the site scan option	6
2.5	MMM site scan setup window	7
2.6	MMM site scan warning dialogue	7
2.7	MMM site scan labelling conditions window	8
2.8	MMM site scan save dialogue window	8
2.9	MMM menu showing MISHAP	10
2.10	MISHAP distance distribution window	11
2.11	MISHAP distance distribution load interface	11
2.12	MISHAP distance distribution window showing a loaded distance distribution	12
2.13	Flowchart of the MISHAP distance distribution window process	13
2.14	MISHAP distance distribution and PDB creator windows	15
2.15	MISHAP PDB creator load PDB file selection	16
2.16	MISHAP PDB creator with loaded PDB file	16
2.17	MISHAP PDB creator load MMM file selection	16
2.18	MISHAP PDB creator load with PDB and MMM file loaded	17
2.19	MISHAP PDB creator showing 2 proteins loaded with MMM rotamer files	17
2.20	Flowchart of PDB handling within the MISHAP PDB creator	18
3.1	Atom naming convention used in MMM and MISHAP for nitroxide spin labels	22
4.1	MISHAP distance distribution window	30
4.2	MISHAP distance distribution load interface	30
4.3	MISHAP distance distribution window showing a loaded distance distribution	31
4.4	MISHAP distance distribution and PDB creator windows	31
4.5	MISHAP PDB creator load with PDB and MMM file loaded	32
5.1	A typical figure from the <code>HADDOCK_CLUSTERS_PROCESSING</code> script showing HADDOCK score plotted against the RMSD of the structure from the average structure	35
5.2	A typical figure from the <code>HADDOCK_CLUSTERS_TO_PYMOL</code> script showing chain B (in grey cartoon) plotted against the top 4 structures of the top 4 clusters	36

1 General

1.1 Overview

1.1.1 What is MISHAP?

MISHAP is the open-source MMM *in silico* simulated spin label to HADDOCK program. It combines several MatLab scripts, functions, programs and user interfaces as well as a set of additional parameter files for HADDOCK.

MISHAP aims to take spin labelled proteins and the distance distributions between labels from *in silico* (MMM) calculations or experimental data (from pulsed electron-electron double resonance experiments which have been analysed with DeerAnalysis) and pass them to the biomolecular docking program HADDOCK as docking restraints.

1.1.2 Functionality

MISHAP provides the ability to generate HADDOCK compatible PDB files of single chains of amino acids where one or more residues have been substituted for nitroxide spin label attached cysteine residues.

MISHAP can generate distance based docking restraints from pulsed EPR experiments (PELDOR/DEER) when they have been analysed with DeerAnalysis or MMM.

The generated PDB files can be used in conjunction with or without the MISHAP generated distance based docking restraints to help determine the docking of protein complexes.

Currently, MISHAP only supports the generation of models and restraints in proteins. Though the open-source nature will allow for the expansion of MISHAP to accept non-nitroxide spin labels, DNA, RNA and other common ligands.

1.1.3 Interface

The user interface of MISHAP initially consists of a simple installer that installs MISHAP as well as modifying any existing installations of MMM and HADDOCK.

After installation the user interface largely consists of 2 windows. The first handles the distance restraints whilst the second handles the generation of HADDOCK compatible PDB files.

The changes to HADDOCK allow for the user to continue using HADDOCK normally via the command line with the addition of being able to use the residues R1A (MTSL attached cysteine) and IA1 (IA-PROXYL attached cysteine).

1.1.4 Help

MISHAP is designed to be as intuitive as possible whilst still maintaining user control. With this in mind, all options are clearly labelled as well as extensively documented in this document as well as online at morganbye.net/mishap

1.2 Getting started

1.2.1 Getting MISHAP

MISHAP is available online from 2 locations.

MISHAP general releases (for most users) should be downloaded from:

<https://sourceforge.net/projects/mishap/>

MISHAP developer releases (constantly updated and should be considered unstable) can be obtained from the git repository

<https://github.com/morganbye/MISHAP>

Alternatively, you can create a new git folder and pull the repository directly from the command line with:

```
git pull https://github.com/morganbye/MISHAP.git
```

1.2.2 Installing MISHAP

If MISHAP was downloaded from sourceforge then it will come as a ZIP file which includes everything required. Simply unpack the ZIP file into a directory of your choosing. The folder should contain several text files including; `LICENSE.txt`, `README.txt` and `UPDATE.txt`. There will also be `MISHAP.m`; the main MatLab file and 2 directories: `test_files` containing some example files (referenced later) and `_private` which contains all the files associated with MISHAP.

The extracted folder needs to be added to the MatLab path with subfolders to the; this can be done by navigating to the folder in the *Current folder* view, right clicking and selecting

Add to Path → with Selected folders and Subfolders

or by clicking from the main MatLab menu

File → Set Path

in the main MatLab window, then click *Add folder with subfolders* then *Save and Close*.

The MISHAP installer then needs to be run. This is detailed in §2.2, but simply in the MatLab command window type

```
MISHAP
```

This begins the installation of MISHAP.

1.3 Change history

Release v13.08 - BETA

v13.08 shows much improved error handling and allows for MISHAP generated PDBs be reloaded in the PDB creator and have additional spin labels attached.

Release v13.06 - BETA

v13.06 is the first publicly viable release of MISHAP and marks the first build from alpha. It shows far greater error handling, Windows compatibility and an improved graphical user interface. MMM and MatLab programming is complete but significant errors still occur in HADDOCK/CNS.

Build v13.05 - ALPHA

v13.05 is the first functional build of MISHAP that generates distance distribution constraints and MMM rotamer merged PDBs.

2 User interface

2.1 MMM

MMM is an open-source program for the visualisation, inspection and improvement of models of proteins and protein assemblies based on restraints from multiple experimental techniques, though primarily focussing on experimental techniques from electron paramagnetic resonance¹.

2.1.1 Save rotamers from MMM

Perhaps MMM's most useful feature is the ability to calculate the rotamer populations of nitroxide spin labels attached to proteins.

For MISHAP we require these spin label attached protein models. To generate a rotamer model first start MMM from the MatLab command window with the command

```
MMM
```

After the splash screen, load a PDB model using the *File* menu - more information available in the MMM manual §2.1.1.

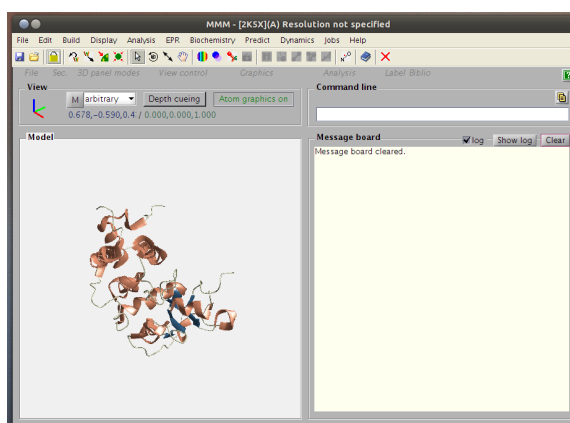


Figure 2.1 Loaded protein model in MMM

The protein model should be visualised in the Model panel whilst the Message board should update.

Next select a residue. To do this at the command line level type

```
select [structure](chain)residue
```

the structure option is not necessary depending upon the file (MMM §1.6). Residue selection can also be done graphically using:

¹MMM - <http://www.epr.ethz.ch/software/index>

2 User interface



Figure 2.2 Select a residue in the command line

Display → Hierarchy (MMM manual §2.3 and 4.1.2)

However, the command line is superior as it allows the selection of multiple residues at once.



Figure 2.3 Message board update showing the select residue(s)

Once a residue is select the message board will update with the selection. If a selection is made incorrectly use:

```
unselect [structure](chain)residue
```

Once happy with the selection we label the selected residues with:

EPR → Site scan/selected residues (MMM manual §2.1.7)

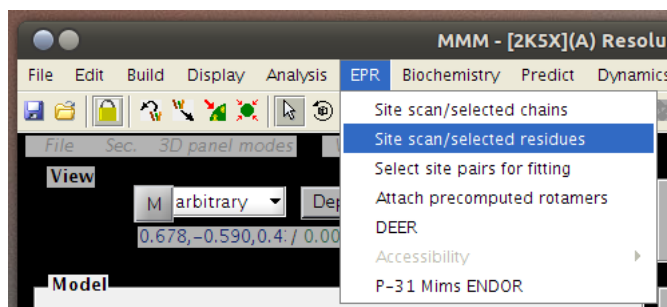


Figure 2.4 MMM menu showing the site scan option

This displays the Site scan setup window (MMM manual §4.1.2). It is imperative that *Save statistics* and *Save PDB rotamers* are selected.

Selecting *Save PDB rotamers* will display a warning message. Ensure that you click Yes.

Clicking *OK* will display the *Set labeling conditions for selected residues* window. Here select the appropriate nitroxide spin label and temperature.

Finally a prompt appears to set the save path of the rotamer analysis files

The rotamer analysis can be tracked in the message board and once complete a prompt appears to ask if the user wishes to view the results in a browser.

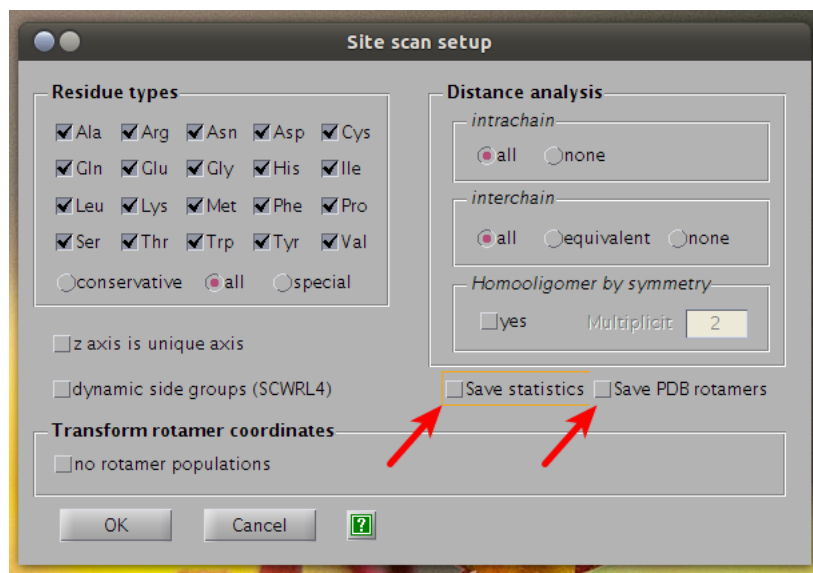


Figure 2.5 MMM site scan setup window

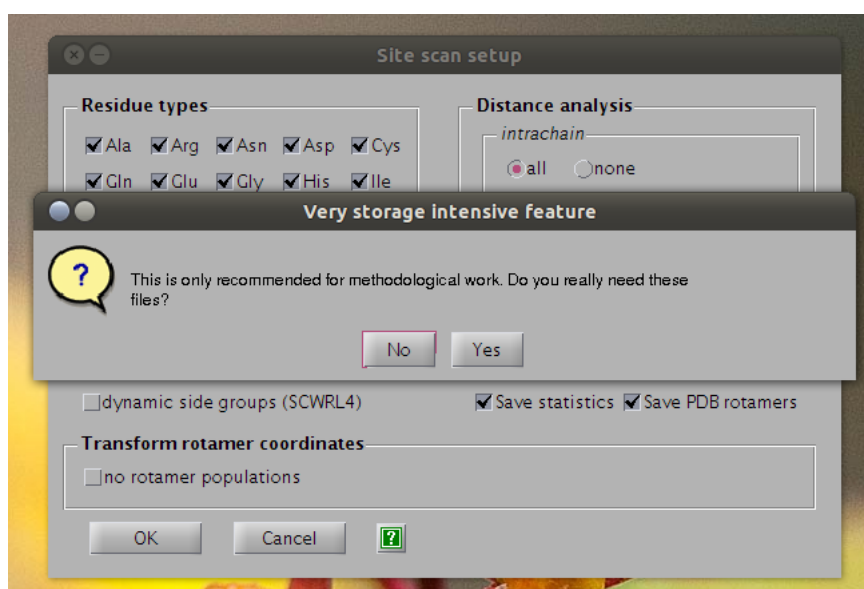


Figure 2.6 MMM site scan warning dialogue

2.2 MISHAP Installer

Installation of MISHAP is kept as much as possible within the MatLab environment for simplicity.

To install MISHAP, extract the MISHAP zip folder. Add the folder with subfolders to the MatLab path; this can be done by navigating to the folder in the *Current folder* view right clicking and selecting

Add to Path → *with Selected folders and Subfolders*

or by clicking from the main MatLab menu or by clicking

File → *Set Path*

in the main MatLab window, then click *Add folder with subfolders* then *Save and Close*.

In the MatLab command window now type

MISHAP

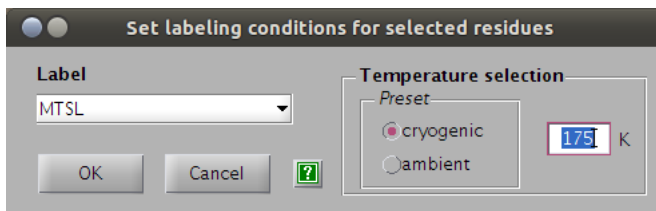


Figure 2.7 MMM site scan labelling conditions window

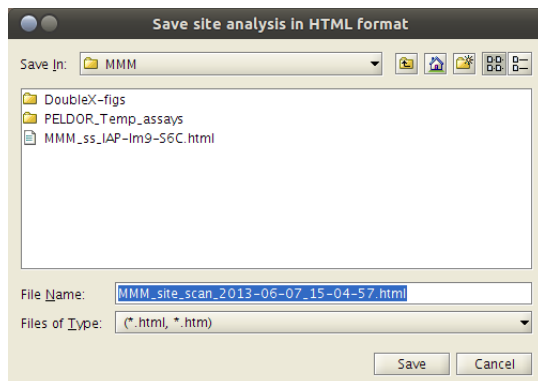


Figure 2.8 MMM site scan save dialogue window

This begins the installation of MISHAP. The installer addresses 3 different programs in turn. First the installer looks to see if MISHAP has previously been installed, if so then the installer aborts and simply loads MISHAP.

If MISHAP has not previously been installed then the installation begins looking first to see if the user has MMM installed (line 36). For safety the installer backs up `MMM_prototype.m` and `MMM_prototype.fig` files to `../MMM_path/old_files` before copying across new versions of `MMM_prototype`, these versions allow you to access MISHAP from MMM.

At a later date this versions will be submitted to MMM and this part of the installer will be removed.

Next the installer asks the user if HADDOCK will be run locally, if so then the local copy of HADDOCK requires some new parameter files. Again existing files are backed up first to `../HADDOCK_path/old_files` before the new ones are copied across.

As HADDOCK is outside the MatLab environment a system search is required and this may take some time. As a safety feature, the search for HADDOCK will time out in 10 seconds and ask the user for the location instead.

The command window output from a typical installation should look something like:

```

1
2      | \ \ | _ _ | / _ _ | | | | \ \ | _ _ \
3      | \ / | | | | ( _ _ | | _ | | / \ | | _ ) |
4      | | \ | | | | \ _ _ \ | _ _ | / \ \ | _ _ /
5      | | | | | | _ _ _ ) | | | | / _ _ \ | |
6      | _ | | | _ _ _ | _ _ _ / | | | | / \ \ \ |
7
8                      by
9
10
11
12
13
14
15
16
17
18
19 =====
20 STARTING INSTALLATION
21 =====
22
23 Installer version - 13.06
24 Build date       - 04th May 2013
25
26 This installer will install
27 MISHAP version  - 13.06
28 Release date    - 04th May 2013
29
30 System          - Linux - 64 bit
31
32 MatLab version  - 7.12.0.635 (R2011a)
33
34 =====
35
36 Finding MMM installation...
37 MMM was located at:
38 /opt/matlab/MyScripts/MMM_2013
39
40 Backing up MMM files to:
41 /opt/matlab/MyScripts/MMM_2013/old_files
42
43 Files backed up!
44
45 Copying across new files...
46 Files successfully copied!
47
48 Installing...
49
50 Configuring...
51 MMM configured!
52
53 =====
54
55 MISHAP parameters are not yet compatible with the HADDOCK webservers,
56 though this should change shortly, MISHAP must be run locally in the mean time
57
58 Are you going to be using HADDOCK locally? Y/N [Y]: y
59
60 Finding HADDOCK installation...
61 detection will abort if nothing is found in 10 seconds...
62 HADDOCK was located at:
63 /opt/haddock
64
65 Backing up HADDOCK files to:

```

2 User interface

```
66 /opt/haddock/old_files
67
68 Files backed up!
69
70 Copying across new files...
71 Files successfully copied!
72
73 Installing...
74
75 Configuring...
76 HADDOCK configured!
77
78 =====
79
80 Writing changes to MatLab...
81 MatLab configured!
82
83 =====
84
85 I hope you find MISHAP useful.
86
87 Documentation can be found at:
88 morganbye.net/mishap
89
90 Please address any feedback to:
91 morgan.bye@uea.ac.uk
92
93 =====
94 INSTALLATION COMPLETE
95 =====
```

With MISHAP installed, MISHAP can now be called from MMM using

Predict → Quaternary → HADDOCK → MISHAP

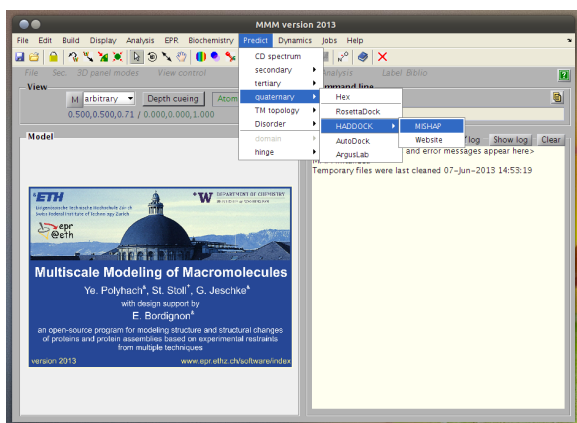


Figure 2.9 MMM menu showing MISHAP

Or without MMM from the command window with the command:

```
MISHAP
```

This will launch the primary MISHAP window, the distance distribution window.

2.3 MISHAP distance distribution window

2.3.1 Loading a distance distribution

Calling MISHAP will launch the primary MISHAP window, the distance distribution window.

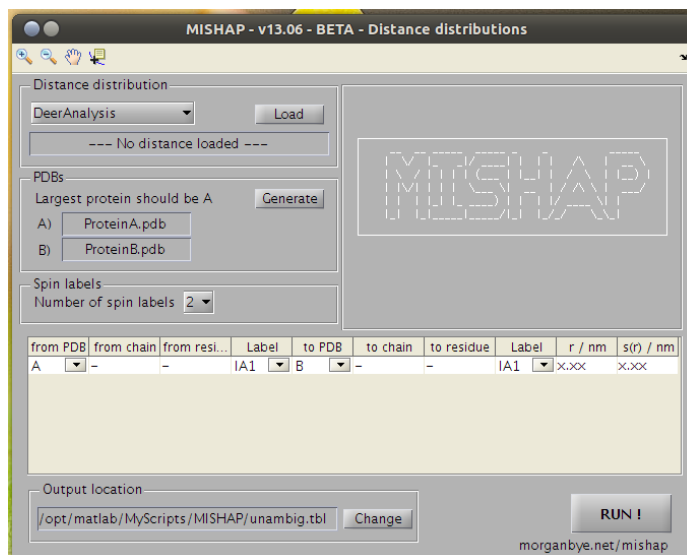


Figure 2.10 MISHAP distance distribution window

From the distance distribution panel select whether the distance distribution is from *DeerAnalysis* or *MMM* and click *Load*.

If a *DeerAnalysis* or *MMM* window is open then MISHAP will try to pull the distance distribution information from the open window. If more than one is open then it will load whichever was opened first. If no window can be found then a user interface for the selection of a previously saved file is opened.

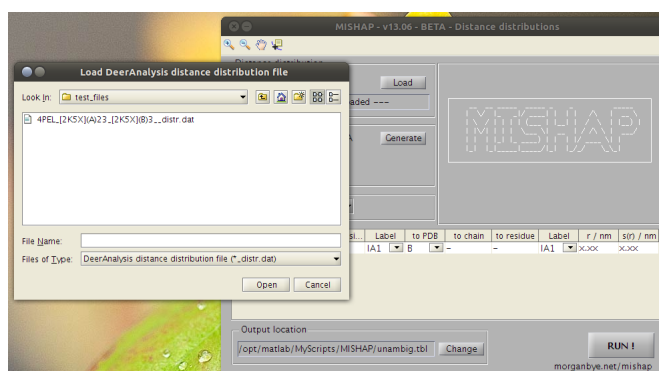


Figure 2.11 MISHAP distance distribution load interface

Once a file is selected MISHAP will automatically plot the distance distribution and search it for any peaks. Peaks are marked with circles. The largest peak is selected as the peak of interest and the peak location is updated into the r column of the table.

A $s(r)$ of 0.5 nm is automatically set and the boundaries are plotted about the distance as dotted lines in the same colour as the peak. At a later date the $s(r)$ will be detected and calculated from the plot using a half height analysis of a Gaussian fit to the peak.

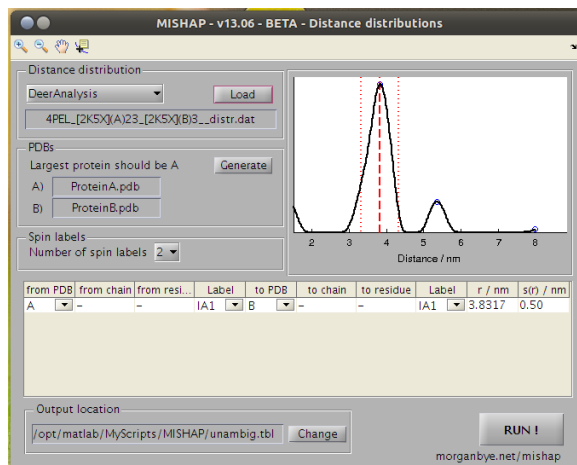


Figure 2.12 MISHAP distance distribution window showing a loaded distance distribution

2.3.2 Multiple spin systems

Currently MISHAP has support for systems with up to 4 nitroxide spin labels. By default it is set for only 2 spins (1 distance), selecting 3 or 4 spin labels adds more distances to the table (3 distances for 3 labels, 6 distances for 4 labels). The distance distribution plot displays each distance in a separate coloured dashed line with the peak's $s(r)$ dotted in the same colour.

2.3.3 Protein information

If the distance distribution tool is being used without the protein PDB creator then the table information needs to be completed with the appropriate information for the chain ID and residue number of the spin label in the PDB.

However, it is recommended to generate the PDBs for HADDOCK with MISHAP before proceeding further. To do so click *Generate* to launch the PDB Creator (see §2.4).

2.3.4 Output location

Once happy with the information displayed in the table finally check the output location box. This is the location and file name that the distance restraints file will be saved as.

By default the file name `unambig.tbl` will be used and the path will be that of the MatLab path when MISHAP was first opened.

If this is not suitable simply click *Change*.

2.3.5 RUN!

When the information in the table is correct and the output location is set click *RUN!*

All information is passed out to the MatLab command window. First the constraint generator checks that the table is complete, then it checks that it can write to the selected location. If

the folder does not exist or the user does not have permission to write to it then they will be prompted.

Finally the file is created. This process should be very fast even on a slow computer, ensure that *RUN!* is only pressed once.

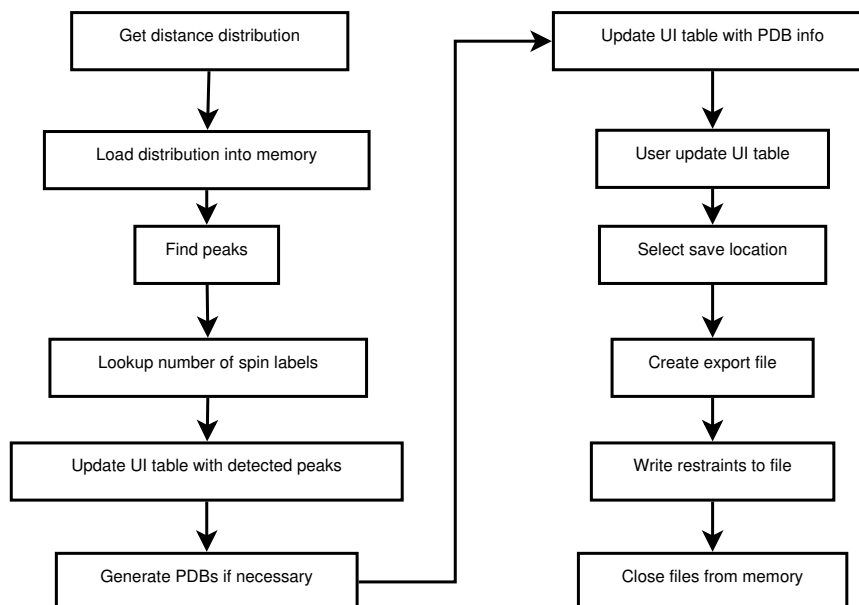


Figure 2.13 Flowchart of the MISHAP distance distribution window process

On the next page is an example output from the command window.

2 User interface

```
1
2      | \ / | _ | / | _ | | | | \ | _ | \
3      | \ / | | | | ( _ | | | | | / \ | | _ ) |
4      | | \ | | | | \ _ | \ | _ | / \ \ | | _ /
5      | | | | | | | _ | | | | | / _ | \ | |
6      | | | | _ | | _ | | | | | / _ | \ | |
7
8
9      / _ | | | | | | | | | | | | | |
10     | | | | _ | | | | | | | | | | | | | |
11     | | | | / _ | \ | \ | | | | | | | | | |
12     | | | | ( ) | | | | \ _ | | | | | | | | | |
13     \ _ | \ _ | / | | | | | | | | | | | | | |
14
15
16
17     / _ | / _ | \ _ | \ _ | \ _ | \ _ | \ _ | \ _ |
18     | ( | | | _ | / | | | | _ | / | | | | ( | | | | ( | |
19     \ _ , | \ _ | | | | | | | | | | | | | | | | | |
20     _ / |
21     | _ /
22
23 =====
24 STARTING Distance Constraint Creation
25 =====
26
27 MISHAP version      - 13.06
28 Release date      - 16th May 2013
29
30 System             - Linux - 64 bit
31
32 MatLab version    - 7.12.0.635 (R2011a)
33
34 =====
35
36 Checking the data table...
37 Checking complete
38
39 Setting label variables...
40
41 Saving file...
42 File saved as
43 /opt/matlab/MyScripts/MISHAP/test_files/unambig.tbl
44
45 =====
46
47 MISHAP - distance constraint generator completed in 0.00589 seconds
48 Thank you for using MISHAP
```


2.4 MISHAP PDB creator window

The MISHAP PDB creator window is launched from the MISHAP distance distribution window by clicking *Generate*.

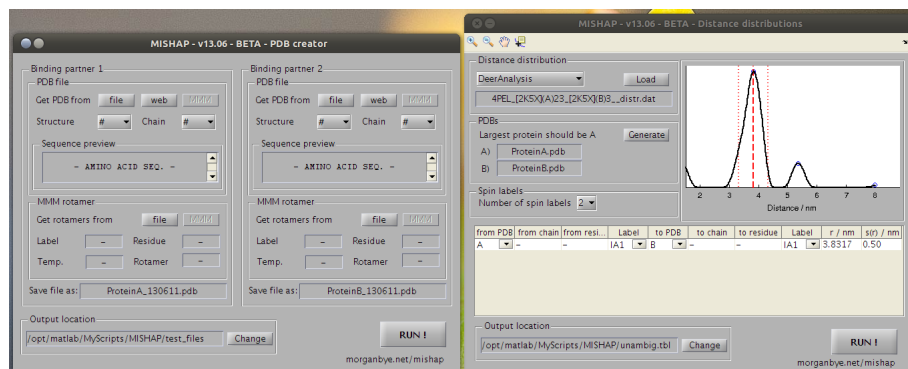


Figure 2.14 MISHAP distance distribution and PDB creator windows

The purpose of the PDB creator is 2 fold. First is to ensure that PDB files are generated that are compatible with HADDOCK.

Unfortunately, the PDB format was first created in the 1970's and has been extensively developed, without any strict enforcement of the PDB format for many years meaning that it is quite common for PDB files generated in one program to not be compatible with a different PDB handling program. MMM version 2013 after a major rewrite of the PDB file writing functions now conforms fully. HADDOCK will only accept PDB files in a very explicit format, which differs significantly from the PDB standard².

The second purpose of the PDB creator is to combine a PDB file supplied from a local source, from the PDB website (using an access number) or from a loaded MMM model with a MMM rotamer analysis.

MMM provides some basic functionality for this but we require a single PDB file where an amino acid residue has been mutated to a spin label attached cysteine residue. Alternatively, a set of PDB files and a list file that can be inserted into HADDOCK as an ensemble of starting conformations - similar to an NMR ensemble.

2.4.1 Load a PDB file

For the first protein select the PDB file location. If the PDB is a local file click *file* this will open a user interface for selecting the file from your computer.

The PDB can also be obtained directly from rcsb.org in which case select *web* and enter the accession number when prompted.

If MMM is open, then the current model can be used by clicking *MMM*.

There will be a delay as the file is processed. Once processed the *Structure* box will update with the PDB accession code, the *Chain* dropdown box will update with the chains found in the PDB and the sequence of the selected chain will be shown in the *Sequence preview*.

²for more information see <http://www.nmr.chem.uu.nl/haddock/pdb.html>

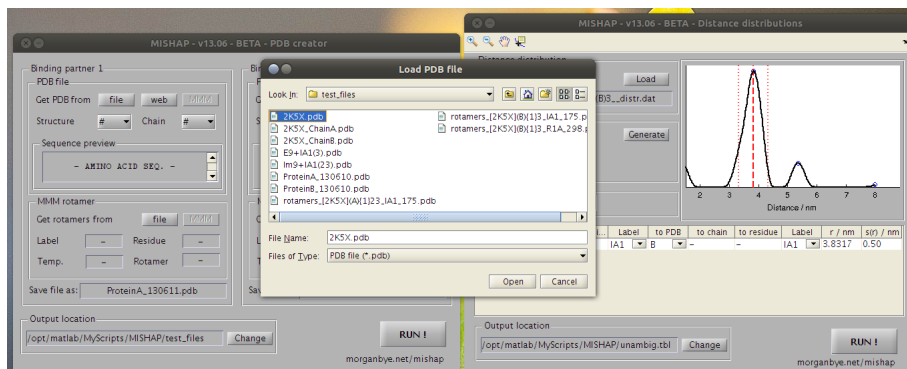


Figure 2.15 MISHAP PDB creator load PDB file selection

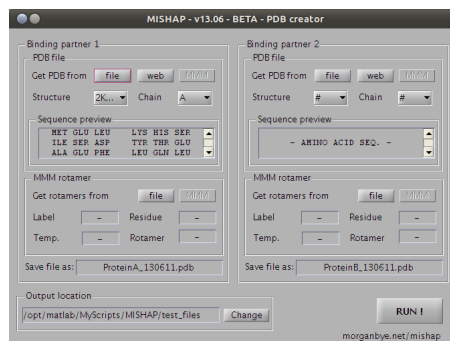


Figure 2.16 MISHAP PDB creator with loaded PDB file

2.4.2 Load a MMM rotamer file

For the first protein select the MMM rotamer file location. If the PDB is a local file saved during the MMM site scan then click *file* this will open a user interface for selecting the file.

In a later release, it will be possible to click *MMM* and pull the rotamer information from MMM.

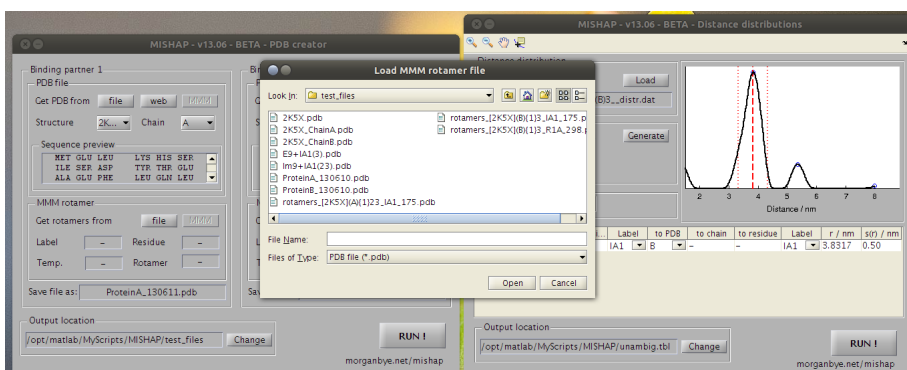


Figure 2.17 MISHAP PDB creator load MMM file selection

It will take a little time to process the rotamer information. Once processed the *Label* box will update with the 3 character amino acid residue code for the spin label, the *Temp.* box will update with the labeling temperature used in the site scan, the *Residue* box will update with the amino acid residue number of the labeling site and the *Rotamer* will update with the selected rotamer number from the MMM rotamer analysis.

By default, the *Rotamer* box will default to rotamer 1 though this can be changed to any rotamer. If a rotamer is selected that is outside the range of available rotamers then the box will automatically default back to rotamer 1. If an ensemble style rotamer analysis is desired

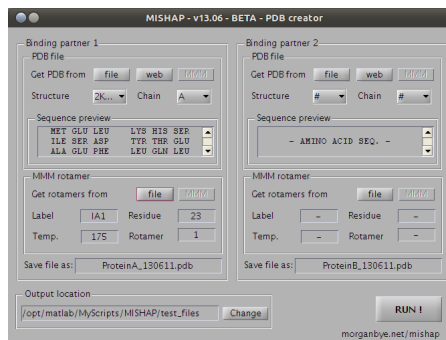


Figure 2.18 MISHAP PDB creator load with PDB and MMM file loaded

type in this box:

all

2.4.3 Output location

By default the protein PDB will be saved as:

Protein [A/B] _ [YYMMDD] . pdb

These can be changed freely with any operating system valid name. Do not include `.pdb` in the file name, otherwise you will generate a file called `ProteinA_YYMMDD.pdb.pdb`

The output path is defined in the *Output location* box, the path will be that of the MatLab path when MISHAP was first opened.

If this is not suitable simply click *Change*.

2.4.4 RUN!

When happy with the settings for *Binding Partner 1* the PDB creator can be run or a PDB and MMM rotamer analysis can also be loaded for *Binding Partner 2*.

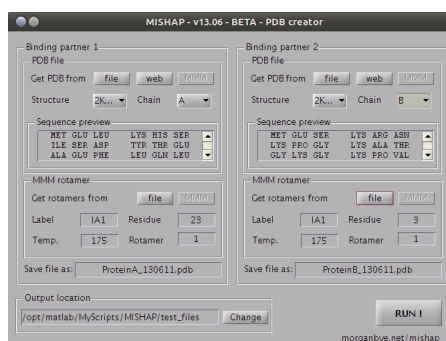


Figure 2.19 MISHAP PDB creator showing 2 proteins loaded with MMM rotamer files

When all the information is loaded and set click *RUN!*

All information is passed out to the MatLab command window. *Binding Partner 1* is addressed before *Binding Partner 2* but in both the creator checks that a PDB and MMM rotamer analysis is loaded, then it checks that it can write to the selected output location. If

2 User interface

the folder does not exist or the user does not have permission to write to it then they will be prompted.

Finally the creator merges the selected MMM rotamer(s) with the PDB file and the output file is created.

Briefly, the creator essentially splits the PDB file before and after the select residue. It then pastes in the rotamer residue and merges the residues before the residue, the rotamer residue and residues after the rotamer residue with appropriate linkers. All hydrogen atoms are then removed as these are prone to causing problems in HADDOCK and CNS. Finally, every atom is renumbered in the merged PDB so that read sequentially. The file is finally written out line by line with space padding so that each line is exactly 80 characters.

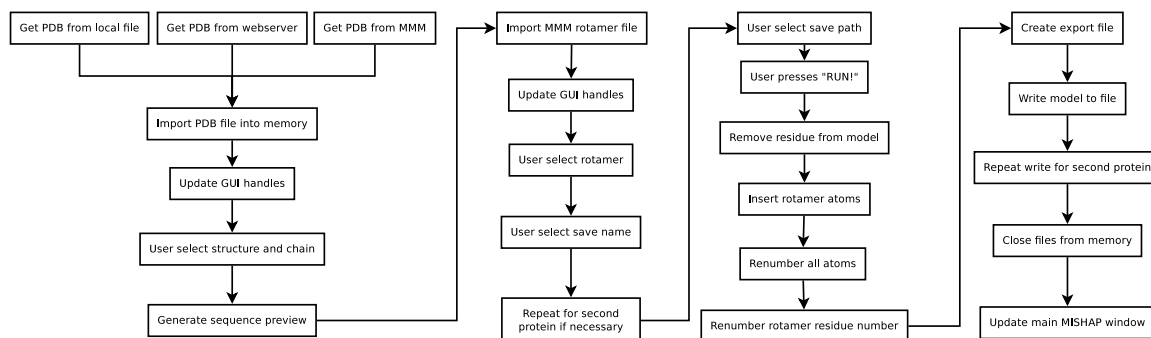


Figure 2.20 Flowchart of PDB handling within the MISHAP PDB creator

On the next page is an example output from the command window.

```

1
2      | \ / | _ _ | / _ _ | | | | \ / | _ _ \
3      | \ / | | | | ( _ _ | | _ | | / \ | | _ ) |
4      | | \ | | | | \ _ _ \ | _ _ | / \ \ | _ _ /
5      | | | | _ | | _ _ _ ) | | | | / _ _ _ \ | |
6      | _ | | | _ _ _ | _ _ _ / | _ | | / / \ \ |
7
8
9      | _ _ \ | _ _ \ | _ \ / _ _ | | |
10     | | _ ) | | | | | ) | | | _ _ _ _ _ _ | | _ _ _ _
11     | _ _ / | | | | _ < | | | | ' _ / _ \ / _ \ | ' _ |
12     | | | | | _ | | | ) | | | _ _ | | | _ / ( | | | | ( ) | |
13     | _ | | _ _ _ / | _ _ _ / \ _ _ _ | | \ _ _ \ _ _ , \ \ _ _ _ / | _ |
14
15 =====
16 STARTING PDB Creation
17 =====
18
19 MISHAP version      - 13.06
20 Release date       - 15th May 2013
21
22 System             - Linux - 64 bit
23
24 MatLab version     - 7.12.0.635 (R2011a)
25
26 =====
27 Binding partner 1
28 =====
29
30 Finding the selected PDB...
31 Loading file...
32 /opt/matlab/MyScripts/MISHAP/test_files/2K5X.pdb...
33 Using structure 2K5X and chain A
34 PDB loaded!
35
36 Finding the selected MMM rotamer...
37 Loading file...
38 /opt/matlab/MyScripts/MISHAP/test_files/rotamers_[2K5X](A){1}23_IA1_175.pdb
39
40 Rotamers loaded!
41
42 Setting labelling conditions:
43 Label               - IA1
44 Temperature         - 175 K
45 Attaching to residue - 23
46 Using rotamer       - 1
47
48 Adding rotamer to PDB...
49 Rotamer successfully added!
50
51 Saving file...
52 File saved as
53 /opt/matlab/MyScripts/MISHAP/test_files/ProteinA_130610.pdb
54
55 =====
56 Binding partner 2
57 =====

```

2 User interface

```
58
59 Finding the selected PDB!
60 Loading file...
61 /opt/matlab/MyScripts/MISHAP/test_files/2K5X.pdb
62 Using structure 2K5X and chain B
63 PDB loaded!
64
65 Finding the selected MMM rotamer
66 Loading file...
67 /opt/matlab/MyScripts/MISHAP/test_files/rotamers_[2K5X](B){1}3_IA1_175.pdb
68 Rotamers loaded!
69
70 Setting labelling conditions:
71 Label - IA1
72 Temperature - 175 K
73 Attaching to residue - 3
74 Using rotamer - 1
75
76 Attaching rotamer to PDB...
77 Rotamer successfully added!
78
79 Saving file...
80 File saved as
81 /opt/matlab/MyScripts/MISHAP/test_files/ProteinB_130610.pdb
82
83 =====
84
85 MISHAP - PDB Creator completed in 0.79016 seconds
86 Thank you for using MISHAP
```

3 HADDOCK configuration

HADDOCK (the high ambiguity data driven biomolecular docking program) is an approach that makes use of biochemical and biophysical interaction data such as chemical shift perturbations from NMR titration experiments, mutagenesis data or bioinformatic predictions to study protein complexes.

Traditionally, most approaches to protein complex formation relied upon a combination of energetics and shape complementarity rather than experimental data. The previously mentioned experiments are presented to the HADDOCK program as ambiguous interaction restraints (AIRs) to drive the docking process. More formally, these are defined as an ambiguous distance between all residues shown to be involved in the interaction.

Since the original 2003 JACS publication HADDOCK has been extended to deal with a large variety of data and complexes including protein-DNA, protein-RNA, protein-oligosaccharides and protein-ligand complexes. HADDOCK has consistently performed well in the blind protein-protein complex CAPRI competitions¹.

HADDOCK remains a command line tool requiring the generation and configuration of many files before a run can commence. HADDOCK consists primarily as a collection of python scripts derived from ARIA written by Michael Nilges and Jens Linge and makes use for CNS as structure calculation.

3.1 Spin label definitions

HADDOCK being designed to accept input data from several different sources already has the facility to use several unnatural amino acids such as phosphorylated CYS, HIS, SER, THR, several methyl variants such as mono-, di- and tri- methyl lysine, to name but a few.

Thus if correctly defined it is perfectly possible to add MTSL and IA-PROXYL labelled cysteines as an unnatural, but allowed, amino acid residue.

For an amino acid to be defined in HADDOCK and the underlying structural calculation program, CNS, requires explicit definitions of all atoms, charges, bonds, bond lengths and allowed bond angles (both dihedral and improper dihedral). When combined this gives the modelling software not only the atomic position and bonding of the atoms, but also allows for the rotation of bonds - which is essential in the final energy optimisation stage of HADDOCK before the final energy optimisation in water.

¹<http://capri.ebi.ac.uk>

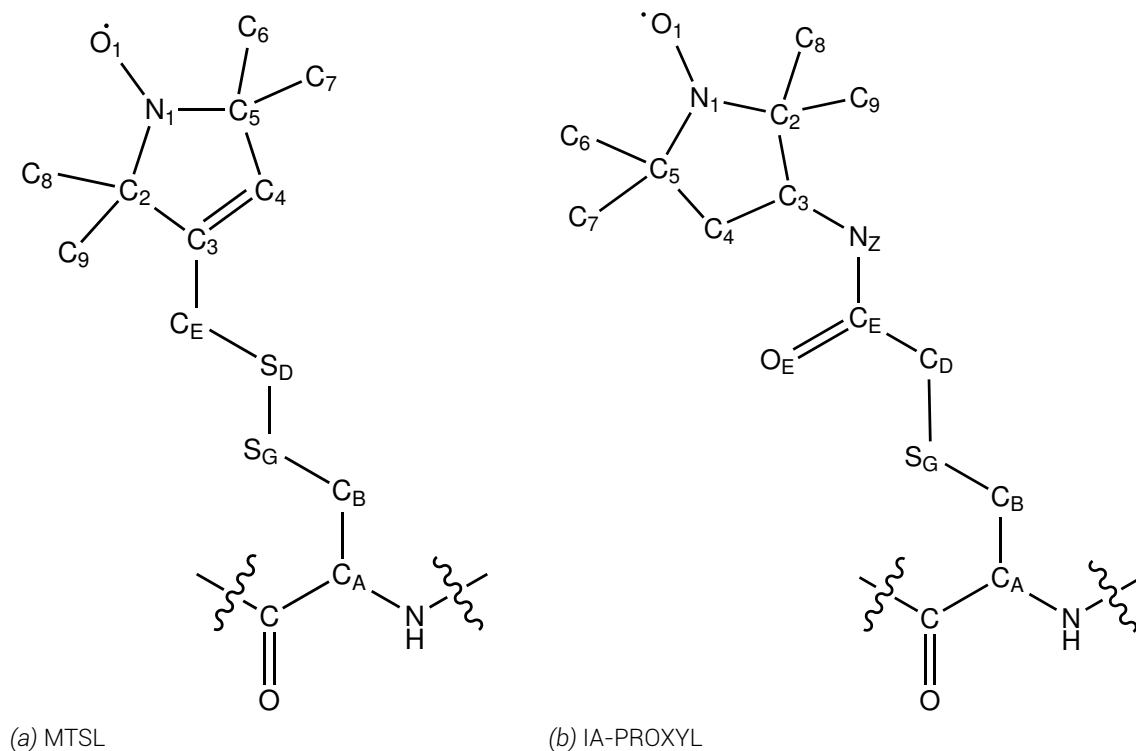


Figure 3.1 Atom naming convention used in MMM and MISHAP for nitroxide spin labels

3.2 Protocols

With MMM there is an increasing amount of published data that shows that rotamer population prediction from a rotamer library accurately reflects the experimental situation. It is not known whether the bond rotation allowed in HADDOCK/CNS reflects well with that seen experimentally. Previously, the authors of HADDOCK when performing paramagnetic relaxation enhancement studies take the spin label and during the early stages of docking spawn 4 structures with the spin label in different conformations. Docking is then performed in an ensemble style analysis and it is hoped that during the allowed side chain freedom stage of energy minimisation that starting from 4 random positions will find the overall position(s) and not get caught in a local energy minima.

To this end, MISHAP provides the ability in the PDB creator to not only create a single PDB file but automatically generate an ensemble folder of PDB files, with the appropriate HADDOCK file list file, of rotamers from MMM with populations of greater than 1 %. Whilst a typical MMM analysis can generate up to 100 rotamers on a greatly solvent exposed site from the rotamer library most will have populations of below 1 %. In more rigidly located spin label positions the rotamer population quickly reduces in number with the remaining allowed rotamers covering a large proportion of the population, in some cases above 50 %, in these cases an ensemble analysis may not be necessary and a single rotamer PDB file should be sufficient for an accurate docking simulation.

3.3 Topology parameters (toppar)

HADDOCK contains a series of files which contain all of the necessary parameters for amino acid residues, ligands, DNA, etc. These parameters largely are default parameters for CNS that have been extensively modified and added to for the purposes of HADDOCK. HADDOCK creates CNS scripts and then calls CNS in order to perform docking.

Note that HADDOCK must make use of topology version 5.4 for many of the features to work. HADDOCK however by default uses version 5.3 - to ensure as broad compatibility as possible the provided MISHAP files provide both 5.3 and 5.4 topology files. The following files from the `HADDOCK/toppar` folder require updating in order for HADDOCK to accept nitroxide spin labels:

topallhdg5.3.pep allows the labels to exist in a peptide

topallhdg5.3.pep-noCter allows the labels to exist in a peptide where the C-terminus does not exist

topallhdg5.3.pep-noNter allows the labels to exist in a peptide where the N-terminus does not exist

topallhdg5.3.pro defines the mass and charge of each atom, as well as each bond, angle, improper angle, dihedral angle and hydrogen bond acceptor/donor site.

topallhdg5.3.pro-caro defines the mass and charge of each atom, as well as each bond, angle, improper angle, dihedral angle and hydrogen bond acceptor/donor site.

topallhdg5.4.pep allows the labels to exist in a peptide, and defines how the peptide would bind to surrounding amino acid residues

topallhdg5.4.pep-noCter allows the labels to exist and bond to it's neighbours in a peptide where no C-terminus exists

topallhdg5.4.pep-noNter allows the labels to exist and bond to it's neighbours in a peptide where no N-terminus exists

topallhdg5.4.pep-noter allows the labels to exist and bond to it's neighbours in a peptide where neither a N nor C -termini exists

topallhdg5.4.pro defines the mass and charge of each atom, as well as each bond, angle, improper angle, dihedral angle and hydrogen bond acceptor/donor site

Whilst it is possible to create these topology parameter files by hand methodically, several tools exist that can speed up the process. For example the PRODRG tool² has a free web server³ that can be used to generate CNS topology files. This tool is only designed to create ligands and so to incorporate a structure as an amino acid requires some modification.

3.4 Run settings

`run.cns` is the primary file that HADDOCK runs a docking experiment from. When a run is first generated it is imperative to change this file to reflect not only your own system but also to ensure the correct settings are used.

²Acta Cryst. (2004). D60, 1355-1363

³<http://davapc1.bioch.dundee.ac.uk/prodrng/>

3.4.1 CNS address

Despite what is set in the `new.html` when the run is generated the CNS address often does not update and is often defaulted to:

```
/sw/nmr/cns_solve_1.2-para/mac-intel-darwin/bin/cns
```

This can be edited by hand or with the command:

```
sed 's, '/sw/nmr/cns_solve_1.2-para/mac-intel-darwin/bin/cns', '/path/to/cns_solve_1.3/intel-x86_64bit-linux/bin/cns',' run.cns > tmp ; mv tmp run.cns
```

Where the `sed` command searches for the string in the first quote and replaces it with the second string. We then need to save to a temporary file and overwrite the original file with the temporary one.

3.4.2 Topology selection

HADDOCK will always by default use topology parameters v5.3, we require HADDOCK to use version 5.4 when using CNS v1.3.

Update the version selection with:

```
sed 's, 'topallhdg5.3', 'topallhdg5.4', ' run.cns > tmp ; mv tmp run.cns
```

3.4.3 Parallelisation

HADDOCK by default assumes only a dual cored processor. If you are using it upon a cluster or server ensure that you update the `run.cns` file to take this into account; replacing 'XXX' with the number available on your system.

```
sed 's, 'cpunumber_1=2', 'cpunumber_1=XXX', ' run.cns > tmp ; mv tmp run.cns
```

The number of CPU cores on a UNIX system can be easily found with the command:

```
grep "cpu cores" /proc/cpuinfo
```

or with the terminal command (available in most Linux distributions):

```
lscpu
```

3.4.4 Initial seed number

HADDOCK uses a random number generator to begin the structure formation in the rigid body stage. However, this number generator requires a starting point and should be recorded for reproducibility. It is easiest to set this as a fixed number, though a random number generator could be piped into the argument.

```
sed 's,'iniseed=917','iniseed=999',' run.cns > tmp ; mv
tmp run.cns
```

3.4.5 MD time step

The HADDOCK protocol uses a process of simulated heating and cooling for energy optimisation and minimisation. By default HADDOCK uses an molecular dynamics (MD) time step of 2 *fs*, however, docking runs will often fail with this setting. With such a time step, atoms that are moving toward each other in one time step can often jump on top of each other, when HADDOCK then calculates the energy there is a massive potential energy in these atoms. Entropy resolves this in the next time step by moving the atoms far apart, often outside the calculation co-ordinate space (> 99,999 Å). This causes the structure to fail, if more than 20 % of a run's structures fail then HADDOCK will abort the run.

By decreasing the time step reduces the risk of this scenario, however, to ensure a good fit the total simulation cannot change, thus the number of steps increases, which increases the experiment time.

```
# Change MD time step
sed 's,'timestep=0.002','timestep=0.0005',' run.cns > tmp
; mv tmp run.cns

# Increase MD iterations to maintain total simulation
time
sed 's,'initiosteps=500','initiosteps=2000',' run.cns >
tmp ; mv tmp run.cns sed 's,'cool1_steps=500','
cool1_steps=2000',' run.cns > tmp ; mv tmp run.cns
sed 's,'cool2_steps=1000','cool2_steps=4000',' run.cns >
tmp ; mv tmp run.cns
sed 's,'cool3_steps=1000','cool3_steps=4000',' run.cns >
tmp ; mv tmp run.cns
```

3.4.6 Annealing temperature

Energy optimisation in HADDOCK is conducted by raising the temperature in the system to a high value and then allowing it to cool *in vacuo*. Reducing the time step described in the previous section can help. However, if greater than 20 % of structures fail during it1 (energy optimisation) step then HADDOCK will abort and solvent optimisation will not occur. First, change the initial seed number and restart the run (this will require deletion of all it0 files, for the creation of new structures). If however, the runs continue to fail then the structures have too much energy and will fall to same problems discussed in the MD time step section.

The best solution is to lower the temperature that the system is raised to in each energy optimisation step. This again is defined in `run.cns` and can be quickly changed using the following commands.

3 HADDOCK configuration

```
sed 's,'tadhigh_t=2000','tadhigh_t=500',' run.cns > tmp ;  
mv tmp run.cns  
sed 's,'tadinit1_t=2000','tadinit2_t=1000',' run.cns >  
tmp ; mv tmp run.cns  
sed 's,'tadinit2_t=1000','tadinit2_t=500',' run.cns > tmp  
; mv tmp run.cns  
sed 's,'tadinit3_t=500','tadinit3_t=250',' run.cns > tmp  
; mv tmp run.cns  
sed 's,'tadfactor=8','tadfactor=4',' run.cns > tmp ; mv  
tmp run.cns
```

These temperatures can be dropped further, but should only be done with considerable care.

3.4.7 Protocols

The `run.cns` in the run root is not the same `run.cns` file that is called when protocols are run. To ensure that our changes are reflected throughout the entire HADDOCK run we need to copy the `run.cns` root and replace the `run.cns` file in the `protocols` folder.

```
cp run.cns protocols/run.cns
```

3.4.8 Structure generation

Each amino acid chain or ligand is generated as a separate structure, so for a docking experiment the bare minimum we will have is structure A and B, whilst HADDOCK supports up to 6 structures.

For each structure to be used in the docking run requires that the structure generation script be forced to use topology parameters v5.4.

```
sed 's,'hdg5.3','hdg5.4',' generate_A.inp > tmp ; mv tmp  
generate_A.inp
```

Repeating this command but replacing `generate_A.inp` with `generate_B.inp`, ..., etc. for every structure.

Also check that the `generate_X.inp` file has the correct file location, all too often it will default to:

```
/home/abonvin/software/haddock/examples/e2a.pdb
```

3.5 Run analysis

If the HADDOCK run is not being performed upon the Utrecht University webserver then the run needs to be manually analysed (detailed in the documentation⁴ and online tutorial⁵).

⁴<http://www.nmr.chem.uu.nl/haddock/analysis.html>

⁵<http://www.wenmr.eu/wenmr/haddock-web-server-tutorial>

With CNS and HADDOCK programs sourced in your terminal session navigate to the run then

```
cd structures/it1/water
```

In this directory run the HADDOCK tool `ana_structures.csh`

```
$HADDOCKTOOLS/ana_structures.csh
```

Move into the analysis directory

```
cd analysis
```

Now run:

```
cluster_struct RUN_NAME_rmsd.disp 7.5 4 > cluster.out
```

Where `RUN_NAME` is the name of the HADDOCK run, 7.5 is the RMSD distance cutoff to which structures within an RMSD of 7.5 Å are clustered, with a minimum cluster size of 4 structures.

Move back to the water (or other solvent) directory

```
cd ..
```

Now run `ana_clusters.csh` to perform cluster analysis on the structures (this requires ProFit to be defined in the `haddock_configure.csh` file). This script will only analyse the top 4 scoring structures in the cluster analysis, the `best` flag can be removed (but will increase calculation time) or the number changed.

```
$HADDOCKTOOLS/ana_clusters.csh -best 4 analysis/cluster.out
```


4 An example docking experiment

MISHAP comes with the folder `test_files` in the root directory. In this folder the user will find all the files necessary to run their first MISHAP/HADDOCK docking experiment, without performing any PELDOR experiments. It is assumed that the user is proficient in DeerAnalysis and MMM and would be able to follow the instructions in chapter 2 to generate their own files. If the user is not versed in these programs then it is recommended to read the freely available user manuals^{1,2}.

4.1 Example files

The `test_files` folder contains `2K5X.pdb` the published NMR energy minimised structure of colicin E9 DNase bound to its immunity protein, Im9 from rcsb.org. There are also the files `2K5X_ChainA.pdb` and `2K5X_ChainB.pdb` files created with the EPR toolbox's³ PDBSplitter function, which simply splits the original PDB file into separate chains.

Additionally in this folder we have several files with the root `4PEL_[2K5X](A)23_[2K5X](B)3`, these files are saved DeerAnalysis results from a PELDOR experiment between the C23 position of the Im9 to the S3C position of colicin E9 DNase.

Finally we have the files:

```
rotamers_[2K5X](A){1}23_IA1_175.pdb
```

```
rotamers_[2K5X](B){1}3_IA1_175.pdb
```

These are the saved MMM site scan rotamer analysis files of the 2 sites using IA-PROXYL at 175 K.

4.2 MISHAP

Open MatLab and into the command window type

```
MISHAP
```

A splash screen will display briefly whilst MISHAP is loaded.

Initially only the MISHAP Distance Distribution window will be open. In the **Distance Distribution** box, select **DeerAnalysis** from the menu and click the **Load** button.

A file selection window will appear. Navigate to the MISHAP test files folder and select the DeerAnalysis file `4PEL_[2K5X](A)23_[2K5X](B)3__distr.dat`. Click **Open**.

¹http://www.epr.ethz.ch/software/DeerAnalysis2013_manual.pdf

²http://www.epr.ethz.ch/software/MMM_2013_manual.pdf

³morganbye.net/eprtoolbox

4 An example docking experiment

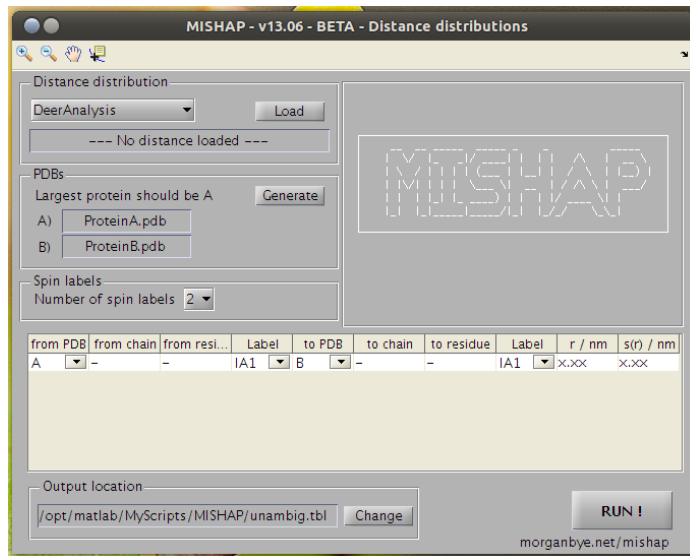


Figure 4.1 MISHAP distance distribution window

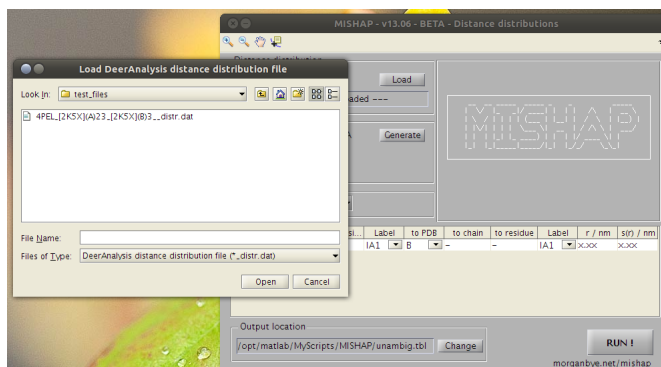


Figure 4.2 MISHAP distance distribution load interface

The distance distribution will be opened and displayed in the box in the top right of the window. The distribution will automatically be scanned for peaks and the best peak selected. Notice that the r / nm column of the table has been updated with the peak information.

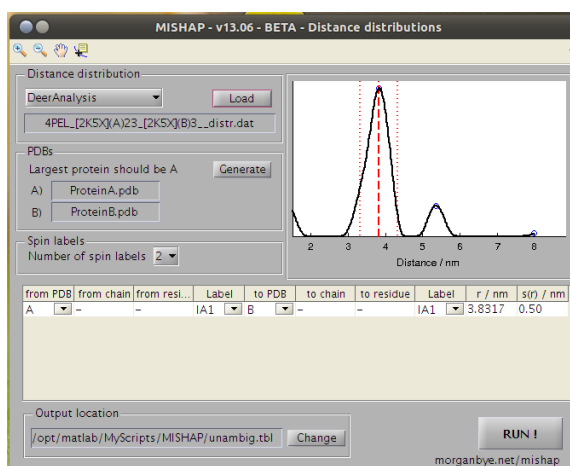


Figure 4.3 MISHAP distance distribution window showing a loaded distance distribution

In the PDBs box click the **Generate** button.

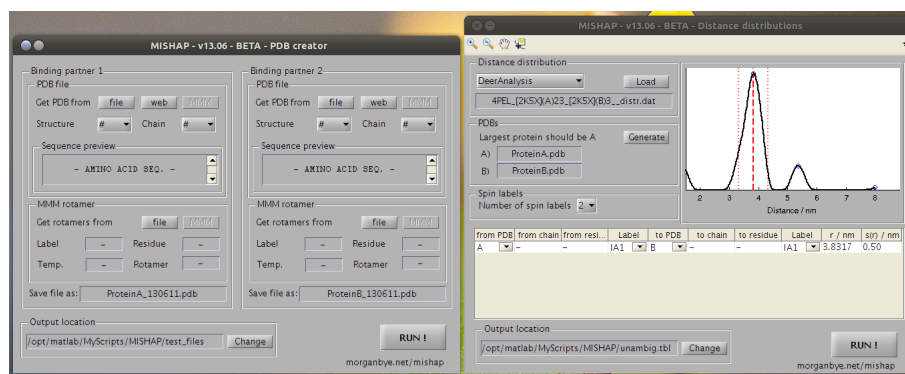


Figure 4.4 MISHAP distance distribution and PDB creator windows

This will open the MISHAP PDB creator window. First only consider the left hand side of the window.

Click the **file** button next to **Get PDB from**, again a file selection window will appear. Navigate to the MISHAP test files folder and select **2K5X.pdb**

After a few seconds the **Sequence preview** will display an amino acid sequence the **Structure** will update with **2K5X**, and **Chain** with **A**

In the **MMM rotamer** box click the **file** button, selecting **rotamers_[2K5X](A){1}23_IA1_175.pdb**

After a few seconds the boxes in **MMM rotamers** will update with the rotamer information that has just been loaded.

Now repeat the process on the right hand side for **Binding partner 2**. This time however, make sure to change the **Chain** with the dropdown menu to **B** and instead select the other **MMM rotamer** file: **rotamers_[2K5X](B){1}3_IA1_175.pdb**

Check that the **Output location** is set to a suitable location, it can be changed by clicking **Change**

Click **RUN!**

4 An example docking experiment

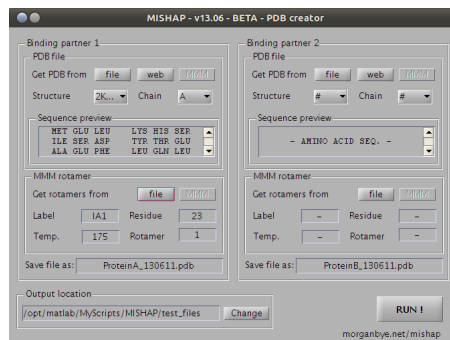


Figure 4.5 MISHAP PDB creator load with PDB and MMM file loaded

The command window will update as the PDBs are generated.

The PDB generation will also update the table in the Distance distribution window.

Now click **RUN!** in the Distance distribution window.

The command window will update as the distance restraints file is generated.

4.3 HADDOCK run generation

Following HADDOCK documentation create a new folder that contains the 2 protein PDBs (from MISHAP), **ambig.tb1** (NMR AIRs file, if available), **unambig.tb1** (from MISHAP) and a **new.html** file.

The new.html file needs to be edited to reflect your personal settings regarding folder location, HADDOCK location, run name, etc. Be sure to source HADDOCK with the MISHAP parameter files.

Open a terminal.

Per HADDOCK documentation source CNS and HADDOCK.

Move to run directory.

Initialise HADDOCK with:

```
haddock2 . 1
```

This will generate the HADDOCK run. Move into the run directory.

Edit **run.cns** as per requirements detailed in §3.4

Copy run.cns into /protocols

```
cp run.cns protocols/run.cns
```

Copy restraints files (**ambig.tb1** and **unambig.tb1**) into /structures/it0, /structures/it1 and /data/distances

Run HADDOCK with:

```
haddock2 . 1
```

4.4 HADDOCK run analysis

Perform the data analysis outlined in §3.5.

Open MatLab, change directory to the run folder. Move to /structures/it1/water.

Run `HADDOCK_CLUSTERS_PROCESSING.m` and `HADDOCK_CLUSTERS_TO_PYMOL.m` scripts if desired (detailed in §5.2).

5 Other tools

5.1 Naccess

Naccess is a program that calculates the exposed surface area of a protein by running a simulated sphere across the surface of a protein. By default the sphere radius is set to 1.4 Å the approximate diameter of a water molecule.

Within the `MISHAP/_private` folder is the `NACCESS` folder, this folder contains `MISHAP.radii`, a special file which allows Naccess to calculate the exposed surface area of a protein which includes the residues R1A/CYM and IA1.

5.2 HADDOCK processing tools

5.2.1 HADDOCK cluster processing

Included is a MatLab script, `HADDOCK_CLUSTERS_PROCESSING.m`, for the processing of HADDOCK clusters. It allows the user to select a cluster or to operate on all the final structures and present them as a figure plotting HADDOCK score against any of the HADDOCK parameters (`bsa`, `dH`, `Edesolv`, `rmsd`, `rmsd-Emin` (default), `ener` (by default this is `Einter`), `ener-Einter`, `ener-Enb`, `ener-Evdw`, `ener-Eelec`, `ener-Eair`, `ener-Ecdih`, `ener-Ecoup`, `ener-Esani`, `ener-Evean`, `ener-Edani`), as shown in figure 5.1. Each structure that completed the docking run is shown as a blue circle and the top 4 scoring structures from each cluster are averaged and shown as closed red circles.

Scores can be saved to the workspace, if the script is called with output variables.

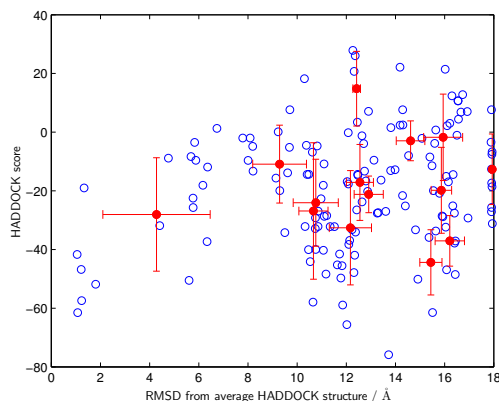


Figure 5.1 A typical figure from the `HADDOCK_CLUSTERS_PROCESSING` script showing HADDOCK score plotted against the RMSD of the structure from the average structure

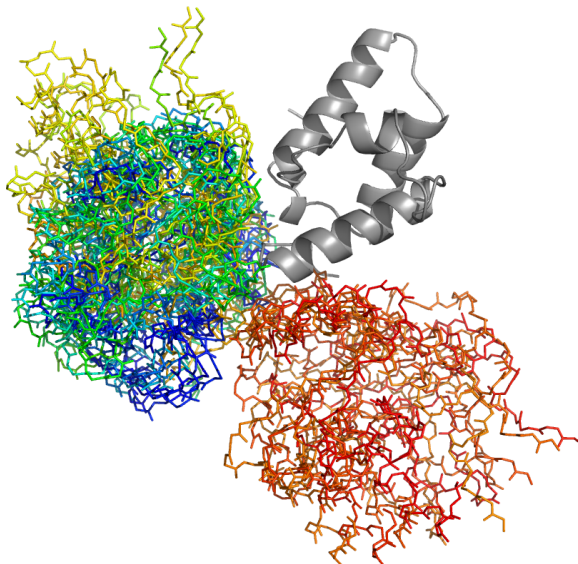


Figure 5.2 A typical figure from the `HADDOCK_CLUSTERS_TO_PYMOL` script showing chain B (in grey cartoon) plotted against the top 4 structures of the top 4 clusters

5.2.2 HADDOCK clusters to PyMOL

PyMOL is a long established program for the three dimensional visualisation of protein structures and their ligands. Due to the nature of the program, PyMOL can easily be called from the command line using Python code.

Included is a MatLab script, `HADDOCK_CLUSTERS_TO_PYMOL.m`, which when called from MatLab by default will take the top 4 structures from the top 4 clusters and align all of the chain B's using an RMSD refinement process. The top result from chain B is displayed as a cartoon model and all the chain A's backbones as colour coded lines, where blue is the best scoring model and red is the 4th structure from the 4th cluster.

This script can be called with different input arguments to adjust the number of clusters and structures from each cluster to be considered. This takes the general form:

```
HADDOCK_CLUSTERS_TO_PYMOL(number_of_clusters , structures)
```

For example,

```
HADDOCK_CLUSTERS_TO_PYMOL(1 , 16)
```

Would present the top 16 structures from only the top scoring cluster.

6 Postface

6.1 License

For a human-readable summary of the legal code below, please visit:

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

This is a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To use MISHAP assumes agreement with this license, and is repeated in LICENSE.txt in the MISHAP root directory.

License

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions

"Adaptation" means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License. "Collection" means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(g) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License. "Distribute" means to make available to the public the original and copies of the Work or Adaptation, as appropriate, through sale or other transfer of ownership. "License Elements" means the following high-level license attributes as selected by Licensor and indicated in the title of this License: Attribution, Noncommercial, ShareAlike. "Licensor" means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License. "Original Author" means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast. "Work" means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work. "You" means an individual or

6 Postface

entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation. "Publicly Perform" means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images. "Reproduce" means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium. 2. Fair Dealing Rights. Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections; to create and Reproduce Adaptations provided that any such Adaptation, including any translation in any medium, takes reasonable steps to clearly label, demarcate or otherwise identify that changes were made to the original Work. For example, a translation could be marked "The original work was translated from English to Spanish," or a modification could indicate "The original work has been modified."; to Distribute and Publicly Perform the Work including as incorporated in Collections; and, to Distribute and Publicly Perform Adaptations. The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. Subject to Section 8(f), all rights not expressly granted by Licensor are hereby reserved, including but not limited to the rights described in Section 4(e).

4. Restrictions. The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(d), as requested. If You create an Adaptation, upon notice from any Licensor You must, to the extent practicable, remove from the Adaptation any credit as required by Section 4(d), as requested. You may Distribute or Publicly Perform an Adaptation only under: (i) the terms of this License; (ii) a later version of this License with the same License Elements as this License; (iii) a Creative Commons jurisdiction license (either this or a later license version) that contains the same License Elements as this License (e.g., Attribution-NonCommercial-ShareAlike 3.0 US) ("Applicable License"). You must include a copy of, or the URI, for Applicable License with every copy of each Adaptation You Distribute or Publicly Perform. You may not offer or impose any terms on the Adaptation that restrict the terms of the Applicable License or the ability of the recipient of the Adaptation to exercise the rights granted to that recipient under the terms of the Applicable License. You must keep intact all notices that refer to the Applicable License and to the disclaimer of warranties with every copy of the Work as included in the Adaptation You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Adaptation, You may not impose any effective technological measures on the Adaptation that restrict the ability of a recipient of the Adaptation from You to exercise the rights granted to that recipient under the terms of the Applicable License. This Section 4(b) applies to the Adaptation as incorporated in a Collection, but this does not require the Collection apart from the Adaptation itself to be made subject to the terms of the Applicable License. You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file-sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works. If You Distribute, or Publicly Perform the Work or any Adaptations or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and, (iv) consistent with Section 3(b), in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). The credit required by this Section 4(d) may be implemented in any reasonable manner; provided, however, that in the case of a Adaptation or Collection, at a minimum such credit will appear, if a credit for all contributing authors of the Adaptation or Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit

required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties. For the avoidance of doubt:

Non-waivable Compulsory License Schemes. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License; **Waivable Compulsory License Schemes.** In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License if Your exercise of such rights is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(c) and otherwise waives the right to collect royalties through any statutory or compulsory licensing scheme; and, **Voluntary License Schemes.** The Licensor reserves the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License that is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(c). Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Adaptations or Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation. Licensor agrees that in those jurisdictions (e.g. Japan), in which any exercise of the right granted in Section 3(b) of this License (the right to make Adaptations) would be deemed to be a distortion, mutilation, modification or other derogatory action prejudicial to the Original Author's honor and reputation, the Licensor will waive or not assert, as appropriate, this Section, to the fullest extent permitted by the applicable national law, to enable You to reasonably exercise Your right under Section 3(b) of this License (right to make Adaptations) but not otherwise. 5. Representations, Warranties and Disclaimer

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING AND TO THE FULLEST EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO THIS EXCLUSION MAY NOT APPLY TO YOU.

6. **Limitation on Liability.** EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination

This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Adaptations or Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above. 8. Miscellaneous

Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License. Each time You Distribute or Publicly Perform an Adaptation, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.

6.2 Third-party software

6.2.1 Software within MISHAP

MISHAP makes use of the following freely available scripts from MatLab Central

Peak detection

fpeak.m by Geng Jun, 2003

Wrapping of strings

linewrap.m by Steve Eddins, 2006

6.2.2 MMM

MMM (Multiscale Modeling of Macromolecular systems) is an open-source program for the visualization, inspection, and improvement of models of proteins and protein assemblies based on restraints from multiple experimental techniques. Authored by Yevhen Polyhach and Gunnar Jeschke, with design support by Enrica Bordignon from ETH Zurich, Switzerland with additional development by Stefan Stoll, University of Washington, USA.

Available from: <http://www.epr.ethz.ch/software>

6.2.3 HADDOCK

HADDOCK (High Ambiguity Driven biomolecular DOCKing) is an information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. Authored by Alexandre Bonvin, Utrecht University, NL.

Available from: <http://www.nmr.chem.uu.nl/haddock/>

With public webservers available at: <http://haddock.science.uu.nl/>

6.3 Acknowledgements

Morgan Bye - primary software writer, testing, documentation

Fraser MacMillan - head of Henry Wellcome Unit for Biological EPR

6.4 Hardware and computation time

The vast majority of testing of MISHAP has been conducted upon a 64-bit Linux system (Ubuntu 12.04 LTS), MatLab R2011a.

This system comprises an Intel Core 2 processor @ $2 \times 2.40 \text{ GHz}$ with 4 GB of RAM.

With this system generation of a distance restraint file is $\sim 5 \text{ ms}$, PDB file load time is $\sim 3 \text{ s}$ (for a $\sim 25 \text{ kDa}$ protein with ~ 2500 atoms), MMM rotamer file load time is $\sim 3 \text{ s}$, the PDB creator running 2 proteins ($\sim 15 \text{ kDa}$) is $\sim 1 \text{ s}$.

These timings will not vary much with newer machines as MatLab by default will not parallelise tasks. If your version of MatLab is set up to use parrallelisation these times will drop significantly.

The MISHAP MatLab interface has been briefly tested upon a Windows XP machine (MatLab 2010a) and is shown to be functional. Though with the specifications of this machine (Intel Celeron @ $1 \times 2.00 \text{ GHz}$ with 1 GB of RAM) loading of PDB files can take as long as 10 s.

HADDOCK runs were performed on a local computing cluster, a 16 core Intel Xeon server node with 32 GB of RAM would complete a run in $\sim 4 - 18 \text{ h}$. Whilst HADDOCK runs are possible on a desktop computer the calculation time will be approximately a week and in which time the machine will be otherwise unusable.

6.5 References

S.J. de Vries, M. van Dijk and A.M.J.J. Bonvin "The HADDOCK web server for data-driven biomolecular docking." *Nature Protocols*, 5, 883-897 (2010).

T.A. Wassenaar, M. van Dijk, N. Loureiro-Ferreira, G. van der Schot, S.J. de Vries, C. Schmitz, J. van der Zwan, R. Boelens, A. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H.R.A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W.F. Vranken, J.F. Doreleijers, G. Vriend, G.W. Vuister, D. Franke, A. Kikhney, D.I. Svergun, R. Fogh, J. Ionides, E.D. Laue, C. Spronk, S. Jurka, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A.M.J.J. Bonvin "WeNMR: Structural Biology on the Grid." *J. Grid. Comp.*, 10, 743-767 (2012).

D. Additional experimental figures

D.1. PELDOR setup experiments

Whilst every effort was made in chapter 6 to find the best experiment conditions for the investigation, each individual PELDOR experiment was further optimised to yield the best possible data.

Before each PELDOR experiment was started a series of small experiments were conducted on each sample. First, a simple field swept echo using a 2 pulse ESEEM experiment with 22 ns $2\pi/3$ pulses with a 380 ns τ_1 , typical for a deuterated protein sample. This allowed the magnetic field to be set such that the maximum amount of signal could be detected, as well as the difference required in the ELDOR pulse frequency. Second, a short T_2 relaxation experiment, again using a 2 pulse ESEEM pulse sequence was conducted to gather the maximum PELDOR observation window. Next a T_1 relaxation time experiment was conducted to determine the minimum possible time between repeats in an experiment.

With the conditions in the primary channel optimised, the initial π pulse of the T_1 relaxation time experiment was replaced by a π pulse in the ELDOR channel. With the spectrometer in transient mode the ELDOR power was increased to give maximum inversion of the echo. With, the ELDOR channel power optimised an ELDOR nutation experiment was conducted to determine optimum ELDOR pulse length, as this is not necessarily the same as the primary channel.

All experiments showed very similar results with only minor variations in field position (within 10 G), T_2 relaxation time (within ~ 1 μs), T_1 relaxation time (within ~ 1 ms), ELDOR pulse power (within 3 dB) and ELDOR pulse length (within 4 ns). Shown here in figure D.1 are the setup experiments for colicin E9 DNase S3C^{IA1}/S80C^{IA1}.

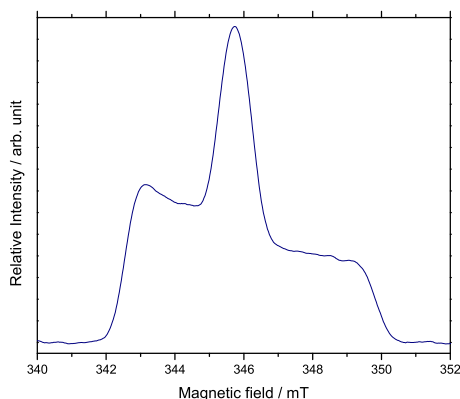
The initial parameter finding experiment consisted of a field swept echo, shown in figure D.1.a, where the centre field position was set corresponding to a $g = 2.022$ for the microwave frequency and the field was swept 15 mT . Through experimentation it was found that it was better to place the observation pulse sequence upon the low field peak, shown here at 343 mT , whilst pumping upon the centre peak, shown here at 346 mT . In strong samples it made little difference which position was selected as the pump position, however with peak samples it was more beneficial to sacrifice overall signal-to-noise in order to have observable PELDOR oscillations.

Figure D.1.b shows the T_2 relaxation experiment which allowed for the determination of the maximum possible observation window. Here the clear deuterium ESEEM modulations can clearly be seen, but despite sample and glycerol deuteration a small amount of proton hyperfine oscillations can be seen on the larger ESEEM oscillations. After 8 μs the signal had decayed to an extent making it indistinguishable from background noise.

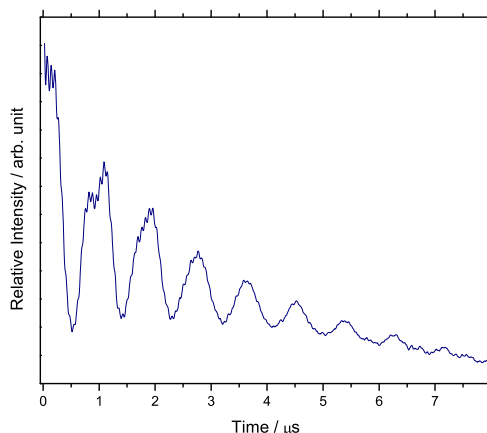
D. Additional experimental figures

Figure D.1.c shows the T_1 relaxation experiment which shows the time required for the system to reach each magnetic equilibrium. The dotted red line represents the zero line. The experiment was deliberately run for a long time as whilst the initial relaxation is fast, complete relaxation takes a long time. The result was fitted using an inverse exponential decay and the resulting relaxation time was multiplied by 2.5 to give the shot repetition time used in the experiment, thereby ensure complete magnetic equilibrium before the next experiment.

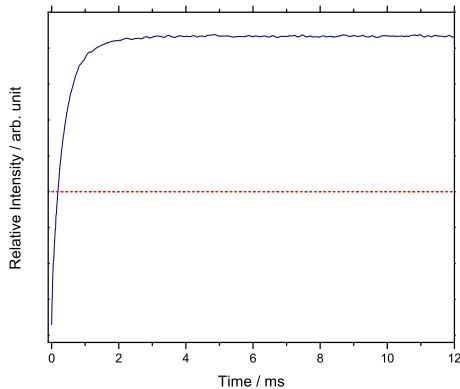
Finally, figure D.1.d shows the ELDOR nutation experiment, where the ELDOR channel power is first optimised with a 32 ns pulse. Then the pulse length is set to 0 ns and in the experiment is increased. The maximum inversion of the ELDOR pulse can be seen here at 34 ns.



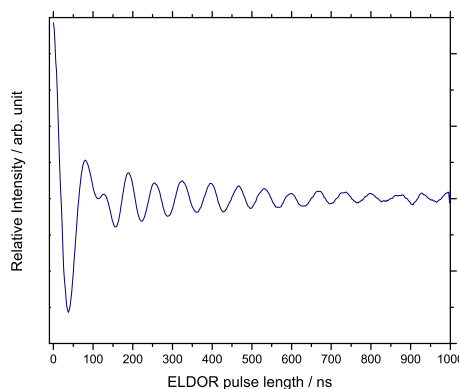
(a) Field swept echo



(b) T_2 experiment



(c) T_1 experiment, with zero intensity as a red dashed line



(d) ELDOR nutation experiment

Figure D.1 Initial parameter finding experiments for colicin E9 DNase S3C^{IA1}/S80C^{IA1}. All experiments: 50 K, 4 scans, 4 shots/point, 1024 points, 9.7047 GHz, 6.0 dB attenuation, 20 MHz video bandwidth, 45 dB video gain a) 12 mT sweep width, 10000 μ s shot rep. time b) 345.70 mT c) 50000 μ s shot rep. time, 345.70 mT, 51 dB video gain d) 10000 μ s shot rep. time, 14.0 dB ELDOR attenuation, 345.70 mT. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

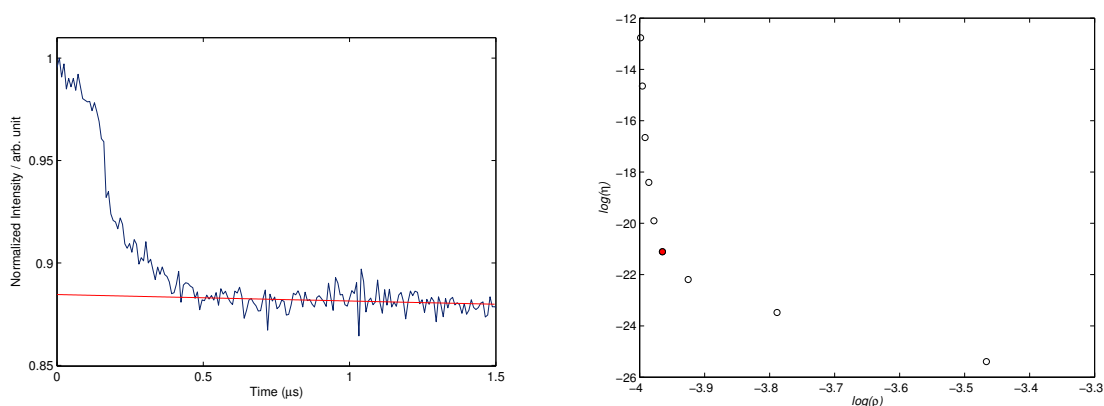
D.2. PELDOR experiments

In the main body of text, chapter 7 displays PELDOR experiments in a 4 panel system. In this system the first panel shows a representation of the experiment superimposed onto the protein structure to aid with visual understanding of that experiment in relation to the others. However, the quality and confidence of the final distance distribution, $p(r)$, from the form factor, $F(t)$, is subject to the quality of the Tikhonov regularisation.

Remember that extraction of the distance distribution is an ill-posed problem where small errors in fitting the form factor can generate large errors in the distance distribution. Tikhonov regularisation requires the addition of extra constraints to stabilise the solution, these constraints are defined to be non-negative (as a distance cannot be negative) and must give a smooth distribution. The subspace of possible values can be scanned to give data that is smooth, without being over-smoothed and which still yields a good fit to the original data. Plotting the log of each parameter in the result of the Tikhonov regularisation yields a characteristically shaped L-curve. At the steep section of the curve a small increase in the smoothing parameter yields a large increase in the quality of the fit. In the flat section of the curve the data is over-smoothed and results in a poor fit to the original data. The best compromise then is at the discontinuity of the L-curve. This appendix shows the Tikhonov regularisation L-curve for each experiment as well as the point selected for the distance distribution.

Each experiment has an additional figure where the background trace is shown superimposed on top of the raw data. This allows for quick visual inspection of the strength of the dipolar interaction. In the case of strong signals PELDOR oscillations deviate strongly from the background decay, in weak signals the PELDOR oscillations barely show movement from the background decay.

D.2.1. Colicin E9 DNase S3C^{IA1}/S80C^{IA1}



(a) PELDOR experiment results with background

(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.2 PELDOR experiment results for colicin E9 DNase S3C^{IA1}/S80C^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S3C^{IA1}/S80C^{IA1}, in red the background data, from running colicin E9 DNase S3C^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 276 points, 6.0 dB attenuation, 20 MHz video bandwidth, 57 dB video gain and 10000 μs shot rep. time, ~9.70 GHz, 345.70 mT. $\tau_1 = 380$ nm, $\tau_2 = 2200$ nm, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

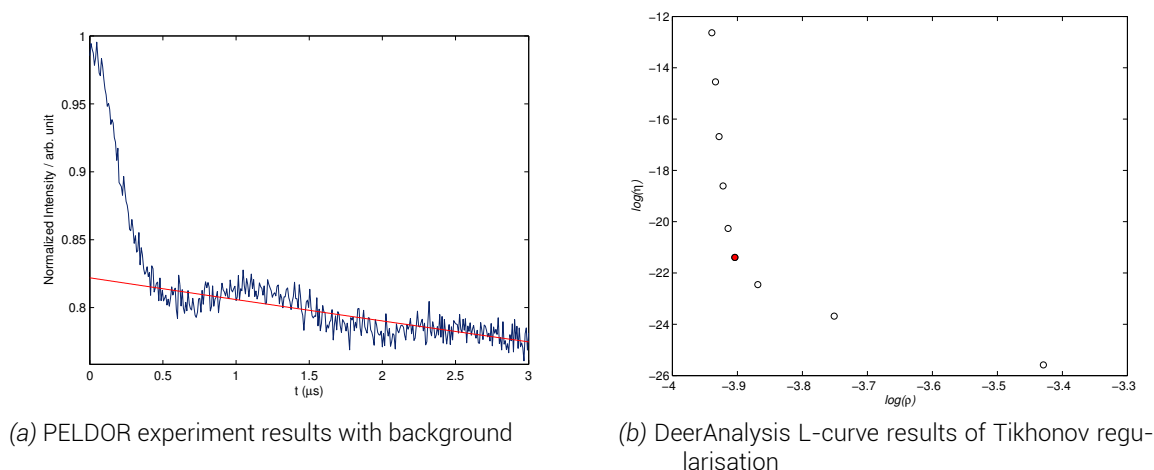
D.2.2. Colicin E9 DNase S30C^{IA1}/S80C^{IA1}

Figure D.3 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S30C^{IA1}/S80C^{IA1}, in red the background data, from running colicin E9 DNase S30C^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 193 points, 10.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μs shot rep. time, ~9.69 GHz, 346.00 mT. $\tau_1 = 380$ nm, $\tau_2 = 4160$ nm, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

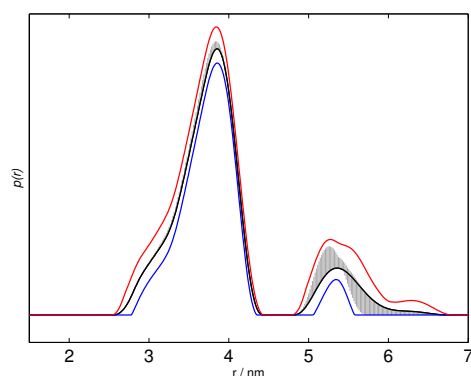
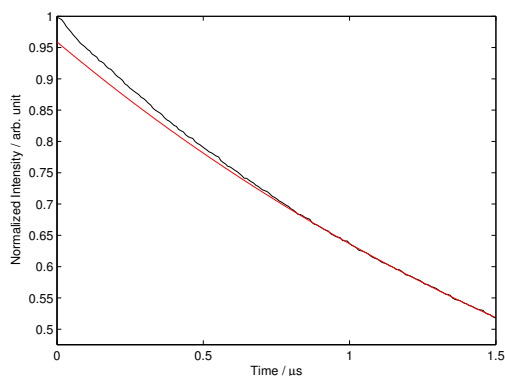
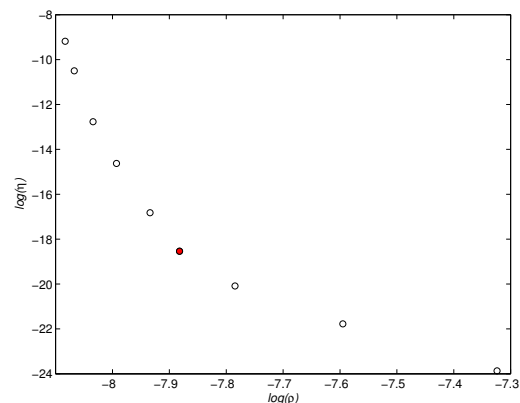


Figure D.4 The resulting distance distributions for the colicin E9 DNase S30C^{IA1}/S80C^{IA1} PELDOR experiment when the background subtraction time is moved. The black line represents the distance distribution presented in the main text, ending data consideration at 3000 ns. Using DeerAnalysis's Tikhonov validation functionality the fitting window was moved to consider the difference in distance distributions when data consideration was moved between 2500 and 3500 ns. The grey shaded areas represent the certainty area of the Tikhonov regularisation, with red representing upper bounds and blue representing lower bounds (2 standard deviations).

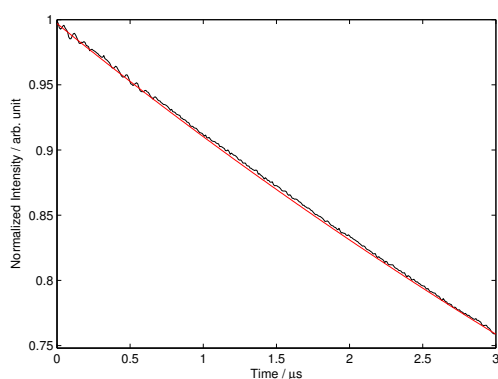
D.2.3. Im9 S6C^{IA1}/C23^{IA1}

(a) PELDOR experiment results with background

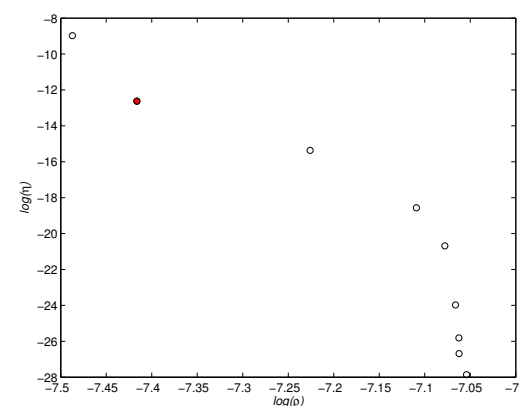


(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.5 PELDOR experiment results for Im9 S6C^{IA1}/C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of Im9 S6C^{IA1}/C23^{IA1}, in red the background data, from running Im9/C23^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 193 points, 10.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μs shot rep. time, ~9.69 GHz, 346.00 mT. $\tau_1 = 380$ nm, $\tau_2 = 4160$ nm, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.4. Colicin E9 DNase S3C^{IA1} - Im9 C23^{IA1}

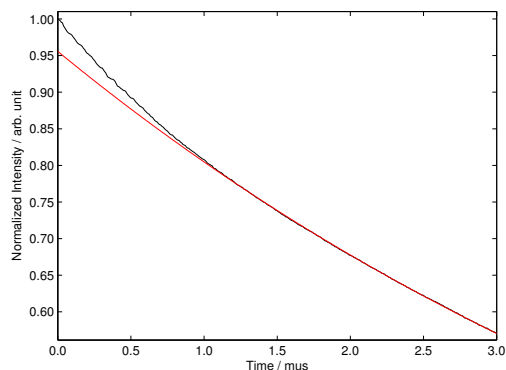
(a) PELDOR experiment results with background



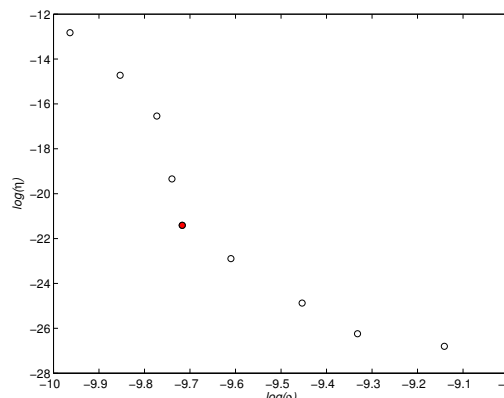
(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.6 PELDOR experiment results for colicin E9 DNase S3C^{IA1} - Im9 C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of Im9 C23^{IA1}, in red the background data, from running colicin E9 DNase S3C^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 414 points, 5.5 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μs shot rep. time, ~9.69 GHz, 345.60 mT. $\tau_1 = 380$ nm, $\tau_2 = 3440$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.5. Colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23S



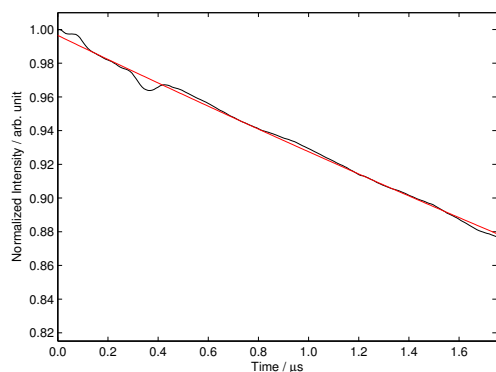
(a) PELDOR experiment results with background



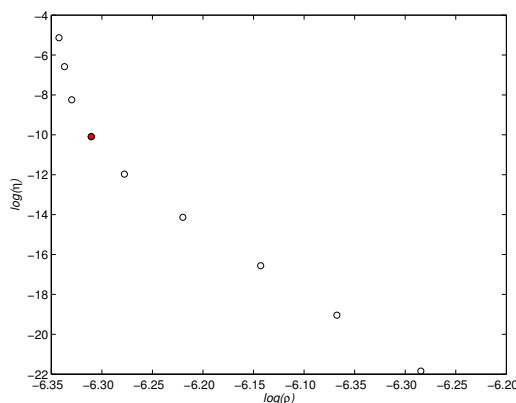
(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.7 PELDOR experiment results for colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23S a) shows in navy the raw 4 pulse PELDOR data of Im9 S6C^{IA1}/C23S, in red the background data, from running colicin E9 DNase S3C^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 20 MHz video bandwidth and 10000 μs shot rep. time, ~9.69 GHz, ELDOR pulse step 8 ns, no nuclear modulation. Data points, magnetic field, attenuation, video gain and delay times were optimised on a per experiment basis suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.6. Colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}



(a) PELDOR experiment results with background



(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.8 PELDOR experiment results for colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S3C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}, in red the background data, from running colicin E9 DNase S3C^{IA1} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 602 points, 3.5 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μs shot rep. time, ~9.74 GHz, 347.30 mT. τ₁ = 380 nm, τ₂ = 2580 nm, ELDOR pulse step 4 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

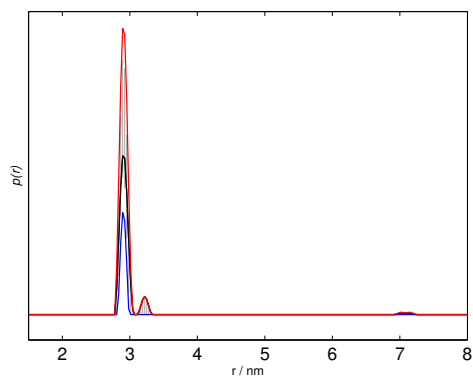
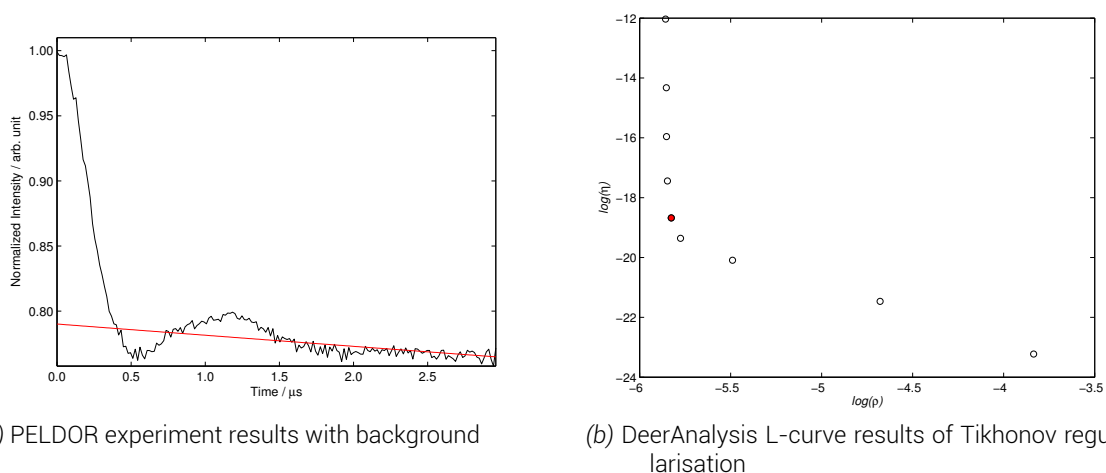


Figure D.9 The resulting distance distributions for the colicin E9 DNase S3C^{I A1} - Im9 S6C^{I A1}/C23^{I A1} PELDOR experiment when the background subtraction time is moved. The black line represents the distance distribution presented in the main text, ending data consideration at 1750 ns. Using DeerAnalysis's Tikhonov validation functionality the fitting window was moved to consider the difference in distance distributions when data consideration was moved between 1000 and 2000 ns. The grey shadder areas represent the certainty area of the Tikhonov regularisation, with red representing upper bounds and blue representing lower bounds (2 standard deviations).

D.2.7. Colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}

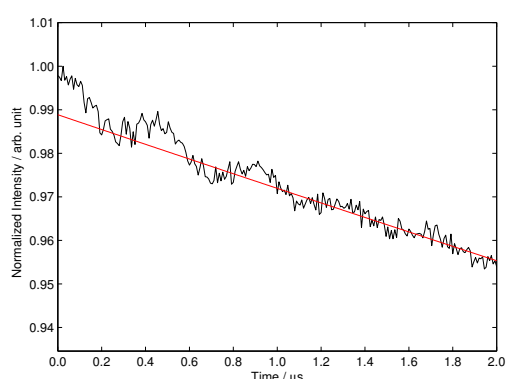


(a) PELDOR experiment results with background

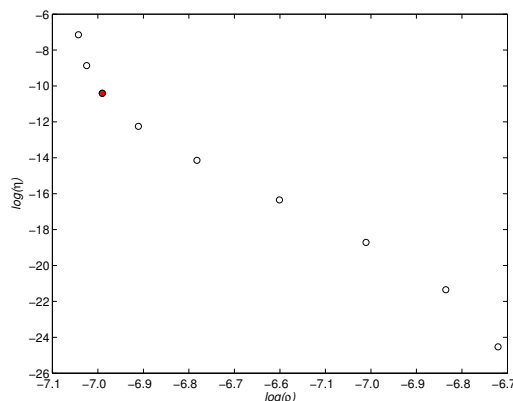
(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.10 PELDOR experiment results for colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S30C^{R1A} - Im9 C23^{R1A}, in red the background data, from running colicin E9 DNase S30C^{R1A} under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 253 points, 4.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μs shot rep. time, ~9.66 GHz, 344.70 mT. $\tau_1 = 380$ nm, $\tau_2 = 3000$ nm, ELDOR pulse step 12 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.8. Colicin E9 DNase E66C^{IA1} - Im9 C23^{IA1}



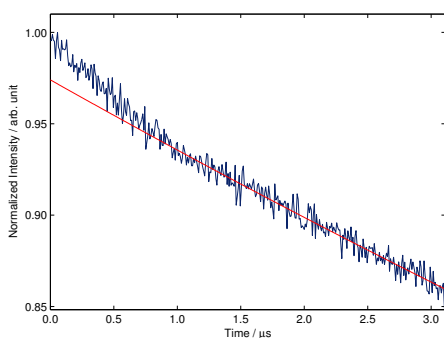
(a) PELDOR experiment results with background



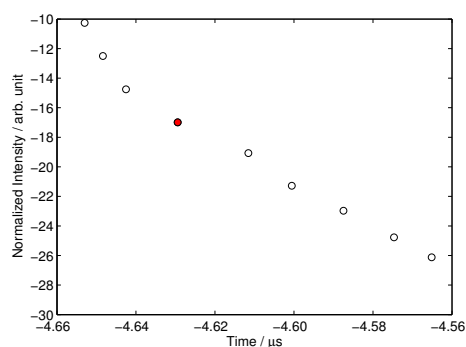
(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.11 PELDOR experiment results for colicin E9 DNase E66C^{IA1} - Im9 C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase E66C^{IA1} - Im9 C23^{IA1}, in red the background data, from DeerAnalysis homogenous background fitting model. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 352 points, 6.0 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μs shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ nm, $\tau_2 = 2600$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.9. Colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23S



(a) PELDOR experiment results with background



(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.12 PELDOR experiment results for colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23S. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23S, in red the background data, from DeerAnalysis homogenous background fitting model. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 500 points, 5.5 dB attenuation, 20 MHz video bandwidth, 60 dB video gain and 10000 μs shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ nm, $\tau_2 = 4160$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

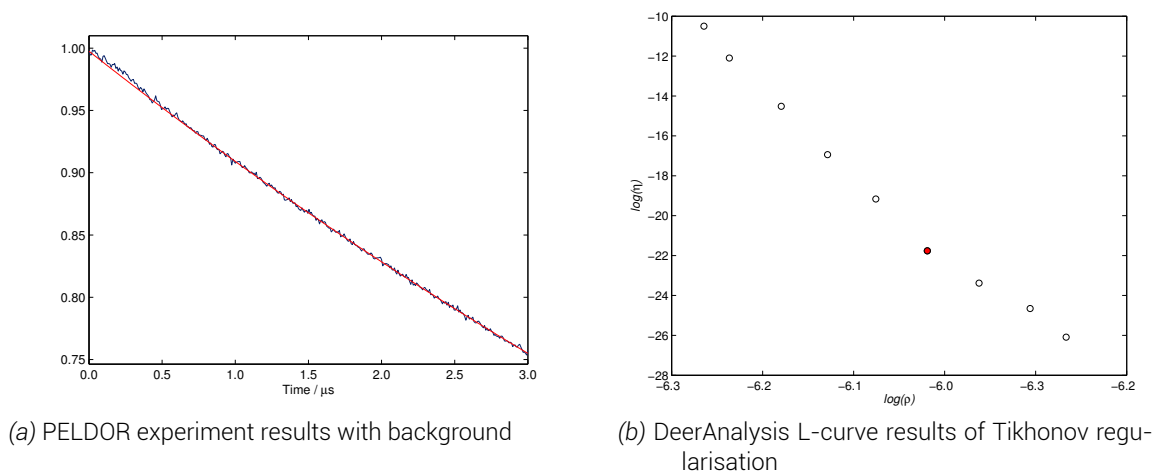
D.2.10. Colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}

Figure D.13 PELDOR experiment results for colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase E66C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}, in red the background data, from DeerAnalysis homogenous background fitting model. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 66 dB video gain and 10000 μs shot rep. time, ~9.71 GHz, 348.00 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

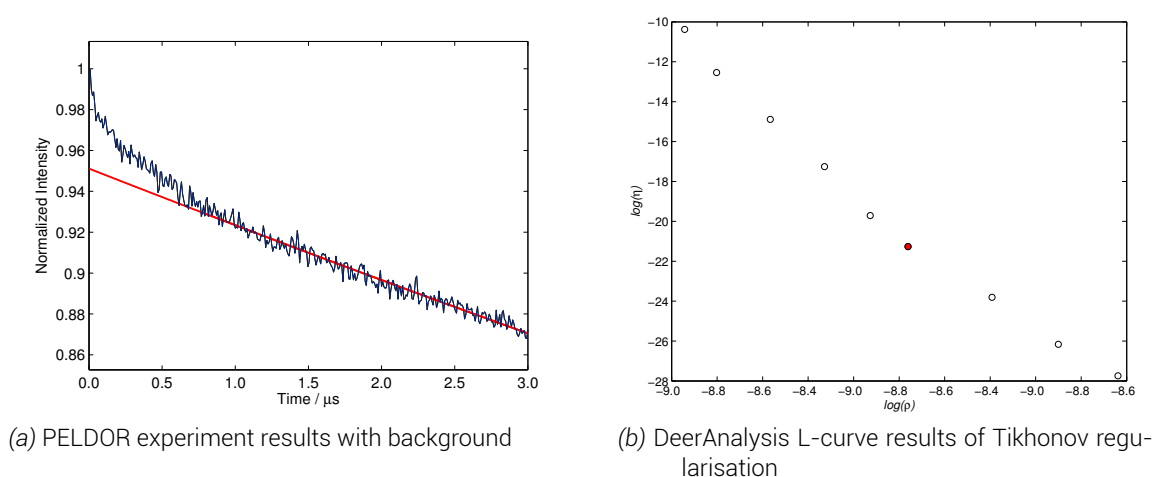
D.2.11. Colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}

Figure D.14 PELDOR experiment results for colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}, in red the background data, from running Im9 C23^{IA1} with unlabelled colicin E9 S3C/S80C under identical experimental parameters. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μs shot rep. time, ~9.68 GHz, 342.60 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

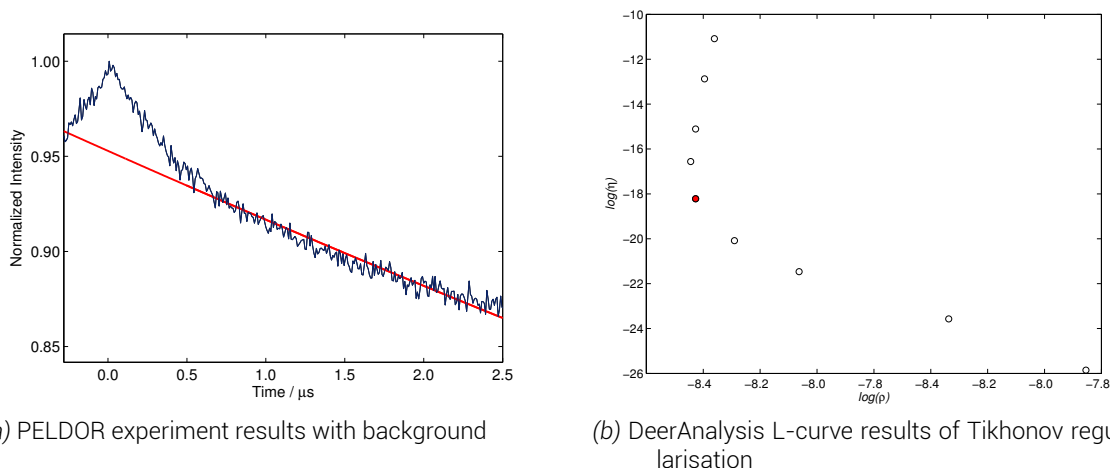
D.2.12. Colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S

Figure D.15 PELDOR experiment results for colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S, in red the background data, from running colicin E9 DNase S3C^{IA1} with unlabelled Im9 under identical experimental parameters. Here the fit has been artificially extended back to before the zero time using linear extrapolation. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 11.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.71 GHz, 346.30 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

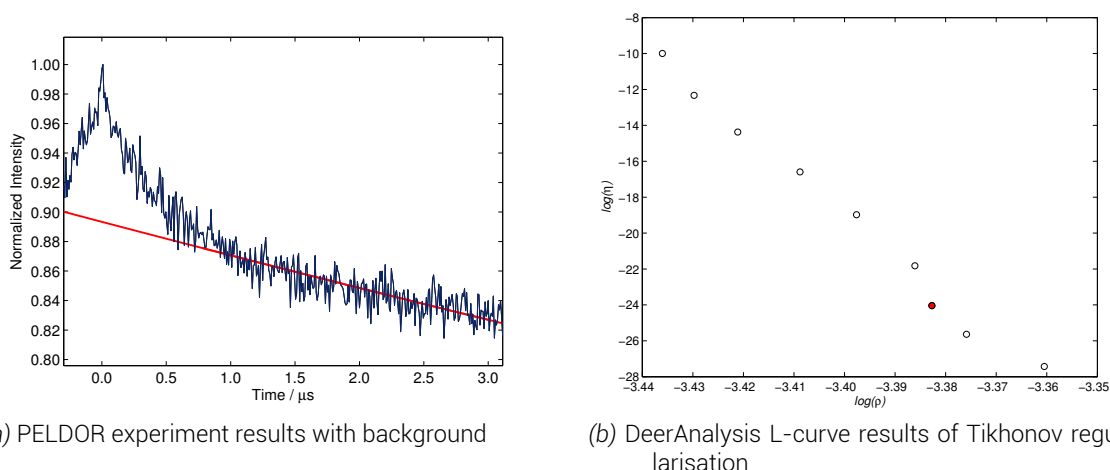
D.2.13. Colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}

Figure D.16 PELDOR experiment results for colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S3C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}, in red the background data, from running colicin E9 DNase S3C^{IA1} with unlabelled Im9 under identical experimental parameters. Here the fit has been artificially extended back to before the zero time using linear extrapolation. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 12.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, ~9.70 GHz, 342.60 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

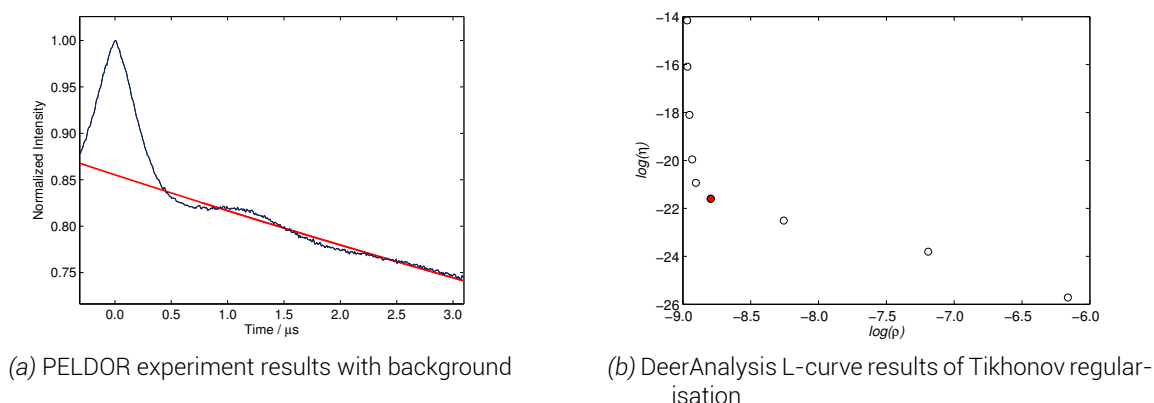
D.2.14. Colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}

Figure D.17 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 C23^{IA1}, in red the background data, from running Im9 C23^{IA1} with unlabelled colicin E9 S3C/S80C under identical experimental parameters. Here the fit has been artificially extended back to before the zero time using linear extrapolation. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 10.0 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μs shot rep. time, ~9.70 GHz, 342.50 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

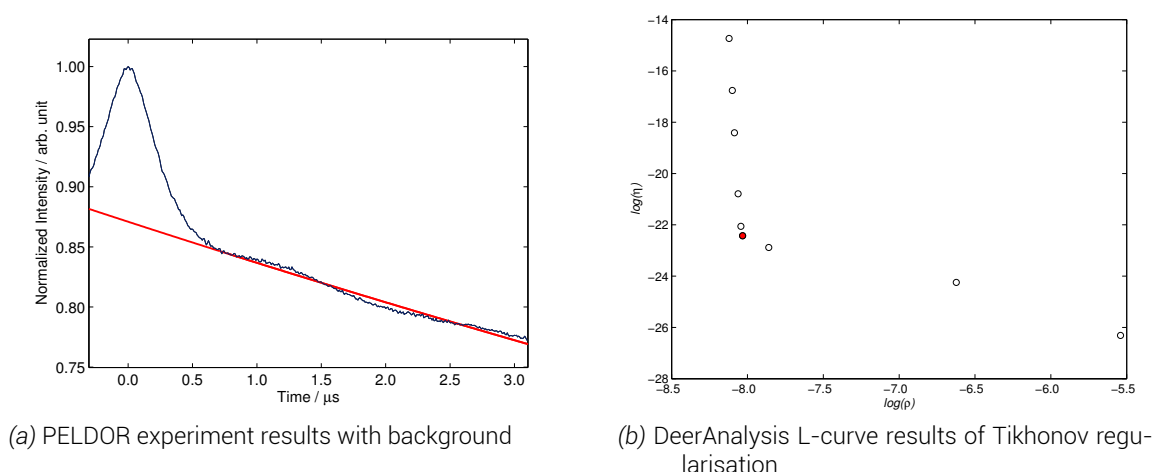
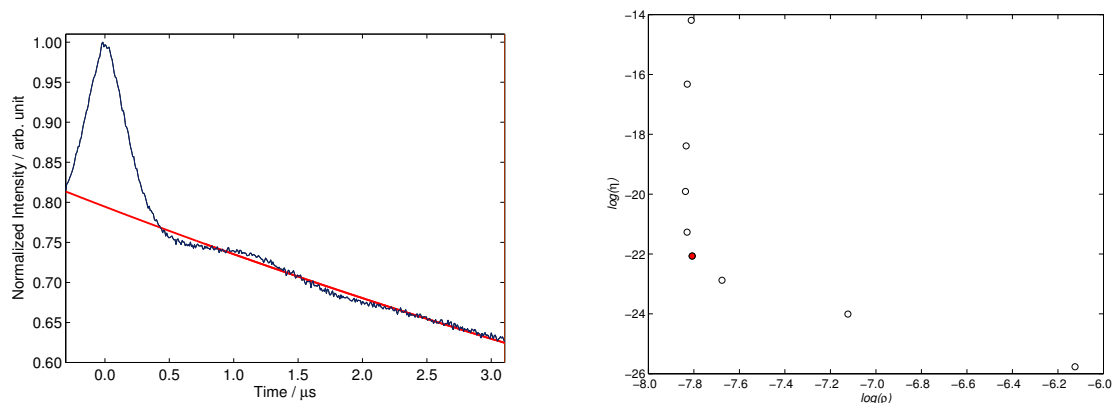
D.2.15. Colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S

Figure D.18 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23S, in red the background data, from running Im9 S6C^{IA1}/C23S with unlabelled colicin E9 S30C/S80C under identical experimental parameters. Here the fit has been artificially extended back to before the zero time using linear extrapolation. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (~ 2500 scans), 4 shots/point, 428 points, 8.0 dB attenuation, 20 MHz video bandwidth, 63 dB video gain and 10000 μs shot rep. time, ~9.79 GHz, 345.50 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μM in D₂O and 50 % glycerol-d₈.

D.2.16. Colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}

(a) PELDOR experiment results with background

(b) DeerAnalysis L-curve results of Tikhonov regularisation

Figure D.19 PELDOR experiment results for colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}. a) shows in navy the raw 4 pulse PELDOR data of colicin E9 DNase S30C^{IA1}/S80C^{IA1} - Im9 S6C^{IA1}/C23^{IA1}, in red the background data, from running Im9 C23^{IA1} with unlabelled colicin E9 S30C/S80C under identical experimental parameters. Here the fit has been artificially extended back to before the zero time using linear extrapolation. b) shows the DeerAnalysis L-curve Tikhonov analysis of the background subtracted data. The data point highlighted in red shows the regularisation parameter used in the final distance distribution. 50 K, 20 hours collection (\sim 2500 scans), 4 shots/point, 428 points, 12.0 dB attenuation, 20 MHz video bandwidth, 69 dB video gain and 10000 μ s shot rep. time, \sim 9.70 GHz, 346.30 mT. $\tau_1 = 380$ nm, $\tau_2 = 3200$ nm, ELDOR pulse step 8 ns, no nuclear modulation suppression. Protein concentration 100 μ M in D₂O and 50 % glycerol-d₈.

E. Consulted materials

References

- [1] C. Altenbach, W. Froncisz, R. Hemker, H. Mchaourab and W. Hubbell, *Biophysical Journal*, 2005, **89**, 2103--2112.
- [2] M. Arakawa, H. Kagi and H. Fukazawa, *Journal of Molecular Structure*, 2010, **972**, 111--114.
- [3] S. J. Barnett, *Physical Review (American Physical Society)*, 1915, **6, 4**, 239.
- [4] G. J. Bartlett, C. T. Porter, N. Borkakoti and J. M. Thornton, *Journal of Molecular Biology*, 2002, **324**, 105--121.
- [5] H. Benedetti, C. Lazdunski and R. Lloubès, *EMB*, 1991, **10 (8)**, 1989--1995.
- [6] M. A. Bernstein, K. F. King and J. X. Zhou, *Handbook of MRI Pulse Sequences*, Elsevier Academic Press, 2004, p. 960.
- [7] R. Boetzel, M. Czisch, R. Kaptein, A. M. Hemmings, R. James, C. Kleantous and G. R. Moore, *Protein Science*, 2000, **9 (9)**, 1709--1718.
- [8] A. Bondi, *The Journal of Physical Chemistry*, 1964, **68**, 441--451.
- [9] A. M. J. J. Bonvin, R. Boelens and R. Kaptein, *Current Opinion in Chemical Biology*, 2005, **9**, 501--508.
- [10] I. R. Booth, *Microbiological Reviews*, 1985, **49 (4)**, 359--378.
- [11] P. P. Borbat and J. H. Freed, *Methods in enzymology*, 2007, **423**, 52--116.
- [12] C. M. Bowman, J. E. Dahlberg, T. Ikemura, J. Konisky and M. Nomura, *Proceedings of the National Academy of Sciences*, 1971, **68 (5)**, 964--968.
- [13] V. Braun, S. I. Patzer and K. Hantke, *Biochimie*, 2002, **84**, 365--380.
- [14] C. J. Camacho, *Proteins: Structure, Function, and Bioinformatics*, 2005, **60**, 245--251.
- [15] E. Cascales, S. K. Buchanan, D. Duche, C. Kleantous, R. Lloubes, K. Postle, M. Riley, S. Slatin and D. Cavard, *Microbiol. Mol. Biol. Rev.*, 2007, **71 (1)**, 158--229.
- [16] D. Cavard and C. Lazdunski, *Journal of Bacteriology*, 1990, **172 (2)**, 648--652.
- [17] D. Cavard, *FEMS Microbiology Letters*, 1994, **116**, 37--42.
- [18] D. Cavard and C. Lazdunski, *FEMS Microbiology Letters*, 2006, **12 (4)**, 311--316.
- [19] K.-F. Chak and R. James, *Journal of General Microbiology*, 1986, 61--71.
- [20] S. T. Cole, B. Saint-Joanis and Pugsley, *Molecular & General Genetics*, 1985, **198 (3)**, 465--472.
- [21] F. H. J. Cornish, *Proceedings of the Physical Society*, 1965, **3**, 427.
- [22] S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius and J. P. Rosenbusch, *Nature*, 1992, **358**, 727--33.
- [23] W. A. Cramer, J. R. Dankert and Y. Uratani, *Biochimica et Biophysica Acta*, 1983, **737 (1)**, 173--193.
- [24] L. J. Curtis, in *Atomic structure and lifetimes : a conceptual approach*, Cambridge University Press, 2003, p. 74.
- [25] M. D. Curtis and R. James, *Molecular Microbiology*, 1991, **5 (11)**, 2727--2733.
- [26] S. J. de Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar and A. M. J. J. Bonvin, *Proteins: Structure, Function, and Bioinformatics*, 2007, **69**, 726--733.
- [27] C. A. Dennis, H. Videler, R. A. Pauptit, R. Wallis, R. James, G. R. Moore and C. Kleantous, *Biochemical Journal*, 1998, **333**, 183--191.
- [28] C. Dominguez, R. Boelens and A. M. J. J. Bonvin, *Journal of the American Chemical Society*, 2003, **125**, 1731--1737.
- [29] D. Duché, A. Frenkian, V. Prima and R. Lloubès, *Journal of bacteriology*, 2006, **188**, 8593--600.
- [30] T. Eaton and R. James, *Nucleic Acids Research*, 1989, **17 (4)**, 1761.

E. Consulted materials

- [31] L. P. Ehrlich, M. Nilges and R. C. Wade, *Proteins: Structure, Function, and Bioinformatics*, 2005, **58**, 126--133.
- [32] M. Feldgarden and M. A. Riley, *Evolution*, 1999, **53 (4)**, 1019--1027.
- [33] K. Furić and V. Volovšek, *Journal of Molecular Structure*, 2010, **976**, 174--180.
- [34] C. Garinot-Schneider, A. J. Pommer, G. R. Moore, C. Kleanthous and R. James, *Journal of Molecular Biology*, 1996, **260**, 731--742.
- [35] A. Giannoulis, R. Ward, E. Branigan, J. H. Naismith and B. E. Bode, *Molecular Physics*, 2013, **111**, 2845--2854.
- [36] S. C. Gill and P. H. von Hippel, *Analytical Biochemistry*, 1989, **182 (2)**, 319--326.
- [37] D. M. Gordon, M. A. Riley and T. Pinou, *Microbiology (Reading, England)*, 1998, **144 (Pt 8)**, 2233--40.
- [38] G. Guihard, P. Boulanger, H. Bénédicti, R. Llobés, M. Besnard and L. Letellier, *Journal of Biological Chemistry*, 1994, **269 (8)**, 5874--5880.
- [39] D. A. Haas, T. Sugano, C. Mailer and B. H. Robinson, *The Journal of Physical Chemistry*, 1993, **97**, 2914--2921.
- [40] E. Hann, N. Kirkpatrick, C. Kleanthous, D. A. Smith, S. E. Radford and D. J. Brockwell, *Biophysical Journal*, 2007, **92**, 79--81.
- [41] R. E. Harkness and V. Braun, *Journal of Biological Chemistry*, 1989, **264 (11)**, 6177--6182.
- [42] H. R. Herschma and D. R. Helinski, *Journal of Biological Chemistry*, 1967, **242 (22)**, 5360.
- [43] N. G. Housden, J. T. S. Hopper, N. Lukoyanova, D. Rodriguez-Larrea, J. A. Wojdyla, A. Klein, R. Kaminska, H. Bayley, H. R. Saibil, C. V. Robinson and C. Kleanthous, *Science (New York, N.Y.)*, 2013, **340**, 1570--4.
- [44] N. G. Housden, S. R. Loftus, G. R. Moore, R. James and C. Kleanthous, *Proceedings of the National Academy of Sciences*, 2005, **102 (39)**, 13849--13854.
- [45] E. J. Hustedt and A. H. Beth, *Annual review of biophysics and biomolecular structure*, 1999, **28**, 129--53.
- [46] R. James, C. Kleanthous and G. R. Moore, *Microbiology-UK*, 1996, **142 (7)**, 1569--1580.
- [47] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger and H. Jung, *Applied Magnetic Resonance*, 2006, **30**, 473--498.
- [48] G. Jeschke, *Macromolecular Rapid Communications*, 2002, **23**, 227--246.
- [49] G. Jeschke and Y. Polyhach, *Physical Chemistry Chemical Physics*, 2007, **9**, 1895--1910.
- [50] A. H. Keeble, A. M. Hemmings, R. James, G. R. Moore and K. C., *Biochemistry*, 2002, **41 (13)**, 10234--10244.
- [51] B. C. Kirkup and M. A. Riley, *Nature*, 2004, **428**, 412--4.
- [52] D. Kivelson, *The Journal of Chemical Physics*, 1960, **33**, 1094.
- [53] C. Kleanthous and D. Walker, *Trends in Biochemical Sciences*, 2001, **26 (10)**, 624--631.
- [54] S. K. Levensgood-Freyermuth, E. M. Click and R. E. Webster, *Journal of Bacteriology*, 1993, **175**, 222--228.
- [55] W. Li, C. A. Dennis, G. R. Moore, R. James and C. Kleanthous, *Journal of Biological Chemistry*, 1997, **272 (35)**, 22253--22258.
- [56] M. Lindgren, G. Eaton, S. Eaton, B.-H. Jonsson, P. Hammarstrom, M. Svensson and U. Carlsson, *Journal of the Chemical Society, Perkin Transactions 2*, 1997, **12**, 2549--2554.
- [57] A. Maton, J. Hopkins, S. Johnson, D. Q. LaHart, M. Warner and J. D. Wright, *Cells Building Blocks of Life*, New Jersey: Prentice Hall, 1997.
- [58] S. F. Mel, A. M. Falick, A. L. Burlingame and R. M. Stroud, *Biochemistry*, 1993, **32 (36)**, 9473--9479.
- [59] B. Michel, *PLoS Biology*, 2005, **3 (7)**, 255.
- [60] M. Monk and R. C. Clowes, *Journal of General Microbiology*, 1964, **36**, 385--392.
- [61] J. A. Moore, *Journal of the History of the Behavioral Sciences*, 1994, **30 (4)**, 427--431.
- [62] K. Mosbahi, C. Lemaître, A. H. Keeble, H. Mobasher, B. Morel, R. James, G. R. Moore, E. J. A. Lea and C. Kleanthous, *Nature Structural Biology*, 2002, **9**, 476--484.
- [63] H. Murata, Ken-ichiro; Tanaka, *Nature Materials*, 2012, **11**, 436--443.
- [64] B. J. Murray and A. K. Bertram, *Physical Chemistry Chemical Physics*, 2006, **8**, 186--192.
- [65] Y. Nakamura, T. Gojobori and T. Ikemura, *Nucleic acids research*, 2000, **28**, 292.
- [66] D. Nelson and M. Cox, in *Lehninger Principles of Biochemistry (4th Edition)*, WH Freeman, New York, 4th edn., 2004, pp. 1073--1074.
- [67] H. Nikaido, *Microbiology and Molecular Biology Reviews*, 2003, **67 (4)**, 593--656.

- [68] T. Ogawa, M. Hidaka, K. Kohno and H. Masaki, *Journal of Biochemistry*, 2009, **145 (4)**, 461--466.
- [69] M. J. Osborne, A. L. Breeze, L. Y. Lian, A. Reilly, R. James, C. Kleanthous and G. R. Moore, *Biochemistry*, 1996, **35 (29)**, 9505--9512.
- [70] G. Otting, *Annual review of biophysics*, 2010, **39**, 387--405.
- [71] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, *Protein Science*, 1995, **4**, 2411--2423.
- [72] P. V. L. Padmavathi and H.-J. Steinhoff, *Journal of molecular biology*, 2008, **378**, 204--14.
- [73] M. W. Parker, F. Pattus, A. D. Tucker and D. Tsernoglou, *Nature*, 1989, **337 (6202)**, 93--96.
- [74] W. Pauli, in *Exclusion principle and quantum mechanics*, Elsevier Publishing, Amsterdam, 1964, ch. Nobel Priz, pp. 27--43.
- [75] C. Perez-Iratxeta and M. A. Andrade-Navarro, *BMC structural biology*, 2008, **8**, 25.
- [76] A. J. Pommer, S. Cal, A. H. Keeble, D. Walker, S. J. Evans, U. C. Kühlmann, A. Cooper, B. A. Connolly, A. M. Hemmings, G. R. Moore, R. James and C. Kleanthous, *Journal of Molecular Biology*, 2001, **314 (4)**, 735--749.
- [77] A. J. Pommer, *Ph.D. thesis*, School of Chemistry, University of East Anglia, 1997.
- [78] A. Popescu and R. J. Doyle, *Biotechnic & Histochemistry*, 1996, **71**, 145--151.
- [79] A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *Journal of the American Chemical Society*, 2010, **132**, 9040--8.
- [80] B. S. Prabhananda and J. S. Hyde, *The Journal of Chemical Physics*, 1986, **85**, 6705--6712.
- [81] J. Pyka, J. Ilnicki, C. Altenbach, W. L. Hubbell and W. Froncisz, *Biophysical journal*, 2005, **89**, 2059--68.
- [82] K. Schaller and M. Nomura, *Proceedings of the National Academy of Sciences of the United States of America*, 1976, **73 (11)**, 3989--3993.
- [83] A. Schweiger, C. Gemperle and R. Ernst, *Journal of Magnetic Resonance (1969)*, 1990, **86**, 70--81.
- [84] K. I. Sen, T. M. Logan and P. G. Fajer, *Biochemistry*, 2007, **46**, 11639--49.
- [85] I. Solomon, *Physical Review*, 1955, **99**, 559--565.
- [86] T. J. Stone, T. Buckman, P. L. Nordio and H. M. McConnell, *Proceedings of the National Academy of Sciences of the United States of America*, 1965, **54**, 1010--7.
- [87] D. G. Thanassi and S. J. Hultgren, *Current Opinion in Cell Biology*, 2000, **12 (4)**, 420--430.
- [88] A. J. Thomson and H. B. Gray, *Current Opinion in Chemical Biology*, 1998, **2**, 155--158.
- [89] M. Ubbink, J. A. R. Worrall, G. W. Canters, E. J. J. Groenen and M. Huber, *Annual review of biophysics and biomolecular structure*, 2002, **31**, 393--422.
- [90] T. Uchimura and P. C. K. Lau, *Molecular & General Genetics*, 1987, **209 (3)**, 489--493.
- [91] E. van den Bremer, W. Jiskoot, R. James, G. Moore, C. Kleanthous, A. Heck and C. Maier, *Protein Science*, 2002, **11 (7)**, 1738--1752.
- [92] A. D. J. van Dijk, R. Boelens and A. M. J. J. Bonvin, *FEBS Journal*, 2005, **272**, 293--312.
- [93] M. Vankemmelbeke, Y. Zhang, G. R. Moore, C. Kleanthous, C. N. Penfold and R. James, *The Journal of biological chemistry*, 2009, **284**, 18932--41.
- [94] A. Volkov, C. Dockter, T. Bund, H. Paulsen and G. Jeschke, *Biophysical journal*, 2009, **96**, 1124--41.
- [95] T. von Hagens, Y. Polyhach, M. Sajid, A. Godt and G. Jeschke, *Physical chemistry chemical physics : PCCP*, 2013, **15**, 5854--66.
- [96] R. Wallis, K. Y. Leung, M. J. Osborne, R. James, G. R. Moore and C. Kleanthous, *Biochemistry*, 1998, **37 (2)**, 476--485.
- [97] R. Wallis, G. R. Moore, R. James and C. Kleanthous, *Biochemistry*, 1995, **34 (42)**, 13743--13750.
- [98] R. Wallis, K.-Y. Leung, A. J. Pommer, H. Videler, G. R. Moore, R. James and C. Kleanthous, *Biochemistry*, 1995, **34 (42)**, 13751--13759.
- [99] R. Wallis, G. R. Moore, C. Kleanthous and R. James, *European Journal of Biochemistry*, 1992, **210**, 923--930.
- [100] R. Wallis, A. Reilly, K. Barnes, C. Abell, D. G. Campbell, G. R. Moore, R. James and C. Kleanthous, *European Journal of Biochemistry*, 1994, **220**, 447--454.
- [101] R. Wallis, A. Reilly, A. Rowe, G. Moore, R. James and C. Kleanthous, *European Journal of Biochemistry*, 1992, **207**, 687--695.
- [102] R. Ward, A. Bowman, H. El-Mkami, T. Owen-Hughes and D. G. Norman, *Journal of the American Chemical Society*, 2009, **131**, 1348--9.
- [103] R. Ward, A. Bowman, E. Sozudogru, H. El-Mkami, T. Owen-Hughes and D. G. Norman, *Journal of Magnetic Resonance*, 2010, **207**, 164--167.

E. Consulted materials

- [104] R. C. Weast and M. J. Astle, *Handbook of Chemistry and Physics*, Boca Raton: CRC Press, 1982, p. E66.
- [105] C. R. Weinhold F Landis, in *Valency and bonding : a natural bond orbital donor-acceptor perspective*, Cambridge University Press, 2005, pp. 715--716.
- [106] G. F. White, L. Ottignon, T. Georgiou, C. Kleantous and G. R. Moore, 2007, **185**, 191--203.
- [107] S. J. Wodak and J. Janin, *Journal of Molecular Biology*, 1978, **124**, 323--342.
- [108] S. D. Zakharov, V. Y. Eroukova, T. I. Rokitskaya, M. V. Zhalnina, O. Sharma, P. J. Loll, H. I. Zgurskaya, Y. N. Antonenko and W. A. Cramer, *Biophysical Journal*, 2004, **87 (6)**, 3901--3911.
- [109] S. D. Zakharov, O. Sharma, M. Zhalnina, E. Yamashita and W. A. Cramer, *Biochemical Society transactions*, 2012, **40**, 1463--8.
- [110] S. D. Zakharov, M. V. Zhalnina, O. Sharma and William A. Cramer, *Biochemistry*, 2006, **45 (34)**, 10199--10207.
- [111] E. Zavoisky, *Journal of Physics USSR*, 1945, **9**, 245--249.
- [112] E. Zavoisky, *Journal of Physics USSR*, 1946, **10**, 197--198.
- [113] E. Zavoisky, *Journal of Physics USSR*, 1945, **9**, 211--216.
- [114] E. Zavoisky, *Journal of Physics USSR*, 1947, **11**, 184--188.
- [115] G. Jeschke, in *Structure and Bonding*, Springer Berlin Heidelberg, 2011, pp. 1--38.

Bibliography

The following books and lecture materials are general source materials that are used non-specifically used throughout this thesis.

E. Bordignon, *Niroxide spectrum analysis - lecture notes*, ETH Zurich

M. Brustolon, E. Giamello, *Electron Paramagnetic Resonance: A Practitioner's Toolkit*, Wiley-VCH, 2009

I. D. Campbell, R. A. Dwek, *Biological Spectroscopy*, The Benjamin/Cummings Publishing Company, 1984 pp. 127-215

W. R. Hagen, *Biomolecular EPR Spectroscopy*, CRC Press, 2009

G. Henson, L. Berliner, *Metals in Biology: Applications of High Resolution EPR to Metalloenzymes*, Springer Science, 2010

P. J. Hore, *Oxford Chemistry Primers: Nuclear Magnetic Resonance*, Oxford University Press, 1995

G. Jeschke, *NMR and EPR lecture notes*, ETH Zurich

A. Kukil, *Molecular Modeling of Proteins (Methods Molecular Biology, Volume 443)*, Springer Link, 2008, chapter 13

G. I. Likhtenshtein, J. Yamauchi, S. Nakatsuji, A. I. Smirnov, R. Tamura, *Nitroxides - Applications in Chemistry, Biomedicine and Materials Science*, Wiley-VCH, 2008

Y. Polyhach, *Modelling of spin labels with rotamer libraries - EPR summer school lecture notes*, ETH Zurich, 2010

A. Schweiger, G. Jeschke, *Principles of pulse electron paramagnetic resonance*, Oxford University Press, 2001

R. T. Weber, *Elexsys E500 User's Manual v1.1*, Bruker Instruments Incorporated, USA, 2000

R. T. Weber, *Elexsys E580 User's Manual v1.0*, Bruker Instruments Incorporated, USA, 2000

J. A. Weil, J. R. Bolton, *Electron Paramagnetic Resonance: Elementary Theory and Practical Applications (2nd edition)*, Wiley-Interscience, 2007

Mailing lists

Academic software is often lacking in official documentation instead favouring a community support system. The following mailing lists were consulted or posted to in the course of this investigation.

CNS bullet-in board - tech.groups.yahoo.com/group/cnsbb/

- subscription from cnsbb-subscribe@yahoogroups.com

HADDOCK - tech.groups.yahoo.com/group/haddock-discuss/

- subscription from haddock-discuss-subscribe@yahoogroups.com

Xplor-NIH help list - nmr.cit.nih.gov

- subscription from xplor-nih-request@nmr.cit.nih.gov

Softography

ChemBioDraw - cambridgesoft.com/software/ChemBioDraw/ Biological / Chemical drawing package	Graphical user interface for \LaTeX
Chromium - google.com/chrome Web browser	MatLab - mathworks.com/products/matlab Technical computing solution
CNSsolve - cns-online.org structural calculation software	Mendeley - mendeley.com Reference manager
DeerAnalysis - epr.ethz.ch/software/ EPR raw data (PELDOR) to distance calculation	MISHAP - morganbye.net/mishap EPR data to HADDOCK conversion tool
Dia - projects.gnome.org/dia Flow chart creation	MMM - epr.ethz.ch/software/ <i>in silico</i> modeling of spin label attachment
EasySpin - easyspin.org EPR spectra simulation	MSMS - mgltools.scripps.edu/packages/MSMS Atomic radius calculations
EPR Toolbox - morganbye.net/eprtoolbox EPR data manipulation toolbox for MATLAB	NACCESS - bioinf.manchester.ac.uk/naccess Protein surface accessibility calculator
FileZilla - filezilla-project.org FTP program	openSUSE - opensuse.org UNIX based operating system
Firefox - mozilla.com/firefox Web browser	OriginPro - originlab.com Graph creation
GENEius - genius.de Optimisation of synthetic genes	PDF-Shuffler - sourceforge.net/projects/pdfshuffler PDF management tool
GitHub - github.com Online code hosting, version management	ProFit - bioinf.org.uk Protein structure comparison program
GIMP - gimp.org Image manipulation program	PyMOL - pymol.org pdb visualiser and figure creation
HADDOCK - nmr.chem.uu.nl/haddock Protein docking software	Roboto - developer.android.com/design/style/typography open-source font used here
Inkscape - inkscape.org Vector image drawing	Shutter - shutter-project.org Screenshot grabber
JabRef - jabref.sourceforge.net Bi \TeX referencing	Thunderbird - mozilla.com/thunderbird Email solutions
\LaTeX - latex-project.org Type setting environment	Ubuntu - ubuntu.com UNIX based operating system
LibreOffice - libreoffice.org Word processor, spreadsheet, presentations	VMware Player - vmware.com Virtual machine environment
LyX - lyx.org	Xepr - bruker-biospin.com/xepr EPR equipment control and data capture
	XeTeX - scripts.sil.org/xetex Font control for \LaTeX

F. Personal

Publications

M. Bye, J. Bradley, C. Bauer, J. Bychowksi, D. Hinderberger, F. MacMillan, *Closed cycle low temperature EPR*, Nature Protocols, *in preparation*

M. Bye, G. Moore, F. MacMillan, *Combining experimental long range EPR distance restraints with docking models*, Nature Structural Biology, *in preparation*

Conference and invited presentations

Combining NMR with EPR distances and in silico calculations for a more complete protein-protein interaction model. Centre for Molecular and Structural Biology Young Researcher's Colloquium, Norwich, UK (2011)

Complex docking models - elucidating protein-protein interactions with EPR. 46th Annual International Meeting of the ESR Spectroscopy Group of the RSC, University of Warwick, Warwick, UK (2013). **Jeol prize talk**

Protein interactions - a complex docking story. Weizmann Institute of Science, Rehovot, Israel (2013)

Complex docking models - elucidating protein-protein interactions with EPR. 55th Annual Rocky Mountain Conference on Magnetic Resonance, Crowne Plaza, Denver, CO, USA (2013)

Conference posters

M Bye and F MacMillan, *Improving protein-protein binding models using EPR*. 5th EF-EPR Summer School on Advanced EPR Spectroscopy, Konstanz, Germany.

M Bawn, M Bye, J van Wonderen and F MacMillan, *What EPR can do for you*. RSC: Molecular Interactions at the interface between chemistry and biology, Norwich, UK.

M Bye, A Stratford, G Moore and F MacMillan, *Combining NMR docking data with EPR distances and in silico calculations for a more complete model of colicin protein-protein interactions*. 44th Annual International Meeting of the ESR Spectroscopy Group of the RSC, University of York, UK.

M Bye, A Stratford, G Moore and F MacMillan, *Combining NMR docking data with EPR distances and in silico calculations for a more complete model of colicin protein-protein interactions*. 45th Annual International Meeting of the ESR Spectroscopy Group of the RSC, University of Manchester, UK. **Poster prize winner**.

F. Personal

M Bye, A Stratford, G Moore and F MacMillan, *Combining NMR docking data with EPR distances and in silico calculations for a more complete model of colicin protein-protein interactions*. How bugs kill bugs: progress and challenges in bacteriocin research, Biochemical Society, University of Nottingham, UK.

M Bye, A Morley and F MacMillan, *Moving beyond traditional structure - elucidating colicin protein-protein complex structures with EPR*. Molecular machineries for ion translocation across biomembranes - COST action CM0902 - training school, Universidade de Lisboa, Lisbon, Portugal

Curriculum Vitae

Personal

Born 15th August 1987

Citizenship British

Education

2009-2013 PhD student supervised by Dr F MacMillan at the University of East Anglia, Norwich, UK

2005-2009 MSci (Hons) Biochemistry and Biological Chemistry (2:1) at the University of Nottingham, UK

2003-2005 A-Levels in Biology (A), Chemistry (A) and Information and Communication Technology (A) at Downham Market Sixth Form College, UK

1998-2003 General Certificates of Secondary Education, 15 A*-C grades (or equivalents) including English, Mathematics and Science at Downham Market High School, UK

Professional memberships

Member - Royal Society of Chemistry, UK - 8 years

Member - Biochemical Society, UK - 2 years

Other Academic

Board of the School of Chemistry - Postgraduate Representative

Schools of Chemistry and Pharmacy Safety Committee - Postgraduate Representative

Demonstration and seminar leader:

- ◇ undergraduate biophysical laboratories
- ◇ undergraduate *in silico* laboratories
- ◇ undergraduate introductory mathematics for scientists
- ◇ 14 - 16 year old school children "Discovery Spectroscopy" open days
- ◇ 9 - 10 year old school children "Discover Chemistry" days

Further information

For further information about the work in this thesis please contact me via the contact form at morganbye.net/contact