Biological NMR: Probing Biomolecular Condensates, Exploration of a Disordered Virus Capsid Protein and Reducing Molecular Size Limitations



James Eaton

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2025

Dedicated to my partner Belia and the rest of my family

Abstract

Solution-state nuclear magnetic resonance (NMR) spectroscopy is capable of characterising atom-specific structural and dynamical information of biomolecules. NMR excels in studying dynamical and liquid-state systems; however, its use in slowly tumbling biomolecules is restricted, due to signal intensity and resolution limitations. In this thesis, NMR was used to characterise biomolecular condensates and a disordered virus capsid protein. This was followed by an exploration of methods to enhance NMR spectral quality for slowly tumbling biomolecules.

Biomolecular condensates are formed through liquid-liquid phase separation and provide cellular compartmentalisation without a phospholipid membrane. The ability of condensates to partition metabolites was determined using an NMR-based method, which showed trends similar to "like-dissolves-like", used extensively to predict partitioning into organic solvents. Additionally, NMR revealed a significant change in the dynamical and chemical properties of water inside condensates. These measurements support the growing evidence that biomolecular condensates act as distinct solvents.

Secondly, the properties of an adenovirus capsid protein, N-terminal truncated protein VI (Δ54-pVI), were explored using NMR. Adenoviruses can cause common respiratory illnesses and are exploited as viral vectors for vaccination. Protein VI is essential for membrane disruption and endosomal escape into a host cell, crucial for adenoviral infectivity. Diffusion NMR showed the highly disordered nature of Δ54-pVI; however, the assignment of chemical shifts revealed regions with amphipathic α-helical propensity, which is thought to enhance interactions with membranes.

Finally, a review of methodologies to combat size limitations in NMR is provided through simulations. The benefits of $^{13}CF_3$ labelling on slowly tumbling biomolecules were also assessed. Experimental measurements on a 395 kilodalton protein showed the $^{13}CF_3$ spin system had favourable relaxation properties. Simulations predicted that the relaxation rate of the slowest relaxing coherence would be largely independent of molecular tumbling. Therefore, $^{13}CF_3$ labelling schemes could extend the applications of NMR in slower-tumbling biomolecules.

Acknowledgements

Firstly, I would like to thank the constant support from my partner Belia, my parents, and my sister. Without your unwavering support, this journey would have been far more difficult.

I would like to thank my supervisor Professor Andrew Baldwin for his tremendous help, feedback and guidance throughout my research career to date. His expertise, scientific rigour and desire for further scientific understanding are a constant source of inspiration and have shaped my development as a scientist.

I have had the pleasure to work with amazing colleagues during my time in the Baldwin group. A special mention goes to Virgínia who has been supportive from the very beginning and whom I have had the pleasure to collaborate with extensively. Thank you for also proof-reading chapters 1-3 of this thesis. I am also indebted to Charlie, who's technical NMR skills have shaped my understanding of practical NMR. I am also grateful to Abi for her ability to always find time for my NMR experiments at the Rosalind Franklin Institute. Thank you to everyone in the Baldwin group for making my time here thoroughly enjoyable.

Thank you also to Gogs, a former Baldwin lab member and my future supervisor at Bind research, for taking the time to proof-read chapter 5 of this thesis. I look forward to working together in the future.

I would like to acknowledge generous funding from Pembroke college (Holger Kramer Memorial Scholarship), the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Partnership Account (DTP) and the European Research Council (ERC, grant agreement No 101002859).

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own work.

James Eaton

Contents

Abstract ii

Acknowledgements iv

Declaration v

List of Tables xiv

List of Figures xvi

List of Equations xxv

List of Acronyms xxvii

1 Introduction 1

1.1 Using solution-state Nuclear Magnetic Resonance (NMR) to probe biology 1

1.2 Thesis outline 4

2 Introduction to NMR 6

2.1 NMR theory and relaxation 6

2.1.1 The Zeeman interaction and chemical shift 6

2.1.2 Fourier transform NMR 8

2.1.3 Relaxation 11

2.1.4 The spectral density 17

2.1.5 Density matrices and operators 19

2.2 NMR imaging experiments 23
2.2.1 Gradient echo experiments 23
2.2.2 Slice selective NMR experiments 25
2.3 Heteronuclear NMR 26
2.3.1 Scalar coupling coherence transfers 28
2.3.2 HSQC 31
2.4 Measuring relaxation in NMR 34
2.5 Measuring molecular translational diffusion using NMR 37
2.6 Measuring chemical exchange and interactions using NMR 39
2.6.1 Chemical Exchange Saturation Transfer (CEST) 39
2.6.2 Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion . . 41
2.6.3 The Nuclear Overhauser Effect and the Nuclear Overhauser Effect Spectroscopy (NOESY) experiment 43
2.7 NMR experiments used throughout this thesis 46

3 Exploration of the solvent environment in biomolecular condensates 47

3.1 Abstract 47
3.2 Introduction to biomolecular condensates 48
3.2.1 Liquid-liquid phase separation 52
3.2.2 Interactions driving phase separation 56
3.3 *In vitro* phase separating systems 57
3.3.1 The DDX4^{N1} model system 57
3.3.2 The EWSR1C model system 63
3.4 Partitioning into biomolecular condensates and organic solvents . . . 68
3.4.1 Introduction to molecular partitioning in solvents 68
3.4.2 NMR-based method to determine partitioning coefficients . . . 71
3.4.3 Metabolite partitioning results in condensates and organic solvents 78

3.4.4 Correlation of metabolite partitioning with metabolite molecular properties . . . 80
3.4.5 Correlation of standardised coefficients with solvent properties 92
3.4.6 Ion partitioning into DDX4^{N1} condensates . . . 93
3.5 Physical and chemical properties of water inside DDX4^{N1} condensates 99
3.5.1 Water as a solvent . . . 99
3.5.2 Biphasic condensate samples . . . 101
3.5.3 Magnetic Resonance Imaging (MRI) experiments . . . 103
3.5.4 Slice selective NMR experiments . . . 106
3.5.5 Chemical shift differences of water inside DDX4^{N1} condensates 108
3.5.6 Comparison of water chemical shift changes to that in organic solvents . . . 110
3.5.7 Chemical shift differences of DDX4^{N1} and metabolites inside DDX4^{N1} condensates . . . 112
3.5.8 Water-protein interactions in condensates revealed by CEST NMR . . . 115
3.5.9 NOESY measurements on DDX4^{N1} biphasic sample . . . 119
3.5.10 Diffusion and relaxation measurements of water inside condensates . . . 121
3.6 Conclusion . . . 127
3.7 Future work . . . 130
3.8 Contributions . . . 132

4 Probing the disordered adenovirus capsid protein VI using NMR 134

4.1 Abstract . . . 134
4.2 Introduction to adenoviruses and protein VI . . . 135
4.3 Δ54-pVI construct disorder revealed using diffusion NMR . . . 140
4.4 NMR backbone assignment of Δ54-pVI . . . 142

4.4.1 Band-selective excitation short-transient (BEST) experiments on disordered proteins . . . 146
4.4.2 NMR protein backbone assignment experiments . . . 149
4.4.3 NMR backbone assignment experimental acquisition and automated peak-picking for Δ54-pVI . . . 153
4.5 Semi-automated assignment of Δ54-pVI . . . 154
4.6 Secondary structure propensity of Δ54-pVI revealed through residue-specific NMR chemical shifts . . . 157
4.6.1 Comparison of experimental Δ54-pVI secondary structure to computational predictions . . . 163
4.7 Mapping the ^{1}H-^{15}N backbone amide correlation peaks assignments of Δ54-pVI from 5^{o}C to 20^{o}C . . . 166
4.8 Chemical exchange experiments . . . 168
4.8.1 CPMG . . . 170
4.8.2 CEST . . . 170
4.9 Δ54-pVI interaction with micelles . . . 171
4.10 Conclusions . . . 173
4.11 Future work . . . 175
4.12 Contributions . . . 176

5 Reducing NMR Limitations for Slowly Tumbling Biomolecules 177

5.1 Abstract . . . 177
5.2 Introduction . . . 178
5.2.1 NMR relaxation and an introduction to RelCalc/SimSeq . . . 180
5.2.2 Evaluation of the current methods to enhance NMR spectral quality in macromolecules using RelCalc and SimSeq . . . 182
5.3 Proposed relaxation interference in $^{13}CF_3$ groups . . . 198
5.4 $^{13}CF_3$ labelling of a 395 kDa protein . . . 200
5.5 NMR relaxation rates for the $^{13}CF_3$ and $^{13}CH_3$ systems . . . 204

5.6 Determination of the optimal 2D NMR experiment for $^{13}CF_3$ labelled HSP16.5 . . . 211
5.7 1D Carbon Spectra . . . 214
5.8 Conclusion . . . 217
5.9 Future work . . . 218
5.10 Contributions . . . 220

6 SpinExplorer: A graphical interface for NMR processing and analysis 221
6.1 Introduction . . . 221
6.2 What SpinExplorer can do . . . 222
6.2.1 SpinConverter . . . 222
6.2.2 Processing . . . 222
6.2.3 Viewing/overlaying spectra . . . 222
6.3 Contributions . . . 223

7 Conclusion 224

Bibliography 227

A Appendices for chapter 3: Exploration of the solvent environment in biomolecular condensates 256
A.1 Protein sequences . . . 256
A.1.1 DDX4^{N1} . . . 256
A.1.2 EWSR1C . . . 256
A.2 Protein expression and purification (DDX4^{N1}/EWSR1C) . . . 257
A.2.1 DDX4^{N1} protein production and purification . . . 257
A.2.2 EWSR1C protein production and purification . . . 259
A.3 Phase separation tests on DDX4^{N1} and EWSR1C . . . 260
A.3.1 DDX4^{N1} phase separation dependence on D_2O . . . 260

A.3.2 EWSR1C phase separation dependence on NaCl 262
A.3.3 EWSR1C phase separation dependence on D_2O 263
A.4 Methods (partitioning experiments) 264
A.4.1 Condensate partitioning method 264
A.4.2 Butanol partitioning method 267
A.4.3 NMR data acquisition for metabolite partitioning measurements 267
A.4.4 Condensate metabolite partitioning NMR data analysis 268
A.4.5 Condensate ion partitioning method 269
A.4.6 NMR data acquisition for the condensate ion partitioning measurements 270
A.5 Methods (biphasic $DDX4^{N1}$ NMR experiments) 271
A.5.1 Creation of a $DDX4^{N1}$ biphasic NMR sample 271
A.5.2 MRI experiments 272
A.5.3 Slice selective NMR experiments 272
A.5.4 CEST experiments 274
A.5.5 NOESY experiment 275
A.6 Condensate volume measurements 276
A.6.1 Metabolite condensate partitioning volume measurements . . . 276
A.6.2 $DDX4^{N1}$ condensate ion partitioning volume measurements . . 277
A.7 Partitioning coefficient formulae derivations 277
A.7.1 Derivation of partitioning coefficient formula used for the biomolecular condensate partitioning experiments 277
A.7.2 Derivation of the partitioning coefficient formula used for 1-butanol partitioning measurements 279
A.8 Partitioning results 281
A.8.1 Partitioning coefficients obtained for the biomolecular condensates and organic solvents 281

A.8.2 Free energy of partitioning measurements obtained for the biomolecular condensates and organic solvents 283
A.9 Seedless optimised pulses for HACA(CO) (figures A.3-A.10) 285
A.10 TOCSY assignment of amino acids (figures A.11-A.17) 293
A.11 NMR Assignment of Nucleotides 300
A.12 Metabolite Chemical Properties 301
A.13 Metabolite partitioning fitting results 303
A.13.1 Results of fitting metabolite properties to metabolite partitioning free energies (grouped properties fitting procedure) 303
A.13.2 Results of fitting metabolite properties to metabolite partitioning free energies (naïve fitting procedure) 304
A.13.3 Fitted coefficients from the naïve fitting procedure 305
A.13.4 Fitted coefficients from the grouped fitting (using χ^2) 306
A.13.5 Bootstrap histograms showing fitted parameter errors for the naïve fitting procedure 307
A.13.6 Bootstrap histograms showing fitted parameter errors for the grouped fitting procedure 310
A.14 Flory-Huggins Phase Separation Theory 312
A.14.1 Calculation of entropy of mixing (ΔS_{mix}) 313
A.14.2 Enthalpy of mixing 317
A.14.3 Free energy of mixing (ΔG_{mix}) 319

B Appendicies for chapter 4: probing the disordered adenovirus capsid protein VI using NMR 320

B.1 Protein sequences 320
B.1.1 pVI 320
B.1.2 Protein VI 320
B.1.3 Δ54-pVI 321
B.2 Methods 322

B.2.1 NMR assignment experiments and processing recorded on Δ54-pVI . . . 322
B.2.2 Experiments to map the assignment of Δ54-pVI from 5^{o}C to 20^{o}C . . . 323
B.2.3 Chemical exchange experiments on Δ54-pVI . . . 324
B.2.4 DDM experiments . . . 327
B.2.5 pH dependence of Δ54-pVI NMR spectra . . . 328
B.3 Chemical shift assignments of Δ54-pVI at 5^{o}C . . . 330

C Appendices for chapter 5: Reducing NMR Limitations for Slowly Tumbling Biomolecules 335

C.1 General expression for a relaxation matrix element $R_{a,b}$ . . . 335
C.2 Relaxation rate expressions . . . 338
C.2.1 Relaxation rate expressions for a ^{15}N-^{1}H spin system . . . 338
C.2.2 Relaxation rate expressions for a ^{13}C-^{19}F spin system . . . 341
C.2.3 Auto-relaxation rate expressions for a ^{13}C-^{1}H$_3$ spin system . . . 344
C.2.4 Relaxation rate expressions for a ^{13}C-^{19}F$_3$ spin system . . . 357
C.3 Methods . . . 358
C.3.1 Pulse sequence simulations . . . 358
C.3.2 Experimental NMR methods and analysis . . . 363
C.4 Justification of the inclusion of external protons for the fitting of ^{13}CH$_3$ relaxation rates . . . 369
C.5 Bootstrap histograms from the global ^{13}CF$_3$ and ^{13}CH$_3$ spin systems fitting procedures . . . 370
C.6 Origin of the 3:1:1:3 quartet peak ratio for ^{13}C in a HSQC on ^{13}CH$_3$ groups . . . 372

List of Tables

4.1 Peak list classification for backbone NMR assignment spectra 154

A.1 Experimental details for determining the D_2O dependence of $DDX4^{N1}$ phase separation 261

A.2 Saturation concentrations (c_{sat}) of $DDX4^{N1}$ at different D_2O percentages 261

A.3 Free energy changes of phase separation ($\Delta\Delta G$) for $DDX4^{N1}$ upon changing the buffer D_2O percentage 262

A.4 Experimental details for determining the NaCl dependence of EWSR1C phase separation 262

A.5 Saturation concentrations (c_{sat}) of EWSR1C at different NaCl percentages 263

A.6 Experimental details for determining the D_2O dependence of EWSR1C phase separation 263

A.7 Saturation concentrations (c_{sat}) of EWSR1C at different D_2O percentages 264

A.8 Free energy changes of phase separation ($\Delta\Delta G$) for EWSR1C upon changing the buffer D_2O percentage 264

A.9 Table of volumes of pellet (V_p) and leftover supernatant (V_{LOS}) for the condensate partitioning experiments 276

A.10 $DDX4^{N1}$ condensate ion partitioning volume measurements 277

A.11 Table 1 (of 2) of partitioning coefficients (PC) obtained for biomolecular condensates and organic solvents 281

A.12 Table 2 (of 2) of partitioning coefficients (PC) obtained for biomolecular condensates and organic solvents 282
A.13 Table 1 (of 2) of partitioning free energies ($\Delta G_{partitioning}$) obtained for biomolecular condensates and organic solvents 283
A.14 Table 2 (of 2) of partitioning free energies ($\Delta G_{partitioning}$) obtained for biomolecular condensates and organic solvents 284
A.15 Metabolite properties table one (of two) 301
A.16 Metabolite properties table two (of two) 302
A.17 Table of fitting results for global metabolite partitioning: grouped properties 303
A.18 Table of fitting results for global metabolite partitioning: non-grouped/naïve properties 304

B.1 Chemical shift assignments of Δ54-pVI (1 of 5) 330
B.2 Chemical shift assignments of Δ54-pVI (2 of 5) 331
B.3 Chemical shift assignments of Δ54-pVI (3 of 5) 332
B.4 Chemical shift assignments of Δ54-pVI (4 of 5) 333
B.5 Chemical shift assignments of Δ54-pVI (5 of 5) 334

C.1 Interaction constant definitions for RelCalc 336
C.2 A table of operators relevant to NMR relaxation mechanisms 336
C.3 Spectral density calculations for different motional models in RelCalc 337
C.4 Table of associated Legendre polynomials 338

List of Figures

1.1 Disorder and dynamics in proteins 2

2.1 The Zeeman interaction and chemical shift 7
2.2 The rotating frame and pulses 8
2.3 Introduction to Fourier transform NMR 10
2.4 Introduction to NMR relaxation 12
2.5 Dipolar and chemical shift anisotropy induced fields 13
2.6 Introduction to longitudinal relaxation 15
2.7 Introduction to transverse relaxation 16
2.8 Introduction to the spectral density function and rotational diffusion 18
2.9 NMR signal intensity and resolution as a function of tumbling time . 19
2.10 Introduction of magnetic resonance imaging (MRI) experiments . . . 24
2.11 Slice selective NMR sequence and explanation 25
2.12 The Fermi contact interaction and scalar coupling 26
2.13 INEPT coherence transfers 29
2.14 Heteronuclear Hartmann-Hahn coherence transfers 30
2.15 Heteronuclear single quantum coherence (HSQC) two-dimensional NMR experiments 32
2.16 Measuring longitudinal relaxation 35
2.17 Measuring transverse relaxation 35
2.18 Schematic of the stimulated echo longitudinal encode-decode diffusion NMR experiment 38

2.19 Introduction to the chemical exchange saturation transfer (CEST) experiment . . . 40
2.20 Introduction to the Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment . . . 42
2.21 Introduction to the Nuclear Overhauser Effect and the NOESY experiment . . . 45

3.1 Examples of biomolecular condensates and an introduction to liquid-liquid phase separation . . . 49
3.2 Solvent dependent properties . . . 51
3.3 Flory-Huggins theory and phase diagrams . . . 54
3.4 Inter/intra-molecular interactions enabling biological phase separation 57
3.5 Introduction to DDX4^{N1} . . . 59
3.6 Factors affecting DDX4^{N1} phase separation . . . 60
3.7 Previous evidence of an alternative solvent environment in DDX4^{N1} condensates . . . 62
3.8 Introduction to EWSR1 and EWSR1C . . . 65
3.9 Factors affecting EWSR1C phase separation . . . 66
3.10 Introduction to condensate partitioning and solubility rules . . . 69
3.11 Biomolecular condensate partitioning method . . . 71
3.12 One-dimensional proton NMR spectra of the nucleotides from the DDX4^{N1} dilute and resuspended condensate samples . . . 73
3.13 Protein-bound amino acids obscure free amino acid signals in proton NMR . . . 74
3.14 HACA(CO) pulse sequence and amino acid assignment . . . 75
3.15 HACA(CO) Cβ decoupling . . . 76
3.16 Amino acid partitioning into DDX4^{N1} condensates quantified using the HACACO sequence . . . 77
3.17 Nucleotide partitioning results into condensates . . . 78

3.18 Amino acid partitioning results into condensates 79
3.19 Chart of all partitioning results into condensates and organic solvents 80
3.20 Fitting procedure to correlate metabolite partitioning with metabolite properties 83
3.21 Fitting procedure to correlate metabolite partitioning with metabolite properties with property grouping 84
3.22 Correlation between metabolite partitioning and metabolite properties using the grouped fitting procedure 89
3.23 Correlation between metabolite property standardised coefficients and solvent/disordered protein properties 92
3.24 Sodium and phosphate ion partitioning results in DDX4^{N1} condensates 96
3.25 Depiction of ways condensate ion partitioning could effect biology . . 97
3.26 Introduction to water as a solvent 99
3.27 How water is altered by the addition of organic solvents 100
3.28 Method of producing a biphasic DDX4^{N1} sample 102
3.29 The properties of DDX4^{N1} condensates to be probed using a biphasic NMR sample 102
3.30 Results of MRI experiments on the DDX4^{N1} biphasic sample 104
3.31 Hardening of the interface of the biphasic DDX4^{N1} sample 105
3.32 Slice selective NMR sequence performance on a biphasic DDX4^{N1} condensate sample 107
3.33 Water chemical shift changes revealed by slice selective water NMR experiments 109
3.34 Comparison of water chemical shift changes in DDX4^{N1} condensates to water in organic solvents 111
3.35 Water suppressed slice-selective NMR experiment and DDX4^{N1} protein chemical shift changes inside DDX4^{N1} condensates 112
3.36 Metabolite chemical shift changes inside DDX4^{N1} condensates 113

3.37 Water-CEST profiles for supernatant and condensate water 116
3.38 Temperature dependence of water-CEST profile and comparison to water chemical shift differences 117
3.39 Water-CEST profile in concentrated BSA 118
3.40 NOESY spectrum of a biphasic DDX4^{N1} sample 120
3.41 NOESY cross peaks of condensate H_2O with protein 121
3.42 Water diffusion differences between DDX4^{N1} condensate and supernatant phases 123
3.43 Water longitudinal relaxation and molecular tumbling differences between DDX4^{N1} condensate and supernatant H_2O 125
3.44 Water transverse relaxation differences between DDX4^{N1} condensate and supernatant H_2O 126
3.45 Biomolecular condensate solvent properties conclusion 128

4.1 Introduction to the human adenovirus 136
4.2 Cellular infection pathway of human adenoviruses 137
4.3 Introduction to protein VI 139
4.4 NMR-based Δ54-pVI diffusion measurements 141
4.5 Dihedral angles, Ramachandran plots and transient secondary structure 143
4.6 Chemical shift dispersion in disordered and structured proteins 144
4.7 Protein NMR coupling constants 145
4.8 Introduction to the HNCO experiment 146
4.9 NMR backbone connectivities through the HNCO and HN(CA)CO experiments 151
4.10 NMR protein backbone assignment experiment coherence pathways . 152
4.11 Strip plots showing connections between two arbitrarily named NH peak correlations and graph matching with protein sequence 155
4.12 ^{1}H-^{15}N backbone amide assignment of Δ54-pVI 157

4.13 Deviations of Δ54-pVI Cα and Cβ chemical shifts from random coil values . . . 159
4.14 NMR determined Ramachandran plot of Δ54-pVI . . . 160
4.15 NMR determined residue-specific secondary structure propensity of Δ54-pVI . . . 161
4.16 Amphipathic nature of α-helical propensity in solution state Δ54-pVI 162
4.17 Comparison of computational secondary structure/disorder predictions for Δ54-pVI . . . 164
4.18 Comparison of the experimental secondary structure propensity of Δ54-pVI with the AlphaFold prediction . . . 165
4.19 Assignment of Δ54-pVI from 5^{o}C to 20^{o}C . . . 166
4.20 Temperature dependent chemical shift perturbations of Δ54-pVI . . . 167
4.21 CPMG dispersion and CEST Z-spectra for Δ54-pVI . . . 169
4.22 Interaction of Δ54-pVI with DDM micelles . . . 172
4.23 Conclusion figure of Δ54-pVI secondary structure propensity and its role in membrane binding and endosomal escape . . . 174

5.1 The spectral density function and transverse relaxation . . . 179
5.2 BMRB and PDB entries versus molecular weight . . . 179
5.3 Background deuteration benefits to NMR . . . 183
5.4 An introduction to relaxation interference in ^{15}N-^{1}H spin systems . . . 185
5.5 Quantification of relaxation interference in ^{15}N-^{1}H spin systems using RelCalc . . . 186
5.6 Simulations of TROSY and non-TROSY pulse sequences on a ^{15}N-^{1}H spin system . . . 187
5.7 Relaxation interference in a $^{13}CH_3$ group . . . 191
5.8 Single quantum carbon auto-relaxation rates in a $^{13}CH_3$ group . . . 192
5.9 A comparison of simulated spectra for a $^{13}CH_3$ group as a function of rotational correlation time . . . 193

5.10 Irreducible basis relaxation rates for a $^{13}CH_3$ group 194
5.11 Delayed decoupling sequences for the study of $^{13}CH_3$ group in slow tumbling systems 196
5.12 Aromatic ^{13}C-F spin system relaxation rates 198
5.13 Illustration of expected relaxation interference within single quantum carbon coherences in $^{13}CF_3$ groups 199
5.14 An overview of the $^{13}CF_3$ labelling method 200
5.15 HSP16.5 protein structure and 1D NMR spectra of $^{13}CF_3$/$^{13}CH_3$ labelled variants 201
5.16 Diffusion of $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 203
5.17 Sequences used to measured single quantum relaxation rates in $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 205
5.18 Spectra from the single quantum carbon relaxation experiment on $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^oC 206
5.19 Spectra from the single quantum fluorine/proton relaxation experiments on $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^oC 207
5.20 Experimental single quantum relaxation rates for $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^oC 208
5.21 Global fitting of single quantum relaxation rates for $^{13}CF_3$ and $^{13}CH_3$ labelled systems 209
5.22 HEHAHA pulse sequence 210
5.23 Comparison of 2D NMR experiments for $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 210
5.24 Simulated relative intensities of HSQC/HMQC pulse sequence for $^{13}CF_3$ and $^{13}CH_3$ spin systems 212
5.25 An evaluation of relaxation rates in the irreducible basis for a $^{13}CF_3$ spin system 213

5.26 Simulated carbon 1D spectra for $^{13}CF_3$ and $^{13}CH_3$ spin systems as a function of τ_c . . . 214
5.27 Experimental carbon 1D spectra for $^{13}CF_3$ and $^{13}CH_3$ HSP16.5 215
5.28 A comparison of HEHAHA and pulse-acquire carbon 1D experiments on $^{13}CF_3$-labelled HSP16.5 . . . 216

A.1 Detailed pulse sequences of the slice selective (**A**) and water-suppressed slice selective (**B**) sequences employed. . . . 273
A.2 Detailed pulse sequences of the CEST (**A**) and NOESY (**B**) sequences employed. . . . 274
A.3 Seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 90^o pulse ($^{13}C_\alpha$ centred) . . . 285
A.4 Seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 180^o pulse ($^{13}C_\alpha$ centred) . . . 286
A.5 Seedless optimisation output for a $^{13}C_o$ selective 180^o pulse ($^{13}C_\alpha$ centred) . . . 287
A.6 Seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 180^o pulse ($^{13}C_o$ centred) . . . 288
A.7 Seedless optimisation output for a $^{13}C_o$ selective 90^o pulse ($^{13}C_o$ centred) 289
A.8 Seedless optimisation output for a $^{13}C_\alpha$ selective 180^o pulse ($^{13}C_\alpha$ centred) . . . 290
A.9 Seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_o$ selective 180^o pulse ($^{13}C_\alpha$ centred) . . . 291
A.10 Seedless optimisation output for a $^{13}C_\beta$ selective inversion pulse ($^{13}C_\alpha$ centred) . . . 292
A.11 TOCSY H_α-C_α assignment for ARG, LYS and GLU . . . 293
A.12 TOCSY H_α-C_α assignment for GLN, ALA and ASN . . . 294
A.13 TOCSY H_α-C_α assignment for ASP, LEU and MET . . . 295
A.14 TOCSY H_α-C_α assignment for HIS, TRP and PHE . . . 296

A.15 TOCSY H_α-C_α assignment for TYR, CYS and SER 297
A.16 TOCSY H_α-C_α assignment for PRO, ILE and VAL 298
A.17 TOCSY H_α-C_α assignment for GLY and THR 299
A.18 The assignments of the nucleotide proton chemical shift resonances used in this paper to measure the partitioning of dA, dG, cAMP, cGMP, ATP, GTP. 300
A.19 Correlation between metabolite partitioning and metabolite properties using the naïve fitting procedure 305
A.20 Correlation between metabolite partitioning and metabolite properties using the grouped fitting procedure (minimising χ^2) 306
A.21 Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure (1 of 3) 307
A.22 Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure (2 of 3) 308
A.23 Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure (3 of 3) 309
A.24 Histogram of bootstrap fitted coefficients for the the winner of the grouped properties fitting procedure (1 of 2) 310
A.25 Histogram of bootstrap fitted coefficients for the the winner of the grouped properties fitting procedure (2 of 2) 311

B.1 ^{15}N CPMG relaxation dispersion pulse sequence 325
B.2 ^{15}N CEST pulse sequence 326
B.3 pH dependence of Δ54-pVI NMR spectra 328

C.1 Pulse sequences used for ^{15}N-^{1}H spectra simulations 359
C.2 Pulse sequences used for $^{13}CH_3$ spectral simulations 361
C.3 HSQC and HMQC pulse sequences used to assess the best 2D NMR experiments for $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 367

C.4 Spline baselining used on 1D carbon experiments on the $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 368

C.5 Comparison of the global fitting of single quantum relaxation rates for $^{13}CH_3$ labelled systems with and without external protons 369

C.6 Histograms of fitting parameters for the global $^{13}CF_3$ relaxation model bootstrapping procedure 370

C.7 Histograms of fitting parameters for the global $^{13}CH_3$ relaxation model bootstrapping procedure 371

List of Equations

2.1 $2B_1\cos(\omega_{\mathrm{rf}}t)\hat{y} = B_1[\cos(\omega_{\mathrm{rf}}t)\hat{y} + \sin(\omega_{\mathrm{rf}}t)\hat{x}] + B_1[\cos(-\omega_{\mathrm{rf}}t)\hat{y} + \sin(-\omega_{\mathrm{rf}}t)\hat{x}]$ . . . 8

2.2 $B_1[\cos(\omega_{\mathrm{rf}}t)\hat{y} + \sin(\omega_{\mathrm{rf}}t)\hat{x}] \rightarrow B_1$ . . . 9

2.3 $$\frac{d}{dt}\begin{pmatrix} M_{x,i} \\ M_{y,i} \\ M_{z,i} \\ E \end{pmatrix} = \begin{pmatrix} -R_2 & \omega_i & -\omega_y & 0 \\ -\omega_i & -R_2 & \omega_x & 0 \\ \omega_y & \omega_x & -R_1 & R_1 M_{eq} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} M_{x,i} \\ M_{y,i} \\ M_{z,i} \\ E \end{pmatrix}$$ . . . 10

2.4 $E = -\frac{\mu_0\gamma_a\gamma_b\hbar^2}{4\pi|r_{ab}|^3}\left[3\frac{(\vec{S}_a\cdot\vec{r}_{ab})(\vec{S}_b\cdot\vec{r}_{ab})}{|r_{ab}^2|} - \vec{S}_a\vec{S}_b\right]$ . . . 13

2.5 $J(\omega,\tau_c) = \frac{\tau_c}{1+\omega^2\tau_c^2}$ . . . 17

2.6 $\mathrm{R}_2 = \gamma^2\left\langle \mathrm{B}_{\mathrm{ind}}^2\right\rangle [J(0,\tau_c) + J(\omega_0,\tau_c)]$ . . . 17

2.7 $\mathrm{R}_1 = 2\gamma^2\left\langle \mathrm{B}_{\mathrm{ind}}^2\right\rangle J(\omega_0,\tau_c)$ . . . 18

2.8 $\frac{d}{dt}\left|\psi(t)\right\rangle = -iH\left|\psi(t)\right\rangle$ . . . 19

2.9 $\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}\hat{H}\rho(t) \quad \hat{H} = [H,\ ]$ . . . 20

2.10 $\hat{I}_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \hat{I}_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \hat{I}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \hat{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ 20

2.11 $\hat{I}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \hat{I}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \hat{I}_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \hat{I}_\beta = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ . . . 20

2.14 $\hat{H}_{rot} = U\hat{H}U^-1 - i\hbar U\frac{dU^{-1}}{dt}$ . . . 21

2.18 $\rho(t) = e^{-i\omega_1 t\hat{I}_y}\rho(0)e^{i\omega_1 t\hat{I}_y} = \cos(\omega_1 t)\rho(0) - i[\hat{I}_y,\rho(0)]\sin(\omega_1 t)$ . . . 22

2.22 $\rho(t) = \cos(\omega t)\hat{I}_x - i[\hat{I}_z,\hat{I}_y]\sin(\omega t) = \cos(\omega t)\hat{I}_x + \hat{I}_y\sin(\omega t)$ . . . 23

2.23 $\omega_0 = \gamma B_{\mathrm{tot}} = \gamma(B_0 + zG)$ . . . 23

2.24 $\hat{H}\hbar = \omega_{0,I}\hat{I}_z + \omega_{0,S}\hat{S}_z + 2\pi J(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y + \hat{I}_z\hat{S}_z)$ . . . 26

2.26 $\hat{H}_{rot}\hbar = \omega_I\hat{I}_z + \omega_S\hat{S}_z + 2\pi J\hat{I}_z\hat{S}_z$ . . . 27

2.29 $\hat{H}/\hbar = \pi J(\hat{H}_x\hat{C}_x + \hat{H}_z\hat{C}_z)$ . . . 31

2.35 $S(t_1, t_2) = e^{(i\omega_H - R_{2,H})t_2} e^{(i\omega_C - R_{2,C})t_1}$ . . . 33

2.37 $I = I_0 e^{-\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D}$ . . . 37

2.38 Hard pulse of duration a $= B_1 \cdot \mathrm{rect}(t/a)$ . . . 39

2.39 $\int_{-\infty}^{\infty} B_1 \cdot \mathrm{rect}(t/a) e^{-i\omega t} dt = B_1 \frac{\sin(\pi a \omega)}{\pi a \omega}$ . . . 39

2.40 $\frac{d}{dt}\begin{pmatrix} I_z \\ S_z \end{pmatrix} = \begin{pmatrix} -R_{1,I} & -\sigma_{IS} \\ -\sigma_{IS} & -R_{1,S} \end{pmatrix} \begin{pmatrix} I_z - I_{eq} \\ S_z - S_{eq} \end{pmatrix}$ . . . 43

2.41 $\sigma_{IS} = \frac{d^2}{10}\left(6J(\omega_I + \omega_S, \tau_c) - J(\omega_I - \omega_S, \tau_c)\right) \quad d = -\frac{\mu_0 \gamma_I \gamma_S \hbar}{4\pi |r_{IS}|^3}$ . . . 43

3.1 $\Delta \mathrm{G_{mix}} = \chi\phi(1-\phi) + \mathrm{RT}\left(\frac{\phi}{\mathrm{r}}\ln(\phi) + (1-\phi)\ln(1-\phi)\right)$ . . . 53

3.2 $PC = \frac{V_i\left(V_p + V_{LOS} + V_{add,p}\right)}{SV_p\left(V_i + V_{add,s}\right)} - \frac{V_{LOS}}{V_p}$ . . . 73

3.3 $\Delta \mathrm{G}_j = \Delta \mathrm{G}_0 + \sum_i \beta_i \frac{x_{i,j} - \overline{x_i}}{\sigma_i}$ . . . 82

3.4 $SSR = \sum_j (\Delta \mathrm{G}_{j,\mathrm{measured}} - \Delta \mathrm{G}_{j,\mathrm{predicted}})$ . . . 85

3.5 Total RMSD $= \sqrt{(\text{Total SSR})/N}$ . . . 86

3.8 $\mathrm{D} = \frac{\mathrm{k_B T}}{6\pi\eta(\mathrm{T})r}$ . . . 122

3.9 $R_1 = \frac{3}{10}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2}{r_{HH}^6}\left(J(\omega_H) + 4J(2\omega_H)\right)$ . . . 124

3.10 $R_2 = R_2^* + R_{2,ex}$
$R_2^* = \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2}{r_{HH}^6}\left(3J(0) + 5J(\omega_H) + 2J(2\omega_H)\right)$ . . . 127

4.1 $\mathrm{R_{h,globular}(nm)} = 0.475\mathrm{N}^{0.29} \qquad \mathrm{R_{h,random\ coil}(nm)} = 0.221\mathrm{N}^{0.57}$ . . . 142

4.2 $S/N = A\frac{(1 - e^{-\mathrm{R}_1 d_1})}{\sqrt{d_0 + d_1}}$ . . . 147

4.5 $\mathrm{R}_{2,eff}(\nu_{\mathrm{CPMG}}) = -\frac{1}{\mathrm{T_{relax}}} \ln\left(\frac{I(\nu_{\mathrm{CPMG}})}{I_0}\right)$ . . . 170

5.1 $\frac{d\rho}{dt} = -i\hat{\hat{H}}_0\rho + \hat{\hat{R}}(\rho - \rho_{eq})$ . . . 180

5.2 $d_{IS} = -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi r_{IS}^3} \qquad c_I = \gamma_I B_0 \Delta\delta_I/3$ . . . 181

5.3 $H_{a,b} = \langle\rho_a|\,\hat{H}\,|\rho_b\rangle = -i\frac{\mathrm{Tr}\left(\rho_a[\hat{H}, \rho_b]\right)}{\mathrm{Tr}(\rho_a\rho_b)}$ . . . 182

5.4 $\frac{d}{dt}\begin{pmatrix} c_{N_+H_\alpha} \\ c_{N_+H_\beta} \end{pmatrix} = \begin{pmatrix} i(\omega_N + \pi J) - R_{N_+H_\alpha} & -R_{N_+H_\alpha, N_+H_\beta} \\ -R_{N_+H_\alpha, N_+H_\beta} & i(\omega_N - \pi J) - R_{N_+H_\beta} \end{pmatrix} \begin{pmatrix} c_{N_+H_\alpha} \\ c_{N_+H_\beta} \end{pmatrix}$ 184

5.5 $\frac{d}{dt}\begin{pmatrix} c_{H_+N_\alpha} \\ c_{H_+N_\beta} \end{pmatrix} = \begin{pmatrix} i(\omega_H + \pi J) - R_{H_+N_\alpha} & -R_{H_+N_\alpha, H_+N_\beta} \\ -R_{H_+N_\alpha, H_+N_\beta} & i(\omega_H - \pi J) - R_{H_+N_\beta} \end{pmatrix} \begin{pmatrix} c_{H_+N_\alpha} \\ c_{H_+N_\beta} \end{pmatrix}$ 184

List of Acronyms and Abbreviations

μL microlitres.

μM micromolar.

μg micrograms.

μs microseconds.

^{o}C degrees Celsius.

Å Angstroms.

ts1 Temperature-Sensitive Human Adenovirus Mutant 1.

1D One-Dimensional.

2D Two-Dimensional.

3D Three-Dimensional.

AD Alzheimer's Disease.

aDMA asymmetric dimethyl arginine.

ADP Adenosine diphosphate.

AI Artificial Intelligence.

ALS Amyotrophic Lateral Sclerosis.

ASD Autism Spectrum Disorder.

ATP Adenosine triphosphate.

ATPS aqueous two-phase system.

AVP Adenovirus Protease.

BEST Band-selective Excitation Short-Transient.

BMRB Biological Magnetic Resonance Bank.

BSA Bovine Serum Albumin.

cAMP cyclic adenosine monophosphate.

CAR Coxsackievirus and Adenovirus Receptor.

CD Circular Dichroism.

CEST Chemical Exchange Saturation Transfer.

cGMP cyclic guanosine monophosphate.

ChAd Chimpanzee Adenovirus.

CheSPI Chemical Shift Secondary Structure Population Inference.

cm centimeters.

CMC Critical Micelle Concentration.

CP Continuous Polarisation.

CPMG Carr-Purcell-Meiboom-Gill.

CRINEPT Cross Relaxation- Enhanced Polarization Transfer.

cryo-EM cryogenic Electron Microscopy.

CSA Chemical Shift Anisotropy.

CSP Chemical Shift Perturbation.

CW Continuous Wave.

dA deoxyadenosine.

DAC Digital-to-Analog Converter.

ddHMQC Delayed Decoupling Heteronuclear Multiple Quantum Coherence.

DDM n-dodecyl-β-D-maltoside.

DDMB n-dodecyl-β-melibioside.

DDX4 DEAD-box helicase 4.

DEST Dark-state Exchange Saturation Transfer.

DEX Dextran.

dG deoxyguanosine.

DIC Differential Interference Contrast.

DIPSI Decoupling In the Presence of Scalar Interactions.

DMSO Dimethyl-Sulfoxide.

DNA Deoxyribonucleic acid.

DPFGSE Double Pulse Field Gradient Stimulated Echo.

dsDNA Double-Stranded Deoxyribonucleic Acid.

DSRCT Desmoplastic Small-Round-Cell Tumour.

DSS sodium trimethylsilyl propane sulfonate.

DSSP Dictionary of Protein Secondary Structure.

E-state Electrotopological-State.

EDTA EthyleneDiamineTetraacetic Acid.

EMBOSS European Molecular Biology Open Software Suite.

ESMFold Evolutionary Scale Modelling Fold.

EWSR1 Ewing Sarcoma breakpoint Region 1.

FET FUS, EWSR1, TAF15.

FID Free Induction Decay.

FLVI1 Friend Leukaemia Virus Integration gene 1.

FRET Förster Resonance Energy Transfer.

FT Fourier Transform.

FuDA Function and Data Analysis.

FUS Fused in Sarcoma.

FWHM Full Width Half Maximum.

g grams.

GARP Globally optimized Alternating-phase Rectangular Pulses.

GHz Gigahertz.

GRAPE Gradient Ascent Pulse Engineering.

GRE Gradient Echo.

GTP Adenosine triphosphate.

GUI Graphical User Interface.

HAdV Human Adenovirus.

HEHAHA Heteronuclear Hartmann-Hahn.

HeLa Henrietta Lacks.

HMDB Human Metabolome DataBase.

HMQC Heteronuclear Multiple Quantum Coherence.

HSP Heat Shock Protein.

HSQC Heteronuclear Single Quantum Coherence.

Hz Hertz.

ICP-MS Inductively Coupled Plasma Mass Spectrometry.

IDP Intrinsically Disordered Protein.

IDR Intrinsically Disordered Region.

INDIANA In-Cell Diffusion Analysis.

INEPT Insensitive Nuclei Enhanced by Polarization Transfer.

IUPAC International Union of Pure and Applied Chemistry.

J Joules.

K Kelvin.

$\mathbf{K}_d$ Equilibrium Dissociation Constant.

kDa Kilodaltons.

kg kilograms.

kHz Kilahertz.

kJ Kilojoules.

L litres.

LFER Linear Free Energy Relationship.

LOS Leftover Supernatant.

LSVM Length Scale-Dependent Viscosity Model.

M molar.

MAGMA Methyl Assignment by Graph Matching.

MCES Maximum Common Edge Subgraph.

MDa Megadaltons.

mg milligrams.

MHz Megahertz.

mL millilitres.

mM millimolar.

mm millimeters.

MQ Multiple Quantum.

MRI Magnetic Resonance Imaging.

mRNA Ribonucleic acid.

ms milliseconds.

MSA Multiple Sequence Alignment.

Mvh Mouse Vasa Homologue.

MW Molecular Weight.

MWCO Molecular Weight Cut Off.

NaPi sodium phosphate.

Nedd4 Neuronally Expressed Developmentally Downregulated 4.

NES Nuclear Export Signal.

NLS Nuclear Localisation Signal.

nm nanometers.

NMR Nuclear Magnetic Resonance.

NOE Nuclear Overhauser Effect.

NOESY Nuclear Overhauser Effect Spectroscopy.

ns nanoseconds.

NUS Non-Uniform Sampling.

ODiNPred Prediction of protein Order and Disorder by evaluation on NMR data.

Pa Pascals.

PAE Predicted Aligned Error.

PAR Poly(adenosine diphosphate ribose).

PARP Poly(adenosine diphosphate ribose) polymerase.

PC Partitioning Coefficient.

PCA Principal Component Analysis.

PCV Packed Cell Volume.

PDB Protein Data Bank.

PEG Polyethylene Glycol.

PLD Prion-Like Domain.

pLDDT Predicted Local Distance Difference Test.

PLS Partial Least Squares Regression.

POTENCI Prediction of Temperature, Neighbour and pH Corrected Shifts for Intrinsically Disordered Proteins.

ppm parts per million.

ps picoseconds.

pVI precursor protein VI.

rad radians.

RelCalc Relaxation Calculator.

RMSD Root Mean Squared Deviation.

RNA Ribonucleic acid.

RRM RNA recognising motif.

s seconds.

SARS-Cov-2 Severe acute respiratory syndrome coronavirus 2.

SBM Solomon-Bloembergen-Morgan.

SEC-MALS Size-Exclusion Chromatography with Multi-Angle Light Scattering.

SMA Simple Moving Average.

SMILE Sparse Multidimensional Iterative Lineshape-Enhanced.

SOFAST band-Selective Optimized Flip-Angle Short-Transient.

SQ Single Quantum.

SSP Secondary Structure Propensity.

SSR Sum of Squared Residuals.

SW Sweep Width.

T Tesla.

TAF15 TATA-box binding protein associated factor 15.

TALOS Torsion Angle Likeliness Obtained from Shift and Sequence Similarity.

tau tubulin-associated unit.

TCEP tris(2-carboxyethyl)phosphine hydrochloride.

TDP-43 TAR DNA-binding protein 43.

TFA TriFluoroacetic Acid.

ThT Thioflavin T.

TOCSY Total Correlation Spectroscopy.

TPPI Time-Proportional Phase Incrementation.

TROSY Transverse Relaxation Optimised Spectroscopy.

UAA Unnatural Amino Acid.

UHPLC Ultra-High-Performance Liquid Chromatography.

UniDecNMR Universal Deconvolution Nuclear Mangnetic Resonance.

uSTA universal Saturation Transfer Analysis.

VIF Variance Inflation Factor.

WATERGATE Water Suppression Through Gradient-Tailored Excitation.

WBR Wangsness-Bloch-Redfield.

WSL Windows-Subsystem for Linux.

WT1 Wilms Tumor gene.

1 Introduction

1.1 Using solution-state Nuclear Magnetic Resonance (NMR) to probe biology

Solution-state nuclear magnetic resonance (NMR) is a non-invasive method to study the structure and dynamics of biomolecules by probing the local electronic environment around different nuclei.[1] It is complementary to techniques such as X-ray crystallography and cryo-electron microscopy (cryo-EM) which, along with NMR, have been revolutionary in structural biology.[2] NMR is able to measure and quantify the dynamics of proteins,[1,3] along with cryo-EM,[4] and more recently artificial intelligence (AI).[5–7]

Biological NMR experiments can be performed on a range of biologically relevant nuclei (e.g. ^{1}H, ^{13}C,^{15}N, ^{31}P, ^{23}Na), where naturally low-abundance ^{13}C and ^{15}N nuclei can be introduced to a biological sample through isotopic enrichment. Furthermore, labelling with non-natural nuclei such as ^{19}F increases the versatility of biological NMR.[8] The first published NMR protein structure was reported in 1985 using only ^{1}H chemical environment information and ^{1}H internuclear distance restraints.[9] This was shortly followed by the pioneering work by Bax *et al.*[10–12] to develop multinuclear protein NMR experiments to obtain ^{1}H, ^{13}C and ^{15}N atomic environments, improving spectral resolution in large proteins and reporting on secondary structure.[13] Combining the multinuclear assignments with ^{1}H internuclear distance restraints allows 3D structure generation.[14–16] Furthermore, NMR can be used to quantify atom-specific rotational,[17,18] translational[19,20] and chemical

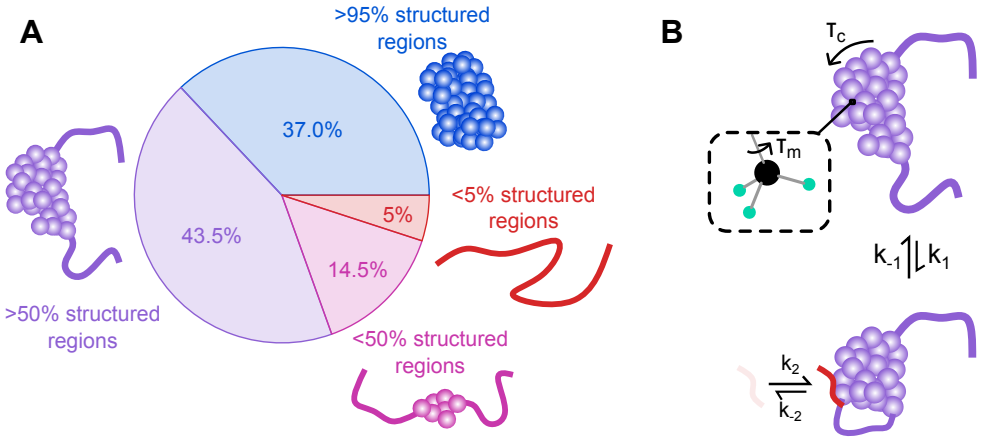


Figure 1.1: An overview of protein structure and disorder along with protein dynamics and interactions on multiple timescales. **A** Percentage of proteins in the human proteome that are expected to be >95% structured, >50% structured (and <95% structured), <50% structured (and >5% structured), and <5% structured. **B** A schematic showing a selection of dynamic processes proteins can undergo: local molecular tumbling with a timescale τ_m, global protein tumbling with a timescale τ_c, protein conformational exchanges between a ground and excited state, and ligand binding to an excited state. To fully understand a protein system it is important to study all of these processes and their associated timescales.

exchange dynamics.[3,21] Therefore, an invaluable atomic resolution structure and dynamic map of the system can be determined. NMR can even be used to study to structural and dynamic changes during protein folding itself.[22,23]

Fully structured proteins make up only 37% of the human proteome, with 58% consisting of both folded domains and intrinsically disordered regions (IDRs), and 5% of the human proteome being intrinsically disordered proteins (IDPs) containing no significant secondary structure (Figure 1.1A).[24] Furthermore, proteins experience dynamic motions across multiple timescales, with global motion, internal motion and conformational changes being key examples (Figure 1.1B).[25,26] Solution-state NMR is uniquely able to quantify all of these processes, as well as the structural information.[1,3,27–32]

In addition, binding of ligands to disordered regions of proteins does not follow the classic lock and key[33] or induced fit[34] binding modes used extensively in folded proteins; instead, the binding is highly non-specific and entropic.[35] The use of solution-state NMR is imperative to quantify the weak multivalent interactions of

ligands to such disordered systems.[36]

NMR can provide structural and dynamic information on the transient excited states of a protein, allowing the development of therapeutics to target this 'invisible' state (Figure 1.1B).[37] However, cryo-EM, and more recently artificial intelligence (AI), are also able to measure/predict protein conformational ensembles. Lower resolution densities in cryo-EM maps are usually interpreted as regions of a protein with significant dynamics. Furthermore, single particle analysis of cryo-EM data enables the structures of different protein conformations in a sample to be determined.[4] Recently, the role of artificial intelligence (AI) in protein biology has exponentially increased following the release of AlphaFold2.[5] Being able to accurately predict the structure of a protein based on its sequence is highly revolutionary. In addition, protein dynamics can be inferred from low AlphaFold pLDDT (predicted local distance difference test) and the predicted aligned error (PAE) maps.[7] AlphaFold2 predicted structures depend highly on multiple sequence alignments (MSAs). By changing the depth of the input multiple sequence alignments to AlphaFold2, plausible alternative conformations of a protein can be predicted and assessed.[6] The databases available on disordered proteins have increased over the last five years.[38] Continuing to advance the knowledge and databases associated with disordered protein regions could further advance the AI revolution for protein dynamics in biology, enabling dynamics to be predicted more accurately and unexplored biological functions to be discovered.[39] Therefore, the combination of NMR, cryo-EM and artificial intelligence methods are instrumental in studying protein dynamics.

NMR has widespread applications in biology, beyond the study of proteins. In fact, it is commonly used to study the structure and dynamics of nucleic acids.[40] In addition, NMR has strong applications in biology outside of studying biological macromolecules due to its highly quantitative nature, with signal intensity proportional to nuclear spin concentration. One example is Magnetic

Resonance Imaging (MRI), commonly used to study proton concentration as a function of position in hospital patients non-invasively.[41] NMR is also widely used in metabolomics to quantify the proportion and concentrations of metabolites in biological samples.[42]

1.2 Thesis outline

In section 1.1, the benefits of using solution-state NMR to probe biological systems were discussed.

Solution-state NMR is used in this thesis to probe biomolecular condensates and characterise a disordered virus capsid protein. Additionally an evaluation of methods to combat molecular size limitations in NMR is provided. An introduction to the NMR concepts and theory used throughout the rest of this thesis is included in chapter 2.

Chapter 3 considers the application of NMR to biomolecular condensates. Biomolecular condensates perform cellular compartmentalisation in the absence of a phospholipid membrane[43] and evidence is emerging that they create an alternative solvent environment to the cytosol.[44] NMR is used to quantify the partitioning of metabolites into condensates and probe the physicochemical properties of water inside condensates. The results add to the field of evidence that condensates behave as distinct solvents.[44–49]

In chapter 4, NMR is used to investigate the residual structure within a construct of the human adenovirus virus capsid protein VI (Δ54-pVI). Protein VI has a multifaceted role in an adenovirus life cycle and is linked to the disruption of endosomal membranes, enabling the virus to enter the host cell cytoplasm.[50–52] The secondary structure propensity within Δ54-pVI is probed using a semi-automated backbone assignment of its chemical shifts. The prevalence of amphipathic α-helical propensity is assessed which could form the basis of interaction with endosomal membranes.

Chapter 5 is concerned with the resolution and intensity limits of NMR in slowly tumbling biomolecules, caused by fast transverse relaxation.[53] A symbolic NMR relaxation rate calculator, RelCalc,[54] is introduced along with a pulse sequence simulator, SimSeq,[55] able to simulate the effects of relaxation in an NMR experiment. RelCalc and SimSeq are used to evaluate the benefits and limitations of the current transverse relaxation optimised spectroscopy (TROSY) experiments.[53,56–60] Following this, the benefits and drawbacks of the relaxation optimised $^{13}CF_3$ spin system are experimentally assessed in a 395 kDa protein complex.

Finally, SpinExplorer (a Python-based graphical user interface) is introduced in chapter 6, able to perform NMR data processing and analysis. A significant proportion of the spectra within this thesis were processed and analysed using this package.

Overall, this thesis explores the application of NMR to biology through the study of biomolecular condensates and a disordered adenovirus capsid protein. The limitations of NMR on slowly tumbling species (such as molecules in biomolecular condensates or full virus capsids) are discussed, followed by an evaluation of the relaxation rates in the promising relaxation optimised $^{13}CF_3$ spin system.

2 Introduction to NMR

2.1 NMR theory and relaxation

2.1.1 The Zeeman interaction and chemical shift

In the absence of a magnetic field, the nuclear spin states of a spin-1/2 particle ($\alpha = \left|\frac{1}{2}\right\rangle$ and $\beta = \left|-\frac{1}{2}\right\rangle$) are degenerate. Upon the addition of a magnetic field, a spin-1/2 particle will split into two energy levels separated by an energy $\Delta E = \hbar\omega_0 = \hbar\gamma B_0$ where γ is the gyromagnetic ratio of the particle, B_0 is the magnitude of the applied magnetic field, ω_0 is termed the Larmor frequency, and $\hbar$ is Planck's constant (Figure 2.1A).[61] This splitting is termed a Zeeman interaction[†], and causes a population difference between the spin states given by the Boltzmann distribution ($N_\beta/N_\alpha = \exp(-\Delta E/(k_B T)$ for $\gamma > 0$ nuclei), where k_B is the Boltzmann constant and T is the temperature in Kelvin. Therefore, a net magnetisation along the direction of the applied magnetic field is produced with a magnitude proportional to the applied field.

Initial experiments to observe the Zeeman interaction in nuclei were performed in the gas phase. In 1933, Otto Stern *et al.*[63,64] performed measurements to determine the magnetic moment of a proton by measuring the deflection of hydrogen molecules in a magnetic field. Isidor Rabi followed this in 1938 and performed the first nuclear magnetic resonance experiment, able to perturb a beam of lithium chloride in a magnetic field by the addition of a perpendicular constant radio-frequency field.[65]

[†]In 1897, Pieter Zeeman published his findings that certain spectral lines of different elements were split into more spectral lines upon the addition of a magnetic field.[62]

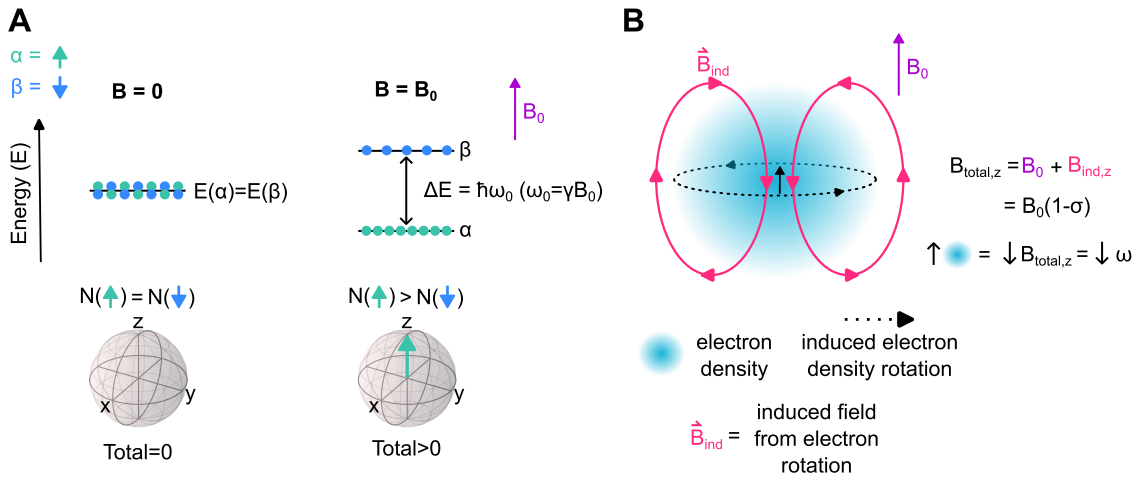


Figure 2.1: The Zeeman interaction and chemical shift. **A** In the absence of a magnetic field, the nuclear spin states of a spin-1/2 nucleus ($\alpha = \left|\frac{1}{2}\right\rangle$, $\beta = \left|-\frac{1}{2}\right\rangle$) are degenerate. However, under the presence of a magnetic field B_0 aligned with the z-axis, the energy gap between the spin states increases to $\hbar\omega_0$, where $\omega_0 = \gamma B_0$. The energy gap creates a net polarisation of the spin states, generating a magnetisation vector along the z-direction. **B** In the presence of an electron density, the external field B_0 induces a rotation of the electron density, which in turn generates an induced field $\vec{B}_{ind}$. Therefore, a nuclear spin located in the centre of the electron density experiences a reduced total magnetic field along the z-dimension ($B_{total,z} = B_0 + B_{ind,z}$).

The frequency of the additional field was kept constant and the amplitude adjusted until interconversion between the α and β states occurred. Similar experiments were not performed on the condensed phases until 1946 when Felix Bloch and Edward Purcell simultaneously demonstrated NMR in water and paraffin, respectively.[66,67]

The alteration of NMR frequency (ω) with electron density in molecules was first discovered in 1950 by measurements on ^{14}N and ^{19}F nuclei whose NMR frequency depended strongly on the chemical compound they were located in.[68,69] This phenomenon is now known as shielding and chemical shift.[70–72] According to Lenz law,[73] electron density placed into a magnetic field (B_0) will rotate to induce a magnetic field (B_{ind}) that opposes the applied field (Figure 2.1B). Therefore, a nuclear spin surrounded by more electron density (shielded) will have a reduced NMR frequency.[61]

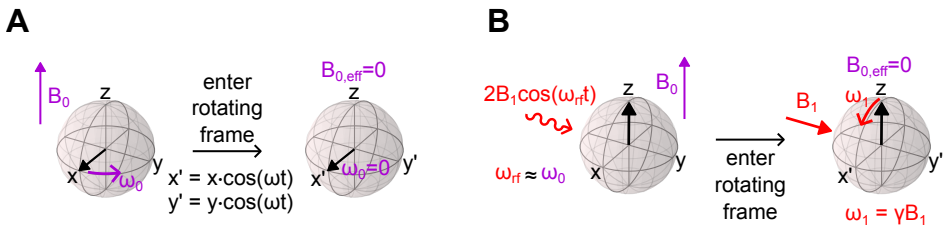


Figure 2.2: An introduction to the rotating frame and pulses. **A** Magnetisation placed in the transverse (xy) plane in an external field B_0 will oscillate at the frequency ω_0 in the laboratory frame. If the x and y axes of the system are rotated at a frequency ω_0, the magnetisation vector will remain stationary in the new frame of reference, called the rotating frame. **B** During an NMR pulse, a linearly polarised radiofrequency field (rf) is applied along the y-axis with amplitude $2B_1$ and frequency ω_{rf} in the laboratory frame. After entering a rotating frame where the x and y dimensions are rotated at a frequency ω_{rf}, the field becomes static in the rotating frame with an amplitude B_1. As $\omega_{rf} \approx \omega_0$, the B_0 field can be neglected. Therefore, the magnetisation vector will rotate about the y-axis with a frequency $\omega_1 = \gamma B_1$.

2.1.2 Fourier transform NMR

A magnetisation vector in the transverse (xy) plane will rotate at the Larmor frequency (ω_0)*. It is possible to enter a rotating frame of reference whereby the magnetisation vector appears stationary and $B_{0,eff}$=0 (Figure 2.2A).

If linearly polarised radio-frequency radiation of amplitude $2B_1$ and frequency ω_{rf} is applied along the y-axis of the system (where $\omega_{rf} \approx \omega_0$), this radiation can be split into two counter-rotating circularly polarised waves (equation 2.1).[75]

$$2B_1 \cos(\omega_{rf}t)\hat{y} = B_1[\cos(\omega_{rf}t)\hat{y} + \sin(\omega_{rf}t)\hat{x}] + B_1[\cos(-\omega_{rf}t)\hat{y} + \sin(-\omega_{rf}t)\hat{x}] \quad (2.1)$$

The second term in equation 2.1 has frequency -ω_{rf} and can be neglected leaving only the first term remaining. Transforming this term into a rotating frame (rotating

*Larmor frequency is named after Joseph Larmor. To explain the splitting of the spectral lines observed by Pieter Zeeman in 1897 upon the addition of a magnetic field,[62] Joseph Larmor hypothesised the splitting was caused by electrons rotating at a frequency proportional to the applied field.[74]

with frequency ω_{rf} about the z-axis) leads to the production of a static field, B_1 (equation 2.2):[75]

$$B_1[\cos(\omega_{rf}t)\hat{y} + \sin(\omega_{rf}t)\hat{x}] \rightarrow B_1 \tag{2.2}$$

As $\omega_{rf} \approx \omega_0$ in the rotating frame, the effective Zeeman field ($B_{0,eff}$) is significantly reduced. In the rotating frame, magnetisation rotates about the z-axis will occur at a frequency $\omega = \omega_0 - \omega_{rf}$. Therefore, under the predominant B_1 field, the equilibrium Zeeman magnetisation vector aligned along the z-axis will rotate in the xz-plane with a frequency ω_1, where $\omega_1 = \gamma B_1$ (Figure 2.2B).[75] If the pulse is applied for a duration $\pi/(2\omega_1)$, the bulk magnetisation vector will be rotated onto the x-axis (Figure 2.3A). If the pulse of radiation is applied for a short duration, a wide variety of radio-frequencies ω_{rf} are produced due to the quantum mechanical time-energy uncertainty relationship.[76,77] Therefore, it is possible to excite multiple nuclei of the same type within a molecule containing different frequencies in one pulsed NMR experiment.

The initial NMR experiments[65–67] involved sweeping over applied fields to determine NMR magnetic moments and are now known as continuous wave methods. However, as explained by the rotating frame concept above, pulses of radio-frequency radiation can alter the bulk magnetisation vector of multiple species simultaneously. In 1950, Erwin Hahn introduced the pulsed NMR experiment whereby a radio-frequency pulse was applied to a sample in a magnetic field B_0.[78] After using a 90^o pulse, the magnetisation oscillates in the transverse plane (with frequency $\omega_0 - \omega_{rf}$)*. The trajectory of magnetisation from two spins in different chemical environments (a,b) following the pulse is shown in Figure 2.3A, with a total signal S(t). For a single species with rotating frame frequency ω_i, the equation of motion of the magnetisation vector $M = (M_{x,i}, M_{y,i}, M_{z,i}, E)$ under the Zeeman

*The rotating frame will be applied throughout the rest of the discussion in this chapter unless otherwise specified.

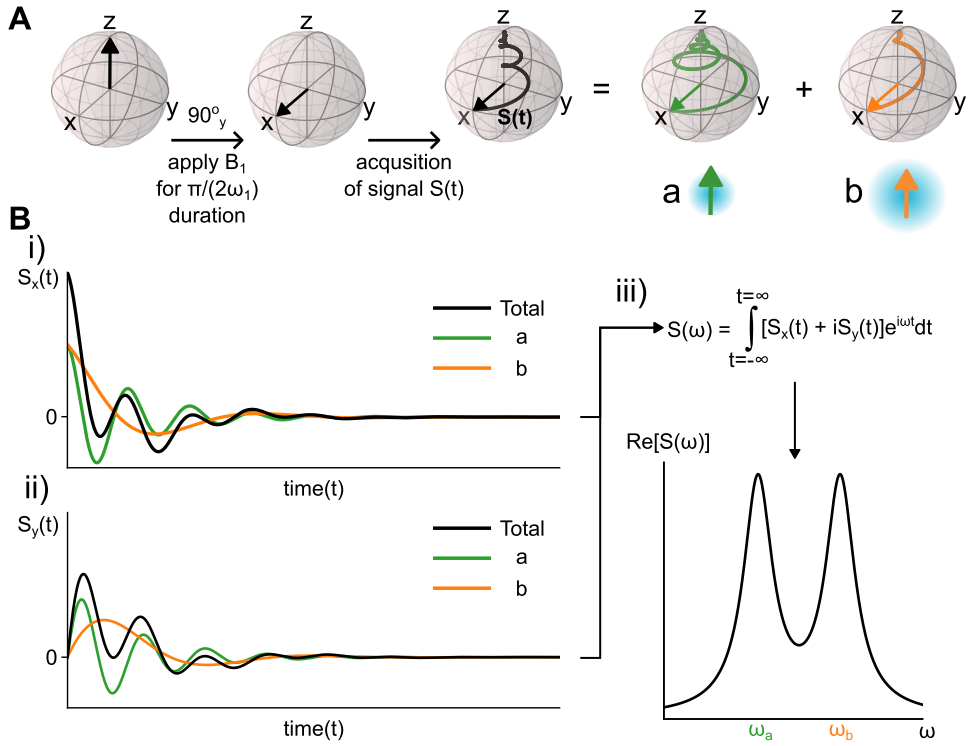


Figure 2.3: An introduction to Fourier Transform NMR. **A** After the application of a 90^o pulse about the y-axis (radio-frequency radiation applied along the y-axis with a duration $\pi/(2\omega_1)$), the magnetisation will be aligned along the x-axis in the rotating frame. The magnetisation vectors of the two species will then begin to oscillate about the z-axis in the rotating frame with a frequencies $\omega_{a/b}$=$\omega_{0,a/b} - \omega_{\mathrm{rf}}$. The combination of these trajectories gives the total signal S(t). **B** If the signal produced along the x/y-axes is recorded, the black traces in (i) and (ii) are observed, known as free induction decays (FIDs). The de-convolved traces from species a and b are also shown. (iii) Combining the time-domain signals along the x/y-axes and performing a Fourier transform enables the production of the NMR spectrum, with peaks corresponding to the species within the system ($\omega_{a/b}$).

interaction and pulses (with frequencies ω_x and ω_y) can be represented as follows:

$$\frac{d}{dt}\begin{pmatrix} M_{x,i} \\ M_{y,i} \\ M_{z,i} \\ E \end{pmatrix} = \begin{pmatrix} -R_2 & \omega_i & -\omega_y & 0 \\ -\omega_i & -R_2 & -\omega_x & 0 \\ \omega_y & \omega_x & -R_1 & R_1 M_{eq} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} M_{x,i} \\ M_{y,i} \\ M_{z,i} \\ E \end{pmatrix} \tag{2.3}$$

where E is the identity, R_1 is the relaxation rate of the system back to M_{eq} along

the positive z-axis, and R_2 is known as the transverse relaxation rate*. This concept was proposed by Bloch in 1946 and is now known as the Bloch equation.[79]

The signal S(t) can induce a current in the receiver coils in the x/y-axes. Plots of the total signals, as well as the de-convolved signals from each environment a/b, in the x-axis ($S_x(t)$) or y-axis ($S_y(t)$) are shown in Figure 2.3Bi/ii. These time-domain plots are known as free induction decays (FIDs).[78] The signals $S_x(t)$ and $S_y(t)$ can be treated as the real and complex parts of the signal S(t). From this, the application of a Fourier transform converts the data from the time domain to the frequency domain, leading to the production of an NMR spectrum, with peaks at the frequencies ω_a and ω_b (Figure 2.3Biii). The development of Fourier transform NMR spectroscopy was pioneered by Richard Ernst[80] in the 1960s and paved the way for the development of pulse sequences and multidimensional NMR (section 2.3).

2.1.3 Relaxation

In Figure 2.3, the observed signals of the resonances decay towards zero over time after the application of a pulse, described by the Bloch equation (equation 2.3) under the rate constants R_2 and R_1. These processes are known as transverse and longitudinal relaxation (Figure 2.4). The first comprehensive theory of relaxation in NMR was established in 1948, known as the Bloembergen-Purcell-Pound theory.[81] Solomon extended this work to produce the Solomon equations for relaxation in two-spin systems in 1955, linking the motion of spins to relaxation and introduced the spectral density function.[82] This was further expanded by Bloembergen and Morgan in 1961 to include the effects of electron spins on nuclei,[83] producing the Solomon-Bloembergen-Morgan (SBM) theory of relaxation. Redfield theory, also known as Bloch-Wangsness-Redfield (WBR) theory, was then created as a more generalised theory of relaxation.[84–86] This section introduces NMR transverse and longitudinal

*The origins of R_1 and R_2 are discussed in section 2.4.

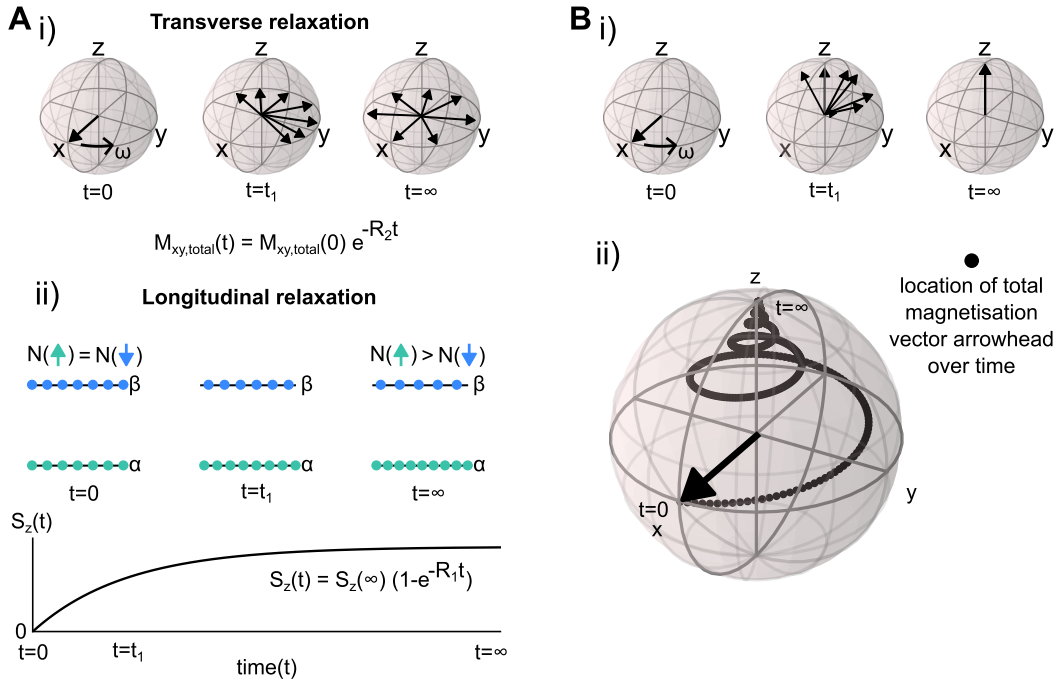


Figure 2.4: A vector model description of transverse and longitudinal relaxation. **A** (i) A vector model depiction of transverse relaxation where magnetisation in the xy-plane becomes de-coherent over time. The process has a rate constant R_2 (equation 2.3) (ii) A schematic showing the process of longitudinal relaxation, the process by which the magnetisation vector returns to thermal equilibrium aligned with the B_0 field after the application of a pulse. The process has a rate constant of R_1 (equation 2.3) **B** (i) An illustration showing how a set of magnetisation vectors evolve under the combined processes of longitudinal and transverse relaxation. (ii) A depiction of the location of the total combined magnetisation vector within a system over time after the application of a pulse, including the effects of transverse and longitudinal relaxation.

relaxation in a qualitative manner. However, the concept of R_1 and R_2 relaxation rates only holds for an isolated spin-1/2 nucleus. Chapter 5 includes a more detailed description of NMR relaxation for non-isolated and scalar coupled spin systems.

Relaxation is caused by the presence of time-dependent fluctuating induced fields ($\vec{B}_{ind}$) within a system. The following subsections introduce the concept of dipolar and chemical shift anisotropy fields.

2.1.3.i Dipolar interaction

Nuclear spins produce an induced field $\vec{B}_{ind}$ around themselves. A molecule containing two spins (a and b), separated by a distance r_{ab}, can interact through

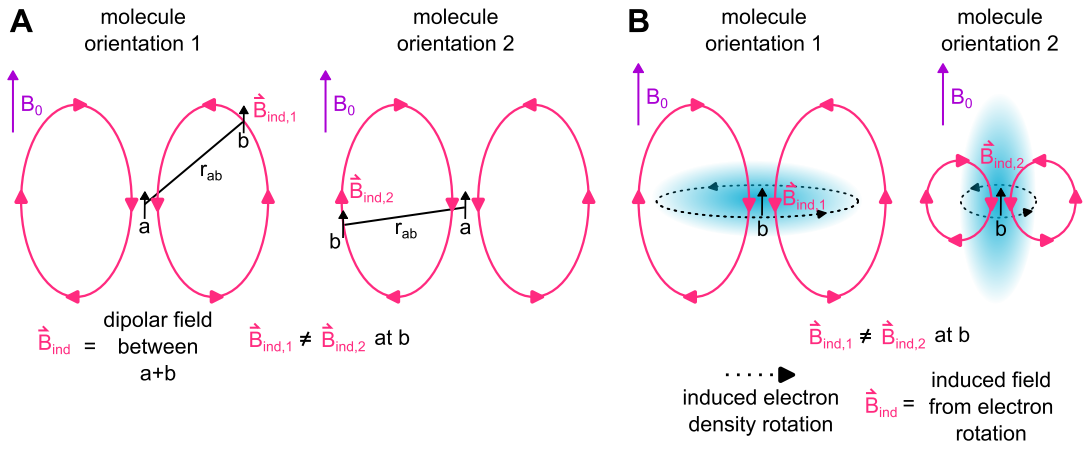


Figure 2.5: The orientational dependence of induced fields $\vec{B}_{ind}$ in a system. **A** Nuclear spins produce an induced field $\vec{B}_{ind}$ around themselves. A molecule containing two spins (a, b) is shown in two different orientations $(1, 2)$, with the induced field at spin b caused by spin a different in each orientation. The fields produced are called dipolar fields, and the interaction between the two spins caused by these fields is termed the dipolar interaction. The internuclear distance is given by r_{ab}. **B** When spin b is placed inside an asymmetric electron density in a field B_0, the induced field at spin b depends on the electron density's orientation with respect to the external field B_0. This behaviour is termed chemical shift anisotropy.

the interactions of these induced fields. The interaction is termed the dipolar interaction with the interaction energy (E) given by:[82]

$$E = -\frac{\mu_0 \gamma_a \gamma_b \hbar^2}{4\pi |r_{ab}|^3} \left[3\frac{(\vec{S}_a \cdot \vec{r}_{ab})(\vec{S}_b \cdot \vec{r}_{ab})}{|r_{ab}^2|} - \vec{S}_a \vec{S}_b \right] \tag{2.4}$$

where μ_0 is the magnetic permeability in a vacuum, $\vec{r}_{ab}$ is the internuclear vector and $\vec{S}_{a/b}$ are the nuclear spin vectors.

The induced field at spin b caused by spin a is orientation dependent (Figure 2.5A) and as will be seen in sections 2.1.3.iv and 2.1.3.iii is capable of causing transverse and longitudinal relaxation.

2.1.3.ii Chemical shift anisotropy (CSA)

As shown in Figure 2.1B, a nuclear spin placed inside electron density will be shielded, leading to the concept of chemical shift. If the electron density is spherical,

the degree of shielding is independent of orientation; however, this is not usually the case for most molecules. The orientational dependence of the induced current by the electron density leads to an orientational dependence in the induced fields $\vec{B}_{ind}$ (Figure 2.5B). This means that the chemical shift will be dependent on orientation, and the non-isotropic component of the chemical shift is termed chemical shift anisotropy (CSA).[87]

In the solution state, the differences in induced dipolar and CSA fields (Figure 2.5) rapidly average to zero and, therefore, do not affect the chemical shift position of a given spin. However, the orientation dependence of the induced fields can cause relaxation, as is described in the following sections (2.1.3.iv and 2.1.3.iii).

2.1.3.iii Longitudinal relaxation

The spontaneous emission rate from an excited state to a lower energy state is proportional to ΔE^3, where ΔE is the energy gap between the two energy levels.[88,89] Given that the energy gaps between states in NMR are small ($\Delta E \approx 4\times10^{-25}$ J for protons in a 14.1 T magnetic field), the rate of spontaneous emission between nuclear spin states in NMR is negligible (Figure 2.6A). Therefore, longitudinal relaxation in NMR relies on stimulated emission caused by fluctuations of induced fields. If the transverse (xy) induced field experienced by a spin is rotating at the Larmor frequency (ω_0) (Figure 2.6Bi), this generates an oscillating magnetic field with the correct frequency to induce stimulated emission between the nuclear spin states and return the system to thermal equilibrium (Figure 2.6Bii).

2.1.3.iv Transverse relaxation

The orientation dependence of induced fields within an NMR sample at an instantaneous time point is shown in Figure 2.7A. Different spins experience different z-components of the induced fields and, therefore, have different instantaneous Larmor frequencies at a given time. The rotational diffusion constant

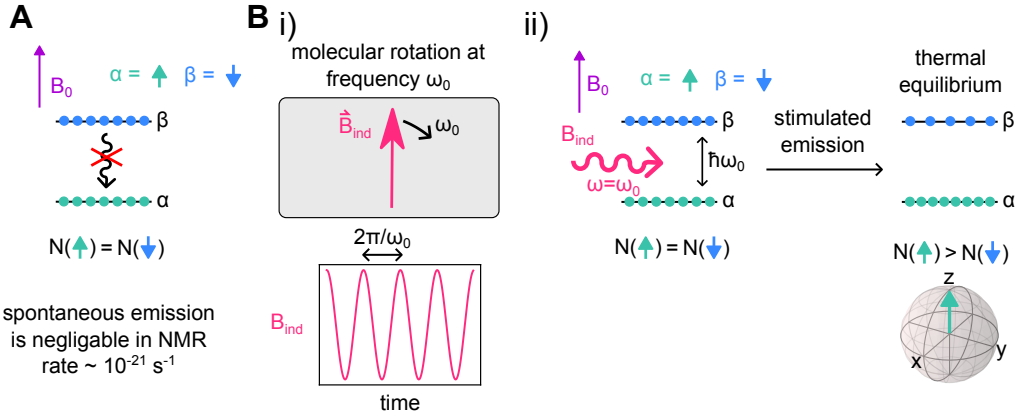


Figure 2.6: An introduction to longitudinal relaxation. **A** The rate of spontaneous emission between nuclear spin energy levels in an excited state system where N_α=N_β is negligible. **B** A spin tumbling with a frequency equal to the Larmor frequency (i) is capable of causing stimulated emission between the α and β spin states of a system (ii), thereby creating a mechanism for the system to return to thermal equilibrium. This is a simplified picture in which tumbling speed is constant. In true solutions molecular tumbling is stochastic. The following animation shows how molecular tumbling at the Larmor frequency ω_0 causes longitudinal relaxation: `https://youtu.be/gVI5icoL61k` (QR code in margin).

(D_{rot}) in a system governs how the orientation of an induced field experienced by a spin can change over time. The rotational diffusion constant can be converted to a rotational correlation time ($\tau_c = 1/(6D_{rot})$), which describes the time taken for a spin to tumble through one radian in a stochastic manner[90] (Figure 2.7B). Spins undertaking fast tumbling have a small τ_c, with spins tumbling slowly possessing a large τ_c.[61,87]

If the spins in the sample are experiencing slow tumbling (larger τ_c), then over a time t_1 the induced fields, and therefore, Larmor frequencies at each position remain roughly constant (Figure 2.7Ci). The difference in Larmor frequency depends on molecular orientation, leading to a large de-coherence in the magnetisation and fast transverse relaxation (larger R_2). On the other hand, if the spins in the sample are experiencing fast tumbling (smaller τ_c), then over a time t_1, the induced fields and Larmor frequencies are averaged, reducing the amount of de-coherence (slower transverse relaxation, smaller R_2) (Figure 2.7Cii). This contribution to transverse relaxation is called the secular contribution.[87]

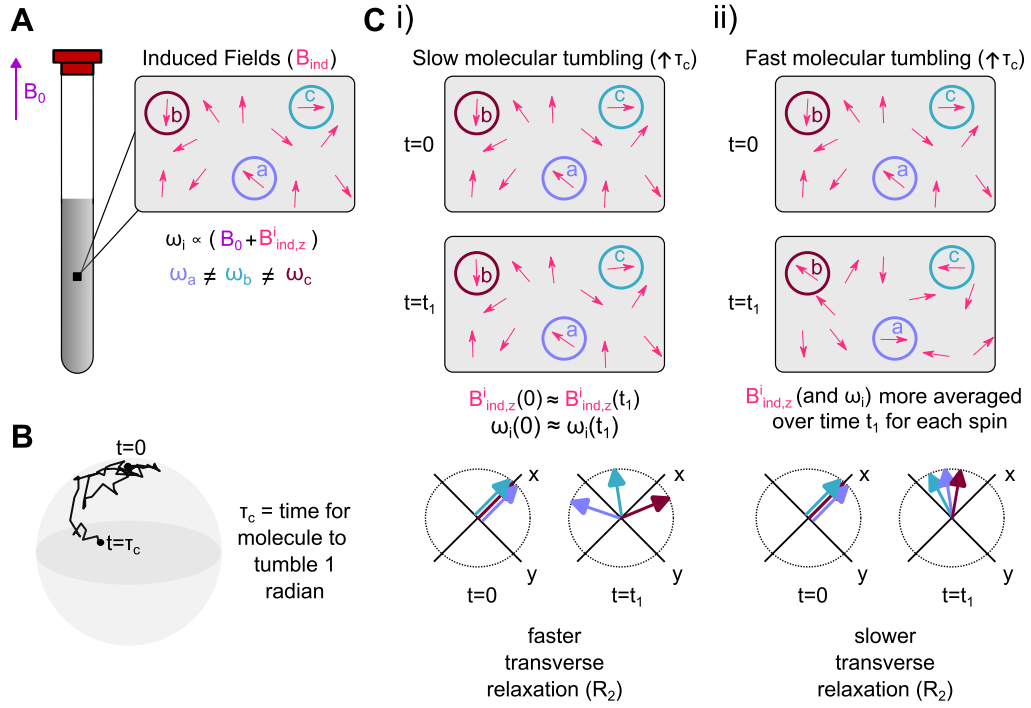


Figure 2.7: An introduction to transverse relaxation. **A** An instantaneous snapshot of the induced fields (B_{ind}) within an NMR sample. The Larmor frequency of a given spin within the system depends on the induced field. **B** An introduction to the concept of rotational correlation time (τ_c), the time a molecule takes to tumble through one radian. **C** Depiction of how slow molecular tumbling (i) and fast molecular tumbling (ii) affect transverse relaxation rates (R_2). The following animation shows how molecular tumbling affects transverse relaxation: `https://youtu.be/k1yvBJR7ihw` (QR code in margin).

An additional contribution to transverse relaxation arises from a similar mechanism as longitudinal relaxation (Figure 2.6). The total magnetisation vector must not exceed its original length and therefore an increase in the z-component of the magnetisation brought about by longitudinal relaxation must also be coupled with a decrease in the overall transverse component of the magnetisation vector (Figure 2.4Bii). This contribution can also be thought of as the lifetime broadening of the magnetisation caused by longitudinal relaxation and is called the non-secular contribution.[87]

2.1.4 The spectral density

Given the importance of molecular rotation on the time dependence of the oscillation of induced fields within an NMR sample, it is important to consider the spectral density function to understand longitudinal and transverse relaxation rates.[61,87]

The spectral density function $J(\omega, \tau_c)$ (equation 2.5)[82] can be thought of as the proportion of induced fields in the system oscillating at frequency ω, where τ_c (rotational correlation time) is the time taken for spins in the system to have rotationally tumbled by one radian.

$$J(\omega, \tau_c) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \tag{2.5}$$

Figure 2.8A shows $J(\omega, \tau_c)$ as a function of ω for systems with different τ_c values. For systems undertaking rapid tumbling (e.g. τ_c=0.01 ns), there are many spins in the system oscillating at all molecular frequencies up to 1 GHz. In contrast, in a system with a large τ_c (e.g. 10 ns), the number of induced fields oscillating at a frequency ω within a system decreases rapidly as ω is increased.

In section 2.1.3.iv, it was established that slow molecular tumbling enhances the transverse relaxation rate R_2. For a nucleus with gyromagnetic ratio γ in the presence of an induced local fluctuating field with a root-mean-squared magnitude of $\left\langle B_{ind}^2 \right\rangle^{\frac{1}{2}}$, the R_2 rate of the nucleus can be expressed as follows:[61,87]

$$R_2 = \gamma^2 \left\langle B_{ind}^2 \right\rangle [J(0, \tau_c) + J(\omega_0, \tau_c)] \tag{2.6}$$

where the terms dependent on $J(0,\tau_c)$ and $J(\omega_0,\tau_c)$ represent the secular and non-secular contributions to transverse relaxation, respectively. In contrast, the longitudinal relaxation rate of the nucleus can be calculated using the following expression:[61,87]

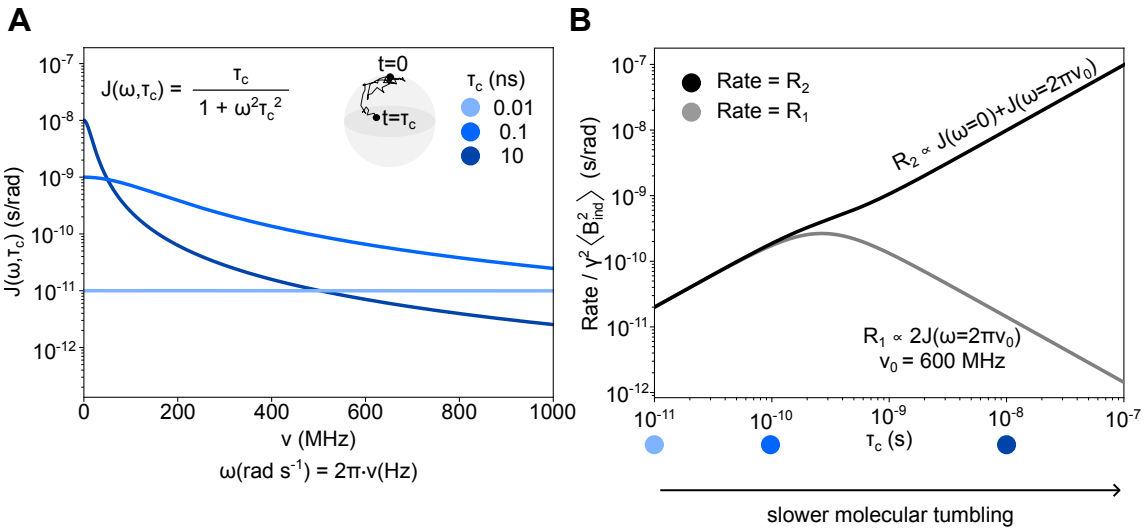


Figure 2.8: Introduction to the spectral density function. **A** The plot shows $J(\omega, \tau_c)$ as a function of ω for systems with different τ_c values. **B** Plots of transverse (R_2) and longitudinal (R_1) relaxation rates as a function of τ_c, where ν_0=600 MHz. Each relaxation rate is normalised by $\gamma^2 \langle B_{ind}^2 \rangle$ and so is in the same units as the spectral density function.

$$R_1 = 2\gamma^2 \langle B_{ind}^2 \rangle J(\omega_0, \tau_c) \tag{2.7}$$

Figure 2.8B shows how the R_1 and R_2 rates depend on rotational correlation time in a 600 MHz spectrometer ($\omega_0 = 2\pi\nu_0, \nu_0 = 600$ MHz). The transverse relaxation rate (R_2) of the nucleus is continuously increased as rotational motion is hindered and τ_c increases. The longitudinal relaxation rate (R_1) passes through a maximum when $\tau_c = 1/(2\pi\nu_0)$ before decreasing again at larger τ_c values. It is also worth noting that the transverse relaxation rate R_2 is always larger than the longitudinal relaxation rate R_1. The increase in R_2 with τ_c has significant consequences for NMR spectra obtained on slowly tumbling systems such as large macromolecular biological complexes, with diminished signal intensity and resolution (Figure 2.9). The problem of transverse relaxation in the NMR of large biomolecules and methods to combat this issue is discussed further in chapter 5.

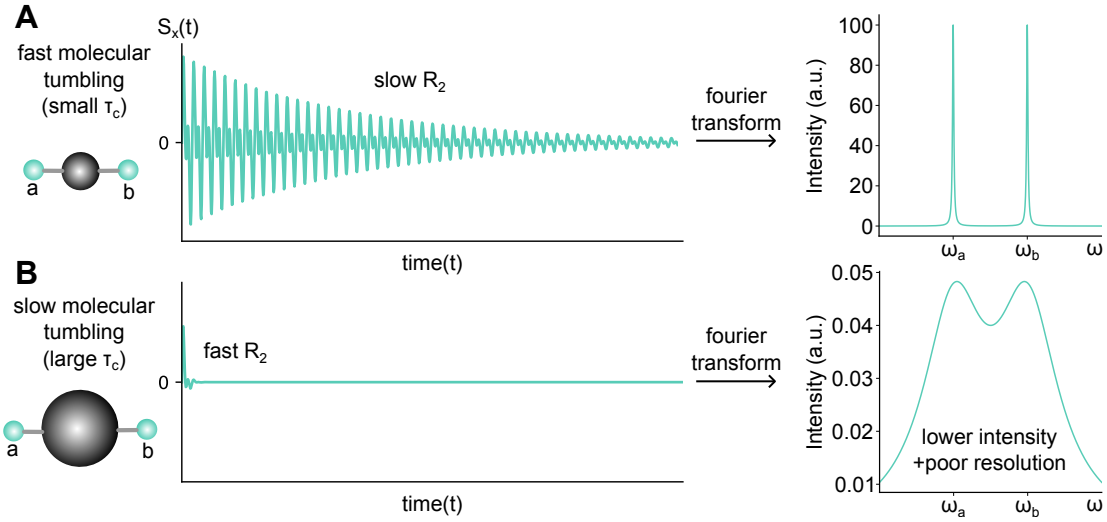


Figure 2.9: A comparison of the NMR spectra obtained from two spins (a/b) in a fast tumbling molecule (**A**) versus a slow tumbling molecule (**B**).

2.1.5 Density matrices and operators

Equation 2.3 represents the equation of motion for magnetisation of an isolated spin-1/2 nucleus in an external magnetic field, published by Bloch in 1946.[79] However, due to the quantum mechanical nature of NMR, it became apparent that using density matrices and operators[91] would provide a more technical route to explain the spin states in NMR and the action of pulses. If a system prepared in state $|\psi(t)\rangle$ is under the presence of an interaction Hamiltonian ($\hat{H}$), the evolution of $|\psi(t)\rangle$ over time is represented through the time-dependent Schrödinger equation:[92]

$$\frac{d}{dt}|\psi(t)\rangle = -i\frac{\hat{H}}{\hbar}|\psi(t)\rangle \tag{2.8}$$

If the density matrix/operator ρ(t) is represented by $|\psi(t)\rangle\langle\psi(t)|$, then the evolution of ρ(t) under the action of the interaction Hamiltonian (H) is given by the Liouville-Von Neumann equation:*

$$\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}\hat{H}\rho(t) \quad \hat{H} = [H, \ ] \tag{2.9}$$

*Derivation in Hore *et al.* 2023.[75]

In the density matrix representation, the x, y, z components of the magnetisation can be represented by the Pauli spin matrices ($\hat{I}_x, \hat{I}_y, \hat{I}_z$),[93] and the identity operator ($\hat{E}$):

$$\hat{I}_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \hat{I}_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \hat{I}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \hat{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.10}$$

When a spin system is described in terms of $\hat{I}_x, \hat{I}_y, \hat{I}_z$ and $\hat{E}$, this is termed the Cartesian basis. However, the same system can be described analogously in the shift basis using the operators $\hat{I}_+, \hat{I}_-, \hat{I}_\alpha$ and $\hat{I}_\beta$. The raising and lowering operators ($\hat{I}_+, \hat{I}_-$) describe the total transverse magnetisation within a system ($\hat{I}_\pm = \hat{I}_x \pm i\hat{I}_y$) and the operators $\hat{I}_{\alpha/\beta} = \frac{1}{2}\hat{E} \pm \hat{I}_z$ represent spins populating the α and β spin-states, respectively.

$$\hat{I}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \hat{I}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \hat{I}_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \hat{I}_\beta = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.11}$$

For an isolated spin-1/2 nucleus (I), the Zeeman interaction can be represented using the following Hamiltonian:

$$\hat{H}_0 = \hbar\gamma_I B_0 \hat{I}_z = \hbar\omega_{0,I}\hat{I}_z \tag{2.12}$$

The action of radio-frequency pulse of frequency ω_{rf} and amplitude B_1 along the y-axis on a spin I in the lab frame is described by the following Hamiltonian:*

*Adapted from equation 2.2.

$$\hat{H}_1(t) = \hbar\omega_1[\cos(\omega_{\mathrm{rf}}t)\hat{I}_y + \sin(\omega_{\mathrm{rf}}t)\hat{I}_x] \tag{2.13}$$

where $\omega_1 = \gamma_I B_1$, with γ_I being the gyromagnetic ratio of spin I. If the frame of reference for the system is changed (as is the case in the rotating frame) according to the similarity transform by unitary transformation U, $\rho_{rot} = U\rho U^{-1}$, then the Hamiltonian in the new frame ($\hat{H}_{rot}$) can be expressed as:*

$$\hat{H}_{rot} = U\hat{H}U^{-1} - i\hbar U\frac{dU^{-1}}{dt} \tag{2.14}$$

To transform from the lab frame to the rotating frame, the rotational transformation $U = e^{i\omega_{rf}\hat{I}_z t}$ is applied. After inserting this into equation 2.14, along with the lab frame B_0 and B_1 Hamiltonians (equations 2.12/2.13), the total Hamiltonian has following form:

$$\hat{H}_{rot} = \hbar\omega_I\hat{I}_z + \hbar\omega_1\hat{I}_y \tag{2.15}$$

where $\omega_I = \omega_{0,I} - \omega_{rf}$. Therefore, when a pulse is applied completely on resonance (ω_I=0), the Hamiltonian in the rotating frame can be expressed as:[75]

$$\hat{H} = \hbar\omega_1\hat{I}_y \tag{2.16}$$

If a system is prepared in state $\rho(0)$, the final state $\rho(t)$ after the application of a static Hamiltonian (such as a radio-frequency pulse about the y-axis) is given by the so-called "sandwich formula":[95]

$$\rho(t) = e^{-i\hat{H}t/\hbar}\rho(0)e^{i\hat{H}t/\hbar} = e^{-i\omega_1 t\hat{I}_y}\rho(0)e^{i\omega_1 t\hat{I}_y} \tag{2.17}$$

This equation can be solved analytically using the following formula:[95]

*Derivation in Cavanagh *et al.* 2007.[94]

$$\rho(t) = e^{-i\omega_1 t\hat{I}_y}\rho(0)e^{i\omega_1 t\hat{I}_y} = \cos(\omega_1 t)\rho(0) - i[\hat{I}_y, \rho(0)]\sin(\omega_1 t) \tag{2.18}$$

The total wavefunction for a spin-1/2 nucleus is represented as $|\psi\rangle = c_\alpha |\alpha\rangle + c_\beta |\beta\rangle$, then the density matrix is represented as:[75]

$$\rho = \begin{pmatrix} c_\alpha c_\alpha^* & c_\alpha c_\beta^* \\ c_\beta c_\alpha^* & c_\beta c_\beta^* \end{pmatrix} \tag{2.19}$$

where $c_\alpha c_\alpha^* + c_\beta c_\beta^* = 1$. Each coefficient can be split into a real magnitude and a phase ($c_k = a_k e^{i\phi_k}$). Therefore, the density matrix can be represented as:[75]

$$\rho = \begin{pmatrix} a_\alpha^2 & a_\alpha a_\beta e^{i(\phi_\alpha - \phi_\beta)} \\ a_\alpha a_\beta e^{i(\phi_\beta - \phi_\alpha)} & a_\beta^2 \end{pmatrix} \tag{2.20}$$

The ensemble average of the density matrix can be determined by calculating the ensemble average of each term in equation 2.20. At equilibrium (ρ_{eq}), the ensemble average of the off-diagonal terms in equation 2.20 are zero as the spins have randomised phases.[75] Furthermore, the terms a_α^2 and a_β^2 represent the probabilities of a spin being in the $|\alpha\rangle$ or $|\beta\rangle$ states, given by p_α and p_β, respectively.[75] Therefore, the equilibrium density matrix can be written as:

$$\begin{aligned} \rho_{eq} &= \begin{pmatrix} p_\alpha & 0 \\ 0 & p_\beta \end{pmatrix} = \frac{1}{2}\begin{pmatrix} p_\alpha + p_\beta & 0 \\ 0 & p_\alpha + p_\beta \end{pmatrix} + \frac{1}{2}\begin{pmatrix} p_\alpha - p_\beta & 0 \\ 0 & -(p_\alpha - p_\beta) \end{pmatrix} \\ &= \frac{1}{2}\hat{E} + (p_\alpha - p_\beta)\hat{I}_z \end{aligned} \tag{2.21}$$

The identity term $\hat{E}$ cannot interact or be changed during any NMR experiment and therefore it is often omitted during calculations. Furthermore, the term $(p_\alpha - p_\beta)$ is a scaling factor and is therefore often omitted, leading to $\rho_{eq} = \hat{I}_z$.[75]

After a 90^o pulse is applied to equilibrium magnetisation ($\rho(0) = \hat{I}_z$), with $t{=}\pi/(2\omega_1)$, the magnetisation is rotated onto the x-axis ($\rho(\frac{\pi}{2\omega_1}) = \hat{I}_x$). After the pulse, under the effect of the Zeeman interaction ($\hat{H} = \hbar\omega\hat{I}_z$), magnetisation will oscillate in the rotating frame according to:

$$\rho(t) = \cos(\omega t)\hat{I}_x - i[\hat{I}_z, \hat{I}_y]\sin(\omega t) = \cos(\omega t)\hat{I}_x + \hat{I}_y \sin(\omega t) \tag{2.22}$$

Thereby, in the absence of relaxation (R_1 and R_2) the equation of motions in equation 2.3 are recovered.

2.2 NMR imaging experiments

2.2.1 Gradient echo experiments

During a magnetic field gradient along the z-axis, the Larmor frequency of a spin is proportional to position:

$$\omega_0 = \gamma B_{\text{tot}} = \gamma(B_0 + zG) \tag{2.23}$$

where G is the magnetic field gradient strength, and z is the z-position (Figure 2.10A/B). This finding allowed the development of magnetic resonance imaging (MRI) whereby nuclear spins oscillating at a given frequency informed you of the location of those spins within a sample.[41]

A type of experiment used commonly in MRI is the gradient echo (GRE).[96] The pulse sequence of a gradient echo experiment is shown in Figure 2.10C, and a vector model depiction is included to explain the action of the pulse sequence on spins in the sample. Recording an echo in MRI is beneficial because sampling of both the positive and negative portions of the free induction decay (FID) is obtained, improving the efficiency of the Fourier Transform (FT) and the signal-to-noise ratio

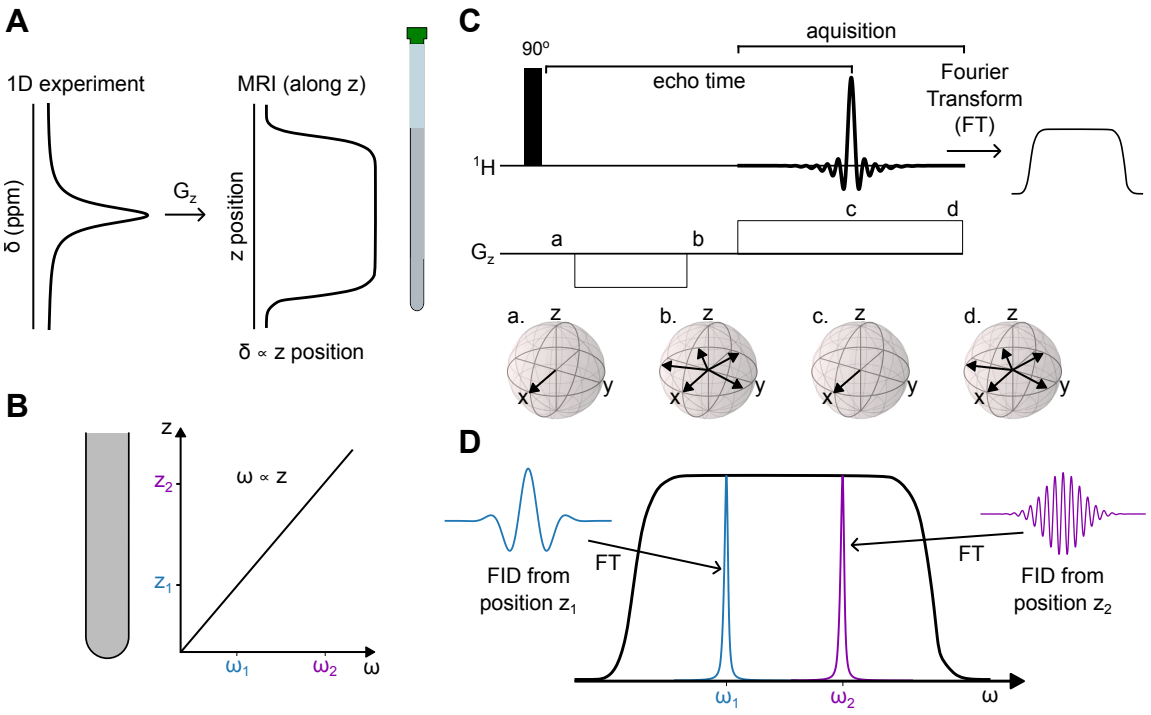


Figure 2.10: An introduction to magnetic resonance imaging (MRI). **A** A schematic showing how magnetic field gradients (e.g. G_z - gradient along the z-axis) can be used to convert chemical shift (δ) information into positional information. **B** After the application of a magnetic field gradient along z (G_z), the Larmor frequency (ω) of identical spins (e.g. water protons) is proportional to the z position. **C** A standard one-dimensional proton gradient-echo (GRE) along the z-axis of a sample.[96] A vector model depiction to explain the action of the pulse sequence on spins in the sample is shown in a-d. **D** An illustration showing how a total MRI spectrum is produced from identical spins at different positions along z having different frequencies (ω). The full MRI spectrum can be reconstructed when all positions are considered and their resulting NMR spectra are added together. The following animation shows how the gradient echo imaging experiment works: `https://youtube.com/watch?v=WKu-GvYDhcU?feature=share` (QR code in margin).

of the resulting spectrum. The resulting MRI spectrum contains information on the intensity of signals in the sample as a function of frequency (ω) and, therefore, z-position (Figure 2.10D). Additional gradients along the x and y dimensions (G_x, G_y) can be applied before the GRE element with their magnitude sequentially changed, leading to x and y positionally dependent phase-encoding of the resultant signal. This is used in a clinical setting to develop 3D images.

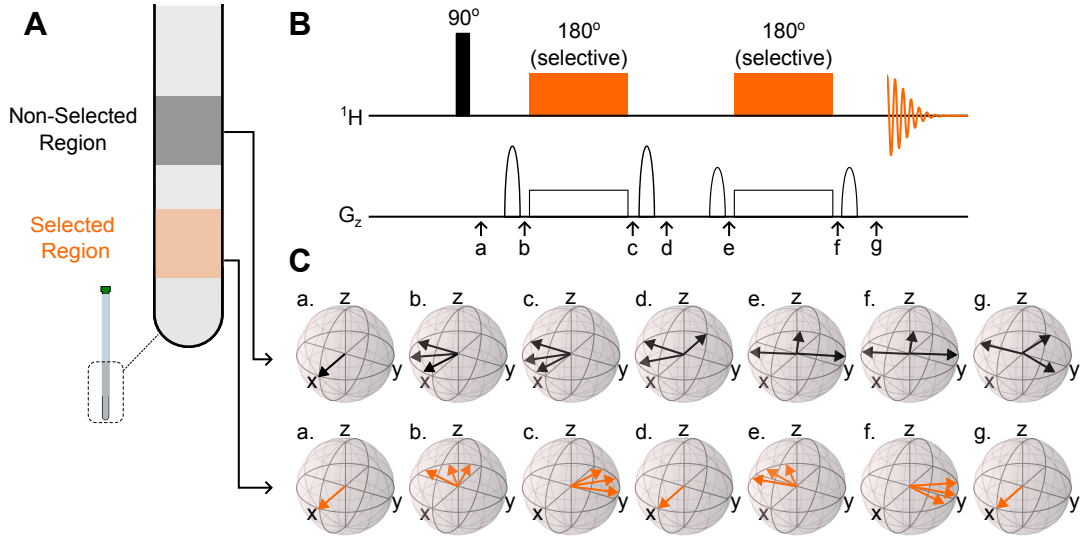


Figure 2.11: Slice selective NMR experiments. **A** An NMR tube showing regions of an NMR sample, where an NMR spectrum is desired for the selected (orange) region only. **B** The pulse sequence of the double pulse field gradient spin echo (DPFGSE) experiment[97] is used to achieve slice selectivity. **C** A vector model depiction of spins in either the selected region (orange arrows) or a non-selected region (grey arrows) at positions a-g in the pulse sequence shown in B. Animation explanation of slice selective experiments: `https://youtu.be/dwfgvanrKRY` (QR code in margin).

2.2.2 Slice selective NMR experiments

The development of magnetic field gradients allowed slice-selective NMR experiments to be performed on a sample. Under the presence of a magnetic field gradient, the Larmor frequency is proportional to position (equation 2.23). Pulses can be designed such that when applied to the sample, they only act upon a particular region of frequencies which, under the presence of a linear magnetic field gradient, is proportion to position (Figure 2.11A). A slice-selective experiment such as the double pulse field gradient spin echo (DPFGSE)[97] can then be used to selectively record the NMR spectrum within the selected region (Figure 2.11B). Spins in the chosen area (orange) are refocussed before acquisition, leading to observable signal (Figure 2.11C). However, spins outside the selected region (grey) are de-phased before the acquisition, so any resulting signal in the FID sums to zero. Therefore, the resulting NMR spectrum only contains signals from the selected region. The same principles described here, apply to slice (or positional)

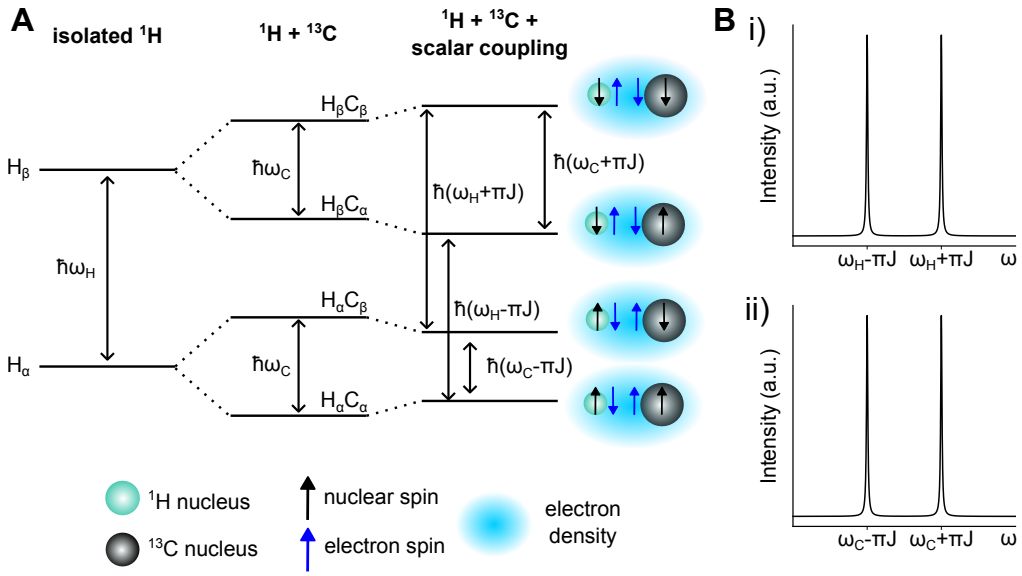


Figure 2.12: The Fermi contact interaction and scalar coupling. **A** Schematic showing how scalar coupling changes the spin energy levels of a chemically bonded ^{1}H-^{13}C spin pair through the Fermi contact interaction. **B** Illustrations of the doublets formed by scalar coupling of the ^{1}H-^{13}C spin pair for (i) the ^{1}H NMR spectrum and (ii) the ^{13}C NMR spectrum (peak intensities not to scale).

selective magnetic resonance spectroscopy experiments used in clinical settings to perform *in-vivo* metabolomics.[98]

2.3 Heteronuclear NMR

In 1951, Erwin Hahn provided an extensive explanation of the modulation of chemical shift due to scalar coupling (J) between two nuclear spins (I/S), including the description of a modified Hamiltonian ($\hat{H}$) under the Zeeman interaction in the presence of scalar coupling:[99]

$$\hat{H}/\hbar = \omega_{0,I}\hat{I}_z + \omega_{0,S}\hat{S}_z + 2\pi J(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y + \hat{I}_z\hat{S}_z) \tag{2.24}$$

where the third term containing the scalar coupling constant (J) is required to explain the splitting of peaks caused by scalar coupling, later attributed to the Fermi contact interaction (Figure 2.12A).

For an isolated proton, splitting of the H_α and H_β spin states occurs under the Zeeman interaction. In the addition of a non interacting ^{13}C nucleus the energy levels are further split into $H_\alpha C_\alpha$, $H_\alpha C_\beta$ and $H_\beta C_\alpha$ and $H_\beta C_\beta$ where transitions of frequency ω_H and ω_C are possible in the system (Figure 2.12A). However, under the presence of the Fermi contact interaction through a single bond, it becomes energetically favourable for the H and C spin states to be opposed to one another due to the favourable orientation of oppositely signed nuclear and electron spin states (Figure 2.12A).[61,100] This leads to a splitting of the energy levels, and therefore, the proton and ^{13}C NMR spectra are split into doublets (Figure 2.12B).

Equation 2.24 is expressed in the laboratory frame; however, it can be converted to the doubly rotating frame for both nuclei using the rotational transformation matrix $U = e^{i(\omega_{rf,I}\hat{I}_z + \omega_{rf,S}\hat{S}_z)t}$ and inserting it (along with equation 2.24) into equation 2.14. This leads to the following doubly rotating frame Hamiltonian:

$$\hat{H}_{rot}/\hbar = \omega_I\hat{I}_z + \omega_S\hat{S}_z + 2\pi J\hat{I}_z\hat{S}_z + \cos((\omega_{rf,I} - \omega_{rf,S})t)(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y) + \sin((\omega_{rf,I} - \omega_{rf,S})t)(\hat{I}_x\hat{S}_y - \hat{I}_y\hat{S}_x) \tag{2.25}$$

where ω_I and ω_S are equal to $\omega_{0,I} - \omega_{rf,I}$ and $\omega_{0,S} - \omega_{rf,S}$.

For the case of a heteronuclear spin system, $\omega_I - \omega_S >> 0$, therefore the trigonometric terms depend on time, rapidly oscillate and average to zero. This leads to the following weak coupling Hamiltonian:[75]

$$\hat{H}_{rot}/\hbar = \omega_I\hat{I}_z + \omega_S\hat{S}_z + 2\pi J\hat{I}_z\hat{S}_z \tag{2.26}$$

2.3.1 Scalar coupling coherence transfers

In scalar coupled systems, due to the presence of the additional scalar coupling term in the Zeeman Hamiltonian (equation 2.24), multidimensional and heteronuclear NMR experiments have been developed by using pulse sequences to transfer magnetisation from one nucleus to another. The following sections outline two common pulse sequences that transfer magnetisation from a high γ nucleus to a low γ nucleus and increase the intensity of the produced NMR spectrum.

2.3.1.i Insensitive nuclei enhanced by polarisation transfer (INEPT)

The insensitive nuclei enhanced by polarisation transfer (INEPT) pulse sequence was published in 1979 as a means of transferring magnetisation from a high γ nucleus (e.g. ^{1}H, ^{19}F) to a low γ nucleus (e.g. ^{13}C, ^{15}N) and increasing the intensity of the produced NMR spectrum.[101] A schematic of an INEPT pulse sequence is shown in Figure 2.13A for the case of a scalar coupled ^{1}H-^{13}C pair. Initially, equilibrium ^{13}C magnetisation is eliminated, followed by a 90^o pulse on ^{1}H about the x-axis. The next part of the sequence (τ/2-180$^o_{x,I+S}$-τ/2) enables coherence transfer from ^{1}H to ^{13}C, with the weak scalar coupling Hamiltonian (equation 2.26) acting during the $\tau/2$ delays. Figure 2.13B shows the proportion of the operators H_y and H_xC_z after the τ/2-180$^o_{x,I+S}$-$\tau/2$ element as a function of the duration τ. A total mixing time (τ) equal to 1/(2J) provides maximum conversion to the desired H_xC_z coherence.[75]

The sequence is then followed by a 90^o_y pulse on ^{1}H and a pulse field gradient along the z-axis to eliminate any remaining H_y magnetisation. The final 90^o_x pulse on ^{13}C converts the magnetisation H_zC_z into observable magnetisation -H_zC_y. The resulting magnetisation will have a magnitude γ_H/γ_C fold larger than that obtained from a single carbon 90^o excitation pulse applied to equilibrium magnetisation, assuming perfect transfer through the pulse sequence. Furthermore, it is worth noting that the operator -H_zC_y produces an anti-phase NMR spectrum, where each multiplet peak has an opposite sign (Figure 2.5C).[75]

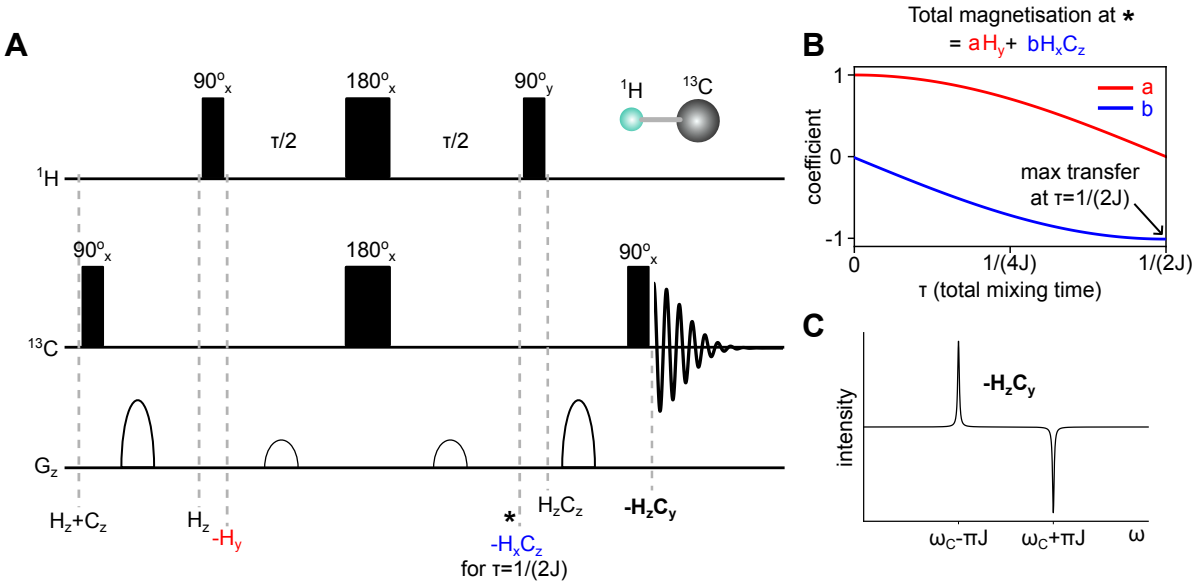


Figure 2.13: An introduction to the INEPT pulse sequence. **A** A schematic of the INEPT pulse sequence. **B** A plot of the magnetisation contributions after the $\tau/2$-$180^{o}_{x,I+S}$-$\tau/2$ section of the pulse sequence as a function of total delay duration τ. Maximum conversion to H_xC_z occurs with a total delay time τ equal to $1/(2J)$. **C** A plot of the anti-phase ^{13}C NMR spectrum obtained on for the -H_zC_y coherence.

2.3.1.ii Heteronuclear Hartmann-Hahn (HEHAHA) transfers

The Hartmann-Hahn experiment was originally described in 1962 to transfer magnetisation between dipolar coupled spins in the solid state, now known as the continuous polarisation (CP) method.[102] The technique was later extended to transfer magnetisation between heteronuclear[103] and homonuclear[104] scalar coupled spins in the solution state.

The heteronuclear Hartmann-Hahn experiment (HEHAHA) relies on the double rotating frame concept.[102] Taking the example of a ^{1}H-^{13}C scalar coupled spin-system, when simultaneous linearly polarised radiation is applied to both ^{1}H and ^{13}C along the y-axis with wave equations of the form $2B_{1,H}\cos(\omega_{\mathrm{rf,H}}t)$ and $2B_{1,C}\cos(\omega_{\mathrm{rf,C}}t)$, respectively, after entering the double rotating frame, the applied fields are now static $B_{1,H}/B_{1,C}$ (Figure 2.14Ai). In the double rotating frame, ^{1}H and ^{13}C magnetisation will oscillate in the xz-plane at frequencies $\omega_{1,H}$ and $\omega_{1,C}$,respectively, where $\omega_{1,H} = \gamma_H B_{1,H}$ and $\omega_{1,C} = \gamma_H B_{1,C}$ (Figure 2.14Ai). A

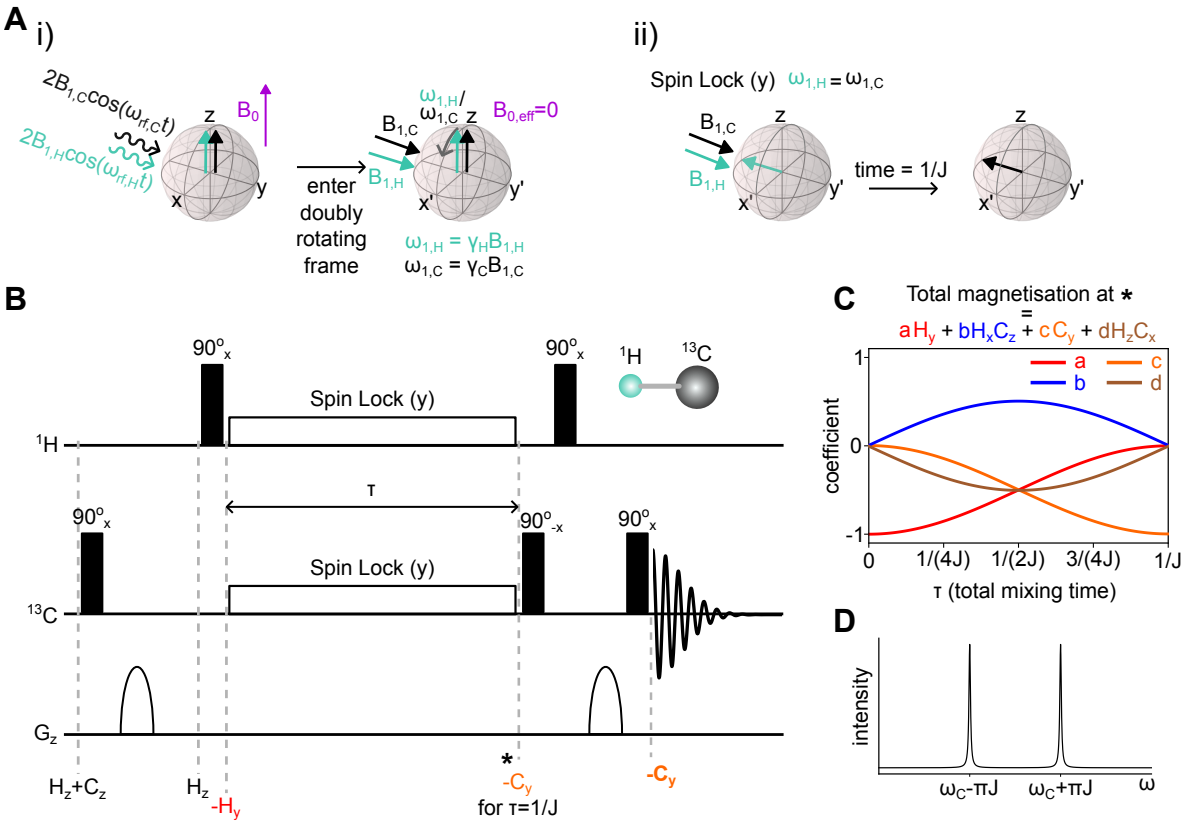


Figure 2.14: An introduction to heteronuclear Hartmann-Hahn (HEHAHA) coherence transfer. **A** (i) Description of the double rotating frame for a ^{1}H-^{13}C spin pair. (ii) If the frequencies $\omega_{1,H}$ and $\omega_{1,C}$ are matched, the system is said to be spin-locked. **B** A schematic of a HEHAHA pulse sequence with a simultaneous spin-lock applied along the y-axis for ^{1}H and ^{13}C. **C** A simulation of the final magnetisation within the system after the spin-lock application as a function of spin-lock duration (τ). **D** Depiction of an in-phase ^{13}C NMR spectrum obtainable using a HEHAHA pulse sequence.

spin-lock is created on -H_y magnetisation if simultaneous ^{1}H and ^{13}C irradiation is applied along the y-axis, with the frequencies $\omega_{1,H}$ and $\omega_{1,C}$ calibrated to be equivalent and equal to ω_1 (Figure 2.14Aii). For a ^{1}H-^{13}C pair, the action of the spin-lock for a duration 1/J will result in conversion from -H_y to -C_y.[105]

Under the duration of an on resonance spin-lock ($\omega_{rf,H/C}$=$\omega_{0,H/C} = 0$) along the y-axis, the doubly rotating frame Hamiltonian is expressed as:

$$\hat{H}_{rot}/\hbar = \omega_1 \hat{H}_y + \omega_1 \hat{C}_y + 2\pi J \hat{H}_z \hat{C}_z \tag{2.27}$$

This Hamiltonian can be expressed in the interaction frame with respect to

$\omega_1\hat{H}_y$ and $\omega_1\hat{C}_y$ using equation 2.14 and the unitary transformation $U = e^{i\omega_1 t(\hat{H}_y+\hat{C}_y)}$ leading to:

$$\hat{H}_{int}(t)/\hbar = 2\pi J[\cos^2(\omega_1 t)\hat{H}_z\hat{C}_z + \sin^2(\omega_1 t)\hat{H}_x\hat{C}_x + \cos(\omega_1 t)\sin(\omega_1 t)(\hat{H}_z\hat{C}_x + \hat{H}_x\hat{C}_z)] \tag{2.28}$$

The $\cos^2(\omega_1 t)$ and $\sin^2(\omega_1 t)$ terms average to $1/2$ when time-averaged, whereas the term $\cos(\omega_1 t)\sin(\omega_1 t)$ averages to zero. Therefore, the Hamiltonian for the scalar coupled system in the interaction frame is given by:[106]

$$\hat{H}/\hbar = \pi J(\hat{H}_x\hat{C}_x + \hat{H}_z\hat{C}_z) \tag{2.29}$$

Figure 2.14B shows a HEHAHA pulse sequence with a spin-lock for duration τ. A simulation of the post-spin-lock magnetisation state under the Hamiltonian in equation 2.29 as a function of delay τ shows that $1/J$ is the optimal mixing time for the spin-lock in a ^{1}H-^{13}C coupled system, converting -H_y into -C_y (Figure 2.14C). After the spin-lock element in the pulse sequence, a series of clean-up pulses/gradients are applied to remove unwanted magnetisation produced as a consequence of imperfect frequency matching in the spin-lock (Figure 2.14A). The final ^{13}C magnetisation state of -C_y can be recorded, and an in-phase doublet is obtained (Figure 2.14D).

2.3.2 HSQC

Multidimensional NMR was first practically implemented for homonuclear correlation spectroscopy (COSY) experiments in 1976,[107] allowing correlations between neighbouring protons to be established. After the conception of the INEPT experiment in 1979,[101] Geoffrey Bodenhausen then published the heteronuclear single quantum coherence (HSQC) two-dimensional NMR experiment in 1980,[108]

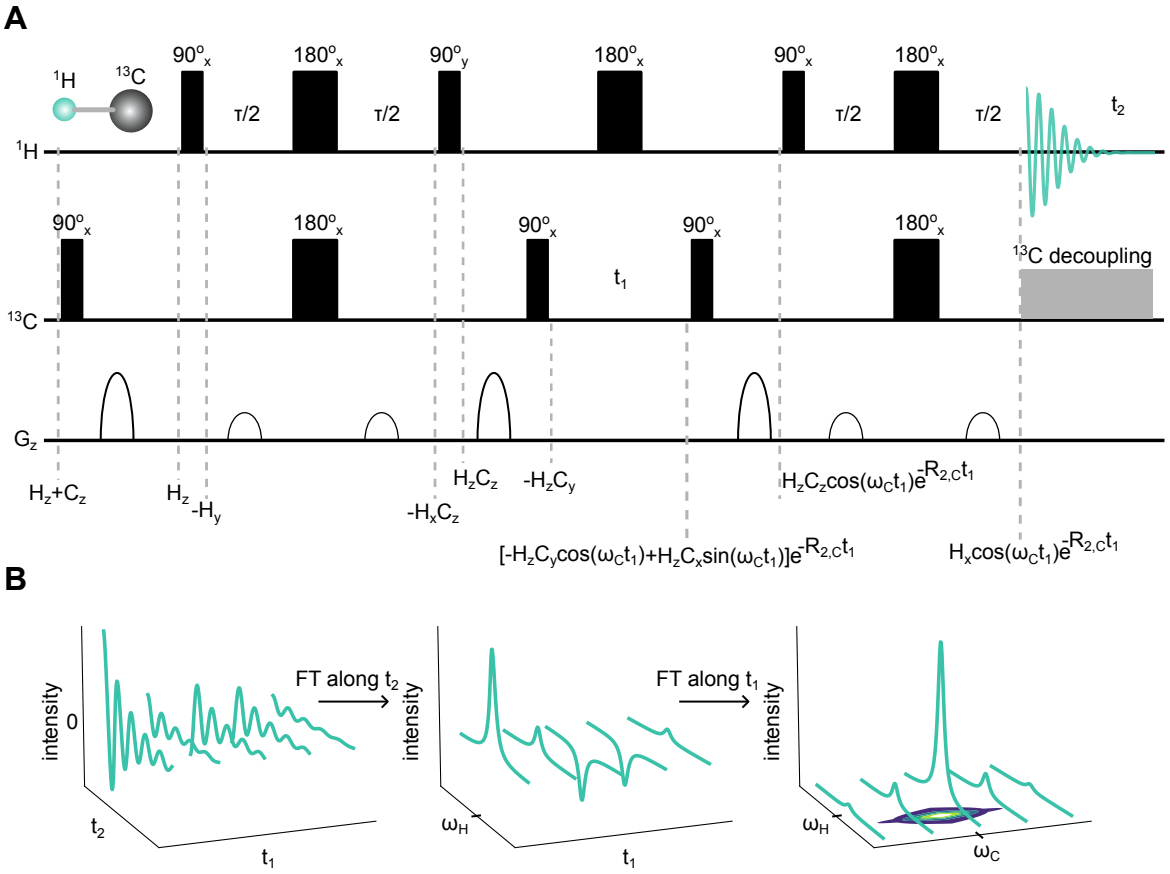


Figure 2.15: The heteronuclear single quantum coherence (HSQC) two-dimensional NMR experiment. **A** A schematic of a HSQC pulse sequence and specified coherences produced throughout the sequence for a ^{1}H-^{13}C scalar coupled system. **B** Recording the FID during t_2 as a function of t_1 delay time followed by successive Fourier transforms along t_2 and t_1 leads to the production of a two-dimensional NMR spectrum with a peak at the coordinates (ω_H,ω_C).

correlating the chemical shifts of two nuclei that are scalar coupled in a single spectrum. A schematic of the HSQC pulse sequence is shown in Figure 2.15A for a ^{1}H-^{13}C scalar coupled system.

All pulses before the time delay t_1 are the same as in Figure 2.13A to produce -H_zC_y. Free precession is allowed during an indirect dimension time delay t_1, with a 180^o pulse on the proton (at 1/2 the t_1 delay) to refocus any scalar coupling evolution that occurs under the Zeeman Hamiltonian (equation 2.24). After the indirect dimension delay period t_1, the magnetisation will have evolved to become:

$$-H_zC_y \xrightarrow{t_1} [-H_zC_y\cos(\omega_c t_1) + H_zC_x\sin(\omega_c t_1)]\, e^{-R_{2,C}t_1} \qquad (2.30)$$

After the t_1 delay, another INEPT element is used to transfer the magnetisation back to the ^{1}H nucleus. Before detection on ^{1}H, the coherence is of the form:

$$H_x \cos(\omega_c t_1) e^{-R_{2,C} t_1} \tag{2.31}$$

Therefore, after the acquisition during time t_2, the magnetisation will be of the form:

$$[H_x \cos(\omega_H t_2) + H_y \sin(\omega_H t_2)] \cos(\omega_c t_1) e^{-R_{2,C} t_1} e^{-R_{2,H} t_2} \tag{2.32}$$

where scalar coupling has been ignored due to the presence of ^{13}C decoupling during acquisition. Treating the H_x and H_y coefficients as real and complex points of the signal S_+, respectively, the signal can be written as:

$$S(t_1, t_2) = \cos(\omega_C t_1) e^{-R_{2,C} t_1} e^{(i\omega_H - R_{2,H}) t_2} \tag{2.33}$$

The pulse 90^o_x pulse on ^{13}C before the t_1 delay can be swapped for a 90^o_y pulse, leading to the following final signal after the acquisition:

$$S(t_1, t_2) = \sin(\omega_C t_1) e^{-R_{2,C} t_1} e^{(i\omega_H - R_{2,H}) t_2} \tag{2.34}$$

The signals from 2.33 and 2.34 can be combined to give quadrature detection in the indirect dimension:

$$S(t_1, t_2) = e^{(i\omega_H - R_{2,H}) t_2} e^{(i\omega_C - R_{2,C}) t_1} \tag{2.35}$$

This procedure of quadrature detection is known as the States method.[109] The experiment is repeated incrementally, increasing t_1 upon each increment.[75] Plots of real signal $\mathrm{Re}[S(t_1, t_2)]$ are shown for different t_1 delays in Figure 2.15B. A Fourier transform is then applied to the data of the form in equation 2.35 along both dimensions t_1 and t_2 to obtain the ^{1}H-^{13}C correlation peak centred at the coordinates (ω_H,ω_C), commonly represented as a contour plot (Figure 2.15B).

2.4 Measuring relaxation in NMR

In sections 2.1.3.iv-2.1.4, transverse and longitudinal relaxation rates (R_2 and R_1) were shown to depend highly on molecular rotations and tumbling of spins within a system. Therefore, information on molecular tumbling can be determined by quantifying the R_1 and R_2 relaxation rates.

Longitudinal relaxation was initially measured using saturation recovery methods as discussed by Bloembergen, Purcell and Pound.[81] The methods involved saturation of the spins using a large radiofrequency field.[81] Then, at varying times after the application of the saturation field, the equilibrium magnetisation was observed using a weak radiofrequency field sweep method to measure the absorption spectrum.[81] By measuring the intensity of the absorption spectrum as a function of time, the longitudinal relaxation rate could be determined. However, longitudinal relaxation rates are typically measured today using the pulsed inversion recovery experiment.[110] The inversion recovery method uses the pulse sequence 180^o-T_{relax}-90^o followed by measurement of the FID (Figure 2.16A). Using a 180^o pulse to invert the z-magnetisation produces a well characterised initial state from which longitudinal relaxation can occur. The experiment is repeated at varied T_{relax} durations and the observed intensities can be fit to equation 2.36.

$$I(T_{\text{relax}}) = I(T_{\text{relax}} = \infty)\left[1 - 2e^{-R_1 T_{\text{relax}}}\right] \tag{2.36}$$

where I are the peak intensities and $I(T_{\text{relax}} = \infty)$ and R_1 are fitting parameters (Figure 2.16B).

NMR spectrometer magnetic fields are not completely homogeneous (Figure 2.17A), and therefore, the Larmor frequencies for a spin depend on its position within the NMR sample.

To quantify the transverse (R_2) relaxation rate, Erwin Hahn published the Hahn-echo NMR experiment in 1950[111] involving a $90^o_x - \tau - 90^o_x - \tau$ sequence. However,

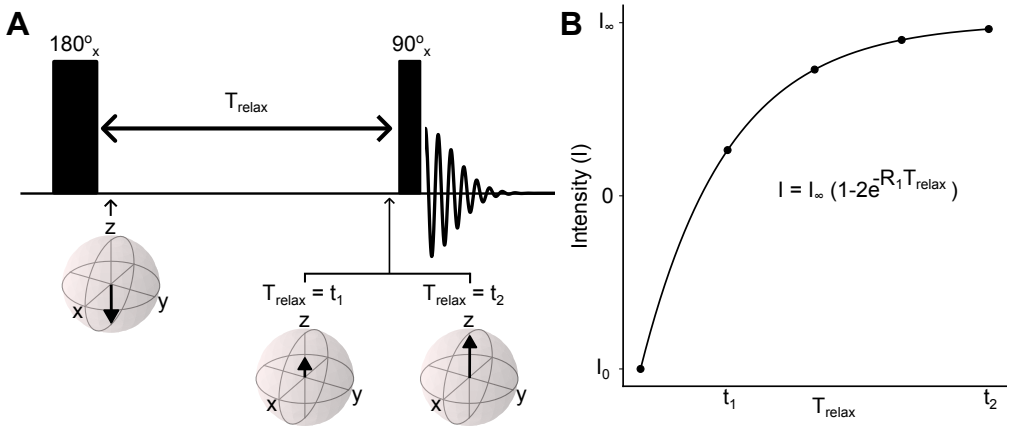


Figure 2.16: The inversion recovery experiment. **A** The inversion recovery pulse sequence with a relaxation delay of T_{relax}. **B** Repeating the inversion recovery experiment at a range of T_{relax} times and fitting the obtained spectral intensities to the equation $I(T_{relax}) = I(T_{relax} = \infty)\left[1 - 2e^{-R_1 T_{relax}}\right]$ enables the determination of the longitudinal relaxation rate (R_1).

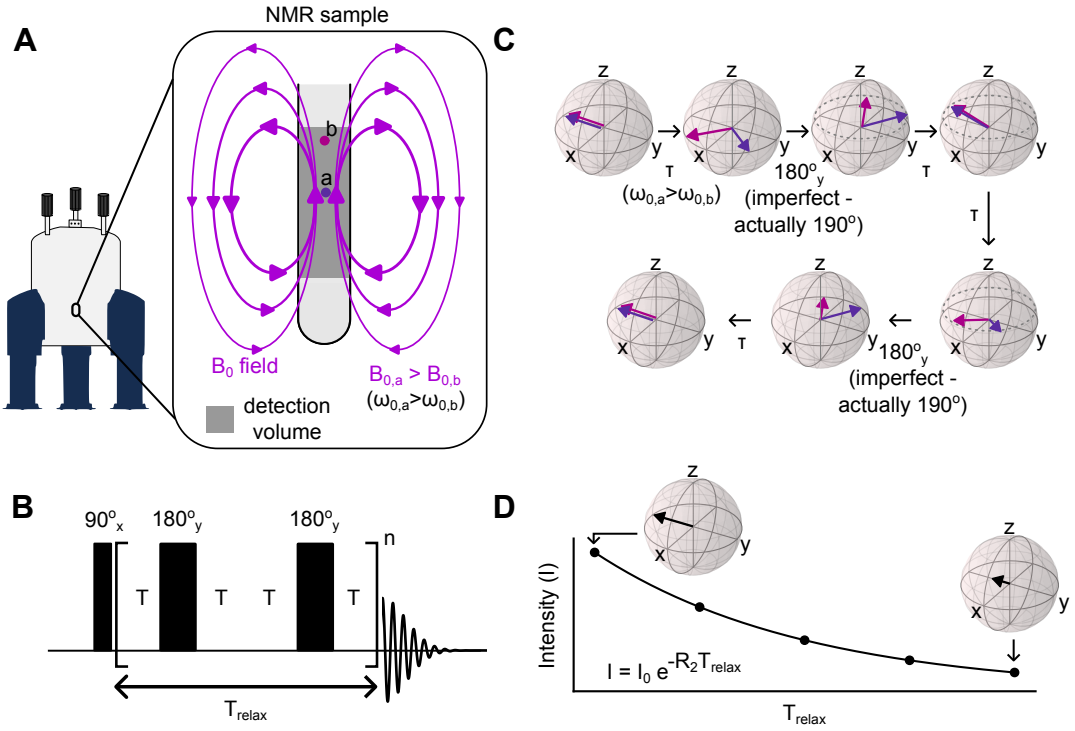


Figure 2.17: Spin-echos and transverse relaxation. **A** A depiction of magnetic field inhomogeneity within an NMR spectrometer. **B** A CPMG pulse sequence,[17,18] with each cycle containing an even number of echos to account for imperfect 180^o pulse calibrations. **C** A vector model description of the CPMG experiment and how it can eliminate any relaxation caused by magnetic field inhomogeneity and reduce the effects of imperfect 180^o pulse calibration. **D** The R_2 relaxation rate of a spin can be determined by repeating the CPMG experiment with increasing T_{relax} durations and fitting the data to a single exponential decay.

only half of the magnetisation can be refocused during the Hahn-echo sequence.[111,112] The Hahn-echo requires the τ delays to be long enough for complete de-phasing in the transverse plane before the 90^o refocussing pulse.[112] However, significant molecular diffusion can occur during the long τ delays which, in the presence of an inhomogeneous B_0 field, can alter the average oscillation frequency of a spin on either side of the 90^o refocussing pulse. Therefore, the presence of molecular diffusion reduces the accuracy of R_2 measurements obtained using this method.

In 1954, Carr and Purcell altered the sequence to $90^o_x - [\tau - 180^o_x - \tau]^N$, known as a Carr-Purcell echo.[17] Swapping the 90^o_x refocussing pulse to a 180^o_x refocussing pulse and increasing the number of cycles (N) whilst shortening τ allowed the effects of molecular diffusion to be minimised.[112] However, such a sequence suffers from an alteration in sign of the refocused magnetisation depending on if an odd or even number (N) of echos were used.[112] In addition, the loss in transverse magnetisation caused by mis-calibration of the 180^o pulse is compounded for every $\tau - 180^o_x - \tau$ cycle, leading to reduced accuracy of the R_2 rate measurement when large numbers of N are used.[112] Therefore, in 1958, Meiboom and Gill[18] modified the sequence to $90^o_x - \left[\tau - 180^o_y - \tau\right]^N$, which is now known as a Carr-Purcell-Meiboom-Gill (CPMG) sequence,[17,18] used for most applications today to measure transverse relaxation (Figure 2.17B). Every echo caused by the CPMG sequence has the same phase and if the number of cycles (N) is even, it also accounts for mis-calibration of the 180^o refocussing pulses (Figure 2.17C).[112]

The CPMG sequence can refocus transverse relaxation caused by magnetic field inhomogeneity (Figure 2.17C). Therefore, any reduction in the signal of a non-coupled nucleus as T_{relax} is increased can be attributed to intrinsic transverse relaxation (Figure 2.17D) in the absence of chemical exchange (section 2.6.2). By repeating the experiment using increasing T_{relax} times (using larger values of N), the data can be fit to a single exponential, with the decay constant equal to R_2 (Figure 2.17D).

2.5 Measuring molecular translational diffusion using NMR

The following section outlines a stimulated-echo-based method to determine the translational diffusion coefficient of a spin in an NMR sample pioneered by Stejskal and Tanner in 1970.[20] A pulse sequence capable of measuring the diffusion coefficient of a spin (stimulated echo longitudinal encode-decode diffusion experiment) is shown in Figure 2.18A. A diffusion delay Δ and the encode/decode gradient durations δ are shown. During the encode/decode gradients, the Larmor frequency (ω) of a nuclear spin is proportional to its z-position in the NMR tube (Figure 2.18B). Without molecular diffusion, the magnetisation vectors at different positions along the z-axis are maximally refocussed (Figure 2.18Ci). Full refocussing in the stimulated echo experiment is not possible as only half the magnetisation after the application of the encode gradient can be rotated onto the z-axis.[20,113]

Nevertheless, diffusion during the delay Δ reduces refocusing during the decode gradient, resulting in diminished signal intensity (Figure 2.18). Therefore, a more considerable reduction in signal intensity during the experiment indicates increased translational diffusion. Stejskal and Tanner derived an equation to determine the loss in signal intensity as a function of the diffusion coefficient (D) of a spin:[19,20]

$$I = I_0 e^{-\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D} \tag{2.37}$$

where γ is the gyromagnetic ratio of the nuclear spin, G is the gradient strength, δ are the encode/decode gradient durations and Δ is the diffusion delay. Practically, to determine the diffusion coefficient (D), the gradient strength G is arrayed, and the subsequent data is fit to equation 2.37.

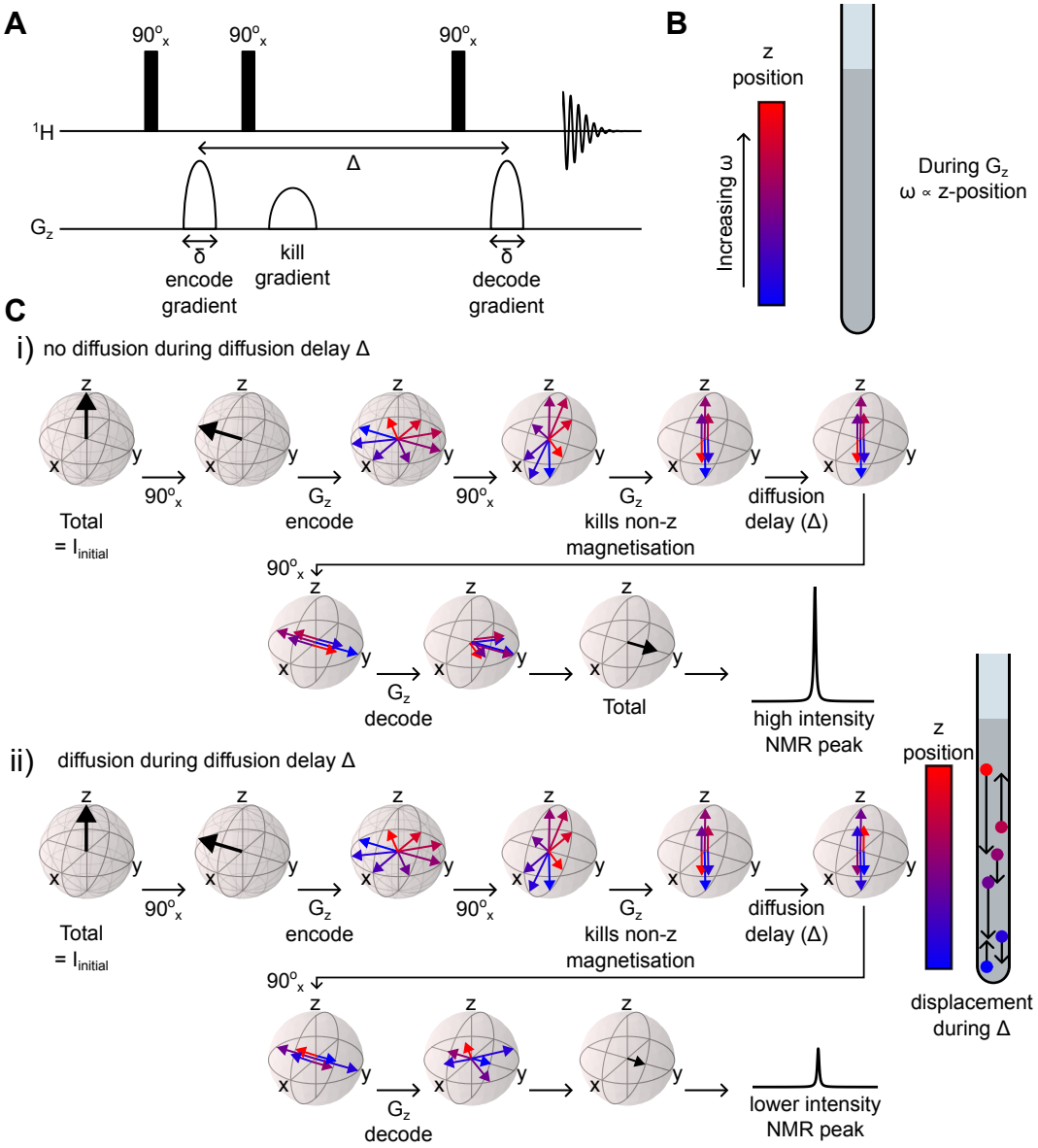


Figure 2.18: The stimulated echo longitudinal encode-decode diffusion experiment. **A** Pulse sequence of a proton stimulated echo longitudinal encode-decode diffusion experiment. A diffusion delay Δ and the encode/decode gradient durations δ are shown. **B** Illustration showing how during the encode/decode gradients, a nuclear spin's Larmor frequency (ω) is proportional to its z-position in the NMR tube. The z-positions, and therefore Larmor frequencies, of the spins in part C are represented using the red-blue colour bar. **C** i) A vector model representation of spins in an NMR sample during the pulse sequence A in the absence of diffusion. ii) A vector model representation of spins in an NMR sample during the pulse sequence A with a high degree of molecular diffusion during the delay Δ. Magnetisation is not refocused effectively in the decode gradient, leading to a reduced signal intensity. Note: the total magnetisation vectors are represented in black. The following animation shows how signal intensity is affected by diffusion in the diffusion experiment: `https://youtube.com/watch?v=RY9tkrIPouU` (QR code in margin).

2.6 Measuring chemical exchange and interactions using NMR

As a technique, NMR is uniquely positioned to measure chemical exchange and inter/intramolecular interactions. The following sections 2.6.1-2.6.3 outline a selection of experiments to measure chemical exchange and/or interactions in a given system.

2.6.1 Chemical Exchange Saturation Transfer (CEST)

The CEST experiment is part of the set of saturation transfer NMR experiments first proposed by Albert Overhauser in 1953.[115] The experiments involve selective excitation of specific chemical shifts and intensity monitoring of peaks not hit by the selective pulse. The pulse profile of a hard rectangular pulse (with constant phase and amplitude) of duration a takes the shape of a rectangular function given by:[116]

$$\text{Hard pulse of duration a} = B_1 \cdot \text{rect}(t/a) \tag{2.38}$$

where $\text{rect}(t/a)$ is equal to 1 for $|t| < a/2$ and is equal to 0 when $|t| > a/2$. The bandwidth of the pulse in frequency space can be viewed by taking the Fourier transform of equation 2.38 giving:[116]

$$\int_{-\infty}^{\infty} B_1 \cdot \text{rect}(t/a) e^{-i\omega t} dt = B_1 \frac{\sin(\pi a \omega)}{\pi a \omega} \tag{2.39}$$

This relationship explains why long-duration pulses have small bandwidths and can selectively excite particular peaks in an NMR spectrum (Figure 2.19A). Long pulse durations (i.e. large a values) lead to a widening of the function $\sin(ax)/(ax)$.

CEST experiments were initially used in MRI applications to study the chemical

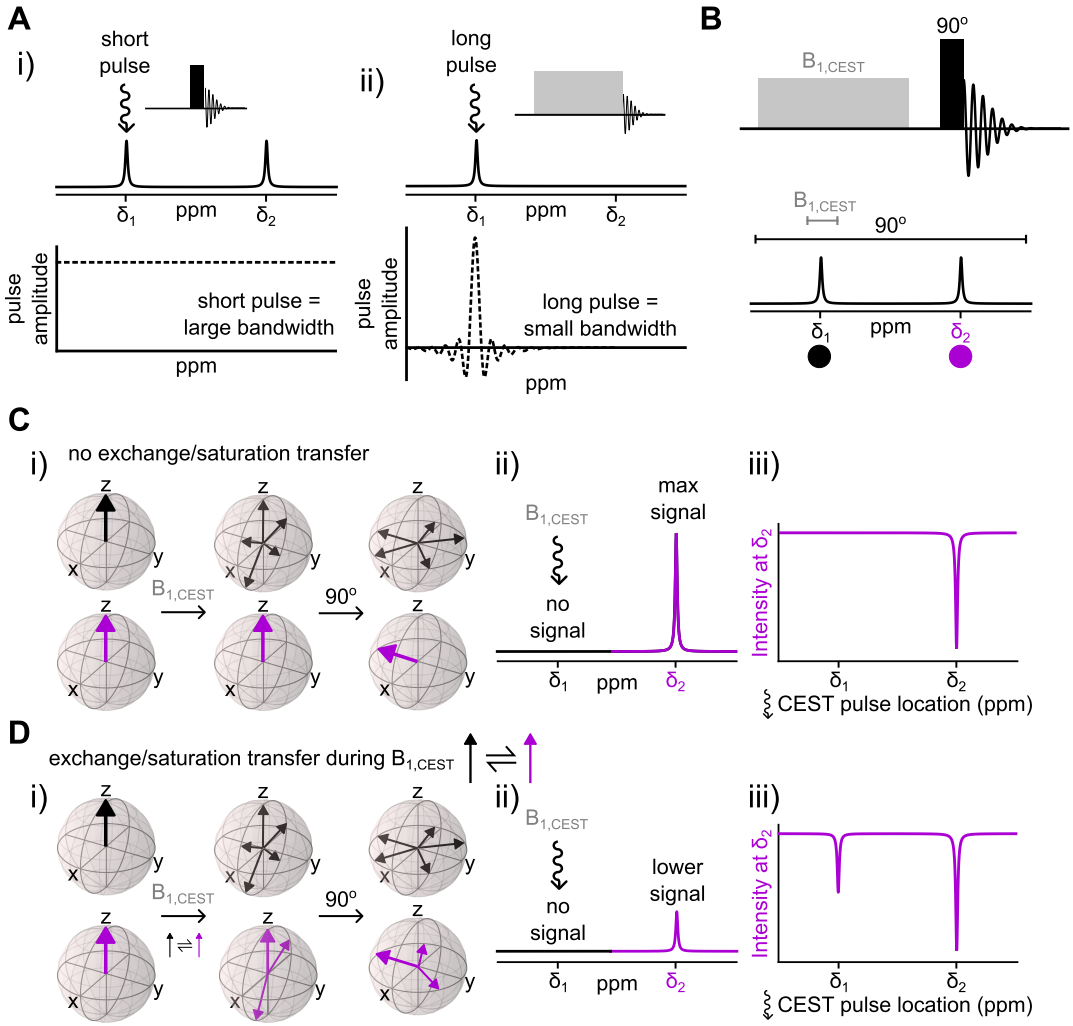


Figure 2.19: An introduction to the Chemical Exchange Saturation Transfer (CEST) experiment. **A** A schematic showing how short (i) and long (ii) NMR pulses centred at δ_1 have large and small chemical shift bandwidths, respectively. **B** A pulse sequence representing a saturation transfer experiment.[114] **C** i) A vector model illustration of spins at chemical shifts δ_1 (black) and δ_2 (purple) during the pulse sequence in B with the CEST pulse centred on δ_1 in the absence of no-exchange/magnetisation transfer between spins. ii) The resulting NMR spectrum gives no signal for the de-phased spins at δ_1 and maximum signal for the non-perturbed spins at δ_2. iii) When this process is repeated iteratively, with the saturation pulse centred on different chemical shifts, a CEST profile can be produced **D** The same as C, but chemical exchange/saturation transfer is occurring during the saturation pulse leading to attenuation of signal at δ_2 when the saturation pulse is centred at δ_1. The following animation is a simplified explanation of how the CEST experiment works: `https://youtube.com/watch?v=8Nq1qeL3BSs?feature=share` (QR code in margin).

exchange of water protons in the presence of contrast agents.[117] A CEST (or saturation transfer) pulse sequence involves a long-duration/short-bandwidth

saturation pulse ($B_{1,CEST}$), followed by a short-duration/large-bandwidth 90^o pulse[114] (Figure 2.19B). The CEST experiment involves repeating the saturation pulse sequence and changing the location of the saturation pulse each time. In the absence of saturation transfer between two spins with chemical shifts δ_1 and δ_2 during the saturation pulse ($B_{1,CEST}$), the intensity of a peak will only be diminished if the saturation pulse is applied directly to it (Figure 2.19C). However, if a saturation pulse is centred on δ_1 and chemical exchange or saturation transfer between spins 1 and 2 occurs during the saturation pulse, the intensity of the peak at δ_2 will also be attenuated (Figure 2.19Di/ii). After repeating the experiment at a range of saturation pulse locations, a CEST profile showing evidence of chemical exchange/saturation transfer between spins 1 and 2 is produced (Figure 2.19Diii). CEST profiles such as those shown in figures 2.19Ciii/Diii are also known as a Z-spectrum.[118]

As well as being used to study the chemical exchange of water with contrast agents in MRI experiments, CEST has been extended to looking at excited state conformers of proteins.[31] The CEST experiment can be used to study slow to intermediate exchange processes, $k_{ex} \lessapprox \Delta\nu$, where $k_{ex} = k_1 + k_{-1}$, with k_1 and k_{-1} being the forward and backward exchange rates between the two species, and $\Delta\nu$ is the frequency difference between the two states in Hz.

2.6.2 Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion

After the development of the CPMG experiment to measure transverse relaxation,[17,18] it was noted that a relaxation dispersion variant of the CPMG experiment could be used to calculate any chemical exchange contribution to the transverse relaxation rate ($R_{2,tot} = R_2 + R_{2,ex}$).[119] At low CPMG frequencies where $\nu_{CPMG} = 1/(4\tau)$, the CPMG elements are not able to effectively refocus magnetisation (Figure 2.20A), thereby leading to reduced signal intensity in

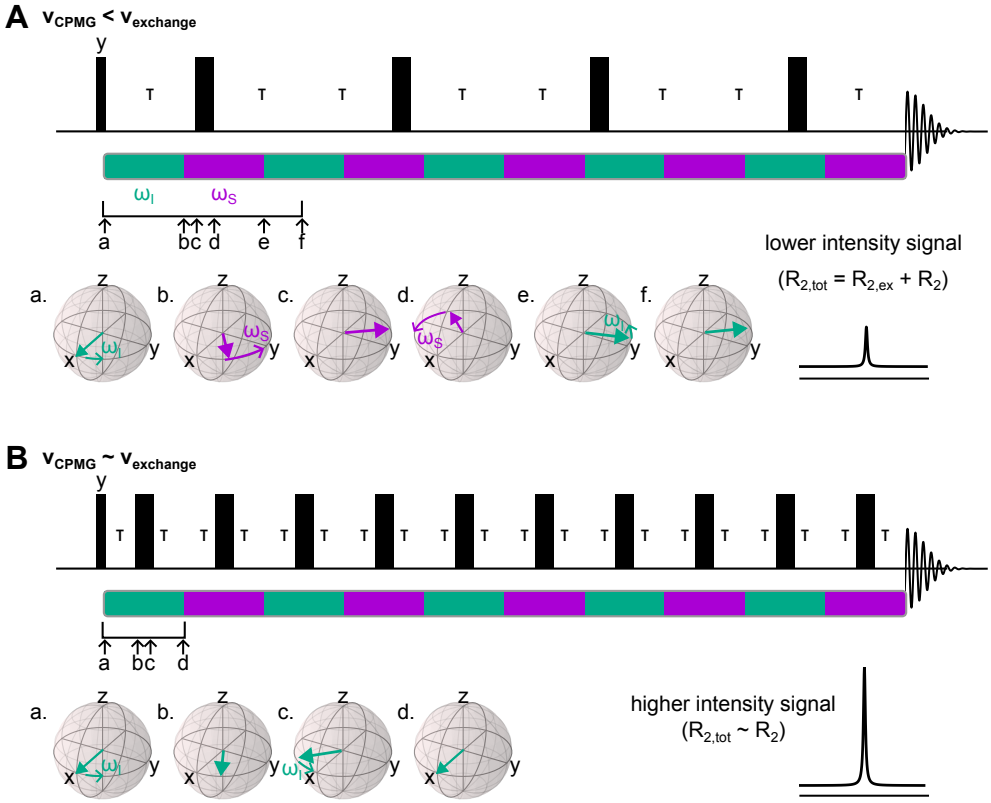


Figure 2.20: An introduction to the Carr-Purcell-Meiboom-Gill (CPMG) experiment applied to a system undergoing chemical exchange between species I and S with frequencies ω_I and ω_S, respectively. **A** At low CPMG frequencies, there is minimal refocusing of the magnetisation during the sequence, leading to lower signal intensity in the resulting NMR spectrum. **B** At large CPMG frequencies, refocusing of the magnetisation during the sequence is more efficient, leading to higher signal intensity in the resulting NMR spectrum. This figure is inspired from Baldwin *et al.*[3]

the resulting NMR spectrum. However, suppose a large CPMG frequency is applied with the same total CPMG duration. In that case, each CPMG element can more successfully refocus the magnetisation, leading to a higher observed intensity in the NMR spectrum (Figure 2.20B). In the limit that $\nu_{CPMG} \rightarrow \infty$, the observed relaxation rate during the total CPMG duration will approach the intrinsic relaxation rate R_2, with any exchange contribution removed. Practically, suppose there is a substantial increase in signal intensity of a peak between the application of a low CPMG frequency and high CPMG frequency. In that case, this shows evidence of chemical exchange. The CPMG experiment can detect chemical exchange between species in the fast exchange regime, where $k_{ex} > \Delta\nu$, where

$k_{ex} = k_1 + k_{-1}$, with k_1 and k_{-1} being the forward and backward exchange rates between two species, and $\Delta\nu$ is the frequency difference between the two states in Hz. However, due to hardware constraints, ν_{CPMG} is not usually increased above 2000 s^{-1}.

2.6.3 The Nuclear Overhauser Effect and the Nuclear Overhauser Effect Spectroscopy (NOESY) experiment

The nuclear Overhauser effect involves the cross-relaxation of magnetisation from one spin I to another S through the dipolar interaction between the spins I and S. The following section is based on the explanations of the NOESY experiment given by J. Keeler.[87]

If the magnetisation of I and S are placed on the negative z-axis, their equation of motion as a function of time will be:

$$\frac{d}{dt}\begin{pmatrix} I_z \\ S_z \end{pmatrix} = \begin{pmatrix} -R_{1,I} & -\sigma_{IS} \\ -\sigma_{IS} & -R_{1,S} \end{pmatrix}\begin{pmatrix} I_z - I_{eq} \\ S_z - S_{eq} \end{pmatrix} \tag{2.40}$$

where σ_{IS} represents the cross-relaxation rate between I_z and S_z which is given by equation 2.41 for two dipolar coupled spins a distance r_{IS} from one another.

$$\sigma_{IS} = \frac{d^2}{10}\left(6J(\omega_I + \omega_S, \tau_c) - J(\omega_I - \omega_S, \tau_c)\right) \qquad d = -\frac{\mu_0 \gamma_I \gamma_S \hbar}{4\pi |r_{IS}|^3} \tag{2.41}$$

In the homonuclear case (where $\omega_I \approx \omega_S$), for small τ_c values σ_{IS} is positive and for large τ_c values σ_{IS} is negative. In a NOESY experiment[120] (Figure 2.21A), the spins I/S (Figure 2.21B/C) have the following form before the mixing time (τ):

$$M_{\text{tot}}(t_1) = -I_z \cos(\omega_I t_1) - S_z \cos(\omega_I t_1) \tag{2.42}$$

In the absence of cross-relaxation between the two spins I/S, the NOESY spectrum will contain two peaks along the diagonal with coordinates (ω_I, ω_I) and (ω_S, ω_S) (Figure 2.21C). The initial rate of magnetisation transfer (where $I_z = -I_{eq}\cos(\omega_I t_1)$, $S_z = -S_{eq}\cos(\omega_I t_1)$) in the system can be represented as:

$$\frac{d}{dt}\begin{pmatrix} I_z \\ S_z \end{pmatrix} = \begin{pmatrix} -R_{1,I} & -\sigma_{IS} \\ -\sigma_{IS} & -R_{1,S} \end{pmatrix} \begin{pmatrix} -I_{eq}[1+\cos(\omega_I t_1)] \\ -S_{eq}[1+\cos(\omega_I t_1)] \end{pmatrix} \tag{2.43}$$

Assuming this rate to be constant over the mixing time (τ), the I_z magnetisation can be calculated using the following integral:

$$\begin{aligned} I_z(\tau) - I_z(0) &= \int_0^\tau [R_{1,I}(1+\cos(\omega_I t_1))I_{eq} + \sigma_{IS}(1+\cos(\omega_S t_1))S_{eq}]\, dt \\ I_z(\tau) + \cos(\omega_I t_1)I_{eq} &= [R_{1,I}(1+\cos(\omega_I t_1))I_{eq} + \sigma_{IS}(1+\cos(\omega_S t_1))S_{eq}]\,\tau \\ I_z(\tau) &= \cos(\omega_I t_1)[R_{1,I}\tau - 1]I_{eq} \\ &\quad + \sigma_{IS}\cos(\omega_S t_1)\tau S_{eq} \\ &\quad + [R_{1,I}\tau I_{eq} + \sigma_{IS}\tau S_{eq}] \end{aligned} \tag{2.44}$$

The third term of the final expression in equation 2.44 has no dependence on t_1 and will be ignored because phase cycling in the NOESY experiment can be used to remove it. The expression for the magnetisation after the mixing time in Figure 2.21C/D assumes no longitudinal relaxation during the mixing period, valid in the regime where $R_1\tau << 1$.

In the case of I and S spins cross relaxing during the mixing time τ in the NOESY experiment, as well as the diagonal peaks shown in Figure 2.21C, there will also be cross peaks at the coordinates (ω_I, ω_S) and (ω_S, ω_I) (Figure 2.21D). Furthermore, if $\sigma_{IS} > 0$ (for small τ_c values), then the cross peaks will have the opposite phase to

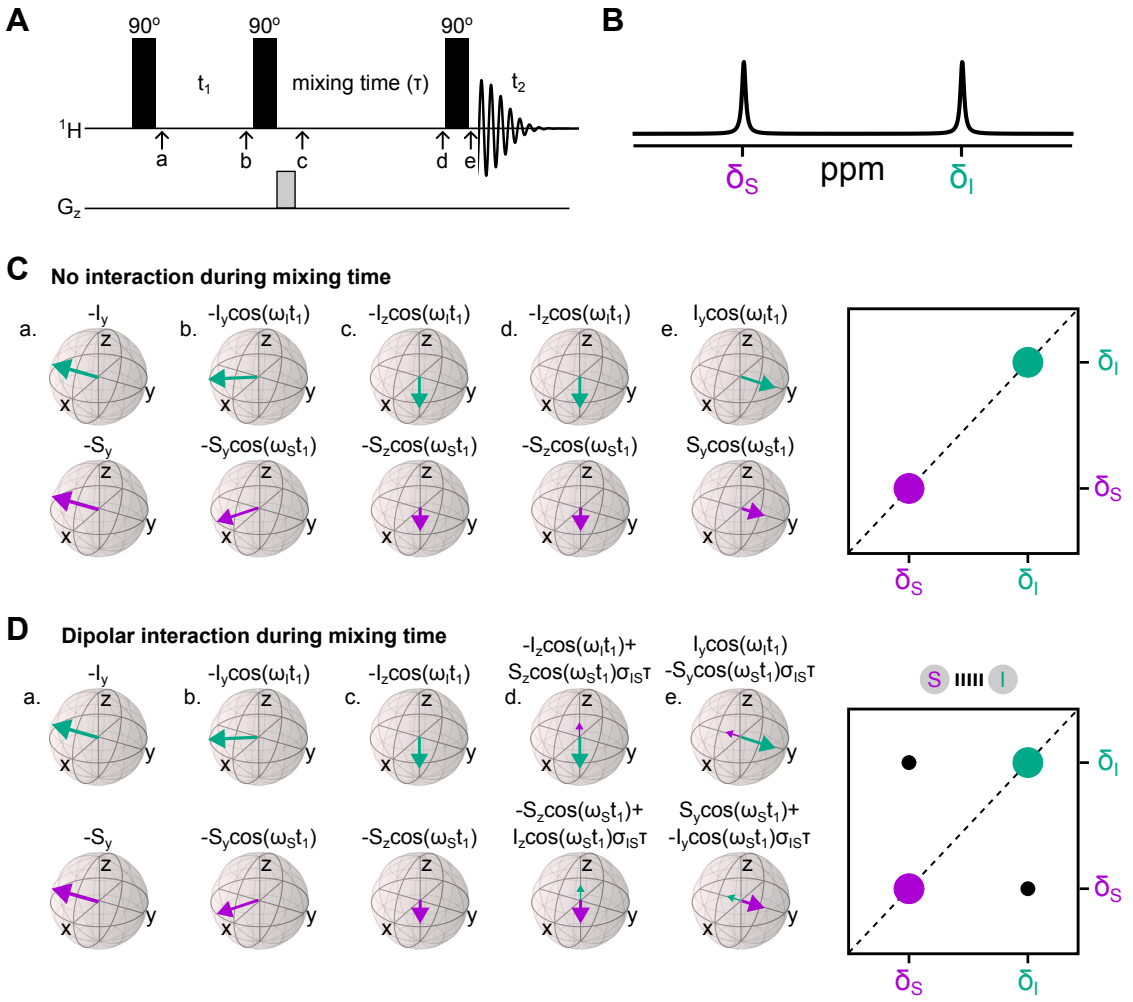


Figure 2.21: An introduction to the Nuclear Overhauser Effect and the NOESY experiment. **A** An example of a ^{1}H-^{1}H NOESY sequence with a mixing time given by τ. **B** An example of a 1D proton spectrum of two spins (I/S). **C** A vector model description of the NOESY experiment without interaction between I and S during the mixing time. For clarity, it is assumed that longitudinal relaxation of I/S is not present during the mixing time. The resulting NOESY spectrum contains two peaks lying on the diagonal. **D** A vector model description of the NOESY experiment where interactions between I and S occur during the mixing time. For clarity, it is assumed that longitudinal relaxation of I/S is not present during the mixing time. The resulting NOESY spectrum contains an additional two cross peaks.

the diagonal peaks. In contrast, if $\sigma_{IS} < 0$ (for large τ_c values), the cross peaks will be of the same sign as the diagonal peaks.

The NOESY experiment has been instrumental in biological NMR; indeed, the first solution state NMR protein 3D structure was produced using ^{1}H-^{1}H distance restraints from NOESY experiments.[27]

2.7 NMR experiments used throughout this thesis

In chapter 3 a variety of NMR experiments are used to study the solvent environment of biomolecular condensates. A combination 1D and 2D experiments are used to measure metabolite partitioning into condensates. Following this, an assortment of NMR experiments introduced in this chapter (gradient imaging, slice selection, relaxation, diffusion, CEST and NOESY) are used to study the physicochemical environment of water inside a model biomolecular condensate. In chapter 4, a host of 3D NMR backbone assignment experiments are used to assess the secondary structure propensity in a disordered protein, followed by CPMG relaxation dispersion and CEST experiments with the aim of identifying chemical exchange within the protein. Finally, within chapter 5, a combination of multidimensional NMR and relaxation experiments are used to quantify the relaxation properties within a $^{13}CF_3$ spin system.

3 Exploration of the solvent environment in biomolecular condensates

3.1 Abstract

Biomolecular condensates are hypothesised to offer an alternative solvent environment in cells, challenging the idea that water is the universal solvent of biology. Solvents are able to interact with and solvate compounds, leading to selective partitioning. Accurate measurement of molecular partitioning into condensates will enhance understanding of their biological function and aid condensate targeted drug design. An NMR-based method was used to measure the partitioning of amino acids, nucleotides and ions into condensates. Trends governing metabolite solubilities were obtained. The solubility rules for each phase-separating system were similar to the concept of "like dissolves like", used extensively to predict solubility in organic solvents. The solvent environment provided by condensates is then explored from the perspective of water. The high concentration of organic matter in the condensate changed the electronic environment of water to a similar extent as organic solvents. This was driven by interactions between aliphatic protons and the water molecules. A 2.5 to 3 fold reduction in the translational and rotational diffusion rate of water was observed in the condensate phase, highlighting an increased effective viscosity. These findings support the growing evidence of an alternative solvent environment inside biomolecular condensates.

3.2 Introduction to biomolecular condensates

Biomolecular condensates, also known as membraneless organelles, are biological assemblies separated from their surrounding environment by surface tension alone[121,122] (Figure 3.1A). They can form reversibly in living cells in response to stimuli.[121] Commonly, they are produced through liquid-liquid phase separation, forming a dense liquid phase containing high concentrations of biopolymers, such as proteins and nucleic acids[43,123] (Figure 3.1B). They can provide cellular compartmentalization, despite their lack of a conventional lipid membrane, and are associated with gene regulation,[124] nucleic acid repair,[125,126] cellular signaling,[123] viral assembly[127] and immune responses.[128,129]

The first observations of a cellular condensate was by Felix Fontana in 1781 with a spot observed in the centre of the nucleus of an eel epithelial cell.[130,131] This was followed in 1835 by descriptions of an exact round body inside the nucleus of a germ cell, later termed the nucleolus.[132–134] This pre-dates the discovery of many well-known membrane-bound organelles, such as the mitochondria and Golgi apparatus.[135,136] However, at the time, the nucleolus's composition, method of formation and function was unknown. Studies in the late 19$^{\text{th}}$ century showed nucleoli were able to fuse and grow, indicating they had liquid-like properties[137] (Figure 3.1C). Furthermore, studies on the protoplasm, an obsolete term for the intracellular medium, showed evidence of additional colloidal “alveoli" containing liquids.[138–140] This was followed by the identification of the Cajal body in 1903, another type of condensate found in cellular nuclei.[141] The emergence of multiple types of condensates within cells indicated their importance and varied functions. In the 1930s, Alexandr I. Oparin stressed the importance of coacervates (multi-component condensates) in biology, where “organic matter became concentrated at different points of the aqueous medium” (in a similar manner to cells).[142,143] This suggested that coacervate formation may have been a crucial stepping stone in the

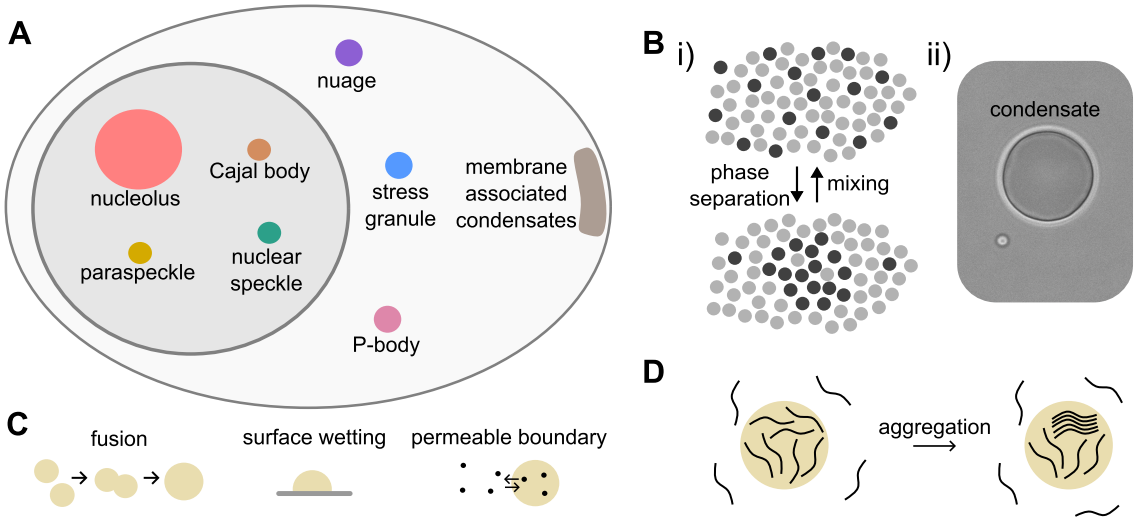


Figure 3.1: An introduction to biomolecular condensates and liquid-liquid phase separation. **A** Examples of biomolecular condensates found within Eukaryotic cells, exhibiting diverse compositions, sizes and locations.[121,144] Figure inspired by Banani, S. F. *et al.*[121] **B** i) An illustration of a two-component system undertaking liquid-liquid phase separation, the method of formation of liquid condensates. ii) A differential interference contrast (DIC) image of an *in vitro* condensate formed from the intrinsically disordered protein (IDP) DDX4^{N1}. **C** Depictions of several defining properties of liquid condensates, separated from the surrounding environment by surface tension alone. **D** Graphical representation of protein aggregation mediated by the formation of liquid condensates.

"evolution of the primary organic substance and in the process of autogeneration of life".[142,143]

More recently, the study of biomolecular condensates has become a rapidly expanding field, with a growing consensus on their vital importance for cellular function. Further evidence has been obtained of their ability to undergo fusion, dripping and wetting in cells, similar to droplets of water. This emphasises the liquid nature of many condensates, with their boundary to the surrounding environment provided solely by surface tension[145–147] (Figure 3.1C). Furthermore, condensates can readily exchange molecules such as proteins with their external environment[148,149] (Figure 3.1C).

A vast quantity of different condensates with differing composition and functions are now known to exist within eukaryotic cells,[121,124,150] with their presence in bacterial cells becoming more evident.[151] It is estimated that around 20% of the

human proteome is functionally involved in biological phase separation[152] and biomolecular condensates have been located in all sections of eukaryotic cells, including the nucleus and cytoplasm, and in contact with lipid membranes[121,150] (Figure 3.1A). Furthermore, it is possible for multi-component phase separating systems to form nested condensates[153] (condensates within condensates) with the nucleolus[154,155] being a prime example.

Membrane-associated condensates are an emerging phenomenon made possible by the wetting ability of liquid condensates. They have been suggested as essential to the formation of tight junctions, allowing cells to be joined together to form tissues.[156,157] Furthermore, structural changes to neuronal postsynaptic densities, a type of membrane-associated condensate, are connected to memory and learning.[158–161]

The link between biomolecular condensates and pathological aggregation causing diseases such as amyotrophic lateral sclerosis (ALS),[162,163] Parkinson's disease[164] and Alzheimer's Disease (AD)[165,166] is becoming apparent. Aggregation of proteins in condensates affects the reversibility of condensate formation,[163] thought to prevent natural cellular function. In addition, recent experiments have shown the potential link between condensate formation reversibility and neurodevelopmental diseases such as autism spectrum disorder (ASD).[167] Condensates also have an ever-emerging link to certain types of cancer, with condensates playing a role in chromatin re-organisation, DNA damage repair, and transcriptional regulation.[168–170]

A model to describe cellular condensates includes scaffold macromolecules essential to their formation, but in addition, client molecules such as other proteins, nucleic acids and metabolites.[121] It is hypothesised that the high concentration of scaffold macromolecules in condensates creates an alternative solvent environment, enabling the selective partitioning of small molecules,[48,49,171,172] ions,[173–175] proteins,[44,176–182] nucleic acids[44,177,183,184] and phospholipids[185] (Figure 3.2iii).

Biomolecular condensates have been shown to cause changes to protein/DNA

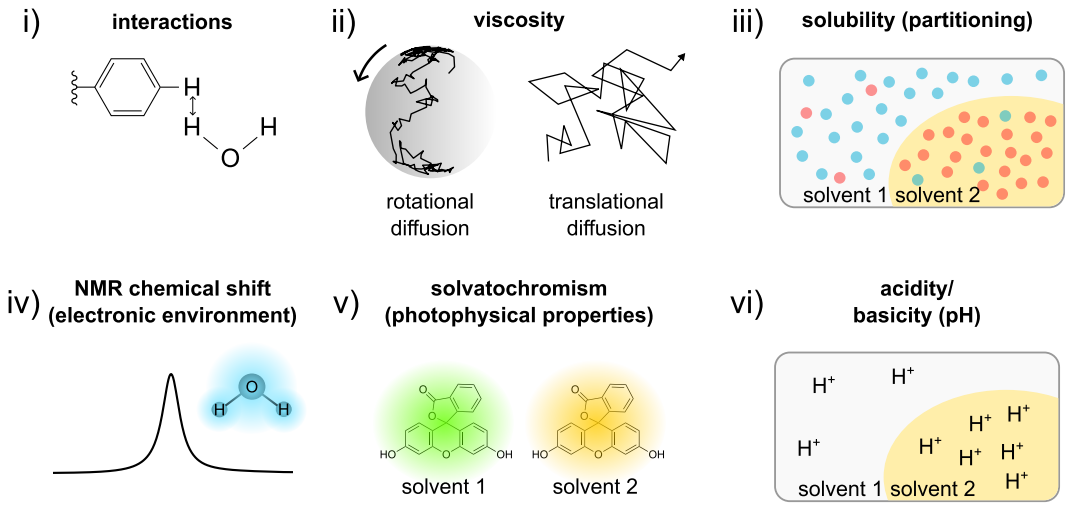


Figure 3.2: A selection of solvent dependent properties. (i) Interactions between molecules are known to be highly solvent dependent, affecting (ii) internal viscosity and rotational/translational diffusion, (iii) the solubility of client molecules, (iv) the local electronic environment of client molecules which can be probed using NMR, (v) photo-physical properties such as electronic energy levels, and finally (vi) the internal acidity/basicity of the system.

folding,[46,177] as well as alter enzymatic activity;[46,186–188] structural rearrangements are likewise observed when biomolecules are dissolved in organic solvents.[189] Dielectic constants (also known as the relative permittivity) describe how well a solvent can screen the Coulomb force between charged particles. A highly researched condensate (DDX4^{N1}) was shown to provide a different internal dielectric constant to water, and instead had a value similar to dimethyl sulfoxide (DMSO) or acetonitrile.[44] Moreover, Raman spectroscopy has been used to show that the material hydrogen bonding properties of water can be altered inside FUS (Fused in Sarcoma) and tau (tubulin associated unit) condensates, with a reduced proportion of tetrahedrally bound water in the condensate phase.[47] It has also been suggested that additional water depletion inside condensates is linked to pathological protein aggregation.[190] Recent studies also point to the emergence of pH gradients produced by biomolecular condensates, with potential links to catalysis[155,175,186,191] (Figure 3.2vi).

Further understanding of solvent properties in biomolecular condensates (Figure 3.2) is essential for determining the extent of their functions inside cells. It would allow the partitioning of drugs for diseases associated with condensates

to be predicted/engineered[48,171] as well as the creation of bespoke drug delivery systems.[192,193] In addition, partitioning inside condensates has been proposed as a method for de-pollution of water.[194]

The liquid and dynamic nature of condensates makes NMR a well-suited method to study them. In this chapter, following a description of liquid-liquid phase separation (section 3.2.1), two *in vitro* phase separating systems, DDX4^{N1} and EWSR1C, are introduced (sections 3.3.1 and 3.3.2). The solvent-dependent properties of these model condensate systems are then explored using a range of NMR based techniques. An NMR-based method is introduced to determine the partitioning of common cellular metabolites and ions into biomolecular condensates (section 3.4). The condensate partitioning results allowed a set of solubility rules for different condensates to be established and compared to those in organic solvents (section 3.4.5). Furthermore, the molecular partitioning rules in each system could be correlated with the solvent itself. Following this, NMR experiments on a biphasic DDX4^{N1} sample are used to assess differences in solvent dependent molecular properties between phases, such as chemical shift, translational/rotational diffusion, as well as intermolecular interactions of water and other condensate components (Figure 3.5). Evidence provided in this chapter adds to the growing literature that biomolecular condensates can provide an alternative solvent environment.[44–49] This challenges the idea that water is the universal solvent in biology. Furthermore, the solubility rules obtained provide a framework for the rational molecular design of drugs targeting biomolecular condensates/solvents.

3.2.1 Liquid-liquid phase separation

The ability of proteins to phase separate was established in the 19th century.[195,196] In 1911, it was observed that phase separation of a mixture of gelatin and gum arabic could be reversed by the addition of salt,[197] indicating buffer composition has an important role in phase separation. Phase separation of this manner was

later termed coacervation in 1930.[198] Complex coacervation, a term to describe phase separation of oppositely charged polymers, was later observed for mixtures of gelatin, gum arabic and nucleic acids. When mixed, all three formed a triphasic system containing coacervates formed within coacervates.[153] In the mid-20$^{\text{th}}$ century, the first descriptions calling the nucleolus a biological coacervate were provided,[199,200] despite being observed over 100 years earlier.[132–134] It was shown to have a temperature-dependent morphology, with its size decreasing as the temperature was raised.[199]

A physical explanation of polymer phase separation was provided by Flory and Huggins in 1942, now known as Flory-Huggins theory.[201,202] The theory treats polymers as a series of statistical segments whose orientations are independent from each other. From the theory, an expression for the free energy of mixing (ΔG_{mix}) of a two-component system (e.g. water and protein) is obtained:*

$$\Delta G_{mix} = \chi\phi(1-\phi) + RT\left(\frac{\phi}{r}\ln(\phi) + (1-\phi)\ln(1-\phi)\right) \tag{3.1}$$

where χ is the interaction parameter between the polymer and solvent, ϕ is the volume fraction of polymer in the system, and r is the number of statistical segments per polymer. R and T represent the molar ideal gas constant and system temperature, respectively. The second term in equation 3.1 represents the configurational entropy change upon mixing and will always favour a system to mix by lowering the free energy. The parameter χ contains information about molecular enthalpy and entropy changes upon mixing ($\chi = \Delta H - T\Delta S$), with strong polymer-polymer interactions stabilising the phase separated state. Often in biological systems, energetic changes upon phase separation are modest in order to remain highly dependent on local stimuli.[122] A plot of ΔG_{mix} against

*Derived in appendix A.14.

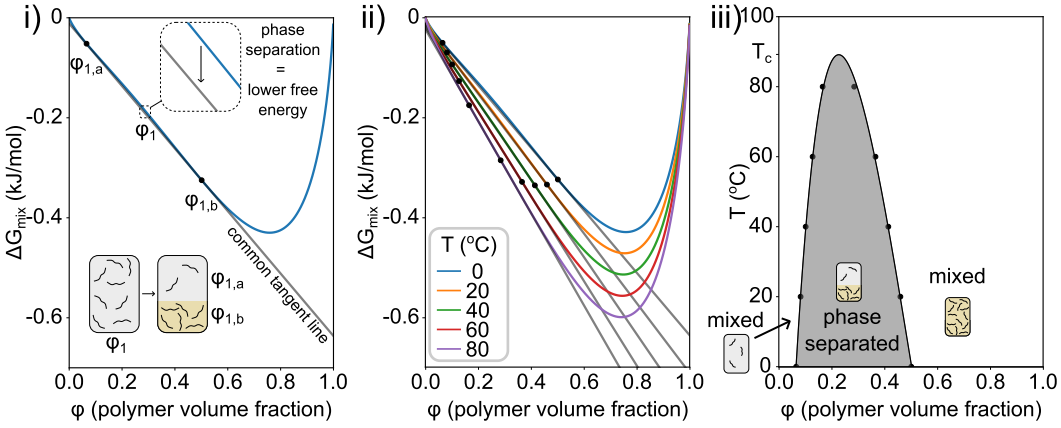


Figure 3.3: An introduction to Flory-Huggins polymer phase separation. (i) A plot of ΔG_{mix} as a function of polymer volume fraction for a binary mixture at 0^oC, where each polymer has r=10 and χ(kJ/mol) = 0.8 + 0.005T(K). The value of χ was chosen to represent a system whose phase separation propensities are modest, as is the case for biological condensate formation which can often be reversed through minor changes to the system. At polymer volume fractions between $\phi_{1,a}$ and $\phi_{1,b}$ (defined by the common tangent line), the system will spontaneously phase separate into two phases containing polymer volume fractions of $\phi_{1,a}$ and $\phi_{1,b}$ in order to reduce the free energy of the system. While the free energy of the dilute phase containing a polymer volume fraction $\phi_{1,a}$ is higher than for the mixed state at ϕ_1, the total free energy of the system is reduced when separating into the two phases. (ii) ΔG_{mix} for the same polymer solution as in (i) at a range of temperatures. The numerically determined common tangent lines (and binodal points) at each temperature are highlighted. (iii) The phase diagram for the polymer solution (with parameters as described above) showing the binodal line defining regions of miscibility for the system. When the temperature is above the critical temperature (T_c), the system remains mixed across all polymer volume fractions.

polymer volume fraction ϕ is shown for hypothetical polymer in solution with r=10 statistical segments and χ(kJ/mol) = 0.8 + 0.005T(K) at a temperature of 0^oC (Figure 3.3i). A system prepared to contain a polymer volume fraction of ϕ_1 will spontaneously phase separate into two phases containing polymer volume fractions of $\phi_{1,a}$ and $\phi_{1,b}$ in order to reduce its free energy (Figure 3.3i). The binodal points of $\phi_{1,a}$ and $\phi_{1,b}$ (defining the regions of instability) were determined numerically by finding two points lying on the same tangent line of ΔG_{mix}. The minimal concentration of protein required for phase separation, termed the saturation concentration (c_{sat}), corresponds to the protein volume fraction of $\phi_{1,a}$. Often, the saturation concentration is used as a proxy for phase separation propensity of a

given polymer, with a smaller value of c_{sat} indicating a higher driving force for phase separation.

The process of finding binodal points (points with a common tangent line) can be repeated at multiple temperatures (Figure 3.3ii) to obtain a phase diagram (Figure 3.3iii) showing regions of miscibility and immiscibility for the system as a function of temperature. The critical temperature ($\mathrm{T_c}$) defines the temperature at which the system is miscible across all polymer volume fractions.

For a given polymer solution, there is no exact analytical solution to determine the binodal points of instability. Instead, numerical approaches to find points with a common tangent line are applied. Recently, analytical approximations to determine the binodal line for a given system have been obtained, but numerical methods are still the preferred option.[203]

Experimental phase diagrams have been produced for a series of *in vitro* models of condensates, with their binodal lines obtained as a function of various physiological properties such as polymer concentration, temperature, ionic strength, pH, pressure and solute concentrations.[44,204–206] By numerically fitting an experimentally determined phase diagram to the polymer solution theory described above, estimates of ΔH, ΔS and N can be obtained for each system, where N is the number of statistical segments per polymer.[44,177] Recently, the determination of biological phase diagrams has been advanced by a rapid microfluidics-based method.[207] By understanding the thermodynamic and molecular contributions to phase separation, it may be possible to enhance or destabilise the process. However, biomolecular condensates are multi-component systems containing multiple proteins, nucleic acids, and salts, and therefore, modelling them as a binary (solvent-polymer) system may not capture their complete behaviour. A multi-component Flory-Huggins theory enables a more complete solution.[208] Recently, the inclusion of salt and ions into the theory has become essential to explain their emerging ability to partition ions.[173]

3.2.2 Interactions driving phase separation

In order for phase separation of a polymer solution to occur, polymer-polymer and water-water interactions are required to outcompete the polymer-water interactions, increasing the Flory-Huggins parameter ($\chi = \Delta H - T\Delta S$) enabling the entropic tendency to mix to be overcome. There are a range of possible polymer-polymer interactions in biological molecules, including cation-anion, dipole-dipole, cation-π, π-π and hydrophobic interactions (Figure 3.4).[44,209] This is in line with many phase-separating IDPs being enriched in amino acids with polar (glutamine, asparagine, serine, threonine), positively charged (arginine, lysine), negatively charged (aspartic acid, glutamic acid), or aromatic (phenylalanine, tyrosine, tryptophan) side-chains.[209,210] Many biomolecular condensates are formed from intrinsically disordered proteins (IDPs) or proteins containing substantial intrinsically disordered regions (IDRs).[211–213] Disorder enhances the number of available contacts for interactions and may be why disordered proteins are more prevalent in biomolecular condensates. However, the exact pattern of amino acids in a given protein sequence has been shown to be essential to its phase separation propensity.[44,210] Furthermore, phase separating biological systems are thought to be minimally stable, with their interactions finely balanced, enabling their phase separation abilities to be regulated through post-translational modifications such as arginine methylation,[44,214] phosphorylation[215] and glycosylation.[216]

Finally, while there is no exact formula to determine if a protein can phase separate, several machine learning models have been developed to predict if a given protein sequence will be able to phase separate.[217–223] It is hoped that these predictive models can identify new protein sequences capable of phase separation with novel properties.

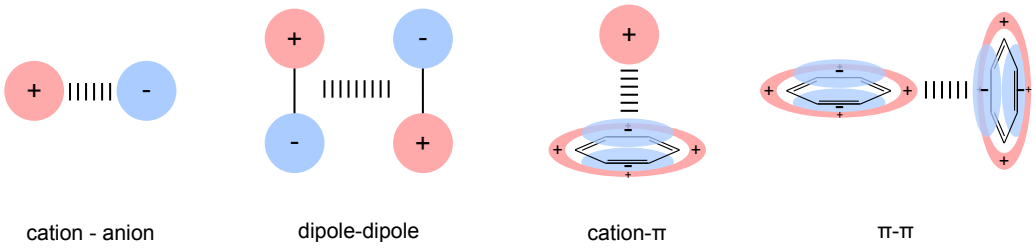


Figure 3.4: Depictions of various inter/intra-molecular interactions capable of stabilizing a phase separated biomolecular condensate.

3.3 *In vitro* phase separating systems

In order to study the solvent properties of biomolecular condensates, the use of two *in vitro* model phase separating systems, DDX4^{N1} (section 3.3.1) and EWSR1C (section 3.3.2), were used. These model systems have differing properties, with DDX4^{N1} being negatively charged, and EWSR1C positively charged at pH 7-8, allowing a comparison of two orthogonal phase separating systems. Furthermore, the EWSR1C (with the addition of polyadenosine DNA (polyA)) and DDX4^{N1} systems are able to phase separate under physiological conditions, unlike other proteins which require the addition of synthetic crowding agents such as polyethylene glycol (PEG) or Dextran (DEX) to induce phase separation.[207,224] Molecular crowding has been shown to change the partitioning, concentrations, and properties of biomolecular condensates.[225] Therefore, the absence of unnatural crowding agents in the *in vitro* phase separation of DDX4^{N1} and EWSR1C enables a more accurate mimic of their intracellular solvent environment.

3.3.1 The DDX4^{N1} model system

3.3.1.i Introduction to DDX4

Germ granules are phase separated bodies present in germ cells, capable of developing into reproductive gametes.[226–228] There are many types of germ granules across a variety of organisms including nuage/chromatoid bodies found in mammals, P-granules in worms, and pole plasm in other non-mammalian species.[144,226,229]

Male germ cells are capable of forming sperm cells through spermatogenesis and germ granules have known functions relating to post-transcriptional mRNA regulation.[226,229,230] Furthermore, germ granules are thought to aid the suppression of somatic differentiation of germ cells, preserving the germline totipotency of an organism.[231]

DDX4, also known as Vasa (Drosophila) or Mouse Vasa Homologue (Mvh, mouse), is a primary scaffold constituent of germ granules, containing a DEAD(Asp-Glu-Ala-Asp)-box ATP dependent RNA helicase flanked by two disordered regions (Figure 3.5Ai).[44,229,232,233] DDX4 is crucial for spermatogenesis. For example, DDX4 knockout in mice (Mvh) has been shown to cause male infertility.[226,234] Germ granules and DDX4 condensates have a strong tendency to absorb RNA from the surroundings, enhancing their ability to sequester and modify gene expression.[44,177,232] Previously, DDX4 was thought to only be expressed in germ cells, however, recent studies have shown that cancerous somatic (non-reproductive) cells are capable of DDX4 expression.[235,236] In cancer cells, DDX4 forms cytoplasmic granules that resemble germ granules, influencing gene expression and promoting tumor growth.[236] Furthermore, the presence of DDX4 in cancerous cells is linked to resistance against anti-cancer drugs.[237] For this reason, DDX4 and its resulting condensates have become a target for cancer therapeutics.[235,236]

Interestingly, $DDX4^{N1}$ (the disordered N-terminus (1-236) of DDX4) maintains the ability to phase separate both *in vitro* and *in vivo* (Figure 3.5Ai).[44] Therefore, $DDX4^{N1}$ is thought to provide the structural architecture of germ granule condensates.[44,226]

3.3.1.ii $DDX4^{N1}$ phase separation behaviour

Interactions driving phase separation in $DDX4^{N1}$ $DDX4^{N1}$ has distinct patterns of charge across its protein sequence (Figure 3.5Aii), highly important to its phase separation properties.[44] A charged scrambled sequence of $DDX4^{N1}$,

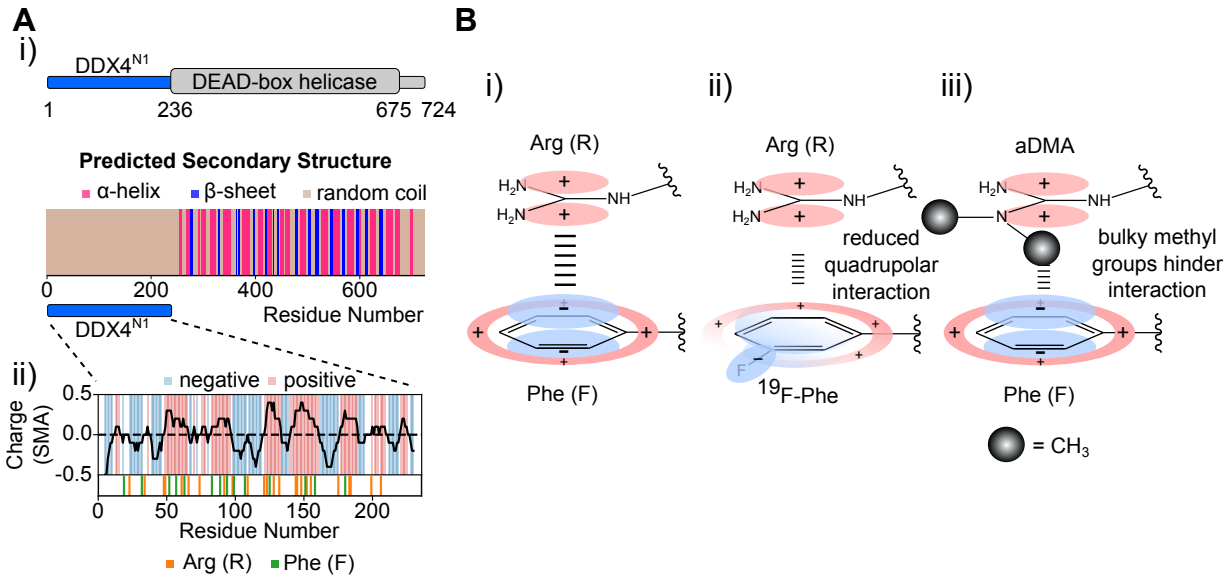


Figure 3.5: An overview of the sequence properties and interactions in the phase-separating protein DDX4^{N1}. **A** i) Schematic of the human protein DDX4, a scaffold component of germ granule condensates, highlighting the disordered N-terminus (DDX4^{N1}). The predicted secondary structure propensity is shown, calculated using AlphaFold.[5,238] ii) A simple moving average (SMA) of the charge across the DDX4^{N1} protein sequence as a function of residue number. A window size of 10 residues was used, and regions of positive and negative charge are represented in red and blue, respectively. Adapted from Nott *et al.*[44]. iii) A plot showing the locations of all arginine (R) and phenylalanine (F) residues within the DDX4^{N1} sequence. **B** i) Depiction of the cation-π interactions between arginine (R) and phenylalanine (F) residues driving phase separation of DDX4^{N1}. ii) Introduction of 3-fluorophenylalanine in place of phenylalanine reduces the quadrupole on the aromatic ring, hindering cation-π interactions and reducing phase separation propensity.[44] iii) Replacement of arginine with asymmetric dimethyl arginine (aDMA) reduces phase separation ability due to the presence of bulky methyl groups disrupting the short-ranged cation-π interactions.[44,209]

where the charged residues in the sequence are scrambled to minimise the presence of blocks of positive/negative charge, is unable to phase separate.[44] This suggests the importance of electrostatic interactions in DDX4^{N1} condensate formation. Furthermore, DDX4^{N1} is rich in arginine and phenylalanine residues (Figure 3.5Aii), enabling cation-π interactions and driving phase separation (Figure 3.5Bi). The net charge on DDX4^{N1} at pH 8.0 is -5.5.[239] Therefore, the attractive interactions favouring phase separation are strong enough to overcome the repulsive electrostatic interactions between monomers. Arginine methylation

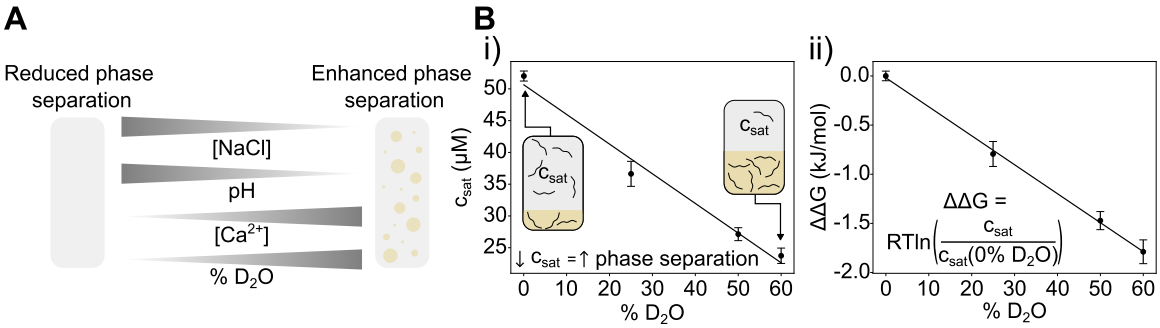


Figure 3.6: Factors affecting DDX4^{N1} phase separation. **A** A schematic of how properties of the buffer affect the phase separation propensity of DDX4^{N1}. Increasing the salt concentration or pH of the buffer reduces phase separation,[44,45] whereas increasing the percentage of D_2O enhances phase separation. **B** i) A plot showing how the saturation concentration (c_{sat}) of DDX4^{N1} in the dilute phase is altered by changing the percentage of D_2O in the buffer. The solutions were prepared to contain a total of 112 μM DDX4^{N1} in a solution containing 20 mM NaPi, 100 mM NaCl, 5 mM TCEP at a pH of 8.0. ii) A plot highlighting the linear dependence of the effect of D_2O buffer percentage on the free energy of phase separation ($\Delta\Delta G$).

and quadrupole reduction of phenylalanine through fluorination decreases the intermolecular interactions between DDX4^{N1} units and reduces its phase separation ability, emphasizing the importance of cation-π interactions for DDX4^{N1} condensate formation (Figure 3.5Bii/iii).[44]

Dependence on ions and pH DDX4^{N1} phase separation has a strong dependence on the concentration of NaCl in the buffer, with a low concentration favouring phase separation and a high concentration hindering it[44,45] (Figure 3.6A). The reduced phase separation ability with increasing salt is hypothesised to be caused by the screening ability of ions in solution, diminishing the attractive electrostatic and cation-π interactions between monomers[44,209] (Figure 3.5). In contrast, the phase separation propensity of DDX4^{N1} is increased upon raising the concentration of multivalent cations such as Ca^{2+} (Figure 3.6A).[240] NMR chemical shift perturbations (CSPs) have indicated the presence of weak binding of Ca^{2+} ions to negatively charged amino acids in the DDX4^{N1} sequence, lowering the overall net negative charge on the protein, reducing electrostatic repulsion between monomers and enhancing phase separation.[240] Furthermore, DDX4^{N1} phase separation is

favoured when the pH of the solution is decreased (Figure 3.6A). This can be explained by the fact that a lower pH reduces the net negative charge on DDX4^{N1} (charge = -5.5 at pH 8.0),[239] thereby minimising electrostatic repulsion between protein chains.

Dependence on D_2O In addition to the strong dependence of DDX4^{N1} phase separation on salt, multivalent cations and pH, the presence of D_2O can also significantly impact its phase separation ability (Figure 3.6B). An increase in the percentage of D_2O in the buffer causes a marked decrease in c_{sat}, thereby indicating enhanced phase separation (Figure 3.6B). Details on the methods and results for D_2O dependence of DDX4^{N1} condensate formation can be found in appendix A.3.1.

Biological phase separation in other systems has also been shown to be enhanced by the addition of small amount of D_2O.[241] The pronounced effect of D_2O percentage on phase separation hints at the role of hydrogen bonding, with D_2O having stronger solvent-solvent interactions than H_2O caused by the lower vibrational energy of an O-D bond.[241,242] This is evident in the increased boiling point (101.42^{o}C)[243] and specific heat capacity (4.301 kJ kg^{-1} K^{-1} at 298 K)[243] of D_2O compared to H_2O (boiling point: 100^{o}C, specific heat capacity: 4.18 kJ kg^{-1} K^{-1} at 298 K).[244] The increased solvent-solvent interactions in D_2O have been hypothesized to lead to stronger hydrophobic interactions between protein chains.[241] Together, these effects increase the Flory-Huggins interaction parameter (equation 3.1), and enhance phase separation.

Biological condensate formation is carefully balanced, so a change as simple as deuteration can be significant. However, this enhancement in phase separation upon addition of D_2O has been utilized for experiments discussed in the remainder of this chapter to increase the volume of condensate phase produced. Therefore, by understanding the factors affecting the phase separation of the model system DDX4^{N1}, the solvent properties of DDX4^{N1} condensates could be studied using a smaller quantity of protein.

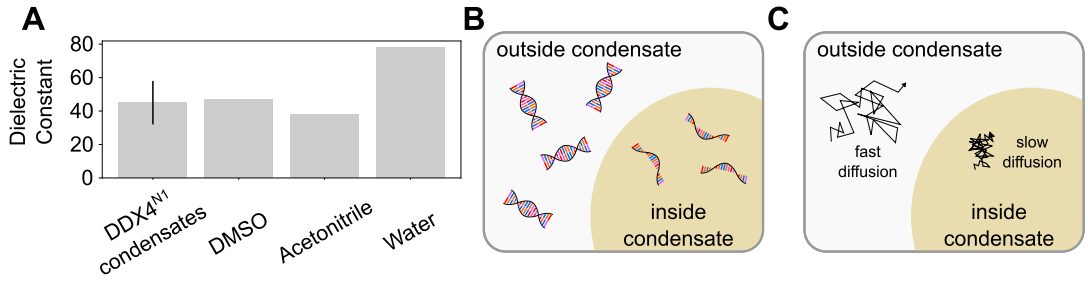


Figure 3.7: Previous evidence of an alternative solvent environment in DDX4^{N1} condensates. **A** The estimated dielectric constant inside DDX4^{N1} condensates is more comparable to the values for the organic solvents DMSO and acetronitrile than it is to pure water.[44] **B** Double-stranded DNA is destabilized inside DDX4^{N1} condensates and unravels into single strands.[177] **C** NMR-based measurements on DDX4^{N1} protein units and small molecules show that translational diffusion is significantly reduced inside the DDX4^{N1} condensate phase.[45]

3.3.1.iii Current evidence of an alternative solvent environment in DDX4^{N1} condensates

DDX4^{N1} has been used extensively as a model system to study the properties of biomolecular condensates, with emerging evidence of its ability to provide an alternative solvent environment. The dielectric constant (relative permittivity) estimations within DDX4^{N1} condensates, obtained from the scaling of the Flory-Huggins parameter with salt, are more akin to organic solvents such as DMSO and acetonitrile than pure water[44] (Figure 3.7A). Furthermore, Förster Resonance Energy Transfer (FRET) measurements have been used to show that DDX4^{N1} condensates have the ability to 'melt' double stranded DNA into its single-stranded counterpart[177] (Figure 3.7B). This unusual ability is hypothesised to be possible due to the unique solvent environment inside the condensates, thereby solvating and stabilising the single stranded system.[177] Finally, NMR-based diffusion measurements on DDX4^{N1} condensates show that the translational diffusion of DDX4^{N1} protein chains and small molecules are noticeably reduced in the condensed phase[45] (Figure 3.7C). These diffusion measurements highlight an apparent change in the viscosity inside the condensates, a key property of a solvent (Figure 3.2ii).

3.3.2 The EWSR1C model system

3.3.2.i Introduction to EWSR1

The protein EWSR1 (Ewing Sarcoma RNA binding protein 1, from the gene EWS breakpoint region 1) is a member of the FET family of proteins. Other members of the family, which are paralogs of one another, include FUS (Fused in Sarcoma) and TAF15 (TATA-box binding protein associated factor 15), with the name FET coming from the first letter of FUS, EWSR1 and TAF15[245–247] (Figure 3.8A). The FET proteins are a set of RNA binding proteins similar in structure and function, found in almost every tissue across a vast range of organisms.[247] They all have RNA recognising motifs (RRMs), allowing them to bind RNA/DNA, significantly impacting mRNA transcription[245,246,248] (Figure 3.8A). Furthermore, their common 'prion-like' and RGG regions are predicted to be highly disordered (Figure 3.8B). EWSR1 is essential for normal cellular function. Indeed, EWSR1 gene knockout in mice results in significant organism defects, including reduced mitochondria function[249] and impaired neuron function.[250] EWSR1 deficient zebrafish embryos produce abnormal mitotic spindles, causing cell death (apoptosis) in the developing central nervous system.[251] Furthermore, meiosis is hindered by the removal of the EWSR1 gene and results in sterility.[252]

In 1921, Dr. James Ewing described the presence of a 'diffuse endothelioma of the bone', now termed Ewing sarcoma, a rare cancer that occurs in children and young adults.[253] The disease predominantly occurs when the EWSR1 gene from chromosome 22 becomes translocated with the Friend Leukaemia Virus Integration 1 gene (FLVI1) from chromosome 11.[247] This translocation causes the expression of EWSR1/FLVI1 fusion protein shown to induce mitotic defects in Henrietta Lacks (HeLa) cell models leading to malignant tumour development.[254] Additionally, translocation of EWSR1 with the transcription factor Wilms tumour gene (WT1), leading to the expression of EWSR1-WT1, is associated with Desmoplastic

small-round-cell tumour (DSRCT), another form of rare and aggressive cancer.[255] The high prevalence of cancers associated with translocation of other genes onto EWSR1, disrupting the expression of wild-type EWSR1, further supports the high importance of EWSR1 for normal cellular function.

When DNA becomes damaged, poly(ADP-ribose) units (PAR) are recruited to the damaged regions.[256] It has been shown that the RGG (Arg-Gly-Gly) repeats of EWSR1 (primarily located in the C-terminus) are highly involved with PAR sensing[257] (Figure 3.8C). It is hypothesised that the ability of PAR to localise EWSR1 increases the local concentration enough to allow it to undertake liquid-liquid phase separation at sites of DNA damage, potentially crucial for the DNA repair process.[257] Indeed, *in vitro* studies of EWSR1 have shown its ability to phase separate under physiological conditions,[210,257] which are thought to be driven by arginine-tyrosine interactions.[210]

The FET proteins are primarily located in the nucleus. However, mislocalisation of FET proteins, including EWSR1, into the cytoplasm of motor neurons, followed by aggregation, is linked with amyotrophic lateral sclerosis (ALS).[258] The FET proteins contain prion-like domains (PLDs) that are likely important for their aggregation behaviour (Figure 3.8A). It is well known that FUS and TAF15 are localised to stress granules under cellular stress and exist as client proteins,[259] despite their ability to phase separate *in vitro*.[210] However, EWSR1 localisation to stress granules is not observed to the same extent,[259] despite EWSR1 overexpression being able to induce similar neurodegeneration as FUS and TAF15 in *Drosophila* models.[258] Therefore, the role of EWSR1 in neurodegeneration and ALS remains elusive.

The C-terminus of human EWSR1 (450-655), termed EWSR1C, is predicted to be highly disordered and used as a model phase separating protein throughout the remainder of this chapter (Figure 3.8Di). It has a charge of +20.6 at pH 8.0,[239] with the charge distribution across the sequence shown in Figure 3.8Dii. Despite the

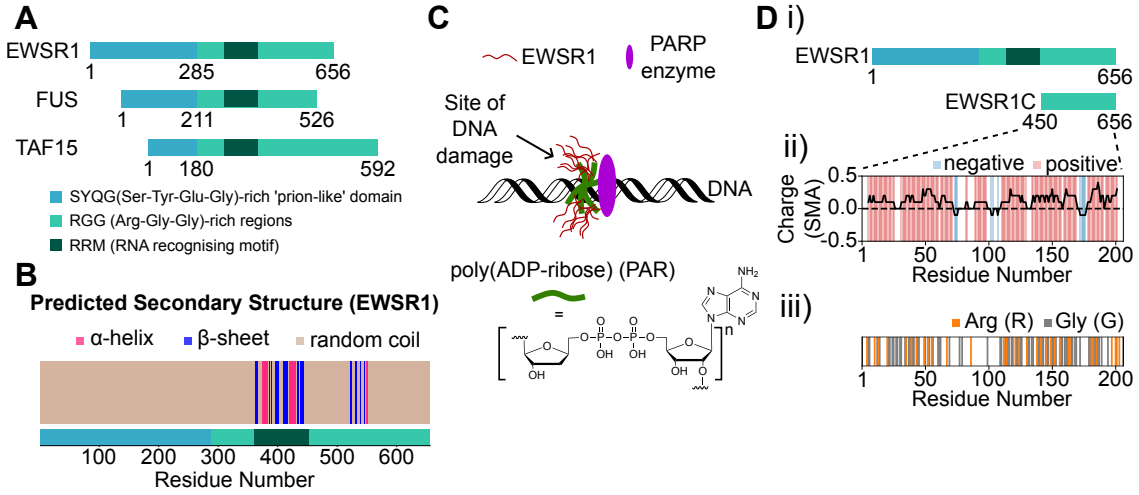


Figure 3.8: An introduction to the phase separating proteins EWSR1 and EWSR1C. **A** An overview of the FET family proteins (FUS, EWSR1 and TAF15) showing their similarity in sequence domains (adapted from Altmeyer *et al.*[257]). **B** The predicted secondary structure of the full-length EWSR1 protein as a function of residue number. The secondary structure was predicted using AlphaFold.[5,238] All predicted α-helical and β-sheet secondary structure is located in the RRM (RNA recognising motif), with all other regions predicted to be significantly disordered. **C** A schematic showing EWSR1 localisation to a site of DNA damage. Nucleic acid-mimicking poly(ADP-ribose) (PAR) is recruited to sites of DNA damage through the PARP enzyme.[257] It has been shown that the RGG-rich regions of EWSR1 interact strongly with PAR, leading to localisation and liquid-liquid phase separation of EWSR1 at the sites of DNA damage.[257] **D** i) A graphic highlighting the C-terminus of EWSR1 (EWSR1C) used as a model phase separating protein in this chapter. ii) A plot of the simple moving average (SMA) charge across the sequence of EWSR1C. A window size of 10 residues was used, and regions of positive and negative charge are represented in red and blue, respectively. The total charge on EWSR1C at pH 8.0 is +20.6.[239] iii) A plot showing arginine and glycine residues within the EWSR1C sequence.

lack of the prion-like domain (PLD), EWSR1C remains highly biologically relevant given the sequence is rich in RGG regions (Figure 3.8Diii), known to interact with PAR units on damaged DNA. Furthermore, there are several sequence mutations associated aggregation-based diseases such as ALS located in the C-terminus of EWSR1 (EWSR1C).[258]

3.3.2.ii EWSR1C phase separation behaviour

The following section discusses the phase separation behaviour of EWSR1C as a function of salt, pH, D_2O and nucleic acids (Figure 3.9A).

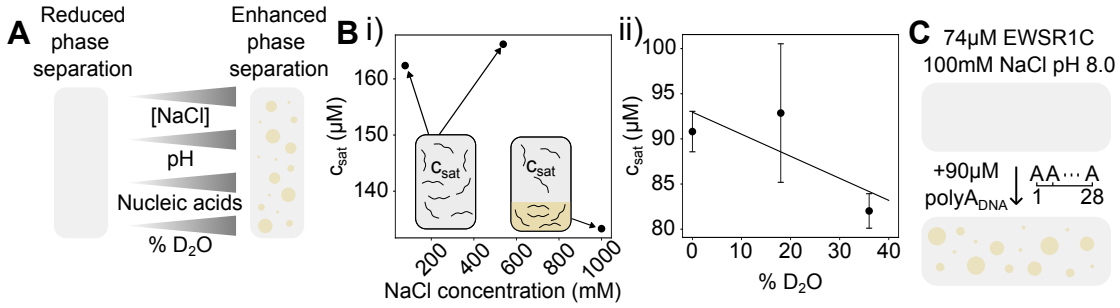


Figure 3.9: Factors affecting EWSR1C phase separation. **A** Schematic showing the factors affecting phase separation of EWSR1C. **B** i) The saturation concentration (c_{sat}) of EWSR1C as a function of NaCl concentration. The EWSR1C concentration was 165 μM and no phase separation was observed for NaCl concentrations less than 550 mM. Note that only one measurement was made at each concentration of NaCl. ii) The saturation concentration (c_{sat}) as a function of buffer D_2O percentage. The EWSR1C concentration was 154 μM with a salt concentration of 1.5M and a buffer pH of 8.0. **C** An illustration showing how the addition of polyA DNA can induce phase separation of EWSR1C.

Dependence on salt The salt dependence of EWSR1C phase separation was measured using the protocol described in appendix A.3.2. High concentrations of NaCl (>550 mM) were required to induce phase separation of EWSR1C (Figure 3.9Bi) at pH 8.0. The high concentration of salt is likely required to screen the repulsive interactions between highly positive protein chains (+20.6). The NaCl dependence of EWSR1C phase separation is in contrast with $\mathrm{DDX4^{N1}}$ (Figure 3.6A), whose phase separation propensity decreases with increasing NaCl concentration. This indicates that $\mathrm{DDX4^{N1}}$ and EWSR1C phase separation occur under different mechanisms.

In contrast, the addition of 1 M KCl to EWSR1 (full length) reduces its phase separation propensity,[210] showing that the phase separation properties of the C-terminal domain of EWSR1 (EWSR1C) are different to that of the full length protein. This altered phase separation dependence on ionic strength is likely attributed to the reduced positive charge of full length EWSR1 (+13.8 at pH 8.0).[239]

Dependence on D_2O The dependence of EWSR1C phase separation on D_2O percentage is shown in Figure 3.9Bii (experimental details and results in appendix A.3.3). The phase separation propensity of EWSR1C shows a small enhancement upon increase in D_2O percentage. This is in contrast to $DDX4^{N1}$, whose phase separation ability is increased drastically upon addition of D_2O (Figure 3.6B). The reason for the difference between the two systems remains elusive, however, it implies that the phase separation abilities in each system have different dependences on hydrogen bonding.

Addition of nucleic acids It was difficult to find a buffer condition to get both $DDX4^{N1}$ and EWSR1C to phase separate, with EWSR1C only being able to phase separate at higher, and less biologically relevant, salt concentrations. However, it is desirable to compare the solvent properties of orthogonal condensate systems placed in buffers of similar composition (NaCl concentration, pH).

To address this issue, phase separation of EWSR1C could be induced under the same buffer compositions as $DDX4^{N1}$ (300 mM NaCl, 20 mM sodium phosphate, 5 mM TCEP, pH 8.0) by the addition of 60 μM of 28mer polyadenosine (polyA) DNA (Figure 3.9C). Nucleic acids have high concentrations inside a cell (approximately 7% by mass[260]); therefore, addition of polyA to induce phase separation was justifiable. PolyA tails of RNA are routinely attached to mRNA molecules to prevent degradation and influence transcription.[261] DNA, rather than RNA, based polyA was used to induce phase separation of EWSR1C due to the fact that RNA (even polyA) has poor stability *in vitro* and is prone to degradation by common RNase enzymes meaning special RNase free buffers would be required.[262] Furthermore, there is a striking similarity of polyA DNA to PAR, known interact with the RGG regions of EWSR1, localising it to sites of DNA damage[257] (Figure 3.8C). Given arginine (R) and glycine (G) make up 50 % of EWSR1C, it is highly plausible that EWSR1C is able to also co-phase separate with PAR.

3.4 Partitioning into biomolecular condensates and organic solvents

Results and methodologies in this section are heavily drawn from Eaton *et al.*[263] (in preparation).

3.4.1 Introduction to molecular partitioning in solvents

Solvent extraction has been used since ancient civilisations to extract and purify molecules for medicinal or cosmetic purposes.[264] The International Union of Pure and Applied Chemistry (IUPAC) definition of solvent extraction is "the process of transferring a substance from any matrix to an appropriate liquid phase".[265] By this definition, the partitioning of client molecules between a cellular condensate and its surrounding environment, causing either enrichment or depletion, is equivalent to solvent extraction (Figure 3.10A).

In the 1800s, scientists began to quantify how molecules are partitioned into organic solvents.[266,267] Following this, extensive study of molecular partitioning into organic solvents has been performed,[268,269] with linear free energy relationships (LFERs) used to compare partitioning between different solvents.[268] The natural logarithm of the partitioning of a molecule into octanol (logP) is used extensively in modern medicinal chemistry, with a "rule of thumb" stating that a drug candidate is likely to be more successful if it has a logP value <5.[270] For a drug candidate to be successful using traditional formulation and administrative techniques (e.g. orally and intravenously) it must be hydrophilic enough to dissolve into blood but hydrophobic enough to pass through the lipid-based cell membrane. Furthermore, a host of computational tools have been produced to predict the partitioning of a molecule into 1-octanol based on its chemical structure, such as ALOGPS[271] and ChemAxon.

Insight into the solvent properties governing molecular partitioning measurements

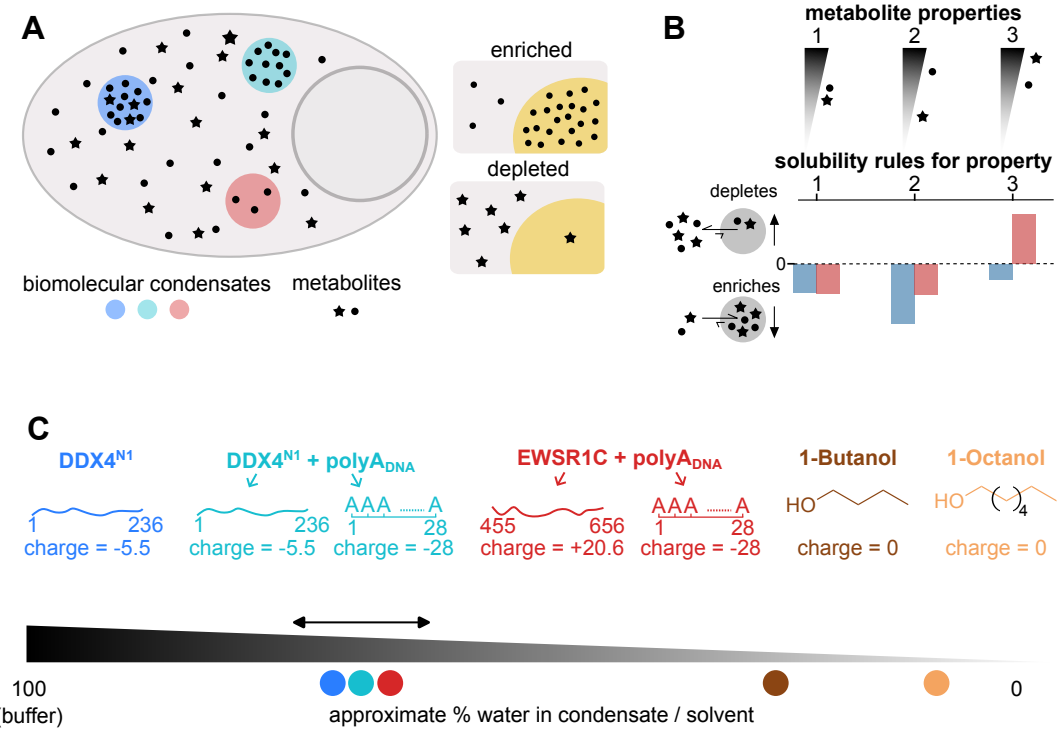


Figure 3.10: An introduction to condensate partitioning, solubility rules and example phase-separating systems (condensates/organic solvents). **A** An illustration of metabolite partitioning in cellular condensates, with a condensate able to selectively enrich or deplete molecules. **B** A schematic outlining how establishing a set of solubility rules (based on metabolite molecular properties) can rationalise metabolite partitioning behaviour. **C** An introduction to the condensate and organic solvent systems whose metabolite solubility rules will be determined and compared in this section. The charges were calculated at pH 8.0.[239] The approximate percentage of water in the condensate/organic solvent phases are shown for comparison.

in polymer/polymer aqueous two-phase systems (ATPS), formed through liquid-liquid phase separation like biomolecular condensates, have also been obtained.[272,273]

This section introduces an NMR-based method to quantify the partitioning of metabolites into biomolecular condensates followed by the determination of the solubility rules governing metabolite enrichment or depletion in each of the phase-separating systems (Figure 3.10B). The partitioning method was used to quantify metabolite partitioning into DDX4^{N1}, DDX4^{N1}+polyA and EWSR1C+polyA condensates at pH 8.0, where DDX4^{N1} and EWSR1C having vastly opposing charges (Figure 3.10C). The total concentration of metabolites inside a cell can reach 200 mM (in *Escherichia coli*), with amino acids and nucleotides contributing

approximately 50% and 15% to the total, respectively.[274] Therefore, understanding the rules governing amino acid and nucleotide partitioning into biomolecular condensates is extremely desirable.

3.4.1.i Previous partitioning measurements into biomolecular condensates

Recently, the partitioning of small molecules (drugs and a selection of metabolites) into biomolecular condensates has been explored significantly using fluorescence and mass spectrometry methods.[48,49,171,185] While these methods are high throughput, the results obtained from these contrasting techniques do not align perfectly.[48] In addition, fluorescence based techniques require a target molecule to be naturally fluorescent, or for a bulky fluorescent tag to be added; thereby changing the chemical properties of the molecule. Furthermore, the quantum mechanical properties of a fluorescent probe are not necessarily identical between a dilute and condensate phases. Solvatochromic shifts could significantly hinder the interpretation of fluorescence intensity-based partitioning methods.

Previous mass spectrometry methods have reported instability of nucleotide phosphates (such as ATP) during analysis, preventing the quantification of their partitioning into biomolecular condensates.[48] Protein digestion before mass spectrometry analysis also prevented partitioning quantification for the majority of amino acids.[48]

Therefore, an orthogonal NMR-based method to determine partitioning into biomolecular condensates is introduced in this section to quantify the partitioning of amino acids and nucleotide metabolites into biomolecular condensates.

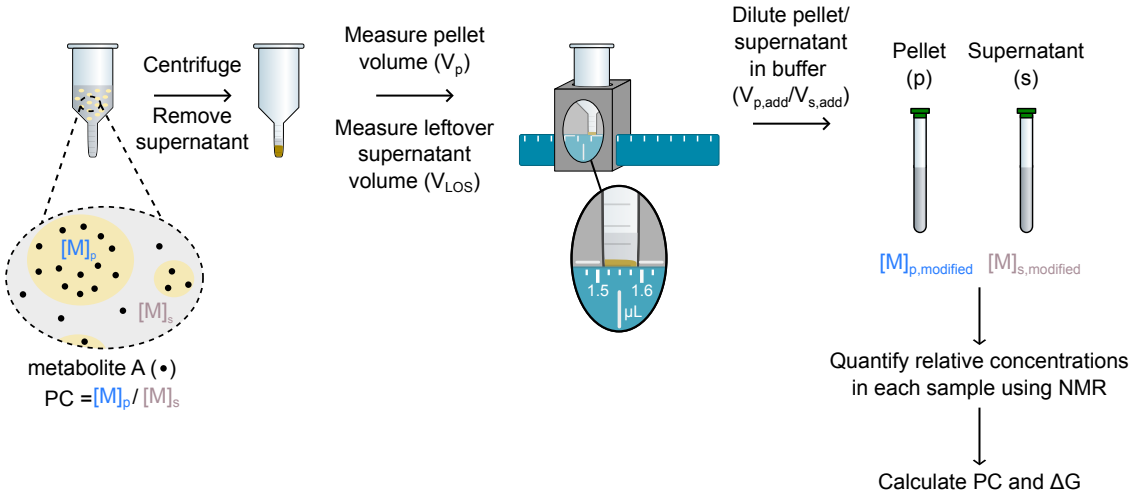


Figure 3.11: The biomolecular condensate partitioning method. A schematic of the procedure used to separate the supernatant and pellet phases from one another and quantify the volume of the pellet phase (and leftover supernatant) using the 'easy-read' pellet measuring device. The supernatant and pellet phases are then diluted in buffer, producing NMR samples from the pellet (p) and supernatant (s) with modified metabolite concentrations $[M]_p$ and $[M]_s$, respectively.

3.4.2 NMR-based method to determine partitioning coefficients

Previous NMR-based methods to determine condensate partitioning have been performed on a biphasic sample of tau,[224] similar to that discussed in section 3.5. However, as shown in Figure 3.36, the intensity of signal observed from metabolites such as cGMP in this manner is small, hindering the exact quantification of partitioning coefficients. Instead, a separation based approach followed by NMR quantification was used.

Traditionally in organic chemistry, partitioning into organic solvents is measured using the "shake-flask" method,[275] whereby a compound is placed into a water-octanol solution followed by extensive shaking. The concentration ratio of a molecule between the phases is then determined, often using spectrophotometry.[276]

The following outlines an overview of the method used to quantify metabolite

partitioning coefficients (PC) into the biomolecular condensates (Figure 3.11A)*. A detailed description of the method is included in appendix method A.4.1. Similar to the "shake-flask" method, a non-phase separated solution of $DDX4^{N1}$/ EWSR1C (20 mM sodium phosphate, 3.5 mM TCEP, pH 8.0, H_2O) was mixed in a 1:2 ratio with no salt buffer (14 mM sodium phosphate, 3.5 mM TCEP, pH 8.0, 70% D_2O) also containing the amino acid/nucleotide metabolites +/- polyA to induce phase separation in packed column volume (PCV) tubes (Figure 3.11)†. After extensive mixing, the samples were then kept at 4^oC for 1 hour before centrifugation. The centrifuge speed of 1500×g was applied for 10 minutes. The speed of 1500×g was chosen as it was the minimum speed that would consistently enable a pellet to form at the bottom of the PCV tubes. This is significantly lower than the 14,000×g used in Ambadi-Thody *et al.*[48] to separate condensate and dilute phases to measure small molecule partitioning. The partitioning coefficient of cGMP into $DDX4^{N1}$ condensates was shown to be within error when increasing the centrifuge speed to 2500×g ($PC_{2500\times g} = 3.29 \pm 0.05$, $PC_{1500\times g} = 3.6 \pm 0.5$). However, the lower speed of 1500×g was chosen for the method in case the independence of measured partitioning coefficient with centrifuge speed was not general across all molecules. The volumes of the biomolecular condensate pellets (V_p) were between 1-5 μL (appendix A.9). Therefore, measuring their volumes quantitatively required the use of the "easy-read" volume measuring device (Figure 3.11). Then, after removing and storing 100 μL (V_i) of supernatant, as much supernatant as possible was removed from the PCV tube and discarded, taking care not to disturb the pellet phase. It was not possible to completely remove all of the supernatant phase, therefore the remaining amount (leftover supernatant V_{LOS}) was also recorded. The pellet (+ leftover supernatant) was then be resuspended using a 250μL ($V_{add,p}$) of

*This method was developed as an equal collaboration between myself and Virgínia Casablancas-Antràs (Baldwin lab).

†PCV tubes are commonly used to measure small cell pellet volumes accurately and were repurposed in this method to measure condensate pellet volumes. Smokers, I. B. A., *et al.*[277] have also utilised PCV tubes to measure condensate volumes.

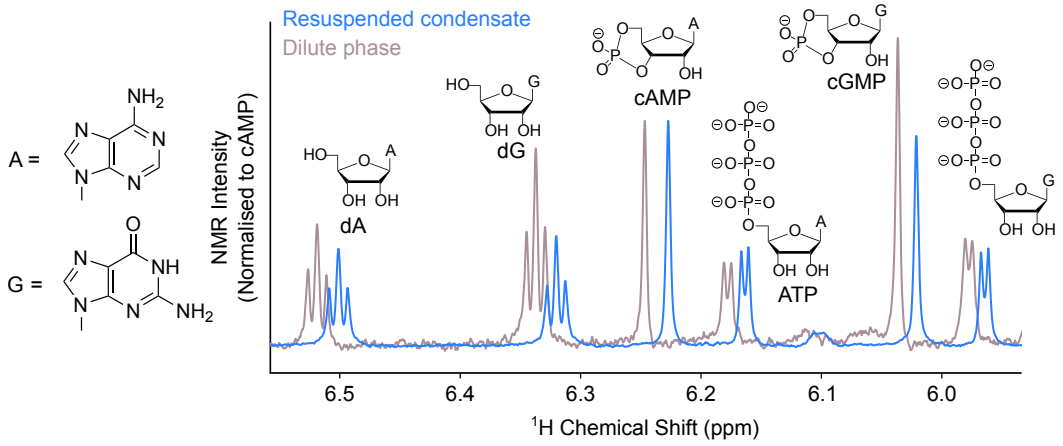


Figure 3.12: Quantification of nucleotide partitioning using proton NMR. The proton NMR spectra of the 1' nucleotide sugar protons obtained from the supernatant (s) resuspended condensate/pellet (p) NMR samples formed from a $DDX4^{N1}$ condensate sample. The spectra are normalised to the intensity of cAMP to show relative trends in partitioning between metabolites. The dilute phase spectrum has been offset for improved clarity.

buffer and transferred to a 3 mm NMR tube. An additional 150μL ($V_{add,s}$) of buffer was added to the stored supernatant (V_i) extracted previously, followed by transfer to a 3 mm NMR tube.

The NMR signal intensity ratio (per unit scan) between the samples gives a signal factor (S) for each metabolite, ensuring the signal intensities are normalised per unit scan. For the nucleotides, the 1' sugar protons from water-suppressed proton NMR spectra were used to calculate the signal factors (Figure 3.12)*. Details on the complexities of determining the signal factor for the amino acids are discussed the next section 3.4.2.i. Once the signal factors were quantified, the partitioning coefficients (PC) were calculated using the following equation (derived in appendix A.7.1):

$$PC = \frac{V_i\left(V_p + V_{LOS} + V_{add,p}\right)}{SV_p\left(V_i + V_{add,s}\right)} - \frac{V_{LOS}}{V_p} \tag{3.2}$$

*The supernatant and pellet NMR samples produced were in very similar buffer conditions (same NaCl concentration, pH) to enable the same T_1 relaxation times for each species in each phase. This enabled the NMR relaxation delays to be shortened whilst remaining quantitative when taking ratios to calculate the signal factor (S).

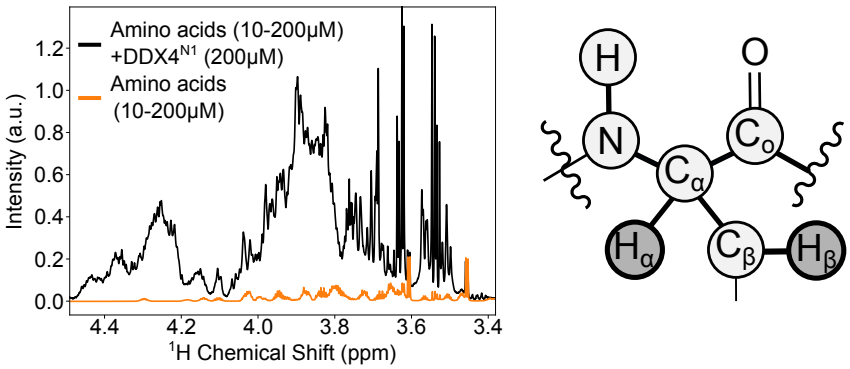


Figure 3.13: Protein-bound amino acid background obscures free amino acid resonances in proton NMR spectra. 1D Proton NMR spectra comparison of a sample containing free amino acids (10-200 μM) + DDX4^{N1} (200 μM) sample, and a second sample containing only the free amino acids. The spectrum is zoomed in on the region containing H_α and H_β resonances.

Following this, the free energy of partitioning ($\Delta G_{partitioning}$) for each system could be obtained using $\Delta G_{partitioning} = -RT \ln(PC)$.

3.4.2.i NMR quantification of relative amino acid concentrations between dilute and condensate phases

Unlike the nucleotides which have proton resonances that do not overlap with any background protein or buffer proton peaks (Figure 3.11B), free amino acids contain very similar proton resonances as the amino acids contained within the protein. Therefore, due to the significant proportion of protein (DDX4^{N1} or EWSR1C) in the supernatant and resuspended condensate NMR samples (Figure 3.11A),* the free amino proton signals are overwhelmed by signals from the protein (Figure 3.13).

To address this issue, ^{13}C, ^{15}N isotopically enriched free amino acids were used in the condensate partitioning experiments, thereby allowing multidimensional NMR experiments (section 2.3) to be used to filter out background signals from unlabelled protein. The HACACO sequence[278] has been used previously to remove background protein signal and observe free amino acid resonances within cells.[280] Here, the original sequence has been optimised through the addition of C_α, C_β

*The protein concentration in the condensate systems can be up to 15 mM.

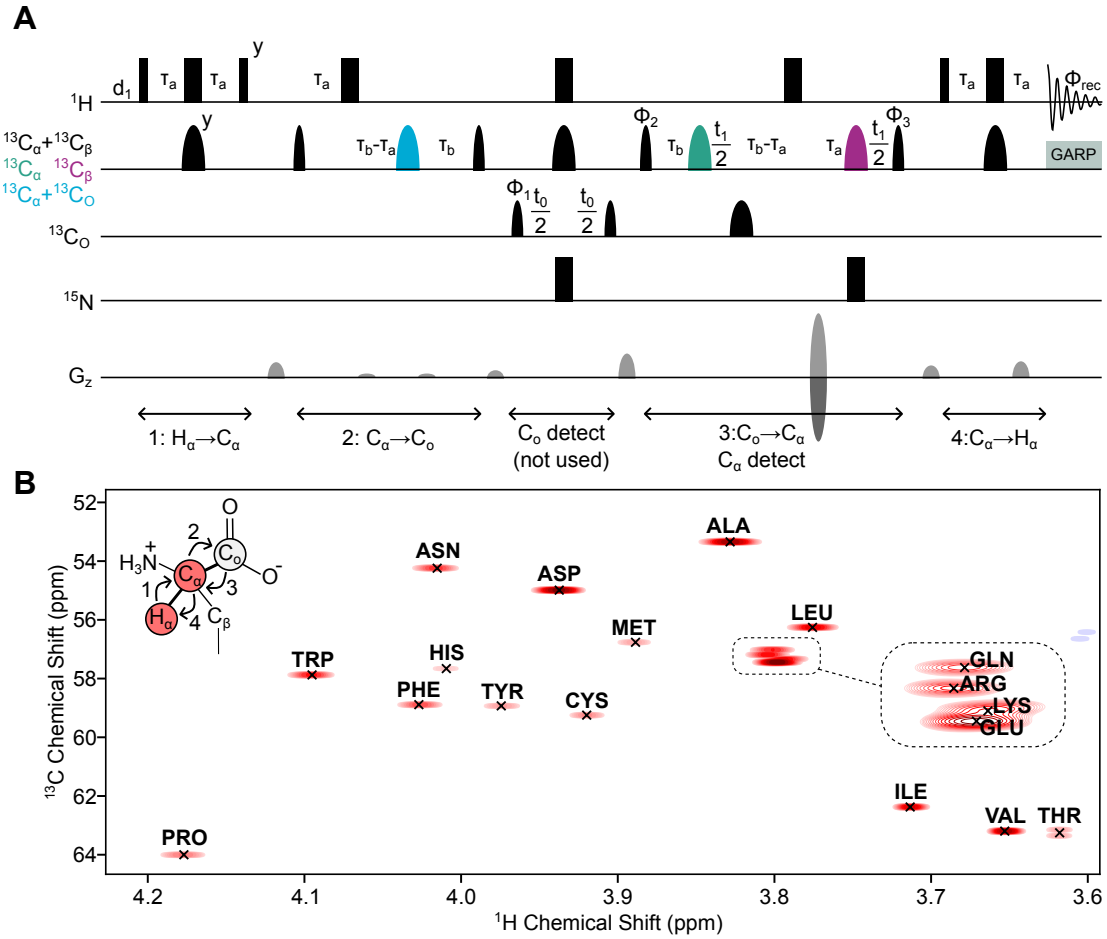


Figure 3.14: Using a HACA(CO) pulse sequence to isolate free amino acid resonances from protein-bound amino acids. **A** The HACA(CO) pulse sequence, adapted from Kay *et al.*[278] to contain optimal control selective C_α (green, 2000 μs), C_β (purple, 5000 μs), C_α+C_β (black, 200 μs) and C_α+C_o (blue, 2000 μs) pulses calculated using Seedless[279] as well as C_β decoupling. The phase cycle used was as follows: Φ_1=(x,-x), Φ_2=(x,x,-x,-x), Φ_3=(x,x,x,x,-x,-x,-x,-x), Φ_{rec}=(x,-x,-x,x,-x,x,x,-x). Delays of $\tau_a = 1/(4J_{HC_\alpha})$ and $\tau_b = 1/(4J_{C_\alpha C_o})$ were used, with values of 1.7 ms and 4.5 ms, respectively. The delays t_0 and t_1 represent the indirect dimension delays for $^{13}C_o$ and $^{13}C_\alpha$, respectively. Proton and nitrogen decoupling 180^o pulses were applied during both indirect dimension acquisitions. Quadrature in the $^{13}C_\alpha$ dimension was obtained in a Rance-Kay (Echo-AntiEcho) manner with TPPI. During proton acquisition, ^{13}C GARP decoupling was applied at a frequency of 22.5 kHz. Further experimental details can be found in appendix A.4.3. **B** An assignment of the H_α-C_α correlations obtained using the pulse sequence in A on a ^{13}C,^{15}N isotopically labelled amino acid mix. The TOCSY-based assignment spectra are shown in appendix A.10.

and C_o selective optimal control pulses, calculated using Seedless[279] (Figure 3.14)*.

*The implementation of the pulse sequence containing optimal control pulses was performed as an equal collaboration between myself and Dr. Charles Buchanan (Baldwin lab).

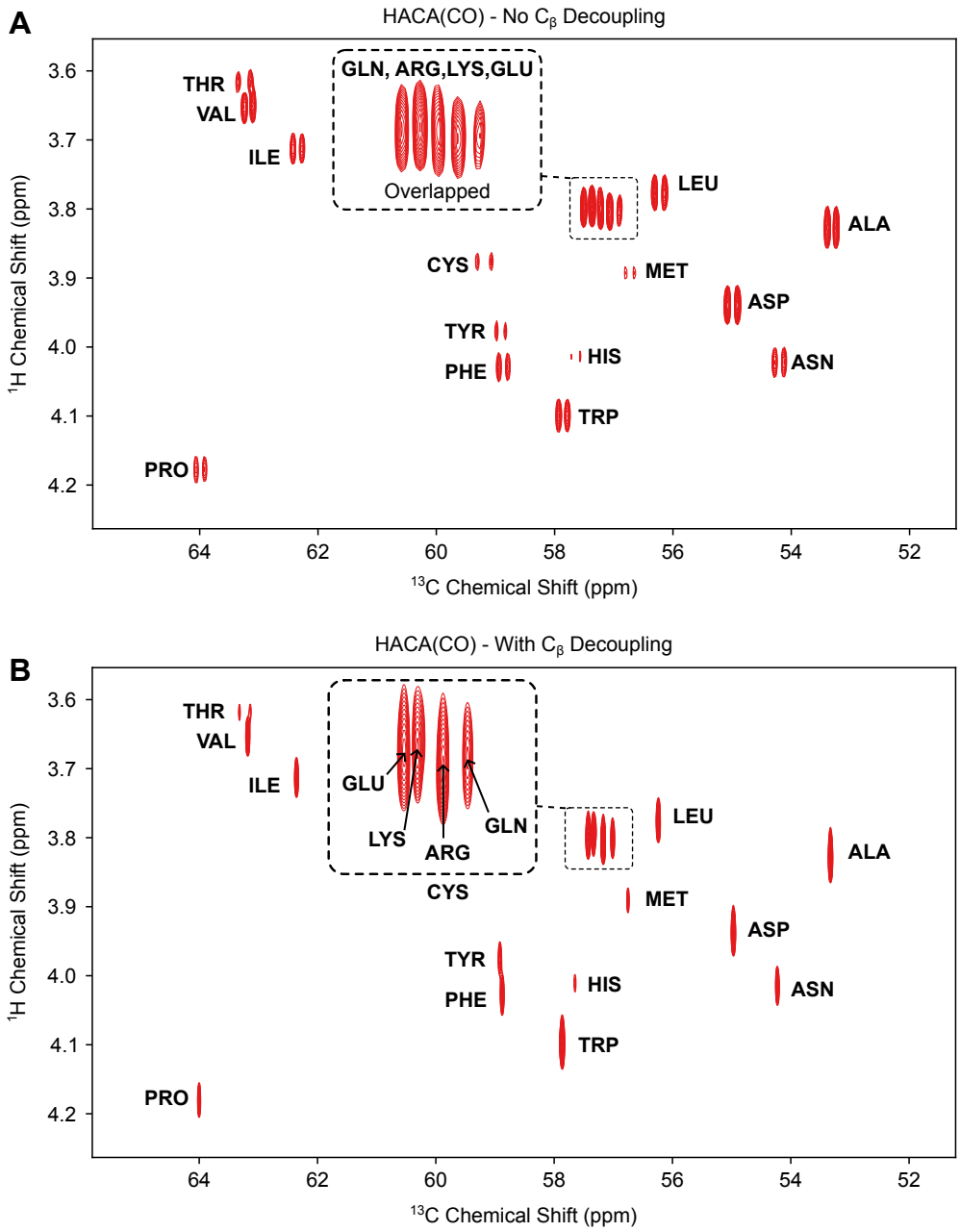


Figure 3.15: Advantages and performance of C_β decoupling. The H_α-C_α correlation spectra obtained using the HACA(CO) sequence in Figure 3.14, without (**A**) and with (**B**) C_β decoupling. Without the C_β decoupling, the peaks corresponding to the amino acids GLN, ARG, LYS, GLU are not fully resolvable. However, the addition of C_β decoupling allows each of these peaks to be resolved separately.

Examples of the chemical shift profiles of the optimal control carbon pulses are shown in appendix A.9. Peak assignments were performed using a Total Correlation Spectroscopy (TOCSY) sequence,[281] with the TOCSY spectra shown in appendix

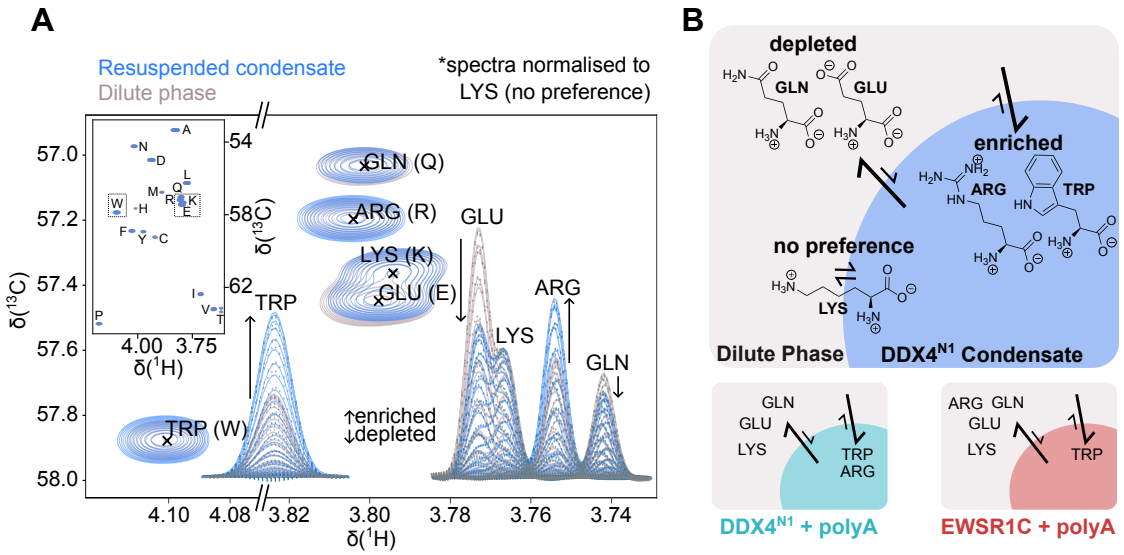


Figure 3.16: Amino acid partitioning quantification into biomolecular condensates. **A** An overlay of the H_α-C_α correlation spectra, obtained using the HACA(CO) pulse sequence, for the resuspended $DDX4^{N1}$ condensate (pellet) and dilute (supernatant) phases. The spectra are normalised to LYS (no preference, PC=1) and highlight the relative trends in partitioning. **B** An illustration showing the partitioning results of a selection of amino acids (TRP, GLN, ARG, LYS, GLU) into $DDX4^{N1}$,$DDX4^{N1}$+polyA and EWSR1C+polyA condensates.

A.10. The combination of many increments in the indirect (C_α) dimension, and C_β decoupling, enabled the GLN, ARG, LYS and GLU peaks to be resolved and isolated (Figure 3.15). However, the presence of C_β decoupling eliminates the magnetisation of glycine (GLY), due to the C_α peak of GLY overlapping with the chemical shifts the C_β selective decoupling was applied to. Therefore, to quantify the partitioning of glycine, a separate experiment without C_β decoupling was applied. Additionally, threonine (THR) and serine (SER) peaks, whose C_β shifts are outside the window of chemical shifts where the C_β selective decoupling was applied, remained as doublets.

The relative intensities of the two-dimensional H_α-C_α peaks in HACA(CO) spectra of the resuspended condensate and supernatant (Figure 3.16) were then used to calculate the partitioning coefficients and free energies of partitioning for the free amino acids into the condensate systems.

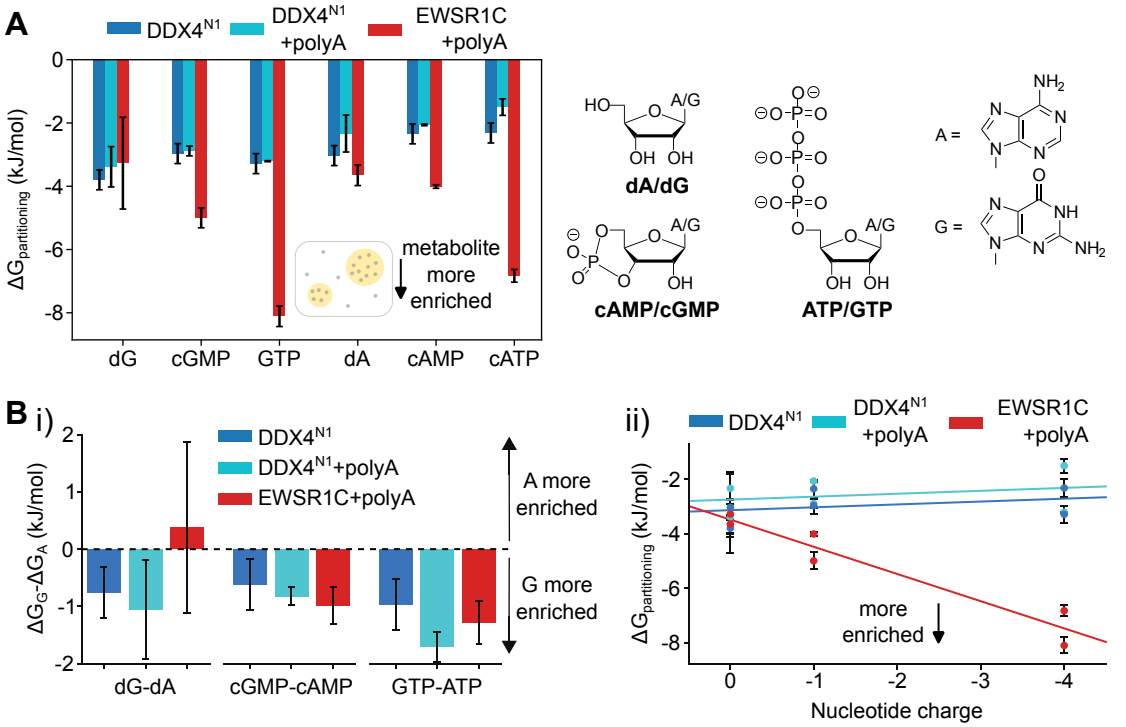


Figure 3.17: Nucleotide partitioning results and initial analysis into the biomolecular condensate systems. **A** Nucleotide partitioning free energies ($\Delta G_{partitioning}$) into DDX4^{N1}, DDX4^{N1}+polyA and EWSR1C+polyA condensates. **B** The difference in free energy of partitioning between guanine and adenine based nucleotides (ΔG_G-ΔG_A) for the condensate systems. On average, guanine based nucleotides are more enriched than adenine based nucleotides. **C** A plot showing the correlations between nucleotide negative charge and $\Delta G_{partitioning}$ in each condensate system. Additional negative charge significantly enhances enrichment into EWSR1C+polyA condensates.

3.4.3 Metabolite partitioning results in condensates and organic solvents

The partitioning free energies ($\Delta G_{partitioning}$) of the nucleotides into the different condensates are shown in Figure 3.17A. Interestingly, all of the nucleotides are enriched in the condensate systems studied, with enrichment into DDX4^{N1}+polyA condensates closely matching that into DDX4^{N1} condensates. The enrichment of guanine based nucleotides (dG, cGMP, GTP) was on average larger than the enrichment of adenine based nucleotides (dA, cAMP, ATP) across the condensate systems (Figure 3.17B). Given the only difference between the bases are the

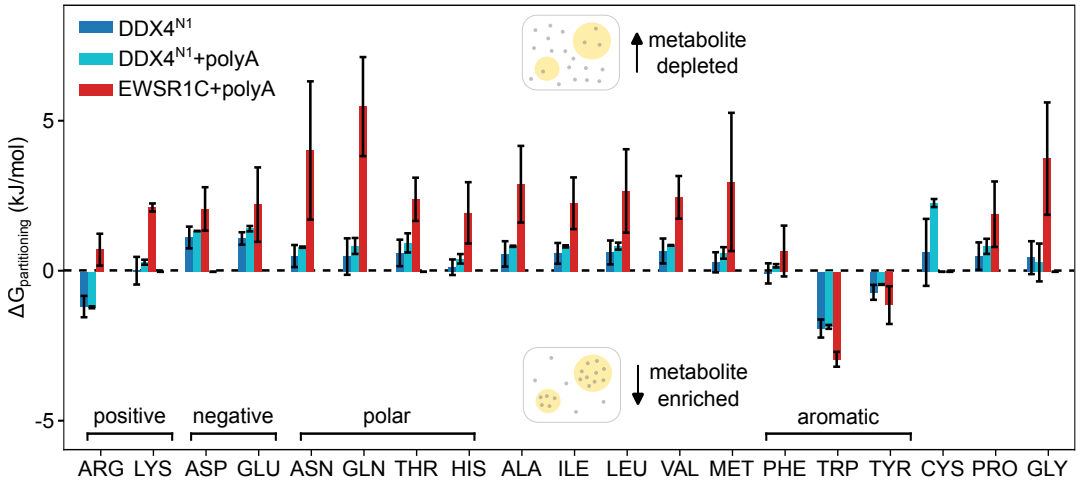


Figure 3.18: Amino acid partitioning results into the biomolecular condensate systems.The amino acid partitioning free energies ($\Delta G_{partitioning}$) into $DDX4^{N1}$, $DDX4^{N1}$+polyA and EWSR1C+polyA condensates. Most amino acids are depleted within the condensates with the exception of arginine (ARG), tryptophan (TRP) and tyrosine (TYR).

substituents on the purine aromatic rings implies that even small changes to a metabolite can significantly effect its partitioning. Furthermore, the enrichment of the nucleotides into EWSR1C+polyA condensates was strongly enhanced in a roughly linear fashion with increasing nucleotide negative charge (Figure 3.17C). However, nucleotide enrichment into $DDX4^{N1}$ condensates was relatively unaltered by adding extra negatively charged units to the nucleotides (Figure 3.17C).

In contrast, the majority of amino acids were depleted inside the condensate systems, with the exception of arginine for $DDX4^{N1}$+/-polyA, and tyrosine/tryptophan for all the condensates (Figure 3.18). The fact that EWSR1C is positively charged (+20.6 at pH 8.0) and $DDX4^{N1}$ is negatively charged (-5.5 at pH 8.0) could provide insight into why positively charged arginine (ARG) is enriched in the $DDX4^{N1}$ condensates, but depleted in EWSR1C-based condensates. While these initial chemical rationalisations provide insight into selective partitioning, a more detailed fitting procedure is required to fully explain the partitioning behaviour. In Figure 3.19, the complete set of nucleotide and amino acid partitioning free energies are shown for the condensates. The condensate partitioning free energies are also compared to those in 1-butanol and 1-octanol to assess the similarities

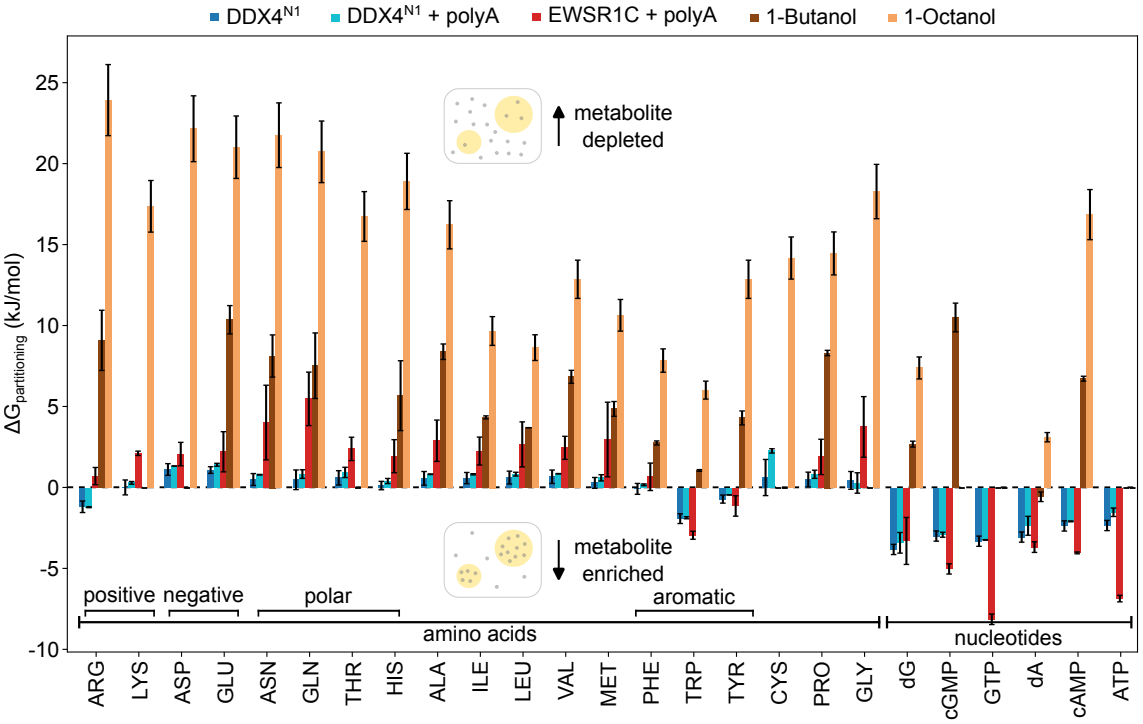


Figure 3.19: Chart of all partitioning results into condensates and organic solvents. The complete set of amino acid and nucleotide partitioning free energies ($\Delta G_{partitioning}$) into the condensates DDX4^{N1}, DDX4^{N1}+polyA and EWSR1C+polyA as well as into the organic solvents 1-butanol and 1-octanol. Raw partitioning coefficients and $\Delta G_{partitioning}$ values are included in appendices A.8.1 and A.8.2, respectively.

and difference in partitioning trends between biomolecular condensates and organic solvents*. In sections 3.4.4 and 3.4.5, metabolite partitioning in each phase separating system (condensates and organic solvents) is correlated with metabolite and disordered protein/solvent molecular properties to obtain a set of rational trends governing selective partitioning.

3.4.4 Correlation of metabolite partitioning with metabolite molecular properties

*The method to quantify metabolite partitioning into 1-butanol is described in appendix method A.4.2. Partitioning free energies for 1-octanol are literature values.[282]

3.4.4.i Method of fitting metabolite partitioning to molecular properties

Computational predictors of molecular partitioning into octanol, such as ALOGPS,[271] often use deep neural networks to predict partitioning based on a fragment-based approach. The neural networks act on vast datasets of partitioning coefficients in octanol; classifying each atom in a molecule through a set of Electrotopological-State (E-state) indices which contain the chemical properties of each atom such as electronegativity, bond order and the functional group it is located in.[283,284] While neural networks provide strong predictability of partitioning in new compounds, they require the establishment of substantial training datasets. Furthermore, insight into trends governing partitioning based on the overall properties of a molecule are often hard to extract and rationalise from a neural network. Instead, a linear regression approach can provide insight into metabolite partitioning rules, allowing comparison between molecular properties governing partitioning into the studied biomolecular condensates and organic solvents. Establishing linear correlations between partitioning free energies and chemical features has been used extensively to rationalise partitioning into organic solvents.[268]

A set of molecular properties was obtained for each of the metabolites in the dataset (extracted from the Human Metabolome DataBase - HMDB[282]). These included the number of aromatic rings, negative charges, positive charges, topological polar surface area (used interchangeably with polar surface area) and the number of rotatable bonds*. The full set of molecular properties for each metabolite are shown in appendix A.12. These molecular properties are used extensively in drug discovery; for example, molecules containing 10 or fewer rotatable bonds and polar surface area no greater than 140 $Å^2$ are predicted to have enhanced oral bioavailability.[285]

*The number of rotatable bonds in a system is related to its conformational entropy.

The dynamic range and mean of each set of molecular properties are not equal; for example, the number of aromatic rings in our metabolite set ranges from zero to two, whereas the polar surface area ranges from 50 Å^2 to 300 Å^2. Therefore, the molecular properties must be standardised to obtain the relative contribution of each molecular property to metabolite partitioning. Standardisation normalises for the dynamic range of each property, with all sets of standardised properties having a mean of zero and a standard deviation of one. To rationalise the metabolite partitioning values in the condensates/organic solvents, it was assumed that the free energy of partitioning could be expressed as a linear combination of contributions from each metabolite property:

$$\Delta G_j = \Delta G_0 + \sum_i \beta_i \frac{x_{i,j} - \overline{x_i}}{\sigma_i} \tag{3.3}$$

where ΔG_j is the free energy of partitioning for metabolite j, $x_{i,j}$ is the value of molecular property i for metabolite j, $\overline{x_i}$ is the mean value of property i across all metabolites, and σ_i is the standard deviation of property i across all metabolites. The fitted standardised coefficients β represent each molecular property's contributions to the free energy of partitioning, with negative values favouring enrichment and positive values favouring depletion. The fitted parameter ΔG_0 represents the partitioning of a hypothetical metabolite with the mean value of all the molecular properties considered.

Generation of combinations of properties to test (models) The following outlines the fitting procedure used to identify the metabolite properties best able to explain the metabolite partitioning data obtained across both the condensates and organic solvents*. The fitting procedure was performed in both a naïve sense, where all 4096 combinations of metabolite properties were allowed (Figure 3.20A),

*This procedure was developed as an equal collaboration between myself and Virgínia Casablancas-Antràs (Baldwin lab).

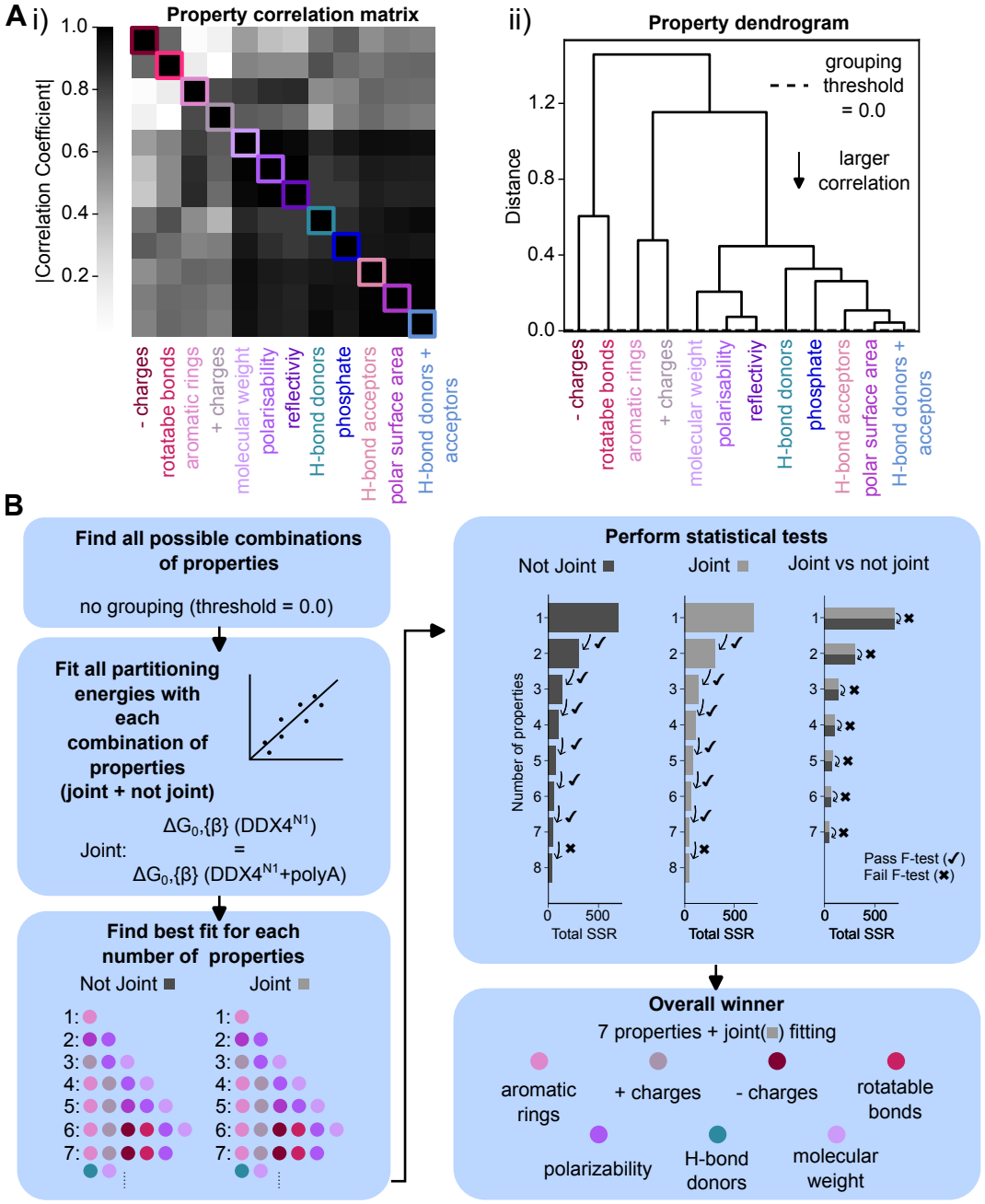


Figure 3.20: Naïve fitting procedure to correlate metabolite partitioning with metabolite properties. **A** A correlation matrix (i) and dendrogram (ii) showing the correlations and grouping between the molecular properties for the amino acid and nucleotide metabolites. **B** A schematic showing the fitting procedure applied to correlate the metabolite partitioning free energies in each phase separating system to molecular properties. An F-test with a p-value of 0.05 was used to justify extra model complexity. The fitting procedure was performed twice with fitted coefficients for $DDX4^{N1}$ and $DDX4^{N1}$+polyA partitioning forced to remain the same (joint), or allowed to differ (not joint). This was to determine if the rules governing partitioning in $DDX4^{N1}$ and $DDX4^{N1}$+polyA were statistically different.

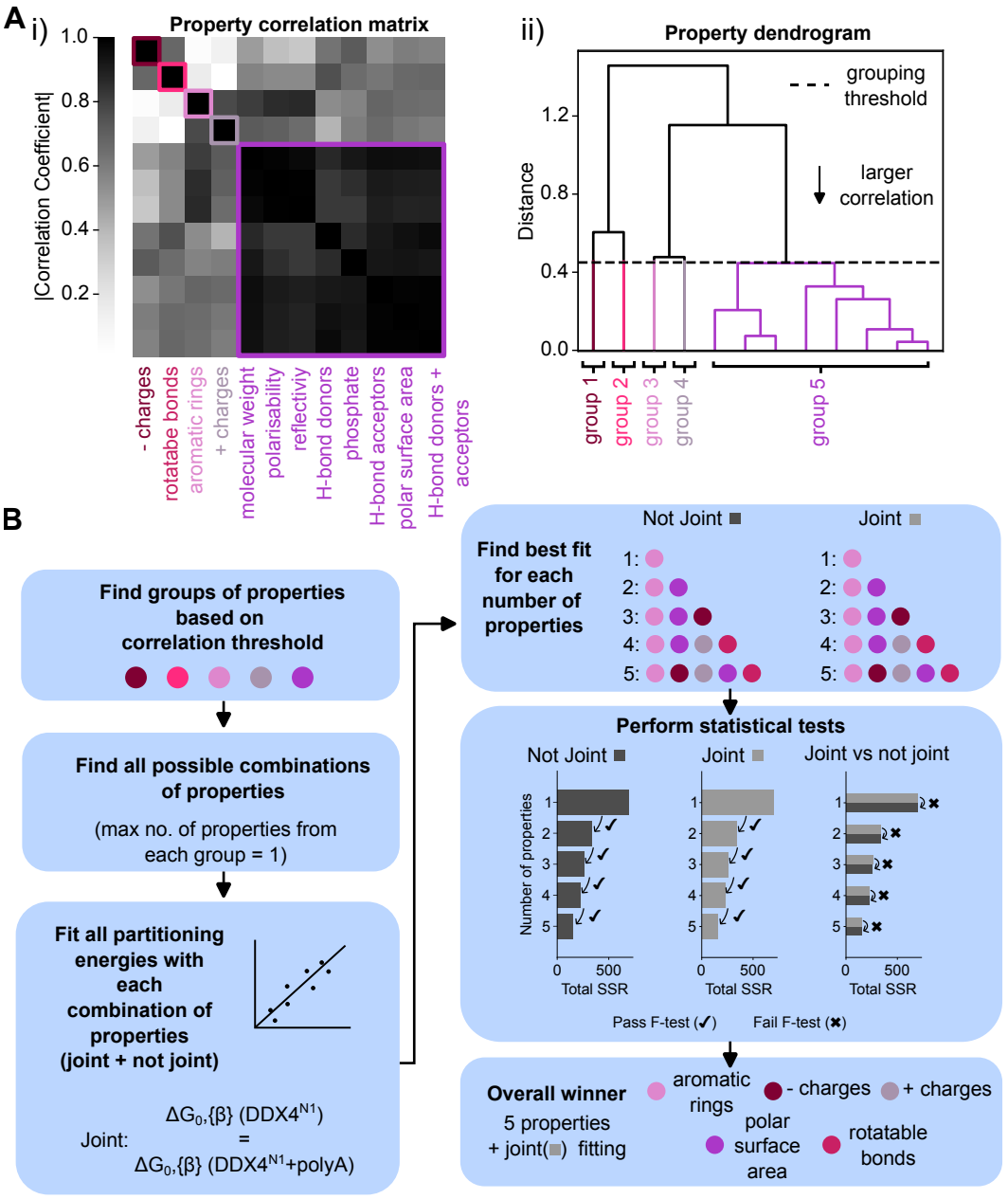


Figure 3.21: Fitting procedure to correlate metabolite partitioning with metabolite properties with property grouping. **A** i) A correlation matrix showing the correlations between the molecular properties for the amino acid and nucleotide metabolites. ii) A dendrogram showing the grouping of metabolite properties after applying the SciPy flat cluster (fcluster) algorithm to the properties with a threshold of 0.45. **B** A schematic showing the fitting procedure applied in order to correlate the metabolite partitioning free energies in a phase separating system to the properties of each metabolite. An F-test with a p-value of 0.05 was used to justify extra model complexity. The fitting procedure was performed twice with fitted coefficients for $DDX4^{N1}$ and $DDX4^{N1}$+polyA partitioning forced to remain the same (joint), or allowed to differ (not joint). This was to determine if the rules governing partitioning in $DDX4^{N1}$ and $DDX4^{N1}$+polyA were statistically different.

and in a grouped manner, with heavily correlated properties prevented from being fitted in the same model. The grouped fitting procedure was used instead of invoking principal component analysis (PCA) or partial least squares regression (PLS) techniques. The grouping of metabolite molecular properties was performed using the SciPy flat cluster (fcluster) algorithm, utilising a distance criterion with a threshold of 0.45, giving a total of 143 possible property combinations containing 1-5 properties (Figure 3.21A).*

Optimisation of each model For each combination of properties, the values of ΔG_0 and β were fitted for each phase-separating system through minimisation of the sum of squared residuals (SSR) using the SciPy Levenberg-Marquardt minimisation algorithm.

$$SSR = \sum_j (\Delta G_{j,\text{measured}} - \Delta G_{j,\text{predicted}}) \tag{3.4}$$

The sum of squared residuals (SSR) obtained for each phase separating system was added together, obtaining the total SSR to determine the most appropriate combination of molecular properties to explain the partitioning data of all systems. The fitted coefficients for $DDX4^{N1}$ and $DDX4^{N1}$+polyA condensates were either forced to remain the same (joint) or allowed to differ (not-joint) to test whether the rules governing partitioning in the two systems were statistically different. For each number of fitted molecular properties, the best fit was that with the minimum total SSR value and, analogously, the minimum total root mean squared deviation (RMSD):

$$\text{Total RMSD} = \sqrt{(\text{Total SSR})/N} \tag{3.5}$$

*Distance thresholds lower than 0.45 suffered from a significant correlation between the fitting parameters thereby reducing the accuracy of the obtained results.

where N is the total number of data points across all datasets. The global SSR was minimised instead of the global χ^{2*} since the partitioning measurements into each condensate and organic solvent had different errors. Therefore, if χ^2 was chosen as the global optimisation parameter, the fitting procedure would preferentially fit the dataset with the smallest errors at the expense of the other datasets. When using the total SSR, the partitioning into each phase-separating system is given equal weight in the fit.

Choosing the winning model to explain the partitioning trends An F-test approach was used to justify additional model complexity to find the most suitable model (molecular property combination), explaining all the small molecule partitioning data across all the phase-separating systems without overfitting the data. More complex models (fewer degrees of freedom) were accepted versus more simple models if the p-value of the F-test was below 0.05. The use of 7 (or 5) molecular properties was statistically justified for the naïve (Figure 3.20B) or grouped (Figure 3.21B) fitting procedures, respectively.[†] Furthermore, using separate fitted coefficients for the DDX4^{N1} and DDX4^{N1}+polyA condensates was not statistically justified in either fitting procedure.

The resultant winner from the naïve fitting procedure produced strong correlations; however, substantial errors[‡] of the fitted parameters revealed significant overfitting (appendix Figure A.19). Furthermore, the fitted coefficients of highly correlated molecular properties (e.g. molecular weight and polarisability) had opposite signs, implicating significant overfitting using this method. Therefore, this method could not provide a meaningful understanding of metabolite partitioning trends in each solvent. In contrast, the errors of the fitted parameters obtained

[*] $\chi^2 = \sum_i (\Delta G_{j,\text{measured}} - \Delta G_{j,\text{predicted}})^2 / \sigma_j^2$ where σ_j is the error of $\Delta G_{j,\text{measured}}$.

[†] Full details on the fitting results are included in appendix A.13.

[‡] Errors were calculated in a bootstrap fashion,[286] where data was re-sampled with replacement to generate synthetic repeat datasets which were fitted separately. The standard deviation of each fitted coefficient across all synthetic datasets was taken as the error.

from the grouped metabolite property fitting procedure were significantly reduced (Figure 3.22), justifying the requirement of metabolite property grouping.

The total Variance Inflation Factor (VIF), a measure of multicollinearity between fitting properties, is calculated by taking each molecular property as a dependent variable and fitting this variable based on linear combinations of the other molecular properties ($x_{i,j} = c + \sum_{i \neq k} m_k x_{k,j}$). The molecular property being fitted is given by x_i, x_k are the other molecular properties in the model, and R_i^2 measures the goodness of fit. The VIF for property i (VIF_i) is equal to $1/(1\text{-}R_i^2)$. The total VIF is the mean of the VIF_i values for different properties i. The total VIF was calculated as 4.7 for the molecular properties in the grouped best-fit model, with values below 5 considered acceptable for linear regression without overfitting.[287] In contrast, the winner of the naïve fitting procedure was a model containing a VIF of 74.8. Therefore, the grouped metabolite properties fitting procedure provided the most reliable method to correlate partitioning to metabolite properties without overfitting.

In a set of molecules covering an even distribution of chemical properties from very hydrophilic to very hydrophobic, the ΔG_0 values would all be close to zero across all the phase separating systems. However, the choice of using common intracellular metabolites, which are inherently hydrophilic, leads to large ΔG_0 values for 1-butanol and 1-octanol (Figure 3.22B). Despite this, the standardised coefficients (β) can still be compared to determine the trends of partitioning between the condensates and organic solvents (Figure 3.22Ci). However, the dynamic range of partitioning energies ($\Delta G_{max} - \Delta G_{min}$) is not constant across the different phase separating systems, larger for the organic solvents than the condensates. Therefore, the standardised coefficients are normalised by dividing by the absolute value of the mean of the standardised coefficients for each system, $\beta_{norm,i} = \beta_i/\overline{\beta}$, to assess the relative contributions to partitioning in each phase separating system.

Justification for using a grouping threshold of 0.45 As described previously, the variance inflation factor (VIF) is a measure of multicollinearity between fitting parameters.[287] A value of 5 was chosen as a cut-off for whether a fit was rejected due to potential multicollinearity.[287] If the threshold was set to 0.4 (rather than 0.45), the winning model had a VIF of 8.94, signifying heavy multicollinearity between the fitting parameters. A threshold of 0.45 was therefore chosen as it was the lowest threshold that gave a final model with a VIF < 5 (VIF=4.7). Furthermore, the winning model at a threshold of 0.45 passed an F-test (p-value=0.05) against the winning models obtained at grouping thresholds of 0.5 or greater.

3.4.4.ii Results of the fitting: solubility trends obtained for the biomolecular condensates and organic solvents

The molecular properties in the best-fitting model were the number of aromatic rings, negative charges, positive charges, polar surface area and the number of rotatable bonds; with joint coefficients for the DDX4^{N1} and DDX4^{N1}+polyA condensates (Figure 3.22). The results show similarities and differences in partitioning trends between the condensates and organic solvents. Aromatic rings favour metabolite enrichment in all the condensate and organic solvent systems, with enrichment into EWSR1C+polyA condensates most pronounced (Figure 3.22Cii). However, polar surface area favours enrichment into the condensates and unsurprisingly favours depletion in the hydrophilic organic solvents. Selectivity between the condensates for the set of metabolites considered arises when considering the effect of metabolite negative charges. Metabolite negative charge drives depletion into the DDX4^{N1}-based condensates, whereas it drives a slight enrichment into EWSR1C+polyA condensates. This behaviour is rational given that DDX4^{N1} is negatively charged and EWSR1C is positively charged. The set of trends governing metabolite partitioning could provide insight into how to create a drug candidate that would be enriched into one

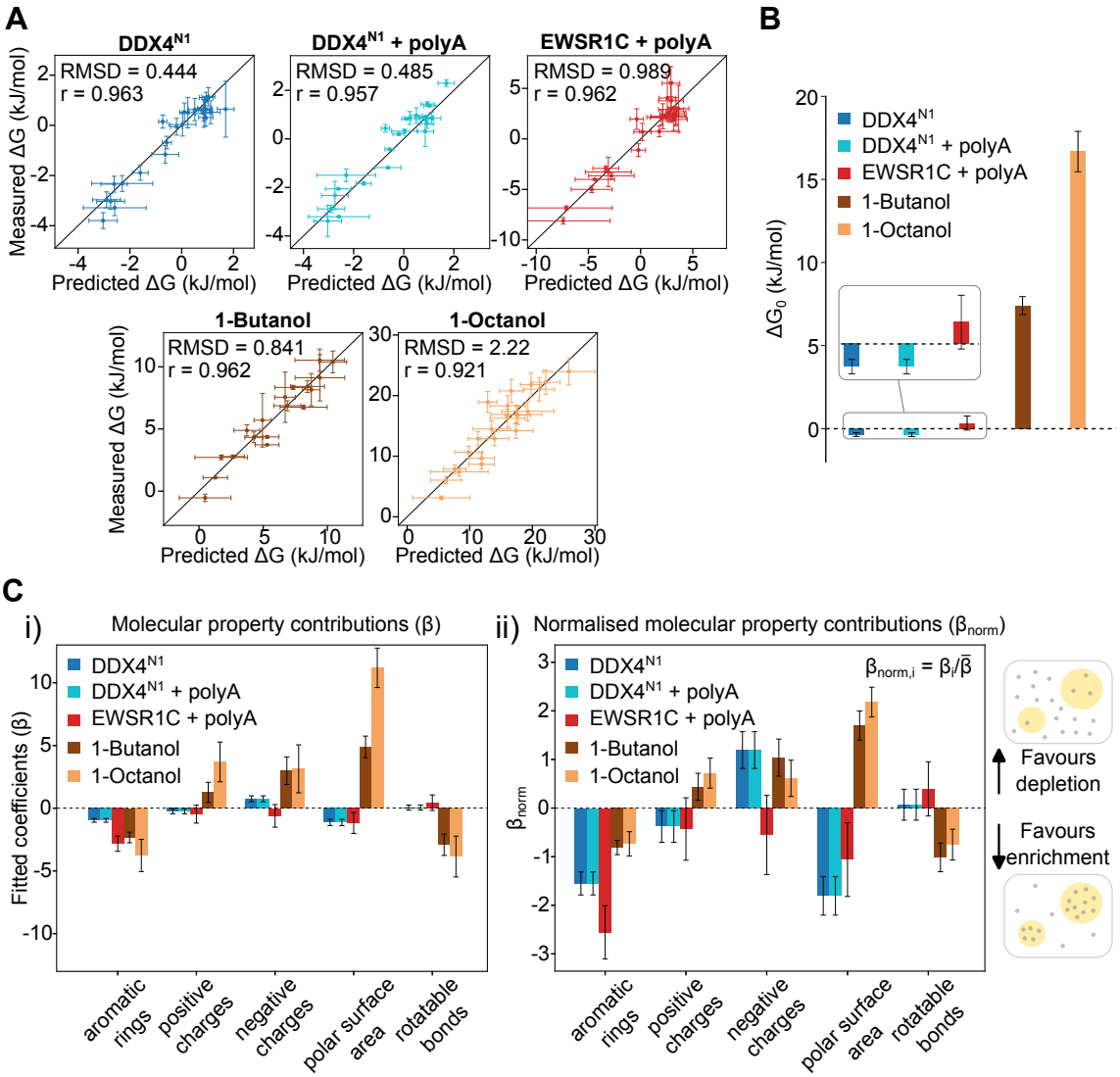


Figure 3.22: Correlation between metabolite partitioning and metabolite properties. **A** Plots showing the correlation between the fitted/predicted and measured partitioning free energies (ΔG) for each of the condensate and organic solvent systems for the overall winner of the grouped fitting procedure. **B** The fitted ΔG_0 values for each phase separating system, representing the free energy of partitioning for a hypothetical molecule whose properties are the mean of the properties for all the amino acids and nucleotides in the dataset. **C** The non-normalised (i) and normalised (ii) standardised coefficients obtained from the fitting procedure, highlighting how metabolite properties affect the partitioning free energy for each condensate/organic solvent. Normalisation allows comparison of the relative contribution of each molecular property to metabolite partitioning in each system.

type of biomolecular condensate but depleted or less enriched in others. This would allow a drug to preferentially target a condensate relevant to a particular disease while causing minimal disruption to other types of condensates. For

example, the partitioning trends predict that a molecule with lots of aromatic rings, polar surface area and negative charges would be preferentially enriched into EWSR1C+polyA condensates over DDX4^{N1} condensates (ATP and GTP are examples of such molecules). FUS (Fused in Sarcoma) and tau (tubulin-associated unit) protein condensate formation have been linked to pathological aggregation causing amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD). The established NMR-based method to determine metabolite partitioning could be applied to determine trends governing molecular partitioning into condensates formed from FUS and tau so that drugs can be created to target these condensates and prevent aggregation.

3.4.4.iii Cross-validation of the winning model

To assess the predictability of the winning model, a 3-fold cross-validation method was applied to each dataset. Each dataset was randomised and then split into 3 groups. Sequentially, one of the groups was removed from the dataset, with the other 2 groups making up a training dataset. The molecular properties in the resultant model were then fitted to the training data. The fitted regression coefficients were applied to the molecules in the test dataset, and the RMSD from their measured values calculated. This process was repeated until all groups had been the test dataset and the RMSD was averaged over the 3 cases. This process was repeated 100 times, randomising the datasets using different seeds before each iteration, so that the average RMSD's over many permutations of possible groups could be obtained, thereby providing a more accurate representation of model predictability.

The RMSD values from the 3-fold cross-validation method were 0.70 kJ/mol (DDX4^{N1}), 0.75 kJ/mol (DDX4^{N1}+polyA), 1.70 kJ/mol (EWSR1C+polyA), 1.66 kJ/mol (butanol) and 3.86 kJ/mol (octanol). These values are slightly larger than the RMSD values from using all data as training data (Figure 3.22A). However, the cross-validated RMSDs are much smaller than the range of the

measured partitioning energies for each solvent: 4.9 kJ/mol ($DDX4^{N1}$), 5.7 kJ/mol ($DDX4^{N1}$+polyA), 13.6 kJ/mol (EWSR1C+polyA), 11.1 kJ/mol (butanol), 20.8 kJ/mol (octanol). Therefore, despite not being able to predict the exact partitioning free energies in each solvent, the model captured the major trends governing the observed molecular partitioning.

3.4.4.iv Justification for minimising RMSD over χ^2

EWSR1C+polyA, butanol and octanol partitioning free energies have substantial dynamic ranges across the dataset used. For these systems, the error in ΔG was heavily dependent on the size of ΔG itself, with higher ΔG values (from molecules being more excluded from the solvent) having a higher error on average. Whilst the errors indicate uncertainty in the exact value of ΔG for heavily excluded molecules, there is significant information contained in these data points in that the molecules are substantially excluded. By using χ^2, these points are effectively excluded from the fit and thereby information on the molecular properties leading to exclusion is lost. Therefore, fitting using χ^2 biases the resultant fit, preferentially fitting molecular properties that cause enrichment into the solvents over those favouring exclusion, leading to poor model predictability.

For example, after fitting the data using χ^2, the resultant model (shown in Appendix Figure A.20) predicts that ATP (a highly polar molecule) would be enriched into butanol ($\Delta G_{predicted}$ = -1.9 kJ/mol). This is not correct as ATP was left out of the fitting procedure for butanol because it was significantly depleted, to the degree that its partitioning coefficient could not be quantified. In contrast, the final model after minimising RMSD correctly predicts that ATP is significantly depleted in butanol ($\Delta G_{predicted}$ = 16.8 kJ/mol). Therefore, despite being non-standard, optimising RMSD was chosen for the fitting procedure so that molecular properties governing both enrichment and exclusion were equally weighted.

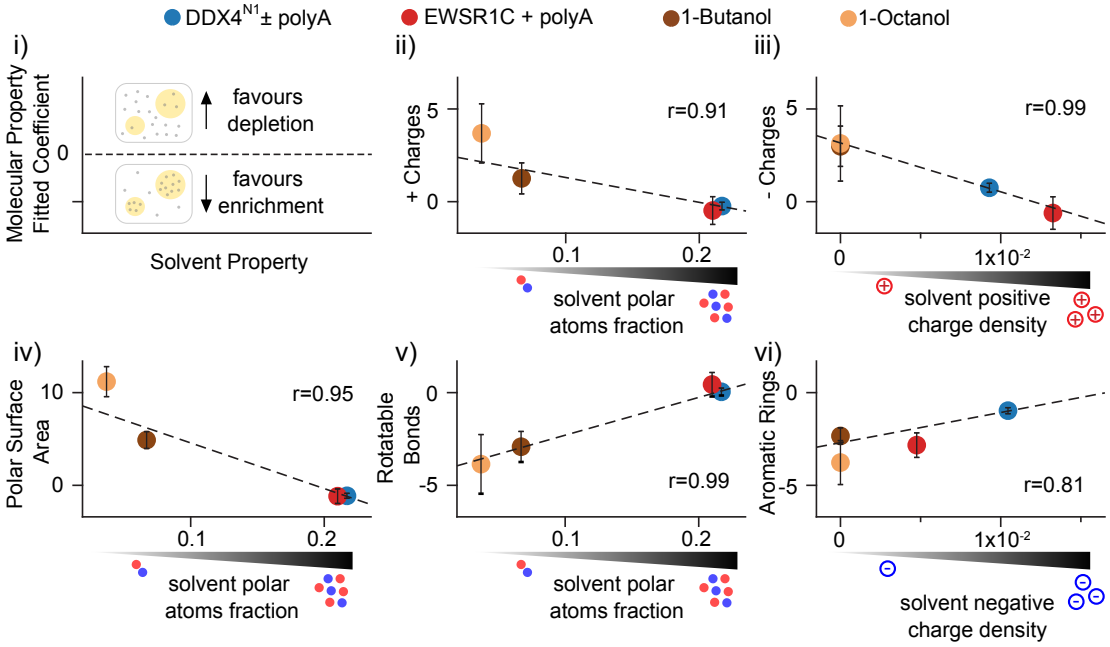


Figure 3.23: Correlation between metabolite property contributions to partitioning and solvent/disordered protein properties. i) A general schematic of a plot of fitted non-normalised standardised coefficient contribution of a molecular property to partitioning versus solvent property. ii)-vi) Linear fits of standardised coefficients of the five molecular properties in the global partitioning fitting model versus properties of the solvent (organic or disordered protein). The solvent properties are defined as follows: polar atoms fraction is the number of heteroatoms in the solvent divided by the total number of atoms, (positive/negative) charge density is the number of (positive/negative) charged atoms in the solvent divided by the total number of solvent atoms. Correlation coefficients (r) represent the correlation between fitted coefficients and predicted fitted coefficients.

3.4.5 Correlation of standardised coefficients with solvent properties

To fully explain the differences in partitioning driving forces between the condensates and organic solvents established in Figure 3.22C, the non-normalised standardised coefficients were fitted against properties of the disordered protein/solvent (Figure 3.23)*. The fitted standardised coefficients of molecular properties driving metabolite partitioning showed strong correlations with solvent properties such as the percentage of polar atoms or the positive charge density. The established correlations between molecular properties governing partitioning and solvent

*This fitting was performed in collaboration with Virgínia Casablancas-Antràs (Baldwin lab)

properties highlight that the selectivity of partitioning into a solvent/condensate is rational and follows chemically reasonable trends.

The main discriminator governing molecular partitioning into DDX4^{N1}+/-polyA and EWSR1C+polyA condensates is negative charges. However, further selectivity could be provided by using other chemical functional groups, such as haloalkanes and nitriles, which are not present in the metabolites. Extending the partitioning trends established here to a more extensive range of molecules/metabolites containing a host of different chemical features would enhance the design of molecules to target specific condensates for drug/therapeutic purposes.

3.4.6 Ion partitioning into DDX4^{N1} condensates

There is a growing interest in the ability of ions to be partitioned into biomolecular condensates, a phenomenon already established in synthetic polymer coacervates.[288,289] Ions are highly prevalent inside cells, with the intracellular concentration of NaCl estimated between 5-15 mM,[260] and intracellular free phosphate is estimated to be on the order of 1 mM.[290] Cells usually expend significant energy to set up concentration gradients across membranes using ion channels.[260] If biomolecular condensates can modulate ion concentrations through partitioning without the expenditure of ATP, this would have profound biological importance. Recently, ion partitioning into condensates has been shown to set up inter-phase electric potentials.[174,291] Therefore, ion partitioning into biomolecular condensates could enable a condensate surface to possess catalytic properties,[174] potentially uncovering a new catalytic role of condensates in cells. Furthermore, ion partitioning coupled with the ability of biomolecular condensates to wet membranes[146] could alter concentration gradients across ion channels.[292] Post-synaptic densities, which are in contact with many ion channels at synaptic junctions in neurons, may utilise this phenomenon to tune chemical signalling across a synapse.

Previous measurements to measure ion partitioning into biomolecular condensates have been predominantly based on ICP-MS (Inductively Coupled Plasma Mass Spectrometry) or similar approaches.[173,174] As an alternative to ICP-MS, the method established in section 3.4.2 was adapted to measure the partitioning of sodium and phosphate ions into DDX4^{N1} condensates using NMR. A similar NMR approach has recently been used to measure ion partitioning into condensates formed from the phase separation of protamine and ATP.[277]

3.4.6.i Adapting the NMR-based partitioning method to determine ion partitioning

The first stages of the method to quantify condensate ion partitioning are identical to that for the metabolite partitioning method (Figure 3.11), up to the measurement of the pellet and leftover supernatant volumes, V_p and V_{LOS}. The condensate and supernatant phases were resuspended/diluted in Ultra-High-Performance Liquid Chromatography (UHPLC)-grade water mixed with 10% D_2O to minimise the addition of extra ions to the sample. Importantly, the condensate/supernatant samples were diluted significantly to 600 μL to prevent phase separation of DDX4^{N1} from re-occurring in the resultant low-salt buffer. Meanwhile, the same liquid handling/tube transfer processes were performed on 600 μL of UHPLC-grade water containing 10% D_2O to act as a control, to take into account the addition of sodium ions from the surrounding environment, tubes or glassware during the method.*.

Broadband ^{23}Na and ^{31}P one-dimensional pulse-acquire NMR spectra were recorded on the supernatant, resuspended condensate and control NMR samples (Figure 3.24A). After quantifying the intensity per unit scan of the peaks in each spectrum (through integration), the intensity per unit scan from the control sample was subtracted from the intensities per unit scan of the supernatant and resuspended condensate signals. In the control sample, there was no observed ^{31}P

*Full details on the method to measure ion partitioning are included in appendix method A.4.5. This method was developed in equal collaboration with Ben Naylor (Baldwin lab).

signal; however, significant intensity was observed in the ^{23}Na spectrum. This justifies the necessary correction of ^{23}Na intensities by subtraction of the intensity from the control sample (Figure 3.24Ai). After quantifying the signal factor (S), the ratio of intensities per unit scan after control intensity subtraction, the partitioning coefficient (PC) and free energies of partitioning (ΔG) were calculated using equation 3.2.

3.4.6.ii Ion partitioning results

The NMR-based ion partitioning method enabled the partitioning coefficients (PC) for sodium and phosphate ions into DDX4^{N1} condensates to be calculated as 1.27±0.05 and 0.71±0.10, respectively (Figure 3.24Bi). These values correlate to free energies of partitioning (ΔG) values of -0.57±0.09 kJ/mol for sodium ions and 0.78±0.32 kJ/mol for phosphate ions at 278 K (Figure 3.24Bii).* Given the charge of DDX4^{N1} is -5.5 at pH 8.0 (used for these partitioning experiments), it is unsurprising that it is enriching positively charged sodium ions while depleting negatively charged phosphate ions. Furthermore, a recent study has shown that ion binding of multivalent and monovalent cations can cause charge inversion of DDX4^{N1}, thereby altering phase transitions.[240] The observed sodium and phosphate ion partitioning likely reports on the attractive and repulsive net interactions between positively charged sodium and negatively charged phosphate ions with DDX4^{N1} chains.

3.4.6.iii Potential biological consequences of ion partitioning for biology

Cells use substantial amounts of energy to establish concentration gradients across membrane-bound compartments as well as across the cell membrane itself. For example, the intracellular concentration of sodium is approximately 10× smaller

*Results obtained in equal collaboration with Ben Naylor and previously deposited in thesis by Ben Naylor.[293]

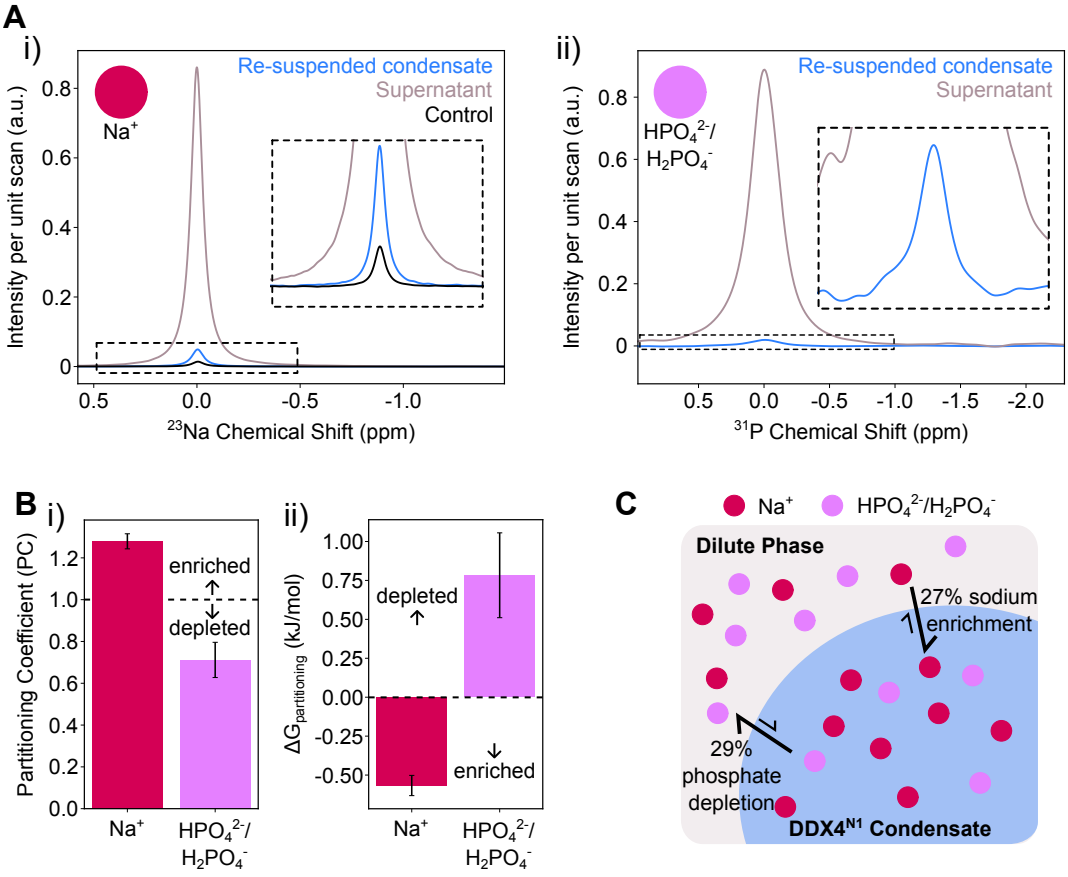


Figure 3.24: Sodium and phosphate ion partitioning results in DDX4^{N1} condensates. **A** Broadband ^{23}Na (i) and ^{31}P (ii) NMR spectra obtained for supernatant and resuspended DDX4^{N1} condensate NMR samples. For ^{23}Na, a control spectrum from a sample containing UHPLC-grade water and 10% D_2O is shown. Each spectrum is normalised per unit scan. Signal factors (S) are calculated from the ratio of the integral of the observed peaks (after subtracting the control integral from the supernatant/resuspended condensate values for ^{23}Na). **B** Plots showing the calculated partitioning coefficients (i) and free energies of partitioning (ii) of sodium and phosphate ions into DDX4^{N1} condensates. **C** An illustration showing the sodium ion enrichment and phosphate ion depletion into DDX4^{N1} condensates.

than the extracellular concentration.[260] Sodium ion pumps utilise ATP-hydrolysis to release free energy (≈6 kJ/mol at 37^{o}C) to maintain this concentration gradient. Assuming that the Na^+ partitioning coefficient of 1.27 into DDX4^{N1} condensates is independent of temperature, this would correlate with a free energy of partitioning of -0.62 kJ/mol at 37^{o}C, approximately 10% of that required to establish the sodium ion gradient across the cell membrane. As DDX4^{N1} condensates are known to wet membranes,[240,294] ion partitioning may be able to change the

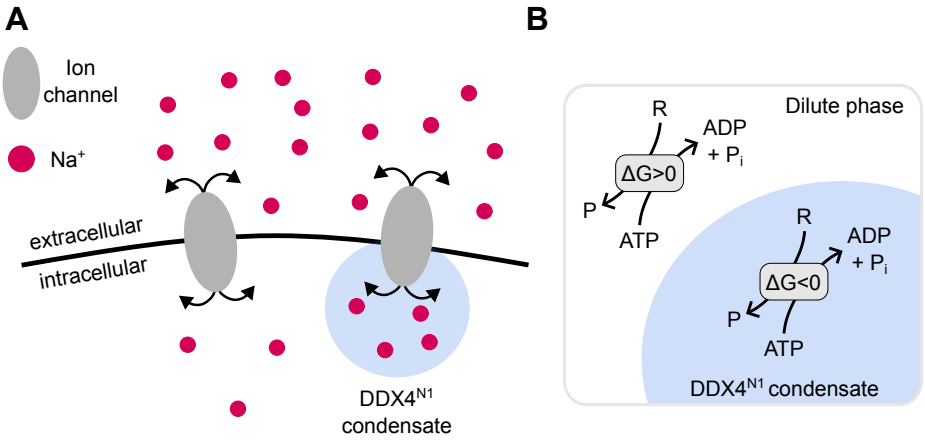


Figure 3.25: Depictions of how condensate ion partitioning could effect biology. **A** The wetting of a DDX4^{N1} condensate to a membrane in the area of an ion-channel, coupled with the partitioning of ions could locally alter the behaviour of ion channels. **B** Phosphate (P_i) partitioning leading to a reduction of phosphate inside DDX4^{N1} condensates is able to alter the free energy of ATP-hydrolysis, with potential consequences for ATP-hydrolysis coupled reactions. R and P represent reactants and products, respectively.

behaviour of ion channels in contact with DDX4^{N1} condensates by altering the local transmembrane ion gradient (Figure 3.25A). Partitioning of cations is likely to extend to multivalent cations like Ca^{2+}. Partitioning of Ca^{2+} into post-synaptic densities could profoundly affect the action of Ca^{2+} pumps at the synapse with consequences for neuron signalling. Broadband ^{43}Ca NMR could be used to measure the partitioning of Ca^{2+} into synaptic densities.

Phosphate partitioning into DDX4^{N1} condensates may also benefit cells regarding reaction coupling. ATP hydrolysis is commonly used as an energy source within a cell:

$$\mathrm{ATP} \rightleftharpoons \mathrm{ADP} + \mathrm{P_i} \tag{3.6}$$

with ADP and phosphate (P_i) being produced. The standard free energy (ΔG°) for this process, with all reactants and products at 1 M concentration, is -31.5 kJ/mol.[295] Intracellular concentrations of ATP, ADP and phosphate (P_i) and can be estimated as 9.6 mM,[274] 0.56 mM[274] and 1 mM,[290] respectively. The partitioning of cAMP and ATP into DDX4^{N1} condensates was calculated to be

2.77±0.38 and 2.73±0.38, respectively. Therefore, it was assumed that ADP also has a partitioning coefficient of 2.7 into DDX4^{N1} condensates. From the partitioning coefficients, intra-DDX4^{N1} condensate concentrations of ATP, ADP and phosphate (P_i) were assumed to be 25.92 mM, 1.51 mM and 0.71 mM. This approximation ignores any competition between partitioning into DDX4^{N1} and other membrane and membraneless organelles. The free energy of ATP hydrolysis can be calculated using:[100]

$$\Delta G = \Delta G^{\ominus} + RT \ln \left(\frac{[P_i][ADP]}{[ATP]} \right) \tag{3.7}$$

where all the concentrations are in mol/L, to give the free energy of ATP hydrolysis in the cytoplasm (no partitioning) as -56.63 kJ/mol at 37^{o}C. In contrast, the free energy of ATP hydrolysis in DDX4^{N1} condensates in cells would equal -57.51 kJ/mol at 37^{o}C. The proposed extra 1 kJ/mol released during ATP-hydrolysis inside cellular DDX4^{N1} condensates is caused by phosphate depletion. However, this calculation is based on the assumption that $\Delta G^{\ominus}$ is unchanged inside the DDX4^{N1} condensed phase.

The energy released from ATP hydrolysis can be harnessed by a cell to allow an energetically unfavourable ($\Delta G>0$) reaction to occur, a process called reaction coupling (Figure 3.25B).[260] The extra free energy released from ATP hydrolysis inside cellular DDX4^{N1} condensates caused by phosphate partitioning could impact intra-condensate chemical reactivity. Potassium-sodium ion channels utilise ATP-hydrolysis to maintain the sodium concentration gradient across a cell membrane. Therefore, a DDX4^{N1} condensate in contact with a potassium-sodium ion pump could enhance the ability of a cell to maintain the desired concentration gradients through extra free energy released from ATP-hydrolysis in the condensate, caused by phosphate partitioning (Figure 3.25). This is in addition to the reduction of the effective sodium concentration gradient across the membrane at the location of the ion channel caused by the enrichment of sodium ions into the DDX4^{N1} condensate.

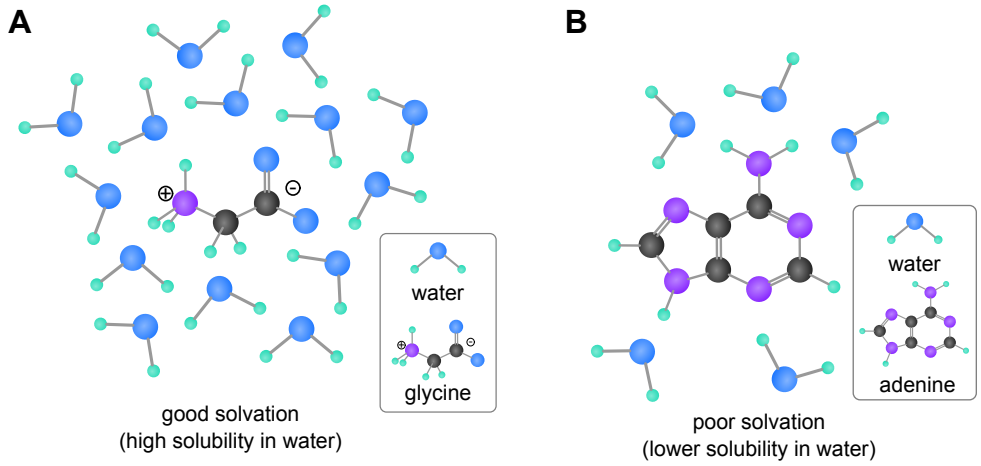


Figure 3.26: Water as a biological solvent. Schematic showing how water has different solvation abilities for the molecules glycine (**A**) and adenine (**B**), therefore affecting their solubility in water.

3.5 Physical and chemical properties of water inside DDX4^{N1} condensates

The previous section outlined trends governing metabolite partitioning into biomolecular condensates. However, molecular partitioning is just one example of a solvent dependent property (Figure 3.2). Given the high abundance of water in biology, measuring its solvent-dependent characteristics inside biomolecular condensates is essential to determine if the chemical and physical environment within the condensates is altered. This section outlines how a biphasic NMR sample of the DDX4^{N1} condensate phase and dilute phase was produced and used to measure the difference in the solvent-dependent properties of water, metabolites and protein between the two phases.

3.5.1 Water as a solvent

Water (H_2O) is often viewed as the universal solvent in biology due to its high prevalence in all forms of life. It has many properties that enable it to function as a suitable solvent in biology. Its high polarity and ability to form hydrogen bonds allow it to solubilise various metabolites, proteins and nucleic acids (Figure 3.26A),

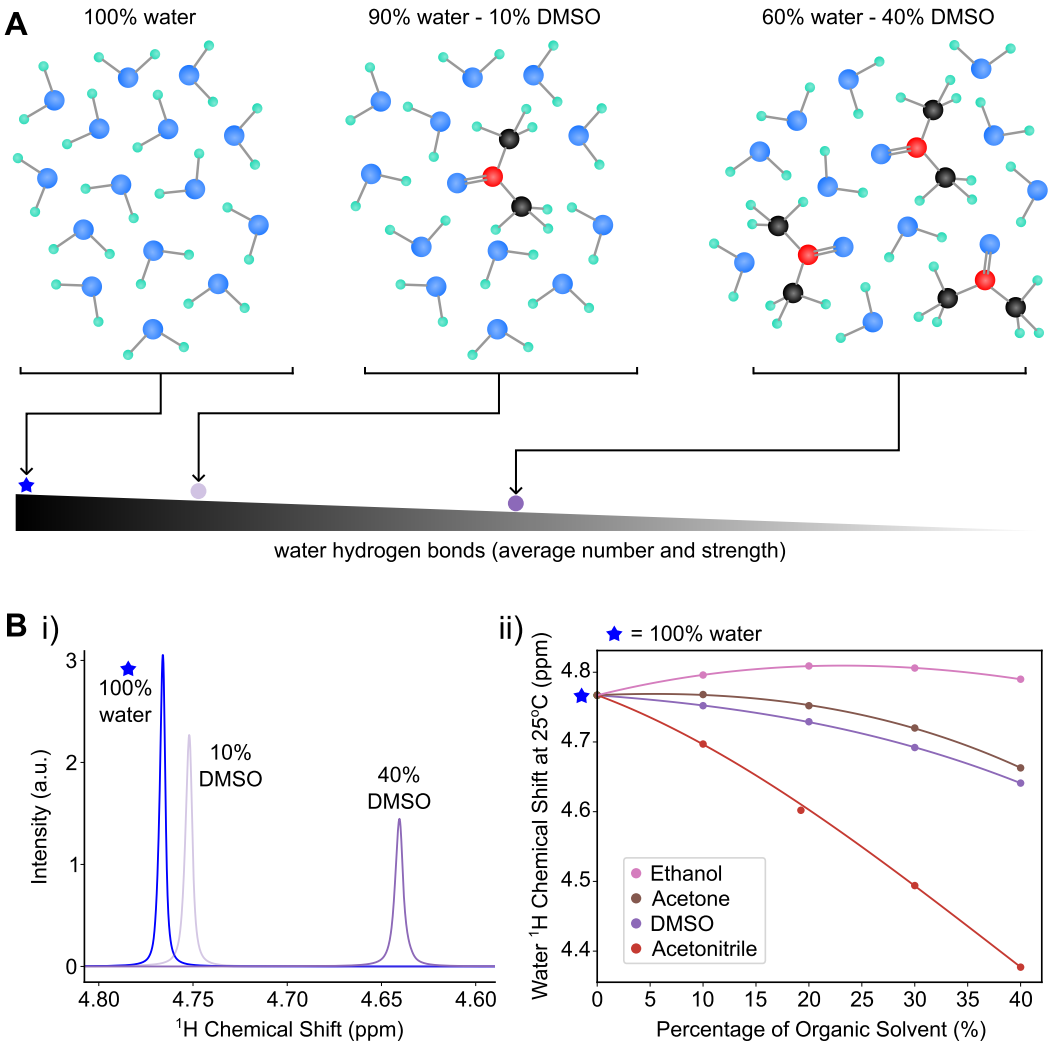


Figure 3.27: The altered chemical environment of water in organic solvents. **A** Pure water contains a strong interconnecting network of hydrogen bonds responsible for its solvent properties. Adding significant quantities of an organic solvent, such as dimethyl sulfoxide (DMSO), reduces the number and strength of the hydrogen bonds water forms. **B** i) The proton NMR spectrum of water for solutions of 100 % water, 90 % water-10 % DMSO and 60 % water-40 % DMSO. Each spectrum was recorded at 25^{o}C and is referenced to DSS (sodium trimethylsilyl propane sulfonate) at 0 ppm. ii) A plot showing how the water chemical shift (at 25^{o}C) changes as a function of organic solvent percentage for the organic solvents ethanol, acetone, DMSO and acetonitrile.

enabling a rich environment for complex biochemistry within the cell. However, not all molecules are highly soluble in water (Figure 3.26B), meaning the concentrations of some molecules are limited in aqueous solutions and the cytosol.

Upon the addition of organic solvents to water, its hydrogen bonding network is disrupted (Figure 3.27A), changing its physical and chemical properties. The

chemical shift of water is a direct measure of the hydrogen bonding ability of water in a solution. Stronger hydrogen bonds will remove electron density from around the H_2O protons, de-shielding it and increasing the observed chemical shift. As increasing percentages (by volume) of dimethyl sulfoxide (DMSO) were added to water, its chemical shift decreased, indicating a reduced quantity and strength of hydrogen bonds in the system (Figure 3.27Bi). The water chemical shift in different percentages (by volume) of commonly used organic solvents (ethanol, acetone, DMSO, acetonitrile) showed the water hydrogen bonding network depends not only on the percentage of organic solvent but also on the solvent type (Figure 3.27Bii).

3.5.2 Biphasic condensate samples

Brady *et al.*[45] performed extensive NMR experiments on a sample containing just the DDX4^{N1} condensed phase. It was shown that DDX4^{N1} remained disordered in the condensed phase, and there were strong intermolecular NOEs involving phenylalanine with several other amino acid residues. Translational diffusion of globular proteins and small molecules such as glucose was also measured to be significantly reduced inside the DDX4^{N1} condensed phase. Furthermore, Pantoja *et al.*[224] produced a biphasic sample of a condensate formed from the tau protein to compare the physicochemical properties between the phases using NMR. Taking inspiration from Pantoja *et al.*[224], a biphasic sample of DDX4^{N1} was produced (Figure 3.28) to compare the physicochemical properties of water, metabolites and DDX4^{N1} protein between the DDX4^{N1} condensed and dilute phases (Figure 3.29).[†] The remainder of this chapter discusses differences of physicochemical properties between the phases obtained using MRI, slice selective NMR, CEST, diffusion and relaxation experiments (introduced in chapter 2).

[†]DDX4^{N1} protein for the production of the biphasic sample was produced as an equal collaboration between myself and Emily Haasz (Baldwin lab), and biphasic sample assembly performed in conjunction with Dr. Charles Buchanan (Baldwin lab) and Emily Haasz.

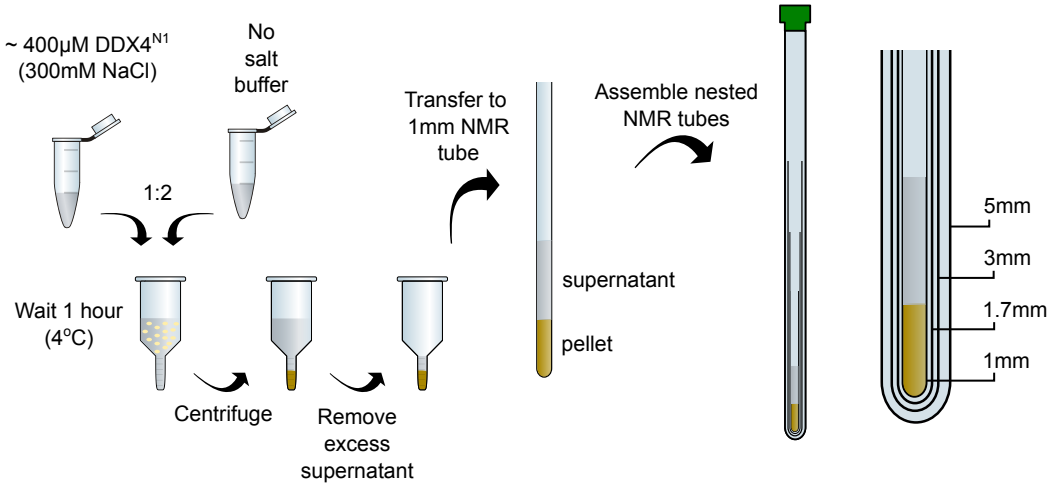


Figure 3.28: Method of producing a biphasic sample of DDX4^{N1}. The condensate phase is centrifuged into a pellet, leaving the dilute phase as the supernatant. The biphasic sample was prepared inside a 1 mm NMR tube to minimise the quantities of DDX4^{N1} protein required. The 1 mm sample tube was subsequently placed inside increasing-sized tubes in a "Russian doll" fashion to record the NMR spectra using an NMR probe designed for 5 mm tubes. Further details in appendix A.5.1.

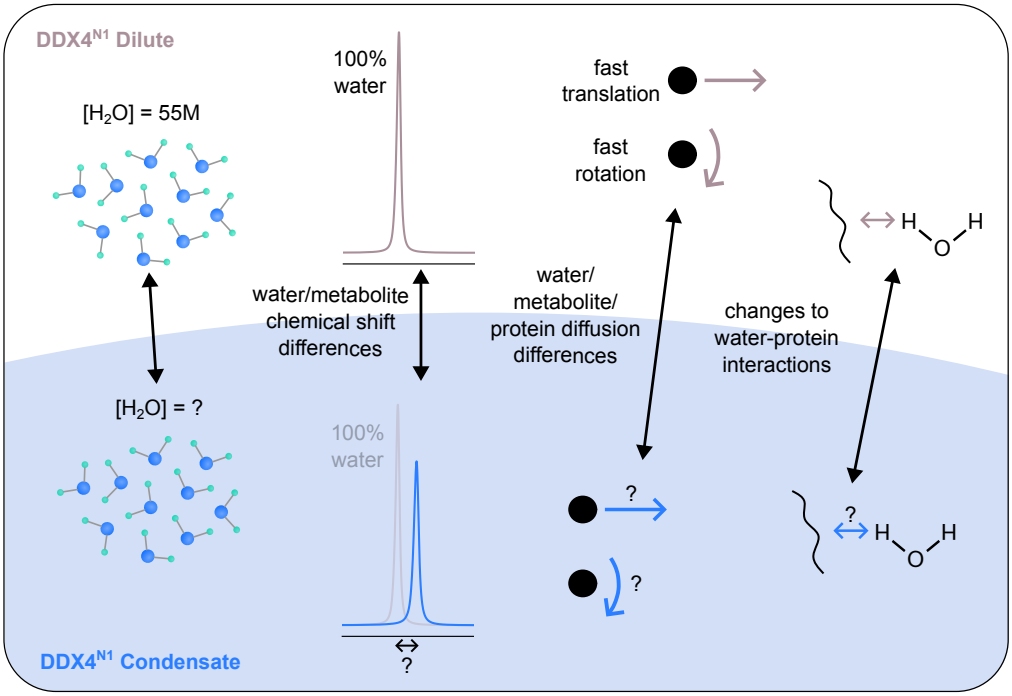


Figure 3.29: An illustration showing the solvent properties of condensates probed in this section using a biphasic NMR sample of DDX4^{N1} condensates. Knowledge of these properties inside DDX4^{N1} condensates will help to understand the alternative solvent environment within condensates.

3.5.3 Magnetic Resonance Imaging (MRI) experiments

Magnetic resonance imaging experiments, as described in section 2.2, were recorded on the biphasic DDX4^{N1} sample to determine the relative proportion of water between the DDX4^{N1} dilute and condensate phases (Figure 3.30A).*

Signal intensity from a given position in an MRI experiment is dependent on the concentration of spins at that position and any transverse relaxation that occurs during the echo time (Figure 2.10C). It was assumed that the majority of the signal intensity from the imaging experiment came from the water protons within the system. Furthermore, the short echo times (3.5 ms) used in the gradient echo imaging experiments ensured that the signal intensity modulation by transverse relaxation during the echo time was minimised. Therefore, it was assumed that the signal intensity from each phase was unaltered by relaxation, allowing the ratio in signal intensities to approximate the difference in water populations between the phases. The fractional water percentages obtained in the DDX4^{N1} condensate phase (between 66% and 76% of the dilute phase) are consistent with a previous estimate of 73% by Brady *et al.*[45] This previous estimate was obtained by comparing the integral of the water peak in an NMR sample containing only DDX4^{N1} condensate phase to the water peak integral in a buffer-only sample.[45]

The resulting spectra showed a significant reduction in water concentration in the pellet (condensate) phase, consistent with the results obtained on a biphasic tau condensate sample[224] and with simulations.[296] Furthermore, previous measurements of the concentration of DDX4^{N1} in the condensed phase were measured to be in the region of 300-400 mg/mL in solutions containing 100 mM NaCl.[45] The large proportion of protein in the condensed phase was accompanied by a reduction in water concentration. The ratio of the intensities of water in the supernatant (s) and condensate/pellet (c) phases was equal to 0.66 at 6^{o}C, marginally increasing to 0.76

*These MRI experiments were recorded in as an equal collaboration between myself and Dr. Charles Buchanan.

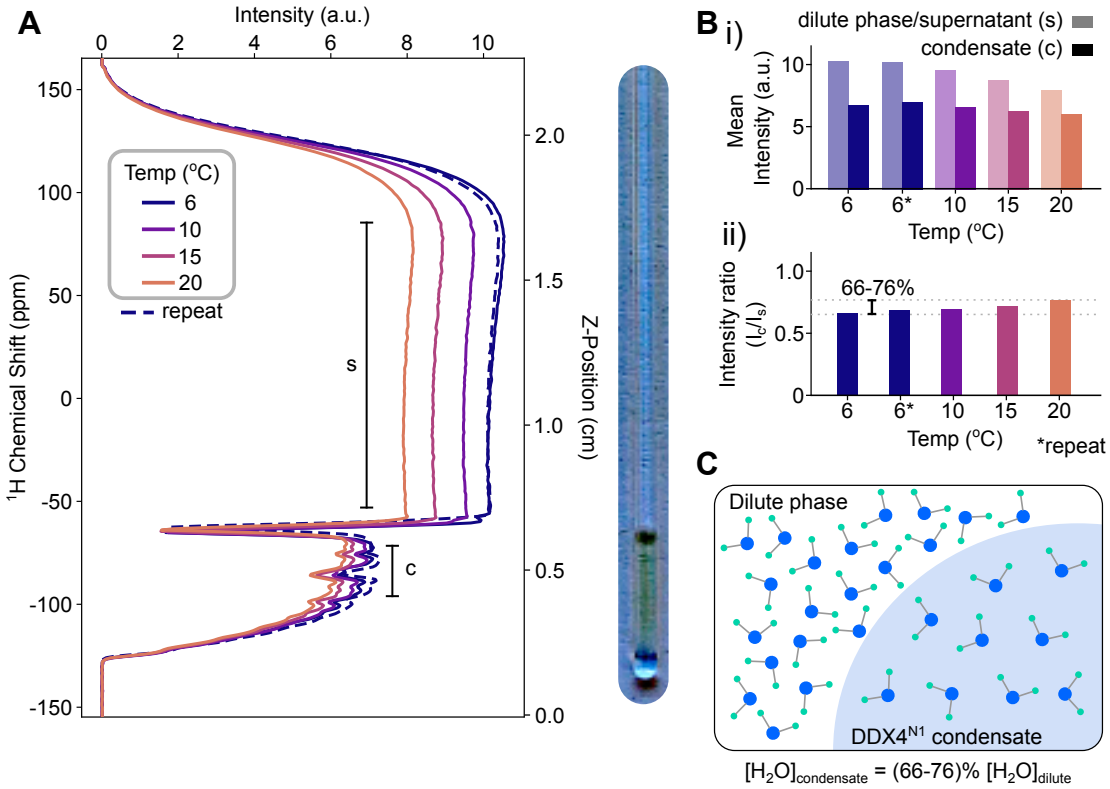


Figure 3.30: MRI spectra recorded on the biphasic sample of DDX4^{N1}. **A** The proton gradient-echo (GRE) MRI spectra recorded on the DDX4^{N1} biphasic sample as a function of temperature. The plots show the intensity of water protons as a function of chemical shift, and therefore z-position, within the sample. After increasing the temperature from 6°C to 20°C, the temperature was reduced to 6°C again, and the measurement was repeated. **B** i) The mean water intensity in the supernatant (s) and condensate (c) portions of the MRI spectra in A as a function of temperature. ii) The ratio of mean water intensity in the condensate and supernatant portions of the MRI spectra (I_c/I_s). **C** An illustration of the results obtained from the MRI spectra showing that the concentration of water in the condensate is between 66% and 76% of the concentration of the dilute phase (assumed to be 55M).

at 20°C (Figure 3.30B). The significant reduction in the water concentration inside the DDX4^{N1} condensate phase, coupled with a large DDX4^{N1} protein concentration, is hypothesised to provide a more organic-like chemical environment (Figure 3.30C).

The inhomogeneity of MRI signal in the condensate phase (Figure 3.30) could be caused by several factors. The first is that the water proton spin density is not homogeneous across the condensate phase. Furthermore, it is possible that the B_0 field is not completely homogeneous across the condensate phase. Due to the high viscosity within the condensate phase, there may be tiny air bubbles trapped

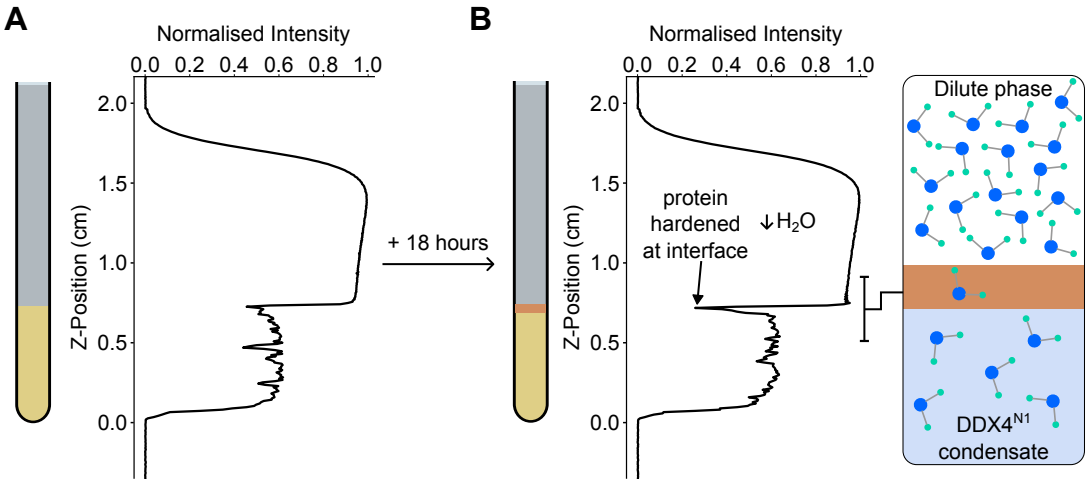


Figure 3.31: Time dependence of the biphasic DDX4^{N1} sample interface hardening. **A** An initial MRI experiment recorded on a fresh biphasic DDX4^{N1} sample showing a sharp interface between the condensate and dilute phase (similar to a Gibbs dividing surface). **B** After 18 hours the MRI signal at the interface of the condensate and dilute phase had changed significantly with a considerable decrease in mobile water signal at the interface due to hardening of the DDX4^{N1} protein. Note: The MRI spectra in A and B were recorded at 6^{o}C and 25^{o}C respectively, with their intensities normalised to improve clarity.

within it, slightly affecting the magnetic susceptibility and MRI signal intensity as a function of position. Finally, the B_1 field may not be completely homogeneous across the condensate phase, with local changes to the dielectric constant across the medium able to produce minor changes in MRI signal intensity.

The signal intensity of water was diminished at the interface between the two phases, coupled with the observation of visible hardening at the interface (Figure 3.30A). A second sample was created, and the MRI signal was monitored to study the time dependence of the loss of NMR signal and hardening at the interface. The normalised MRI spectrum recorded as soon as the DDX4^{N1} biphasic sample was prepared shows no significant drop in signal intensity at the interface between the phases (Figure 3.31A). However, after an additional 18 hours, the signal intensity at the interface had dropped (Figure 3.31B). DDX4^{N1} contains several cysteine residues capable of forming disulphide bridges, especially at high protein concentrations. The reducing agent TCEP was added to the original buffer to minimise the impact of disulphide bridge formation. However, TCEP becomes

less active over time ($t_{1/2}$ ≈72 hours in 0.15 M phosphate buffered saline) allowing disulphide bridge formation. Similar discoveries were made by Brady *et al.*[45], where Thioflavin T (ThT) fluorescence measurements ruled out the formation of amyloid during condensate maturation.

Recent studies on biomolecular condensates have been particularly interested in the behaviour of the interface between the concentrated and dilute phases with suggestions that the interface can act as an inherent catalyst.[297] Furthermore, there is emerging evidence that oxidation of cysteine residues in TDP-43 (TAR DNA-binding protein 43) on the interface of stress granules can induce pathological aggregation of TDP-43.[298]

3.5.4 Slice selective NMR experiments

It was desirable to independently obtain NMR spectra from each phase of the DDX4^{N1} biphasic sample to study the difference in NMR sensitive properties between the two phases. The slice selective NMR experiment (DPFGSE) introduced in section 2.2.2 requires the production of chemical shift-dependent NMR pulses, where under the action of a magnetic field gradient chemical shift is proportional to position (Figure 2.11B).

Seedless[279] is a tool to optimise pulse performance for specialised operations and can be optimised to hit specific chemical shift regions. The Seedless[279] program uses a gradient ascent pulse engineering (GRAPE)[299] algorithm to optimise the pulses by providing phase modulation at constant pulse amplitude to achieve the desired operation. The desired operations to optimise were unitary 180^o_x pulses applied only to the supernatant or pellet phases (Figure 3.32A), inverting any y/z magnetisation in only the selected region. The selective pulses for each selected slice were optimised using a 2000 μs pulse duration, with simulations showing the supernatant selective (Figure 3.32Bi) and pellet selective (Figure 3.32Ci) pulses were highly specific. The positional selectivity of each pulse was demonstrated experimentally through an MRI

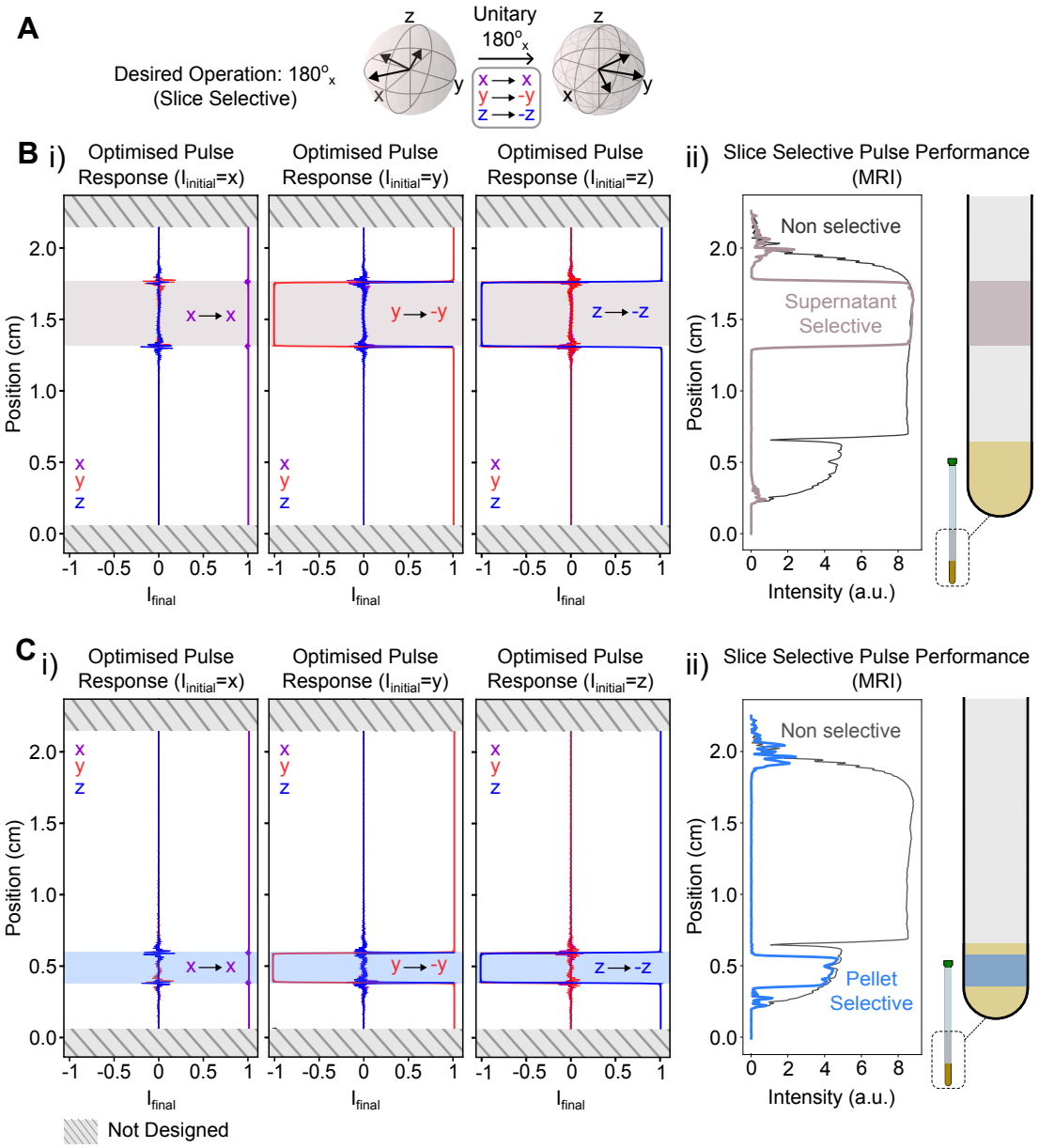


Figure 3.32: Optimisation and performance of supernatant and pellet slice selective pulses. **A** The desired operation was to perform a slice selective unitary 180^o pulse about the x-axis, where all x-magnetisation remains unchanged, and all y/z magnetisation is inverted. The supernatant (**B**i) and pellet (**C**i) slice selective pulses were created using optimal control (Seedless[279]), where the shaded areas denote the desired supernatant/pellet regions to be selected. **B**/**C**ii) The positional selectively of the pulses could be viewed directly on an MRI plot when inserted into the DPFGSE slice selective NMR experiment (Figure 2.11B) containing a gradient-echo (GRE) experiment (Figure 2.10C) added to the end.

spectrum after insertion into the slice-selective NMR experiment (Figure 2.11B) followed by a gradient-echo (GRE) experiment (Figure 2.10C). The results showed

that the pulses achieved high selectivity (Figure 3.32Bii/Cii).* Imperfections at the top and bottom of the sample were assumed to be caused by gradient imperfections when nearing the boundaries of the NMR coil volume.

3.5.5 Chemical shift differences of water inside DDX4^{N1} condensates

The slice selective NMR pulses applied to the DPFGSE slice selective pulse sequence (Figure 2.11B) enabled the 1D proton spectra to be obtained for the supernatant and condensate phases independently (Figure 3.33A). Interestingly, a reduction in the chemical shift of H_2O was observed in the pellet/condensate. This is in agreement with Raman spectroscopy experiments, which showed an altered hydrogen-bonded network of water in FUS condensates.[47] Similarly, Pantoja *et al.*[224] observed a decrease in the chemical shift of the reference standard DSS when inside a tau condensate. A second (smaller) peak was observed in the pellet selective spectrum, aligning with the supernatant H_2O signal. The origin of the second peak in the pellet selective spectrum could include the presence of water in two chemical environments in slow exchange. However, it could be caused by imperfections in the spectrometer gradients, leading to a small quantity of supernatant water being detected. Producing a sample with a larger condensate phase (which would fill the entire NMR coil volume) would enable this distinction to be made.

The decrease in chemical shift indicates enhanced shielding of the H_2O protons, caused by reduced hydrogen bonding strength (Figure 3.33B). Furthermore, the difference in chemical shift between the two phases decreased with increasing temperature (Figure 3.33C). This indicated that with enhanced temperature, the hydrogen bonding network of water in the condensate phase became more similar

*The slice selective pulses were calculated by Dr. Charles Buchanan and implementation of the DPFGSE pulse sequence containing the slice selective pulses was performed in equal collaboration with Dr. Charles Buchanan.

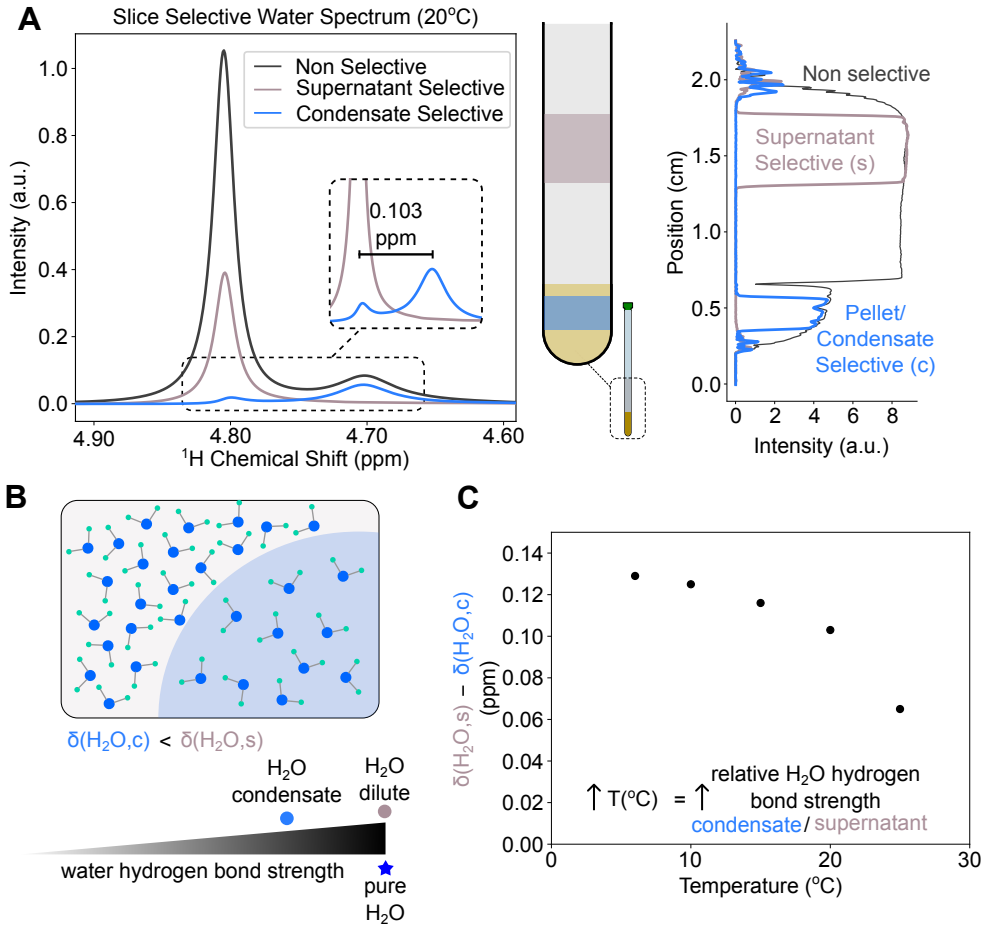


Figure 3.33: Water chemical shift changes in DDX4^{N1} condensates. **A** The NMR spectra of water obtained using non-selective, supernatant selective and pellet selective unitary 180^o pulse in the DPFGSE experiment (Figure 2.11B) at 20^{o}C. Most of the water signal in the condensate (pellet) phase is 0.103 ppm more shielded than in the supernatant phase. **B** The reduction in chemical shift (δ) of water inside the DDX4^{N1} condensate indicates the hydrogen bonding network of water is substantially disrupted inside the condensate. **C** The difference in supernatant and pellet water chemical shift as a function of temperature.

to the dilute phase. Extrapolation of the chemical shift difference evolution with temperature predicts that the difference will reach 0.00 ppm at a temperature of 35^{o}C.

Recent measurements have shown the pH inside other condensate types can be up to approximately 1 pH unit different from the supernatant phase.[155] Measurements performed on the pH dependence of H_2O chemical shift* (buffer components were 20 mM NaPi, 5 mM DSS, 100% D_2O) show that a reduction in the pH from 8.0

*Measurements performed by Ben Naylor (Baldwin lab).

to 7.0 causes only a small (<0.01 ppm) reduction in the observed water chemical shift at 20^{o}C when referenced to DSS. Therefore, assuming the condensate phase's pH is not substantially more than 1 pH unit lower than the dilute phase, the large chemical shift changes observed cannot be explained by pH alone.

The calibrated 90^o pulse durations in the supernatant and pellet phases were 7.88 μs and 8.35 μs at a temperature of 6^{o}C, respectively. A 6% increase in the calibrated pulse duration for the condensate phase is indicative of a decrease in the relative permittivity (ϵ_r), also known as the dielectric constant, with the pulse duration proportional to $1/\sqrt{\epsilon_r}$.[300] However, due to the substantial quantity of glass used in the assembly (Figure 3.28), a quantitative description of the change in ϵ_r between the phases could not be obtained. Bruker produce MicroProbes (1 mm TXI, 1.7 mm TXI) which could be used directly on the fragile 1 mm biphasic sample without the requirement of nested tubes in the future to quantify this more accurately. The decrease in ϵ_r in the DDX4^{N1} condensate phase is in agreement with previous estimates (Figure 3.7A),[44] but could also indicate an enhanced concentration of NaCl or other ions.[301] This would be in agreement with sodium ion partitioning measurements showing enrichment into DDX4^{N1} condensates (Figure 3.24).

3.5.6 Comparison of water chemical shift changes to that in organic solvents

To assess the ability of DDX4^{N1} condensates to act as an alternative solvent, the chemical shift of H_2O in the DDX4^{N1} condensate phase at 25^{o}C was compared to that observed for H_2O in different percentages (by volume) of the organic solvents ethanol, acetone, DMSO and acetonitrile (Figure 3.34A). The MRI results showed previously (Figure 3.30) indicate that at 25^{o}C, the water concentration in the condensate phase

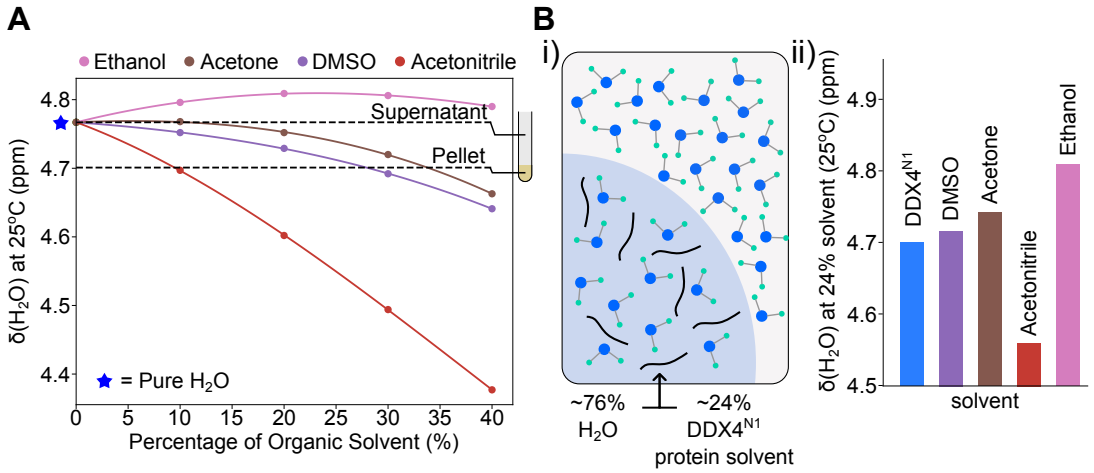


Figure 3.34: Comparison of water chemical shift changes in DDX4^{N1} condensates to organic solvents.**A** A plot of the water chemical shift in different percentages (by volume) of a selection of organic solvents at 25^{o}C. The supernatant and condensate water chemical shifts at 25^{o}C are shown as the dotted lines. **B** i) Illustration showing that the water/DDX4^{N1} percentage volumes in the condensate phase at 25^{o}C are approximately 76% and 24%, respectively, as revealed by the previous MRI results (Figure 3.30). ii) A plot showing water chemical shift in solutions containing 24% solvent by volume for the protein and organic solvents. The results show that DDX4^{N1} protein in the condensate has a similar effect to water per unit volume as DMSO.

is reduced by approximately 24%*. Assuming that the total number density of atoms (number of atoms per unit volume) for each phase is the same, this implies that DDX4^{N1} protein accounts for approximately 24% of the volume in DDX4^{N1} condensates (Figure 3.34Bi). Therefore, the effect of 24% by volume DDX4^{N1} on the water chemical shift could be compared to that in water-organic solvent mixtures containing 24% (by volume) organic solvent (Figure 3.34Bii). The comparison highlights that DMSO has the most similar effect on the water chemical shift per unit volume to DDX4^{N1} condensates. This is in agreement with previous estimations of the dielectric constant inside DDX4^{N1} condensates, predicted to be most similar to DMSO (Figure 3.7A).[44] The similarities between DDX4^{N1} condensates and the organic solvent DMSO on the properties of water indicate that DDX4^{N1} condensates can be regarded as an alternative solvent within cells.

*The MRI measurements in Figure 3.30 were not recorded at 25^{o}C; therefore the values at 25^{o}C are being approximated to the same as the values at 20^{o}C.

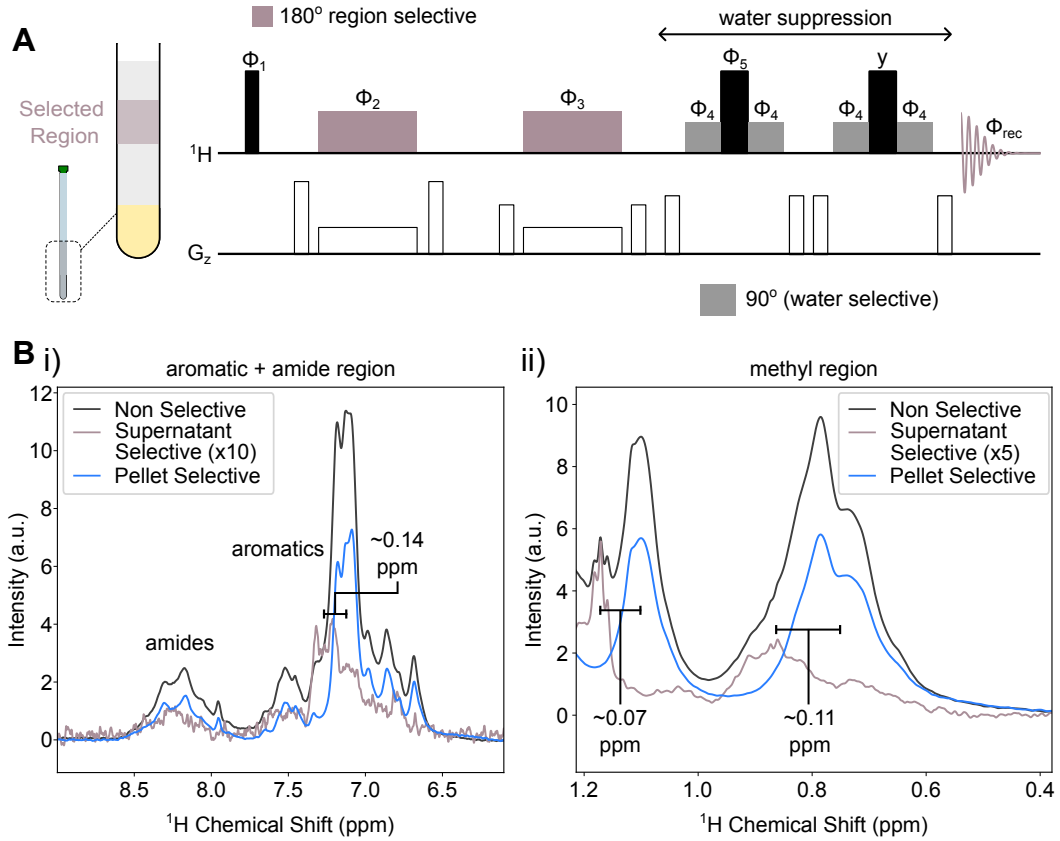


Figure 3.35: Protein chemical shift changes inside DDX4^{N1} condensates. **A** Pulse sequence of a water-suppressed slice selective NMR experiment produced by combining the DPFGSE sequence (Figure 2.11B) with two successive watergate elements. The water selective 90^o pulses were long-duration non-shaped pulses centred on water. Thin and thick black rectangles represent non-selective 90^o and 180^o pulses, respectively. The phase cycle is as follows: Φ_1 = (x,-x,x,-x,x,-x,x,-x), Φ_2 = (x,-y,-x,y,x,y,-x,-y), Φ_3 = x, Φ_4 = (-x,-y,y,x,-x,-y,y,x), Φ_5 = (x,y,-y,-x,x,y,-y,-x), Φ_{rec} = (x,-x,-x,x,x,-x,-x,x). **B** Spectral differences of amide/aromatic (i) and methyl (ii) protons between the supernatant and pellet phases at a temperature of 20^{o}C. Chemical shift differences are displayed as a guide and are not quantitative due to differences in peak shapes between the phases, making quantification difficult. Supernatant spectral intensities are scaled for easier visual comparison to the pellet/non-selective spectra.

3.5.7 Chemical shift differences of DDX4^{N1} and metabolites inside DDX4^{N1} condensates

To establish whether chemical shift differences are also observed for DDX4^{N1} when in the DDX4^{N1} condensate phase, a water-suppressed variant of the DPFGSE slice

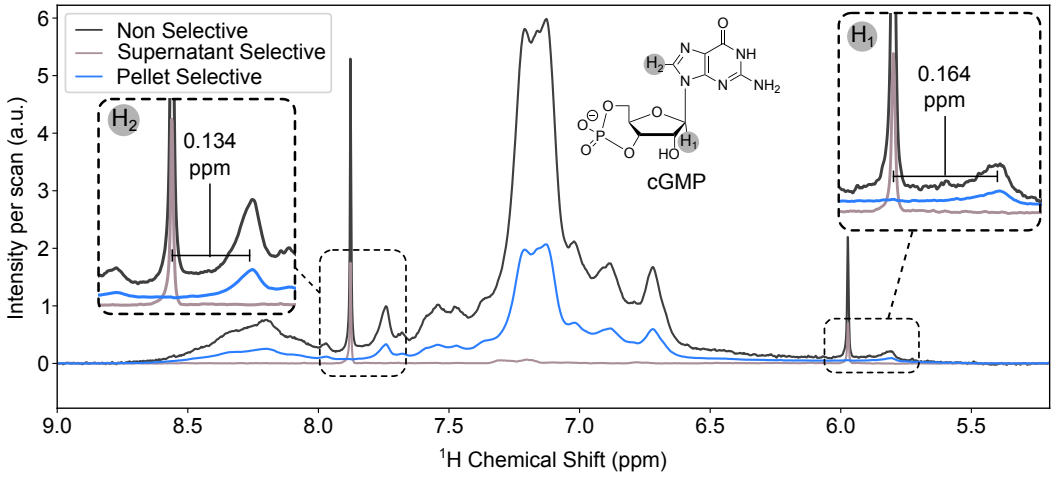


Figure 3.36: Chemical shift differences of protons in the metabolite cGMP (cyclic guanosine monophosphate) in a biphasic DDX4^{N1} condensate, prepared as described in Figure 3.28 with the addition of 10 mM cGMP.

selective experiment was created (Figure 3.35A).* The sequence ends with a double WATERGATE (Water suppression through gradient-tailored excitation)[302] element for removing as much water signal intensity as possible.

Substantial reductions in the chemical shift of the aromatic (Figure 3.35Bi) and methyl group (Figure 3.35Bii) protons were observed at 20^{o}C in the DDX4^{N1} condensate phase, indicating enhanced shielding. However, exact chemical shift differences could not be obtained because the observed peak shapes of the methyl and aromatic proton peaks were altered between the supernatant and pellet phases. The reduced chemical shift of methyl group protons in the condensate phase is evidence of ring-shifts caused by aromatic-methyl group interactions. In contrast to the aromatic and methyl group protons, the amide chemical shifts were altered to a much lesser degree between the phases (Figure 3.35Bi). Unlike the water protons, the chemical shift changes in the aromatic and methyl group protons persisted as the temperature was increased from 6^{o}C to 25^{o}C, indicating that the cause of the shielding effects on these protons was maintained as a function of temperature. In section 3.5.9, protein-protein interactions capable of the observed shielding effects

*The water-suppressed variant of the DPFGSE slice selective experiment was implemented in collaboration with Dr. Charles Buchanan.

inside DDX4^{N1} condensate are discussed.

Measurements on a second biphasic sample, also containing 10 mM of the metabolite cyclic guanosine monophosphate (cGMP), were performed to establish differences in its proton chemical shifts within the DDX4^{N1} condensate phase. cGMP is a highly important secondary messenger in biological systems,[282] that can interact with ion channels to relay extracellular information provided by neurotransmitters into the cell.[260] It has links to photon recognition and vision, as well as allowing smooth muscle cells to relax, increasing flow through a blood vessel.[260] Therefore, any changes to the chemical environment and dynamics experienced by cGMP could have a substantial impact on its function. Additionally, cGMP was determined to be enriched into DDX4^{N1} condensates, with a partitioning coefficient of 3.6±0.5. The chemical shift of two of the protons in cGMP, one on the ribose sugar ring and another on the adenine base, were both more shielded in the DDX4^{N1} condensate phase with respect to the dilute phase to different degrees (Figure 3.36). The resonances of the metabolite from the condensate phase are also substantially broadened, indicating an increased relaxation rate and slower molecular tumbling in the condensed phase.

Altogether, the protons of water, protein (amides, aromatic, aliphatic) and the metabolite cGMP are all shielded to different degrees within the DDX4^{N1} condensate phase, indicating that the protein is providing a solvent environment capable of shielding and increasing the relative local electron density around protons within it. Further measurements such as enzymatic kinetic studies and enzymatic structure determination in the condensate phase are required to link the more shielded solvent environment experienced inside condensates to possible functions biochemical.

3.5.8 Water-protein interactions in condensates revealed by CEST NMR

To determine a potential reason for the water chemical shift reduction inside the $DDX4^{N1}$ condensate phase, CEST experiments (as described in section 2.6.1) were performed.*

Due to the separation of the chemical shifts of H_2O between the biphasic sample phases, the effect of the saturation pulse on the intensity of the condensate (Figure 3.37Ai) and supernatant (Figure 3.37Ai) H_2O peaks could be monitored as a function of saturation (CEST) pulse chemical shift. Normalising the intensities recorded for the condensate and supernatant H_2O with the corresponding reference spectrum (without the application of a saturation pulse) enabled the production of a normalised CEST profile, also known as a Z-spectrum (Figure 3.37B). The normalised CEST profile showed significant signal losses when the saturation pulse was centred on the amide, aromatic and aliphatic protons on the $DDX4^{N1}$. In contrast, the same signal attenuation in the amide, aromatic and aliphatic proton region was not observed for the corresponding supernatant CEST profile. This indicates that there are significant $DDX4^{N1}$ protein - water interactions in the $DDX4^{N1}$ condensate phase (Figure 3.37C), possibly linked to the altered solvent environment experienced by water.

The CEST experiment was repeated at temperatures between 6^oC-20^oC (Figure 3.38Ai) to determine whether the observed protein-water interactions in the condensate phase correlated with the temperature dependent reductions of water chemical shift in the condensate phase that were observed (Figure 3.33). The mean and standard deviation of the intensities of the condensate H_2O CEST profiles for the amide, aromatic and aliphatic regions of the spectra were obtained

*The CEST experiment employed was a coupled CEST-stimulated echo diffusion experiment with the encode/decode diffusion gradients turned off to ensure only the CEST element remained active. More details can be found in appendix method A.5.4

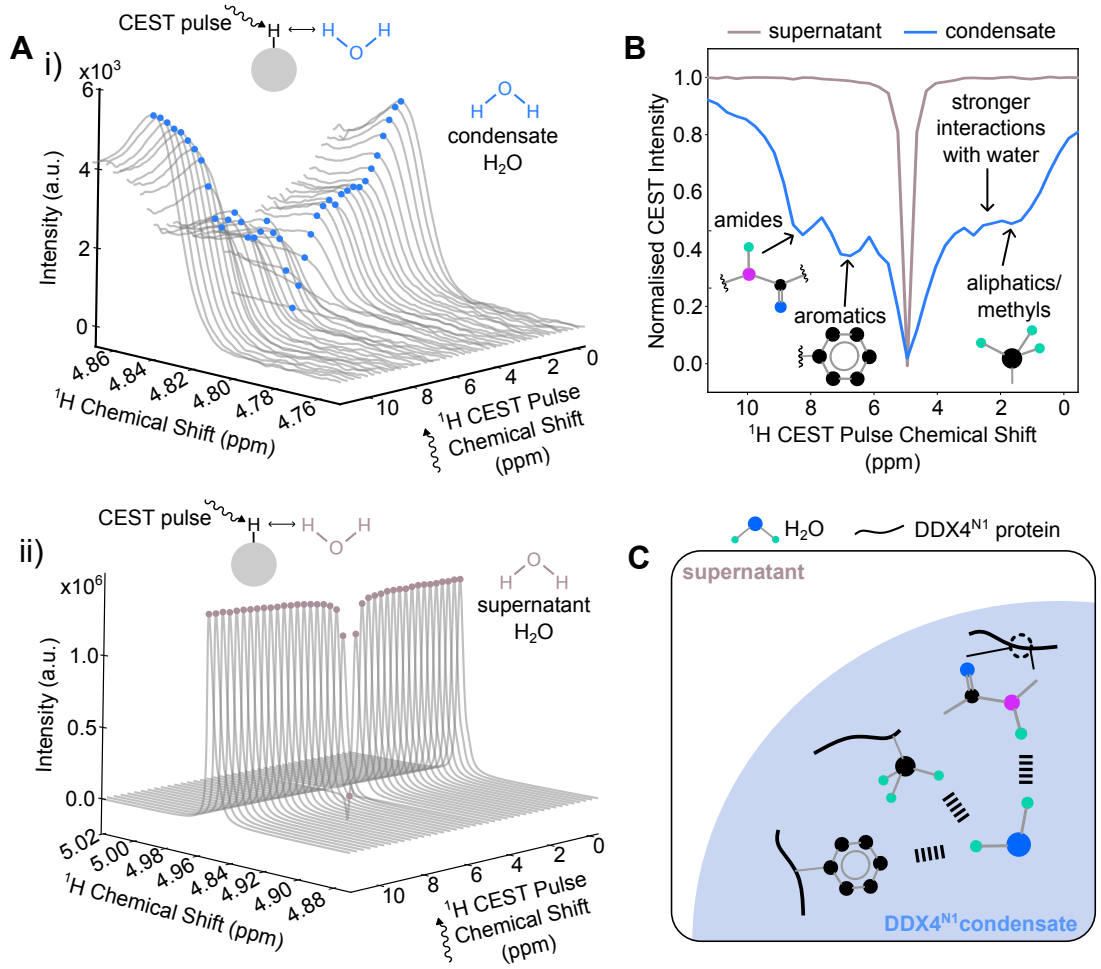


Figure 3.37: Water-CEST profiles for supernatant and condensate water. **A** NMR spectra of condensate H_2O (i) and supernatant H_2O (ii) as a function of the CEST pulse chemical shift obtained on the DDX4^{N1} biphasic sample at 6^{o}C. The dots represent the water intensity at the condensate or supernatant H_2O chemical shifts and generate a CEST profile for each system. **B** An overlay of the normalised CEST profiles (Z-spectra) for supernatant and condensate H_2O. **C** An illustration showing condensate H_2O interacting with amide, aromatic and aliphatic protons from DDX4^{N1}, as determined in B.

and fitted linearly against temperature (Figure 3.38Aii). Interestingly, as the temperature increased, the interaction between the aliphatic protons and the water was diminished most significantly. Furthermore, a correlation between the normalised CEST intensity in the amide, aromatic and aliphatic spectral regions for condensate H_2O and the differences in water chemical shift between the two phases could be established (Figure 3.38B). The interaction of condensate H_2O with aliphatic DDX4^{N1} protons appeared to have the most substantial impact on the

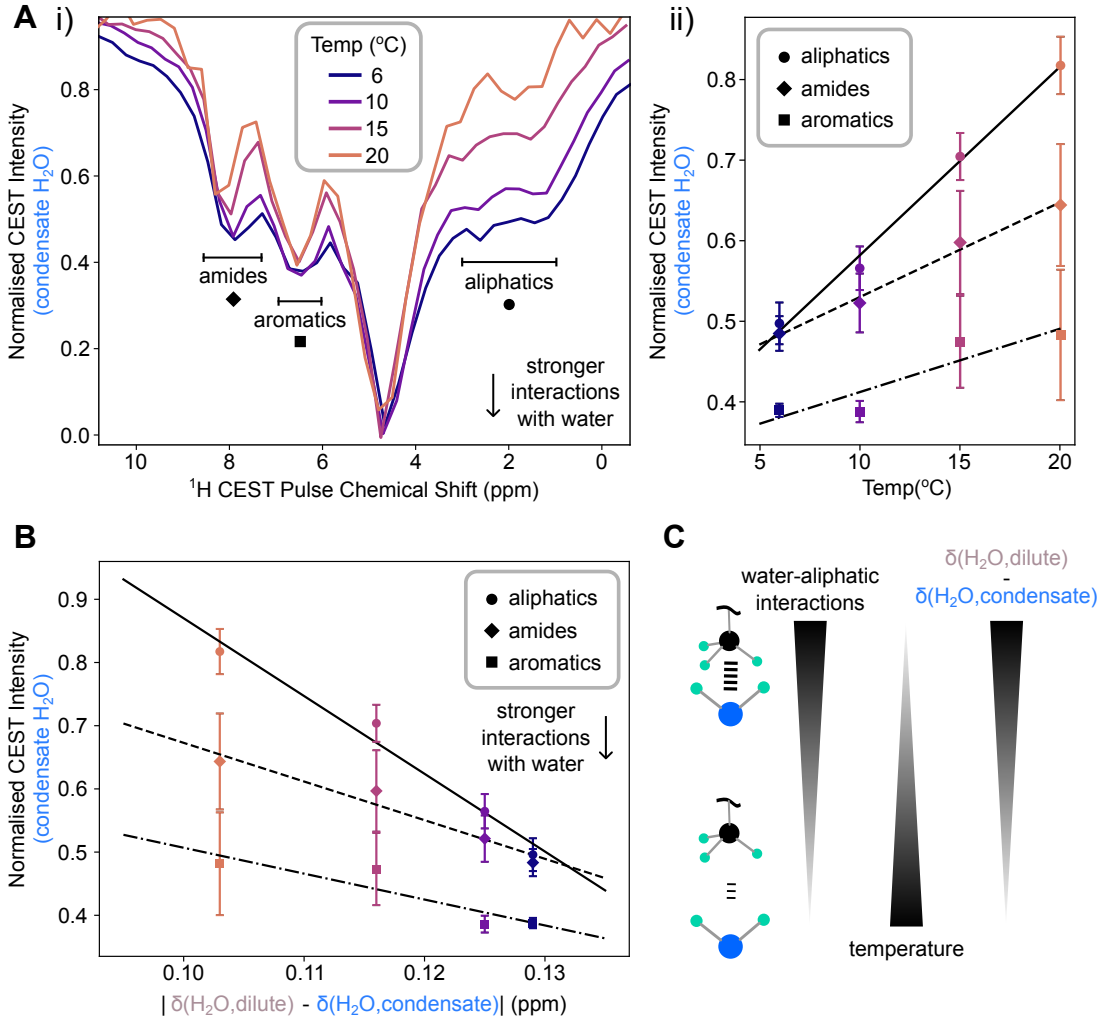


Figure 3.38: Temperature dependence of water-CEST profiles for condensate water.**A** i) An overlay of the condensate H_2O CEST profiles obtained on the DDX4^{N1} biphasic sample as a function of temperature. ii) A plot showing the linear increase of the normalised CEST intensities for the CEST profile's aliphatic, amide and aromatic regions on condensate H_2O as a function of temperature. **B** A plot showing the linear correlation between condensate H_2O normalised CEST intensity in the amide, aromatic and aliphatic spectral regions and the difference in water chemical shift between the dilute and condensate phases. **C** A schematic depicting how temperature, and therefore water-aliphatic interactions, are correlated with differences in water chemical shift between the dilute and condensate phases of the DDX4^{N1} biphasic sample.

solvent environment and chemical shift it experienced (Figure 3.38B/C), implying that water-aliphatic interactions are a major contributor to the shielding of water protons in the DDX4^{N1} condensate phase.

Similar CEST profiles of water showing strong water-protein interactions have

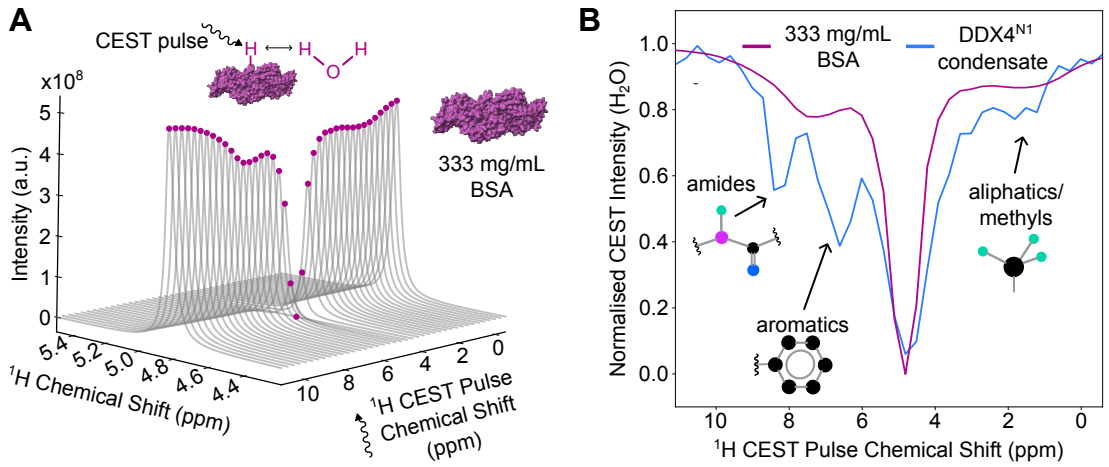


Figure 3.39: Water-CEST profile in concentrated BSA. **A** The intensities of the water peak in a 333 mg/mL sample of BSA as a function of saturation pulse location recorded at 20^{o}C. **B** An overlay of the normalised CEST profile of water in 333 mg/mL BSA and the DDX4^{N1} condensate.

been shown using CEST MRI experiments on the cortex of rat brains.[303] Therefore, it was interesting to question whether the CEST responses observed in Figure 3.37 are purely a result of high protein concentration in the DDX4^{N1} condensate phase. A concentrated sample (333 mg/mL, 5 mM) of Bovine Serum Albumin (BSA) was prepared in a solution containing 10% D_2O to act as a control solution containing high protein concentration in the absence of phase separation. This corresponds to a total amino acid residue molarity of 2.9M, given the number of amino acids per BSA unit is 583. This is similar in magnitude to estimates of the intra-condensate concentration of DDX4^{N1} (300-400 mg/mL, 11.8-15.7 mM) at the same salt concentration as the DDX4^{N1} biphasic NMR sample.[45] Therefore, the intra-DDX4^{N1} condensate concentration of amino acid residues is estimated as between 2.7 and 3.7 M.

The CEST profile of water in the concentrated BSA sample was recorded at 20^{o}C and compared to that observed for the DDX4^{N1} condensate phase. Interactions between H_2O and BSA are apparent but are much less intense than in the DDX4^{N1} condensate. The profiles produced for BSA are visually similar to previous CEST measurements on concentrated BSA.[303] The significant proportion of secondary and

tertiary structure in BSA means that many of the amino acid residues will not be in contact with H_2O, thereby diminishing the observed protein-water interactions per unit concentration. Furthermore, many of the hydrophobic (aliphatic) residues will be buried within the structure, preventing their interactions with the surrounding water molecules. Many cellular biomolecular condensates are formed from disordered proteins.[192,211,213] Perhaps the benefit to biology forming condensates containing disordered proteins is their ability to modify the solvent environment within the condensate more efficiently through a larger number of available interaction sites than would be the case for a structured protein undergoing phase separation.

3.5.9 NOESY measurements on DDX4^{N1} biphasic sample

A ^{1}H-^{1}H NOESY experiment (as explained in section 2.6.3) enables inter- and intra-molecular contacts to be determined within a protein. The water-suppressed NOESY sequence† in Figure 3.40A was applied to the biphasic DDX4^{N1} NMR sample (containing 10 mM cGMP). The experiment was recorded at 25^{o}C using a mixing time τ of 200 ms.

The experiment revealed the presence of cross peaks between the protein proton resonances, specifically stemming from the aromatic protons (Figure 3.40B). This is in agreement with previous NOESY measurements on a sample containing just the DDX4^{N1} condensed phase, whereby strong intermolecular contacts were determined and dominated by the residue phenylalanine.[45] Most NOESY cross-peaks observed had the same sign as the diagonal peak, a hallmark of a slow tumbling system (section 2.6.3). Furthermore, NOESY cross-peaks are observed between the aromatic and methyl group protons, indicating the reduced chemical shift of the methyl group protons (Figure 3.35Bii) could be caused by ring shifts. However, due to the high likelihood of spin diffusion caused by the slow tumbling of DDX4^{N1} in the condensate phase, the NOESY cross-peaks only provide qualitative information

†The NOESY sequence was implemented in equal collaboration with Dr. Charles Buchanan.

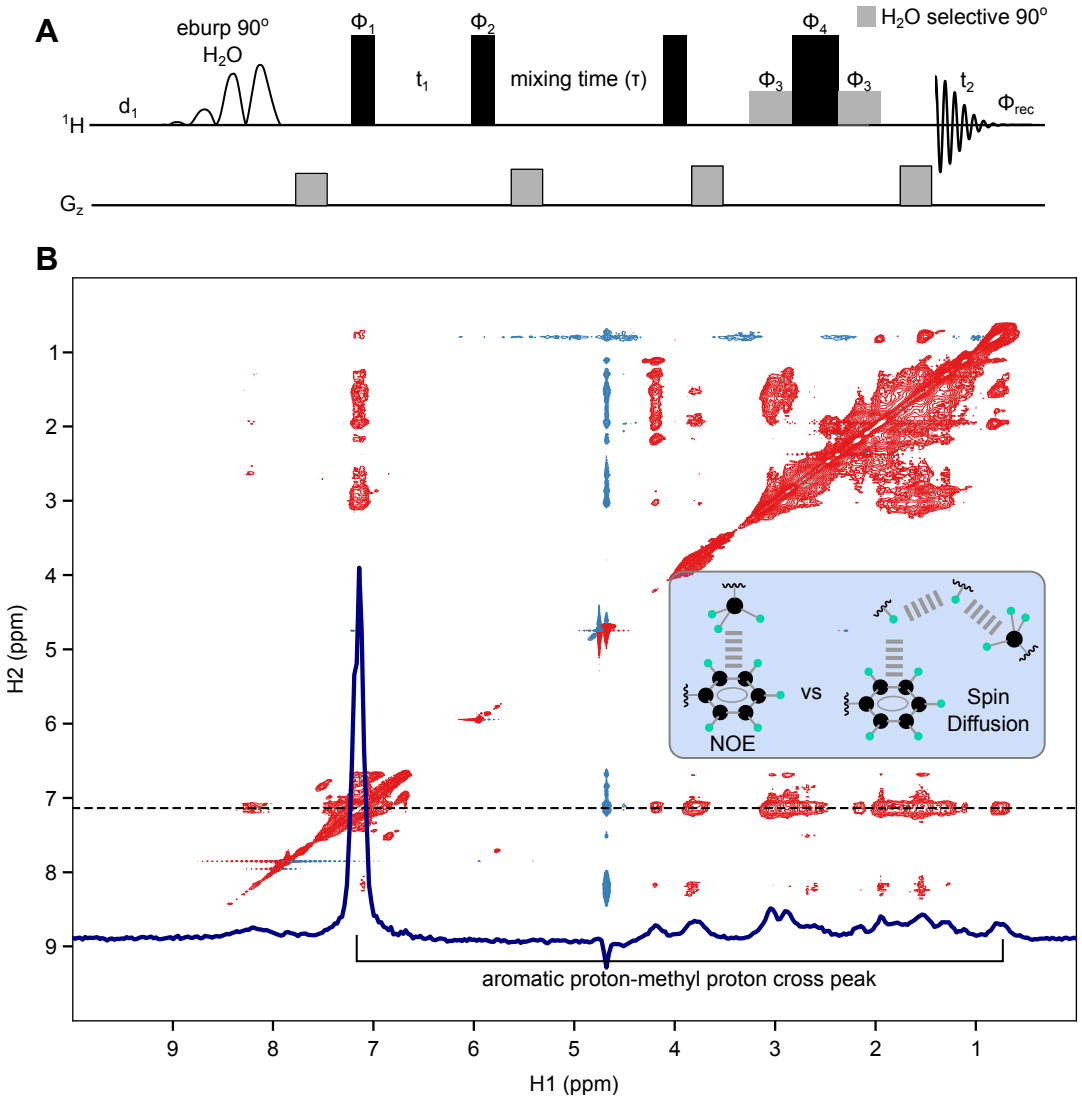


Figure 3.40: NOESY spectrum of a biphasic DDX4^{N1} sample. **A** A ^{1}H-^{1}H NOESY pulse sequence. Thin and thick black rectangles represent non-selective 90^{o} and 180^{o} pulses, respectively. The mixing time (τ) was set to 200 ms. The following phase cycles were employed: Φ_1 = (x,-x,x,-x,x,-x,x,-x), Φ_2 = x, Φ_3 = (-x,-y,y,x,-x,-y,y,x),Φ_4 = (x,y,-y,-x,x,y,-y,-x), Φ_{rec} = (x,x,-x-x,x,x,-x,-x). Further experimental details on the NOESY experiment recorded are included in appendix method A.5.5. **B** The ^{1}H-^{1}H NOESY spectrum recorded on the biphasic DDX4^{N1} sample.

on protein-protein interactions.

Intermolecular NOESY cross-peaks could only be observed between the cGMP proton resonances in the condensate phase. NOESY cross-peaks could not be observed between cGMP and the protein indicating that cGMP is not significantly interacting with DDX4^{N1} in the condensed phase. However, given that cGMP is enriched into DDX4^{N1} condensates (Figure 3.17A) it is likely there will be

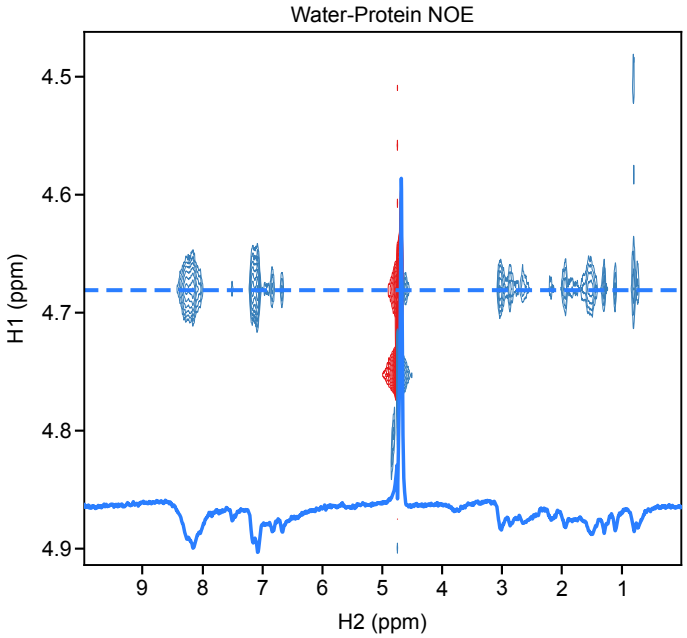


Figure 3.41: A zoom in on the NOESY cross peaks observed for the condensate H_2O resonance. This spectrum is transposed from that shown in Figure 3.40B.

weak multivalent interactions between cGMP and DDX4^{N1}. Additional saturation transfer experiments could be performed in the future to detect this binding and determine its strength.

The NOESY spectrum also contained evidence of cross peaks between the condensate H_2O and the protein (Figure 3.41) despite the application of water suppression during the pulse sequence. The application of water suppression prevents the cross-peaks from being interpreted quantitatively; however, the cross-peaks observed show consistency with the condensate water CEST measurements (Figure 3.38).

3.5.10 Diffusion and relaxation measurements of water inside condensates

The change in the chemical properties of water, such as NMR chemical shift, inside DDX4^{N1} condensates, indicates a change in the solvent environment. However, water has many other solvent-dependent physical properties, such as translational

and rotational diffusion, which need to be quantified to determine the full extent of the altered solvent environment within DDX4^{N1} condensates.

In this light, stimulated echo diffusion experiments (as explained in section 2.5) were recorded on the DDX4^{N1} biphasic sample (Figure 3.42A). The diffusion coefficients (cm^2/s) were calculated for the condensate and supernatant water species by fitting the experimental data to the Stejskal-Tanner equation.[20] There was roughly a 3× reduction of the diffusion coefficient of water within the condensate phase. This is consistent with the reduction in small molecule translational diffusion coefficients previously observed in the DDX4^{N1} condensed phase.[45] In addition, the diffusion measurements were repeated at a range of temperatures between 6^{o}C and 20^{o}C, with the diffusion coefficients showing a linear increase with increasing temperature (Figure 3.42Bi). In the translational Brownian motion of a spherical molecule, the diffusion coefficient can be calculated using the following expression:[304–306]

$$\mathrm{D} = \frac{\mathrm{k_B T}}{6\pi\eta(\mathrm{T})r} \tag{3.8}$$

where $\mathrm{k_B}$ is the Boltzmann constant, T is the temperature (K), η(T) is the dynamic viscosity of the solution (Pa s), and r is the radius of the spherical particle. This formula applies to non-complex systems where motion is unrestricted. A modified form is required in the case of complex liquids, whereby a length scale-dependent viscosity model (LSVM) is included, meaning that the viscosity is a function molecular radius ($\eta = \eta(\mathrm{T}, r)$), increasing as r is increased.[307] The hydrodynamic radii of small molecules in DDX4^{N1} condensates is assumed to same in both phases. Therefore, the relative effective viscosities in each phase of the DDX4^{N1} biphasic sample can be established (Figure 3.42Bii). The relative ratio in effective viscosity between the DDX4^{N1} dilute and concentrated phases equals ≈3 at 6^{o}C, linearly decreasing with increased temperature.

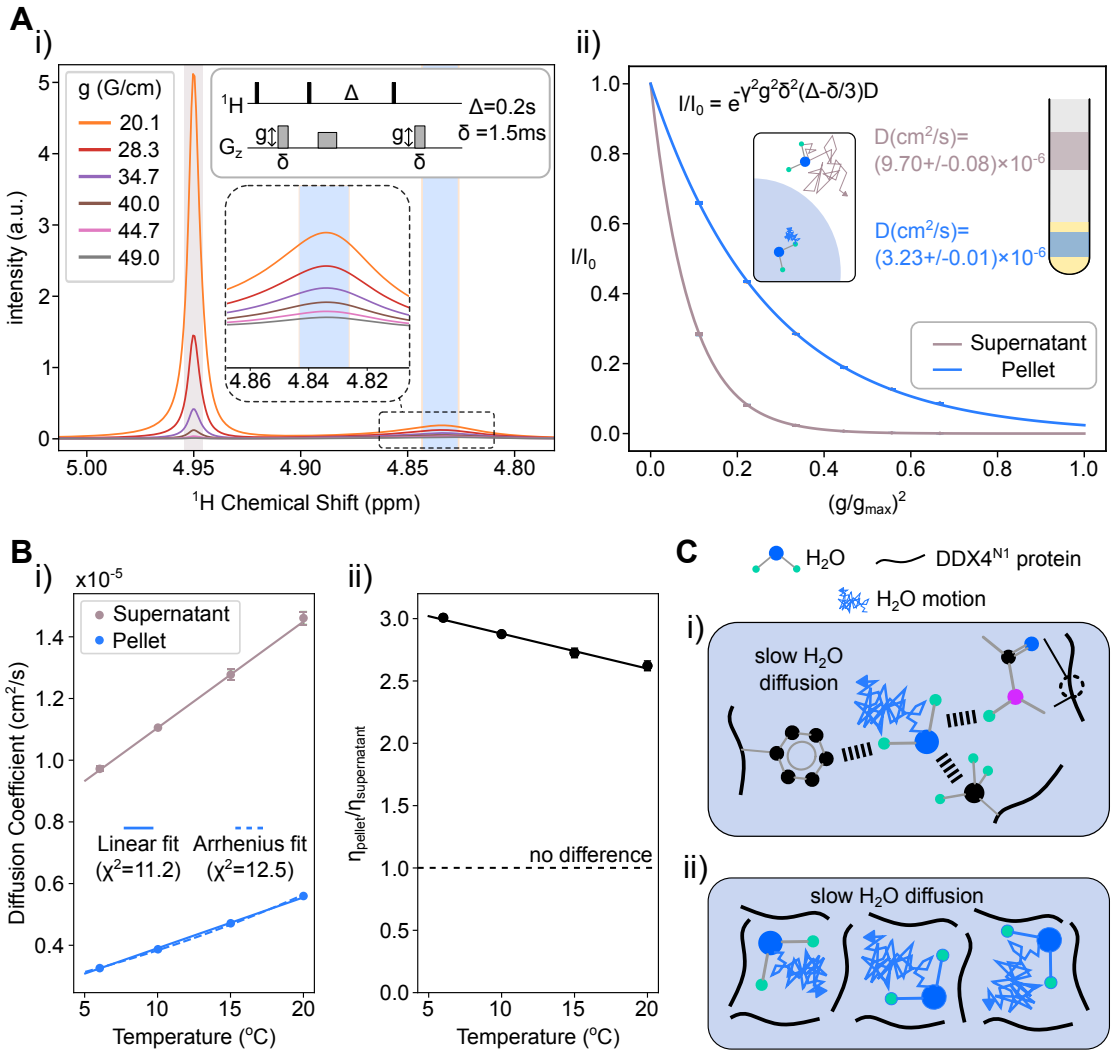


Figure 3.42: Water diffusion differences between DDX4^{N1} condensate and supernatant phases. **A** i) The NMR data for the stimulated echo longitudinal encode-decode diffusion experiment recorded on the DDX4^{N1} biphasic sample at 6^{o}C. A diffusion delay (Δ) of 0.2 s was used, with the encode/decode gradient durations (δ) set to 1.5 ms. ii) A fit of the supernatant/condensate NMR peak intensities vs gradient strength to the Stejskal-Tanner equation.[20] **B** i) Plots of water diffusion coefficients in the pellet/condensate and supernatant phases as a function of temperature. For the condensate phase, linear and Arrhenius fits are shown for the data, with only a linear fit included for the supernatant phase. ii) A plot of the ratio of effective viscosity $\eta(\mathrm{T}, r)$ for water between the pellet and supernatant phases as a function of temperature. **C** Potential causes of reduced water diffusion in DDX4^{N1} condensates through (i) strong water-protein interactions or (ii) restricted diffusion.

One hypothesis for the decrease in diffusion of water in the condensate is enhanced water-protein interactions that lead to an increased effective viscosity for water. The increase in water diffusion coefficient in the condensate phase

as the temperature is raised correlates with reduced water interactions with the protein (Figure 3.38), providing evidence for this mechanism. However, a second plausible reason for decreased water diffusion in a system is restricted diffusion.[308] Previous measurements on polymer coacervates have shown evidence of a porous internal environment.[309] In restricted diffusion, measurements of the diffusion coefficient in a system will depend on the temperature in an Arrhenius-like manner ($D = Ae^{-E_a/(RT)}$).[310] However, fitting the temperature dependence of the condensate water diffusion coefficient to a restricted diffusion (Arrhenius) model led to an increase in the χ^2 metric (a measure of goodness of fit)* when compared to a linear fit (Figure 3.42Bi). Although both linear and Arrhenius fits followed the correct trend. It is entirely possible that both mechanisms, shown in Figure 3.42C, are occurring in the condensate phase to reduce diffusion, with restricted diffusion being a minor contributor. To determine if a small portion of water is undergoing restricted diffusion or not within the DDX4^{N1} condensate phase, INDIANA (IN-cell DIffusion ANAlysis)[311] type experiments could be performed in the future.

In section 2.1.4, it was shown that NMR relaxation measurements can provide insight into molecular tumbling. In this light, water longitudinal relaxation rates (R_1) were measured and compared between the two phases (Figure 3.43A/B). Estimates of the water rotational correlation time (τ_c) in each phase were obtained (Figure 3.43C) by fitting the R_1 rates to the following equation:[82]

$$R_1 = \frac{3}{10}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2}{r_{HH}^6}\left(J(\omega_H) + 4J(2\omega_H)\right) \tag{3.9}$$

This equation represents the dipolar contribution to relaxation between the two H_2O protons. The constants μ_0, $\hbar$, γ_H and ω_H are the magnetic vacuum permeability, Planck's constant, the proton gyromagnetic ratio and the proton Larmor frequency. The spectral density $J(\omega)$ is a function dependent on the

*$\chi^2 = \sum_i (y_{\text{experimental,i}} - y_{\text{fit,i}})^2 / y^2_{\text{error,i}}$

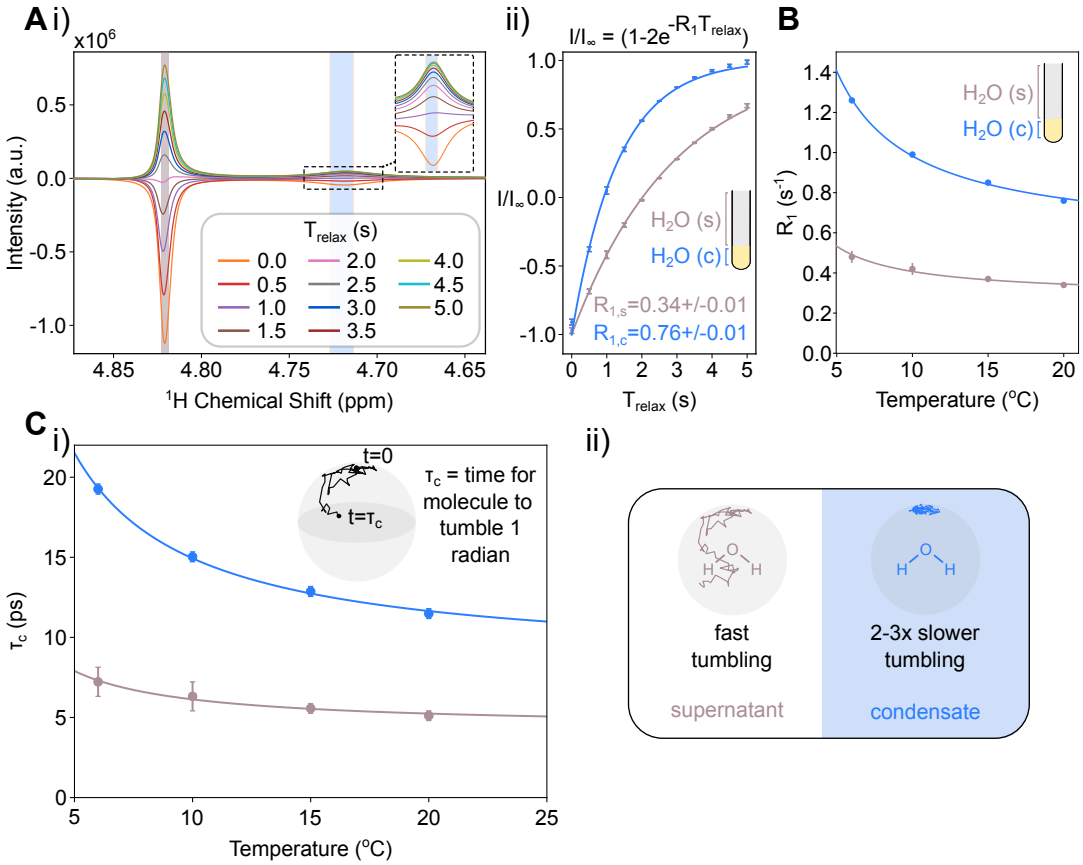


Figure 3.43: Water longitudinal relaxation differences between DDX4^{N1} condensate and supernatant H_2O. **A** i) Inversion recovery longitudinal relaxation data obtained for supernatant (s) and condensate (c) H_2O at 20^{o}C. ii) A fit of the peak intensities in (i) to the equation $I(T_{\text{relax}}) = I(\infty)[1 - 2e^{-R_1 T_{\text{relax}}}]$, enabling the longitudinal relaxation rate (R_1) to be determined for water in each phase. **B** The temperature dependence of the R_1 relaxation rate for the supernatant and condensate water species. Both condensate and supernatant water R_1 relaxation rates appear to follow a $R_1 = a + b/T(^oC)$ behaviour with temperature. **C** (i) Fitted rotational correlation times (τ_c) of water in the DDX4^{N1} condensate and supernatant H_2O phases as a function of temperature. (ii) Schematic showing 2-3× reduced water tumbling rate in the DDX4^{N1} condensate phase.

rotational correlation time as shown in equation 2.5 and r_{HH} is the internuclear proton distance, taken to be 1.53 Å.[312] The rotational correlation time of water (or effective rotational viscosity) in the condensate phase was 2.5-3× larger than in the supernatant phase (Figure 3.43C), matching the change in translational viscosity (Figure 3.42Bii). Furthermore, the rotational correlation time of water in the supernatant phase was similar in magnitude to literature values in pure water (4.4 ps at 275 K).[313]

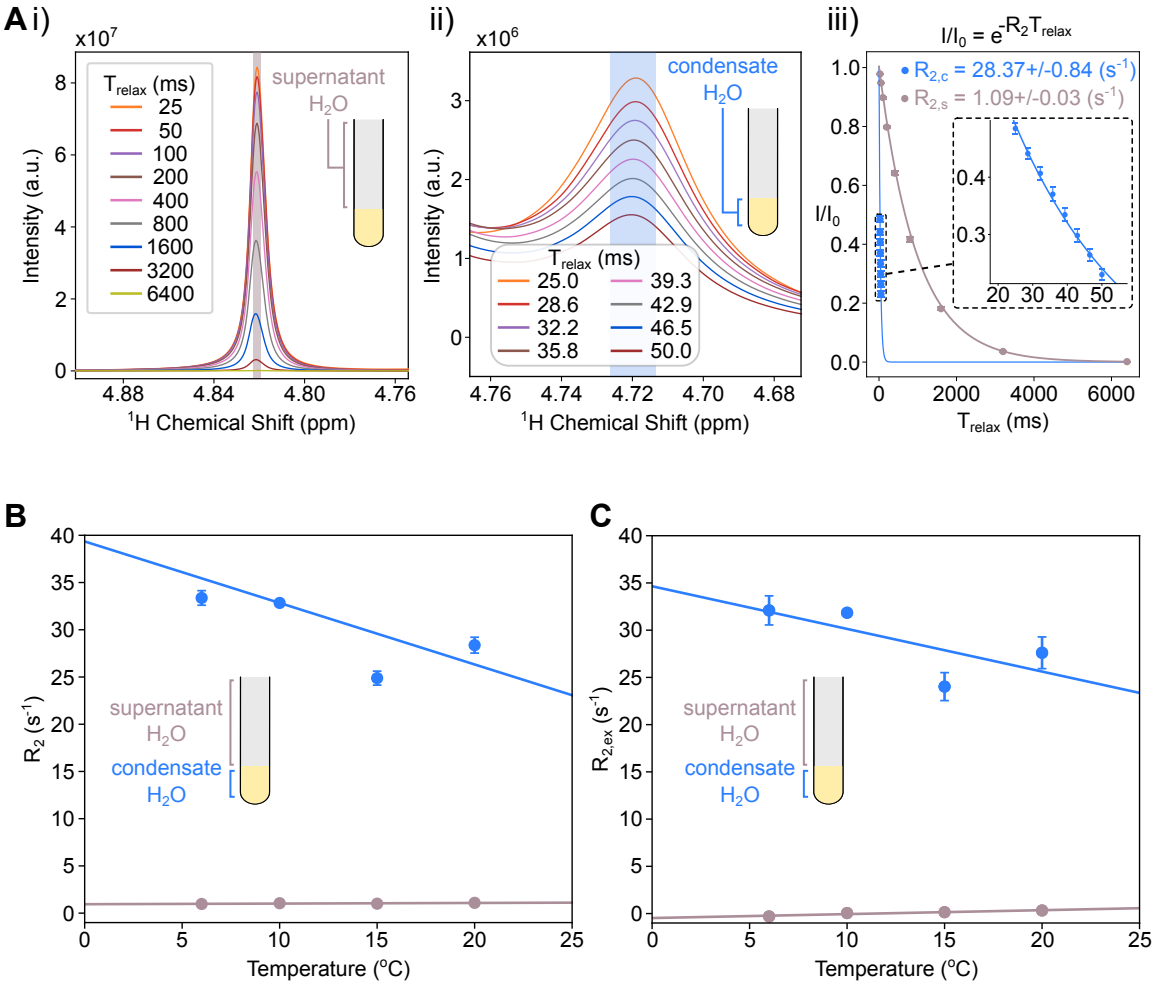


Figure 3.44: Water transverse relaxation differences between DDX4^{N1} condensate and supernatant H_2O. **A** CPMG proton NMR spectra recorded on the DDX4^{N1} biphasic sample using different relaxation delays (T_{relax}) for the supernatant (i) and condensate (ii) water peaks at 20°C. iii) Fitting of the relaxation data to an exponential decay function allowed the extraction of transverse relaxation rates (R_2) for the condensate and supernatant water species. **B** A comparison of the supernatant and condensate water transverse relaxation rates as a function of temperature. **C** The extracted exchange contributions to the observed transverse relaxation rates ($R_2 = R_2^* + R_{2,ex}$).

The CPMG experiment (Figure 2.17) was performed to determine the transverse (R_2) relaxation rates of H_2O in both the supernatant and DDX4^{N1} condensate phases (Figure 3.44A). The observed R_2 relaxation rate of the condensate H_2O was approximately 30× larger than H_2O in the supernatant phase. Using the fitted τ_c values from the R_1 experiments (Figure 3.43C), the intrinsic transverse relaxation rates (R_2^*) could be estimated.[82]

$$
\begin{aligned}
R_2 &= R_2^* + R_{2,ex} \\
R_2^* &= \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2}{r_{HH}^6}\left(3J(0) + 5J(\omega_H) + 2J(2\omega_H)\right)
\end{aligned}
\tag{3.10}
$$

Therefore, the chemical exchange contribution to transverse relaxation ($R_{2,ex}$) could be established and was shown to be primarily responsible for the enhanced R_2 values (Figure 3.44C). This is consistent with the condensate water CEST profiles which show significant water-protein interactions (Figure 3.38).

The 2.5-3× reduction in water translational and rotational diffusion inside a DDX4^{N1} condensate caused by enhanced effective viscosity is indicative of an alternative solvent environment, with viscosity a key property of a solvent. Furthermore, the similarities in the fractional change of rotational and translational motion between the phases is evidence of an unaltered hydrodynamic radius of the water between the phases. The increased effective viscosity inside condensates may reduce the rate of biochemical reactions that are limited by the diffusion of substrates,[260] with the conversion of fumarate to L-malate by the enzyme fumarase being a prime example of a diffusion limited reaction that is highly dependent on viscosity.[314] However, even with the diffusion rate reduced in the condensate, the effect on the overall reaction rate may be balanced by an increase in local concentration through substrate/enzyme partitioning.

3.6 Conclusion

This work established *in vitro* NMR-based methods to determine metabolite partitioning into biomolecular condensates. These methods were then applied to measure the partitioning of nucleotide and amino acid metabolites into two orthogonal biomolecular condensates formed from the disordered proteins DDX4^{N1} and EWSR1C. The metabolite partitioning measurements could be rationalised based on whether a particular molecular property favoured enrichment or depletion

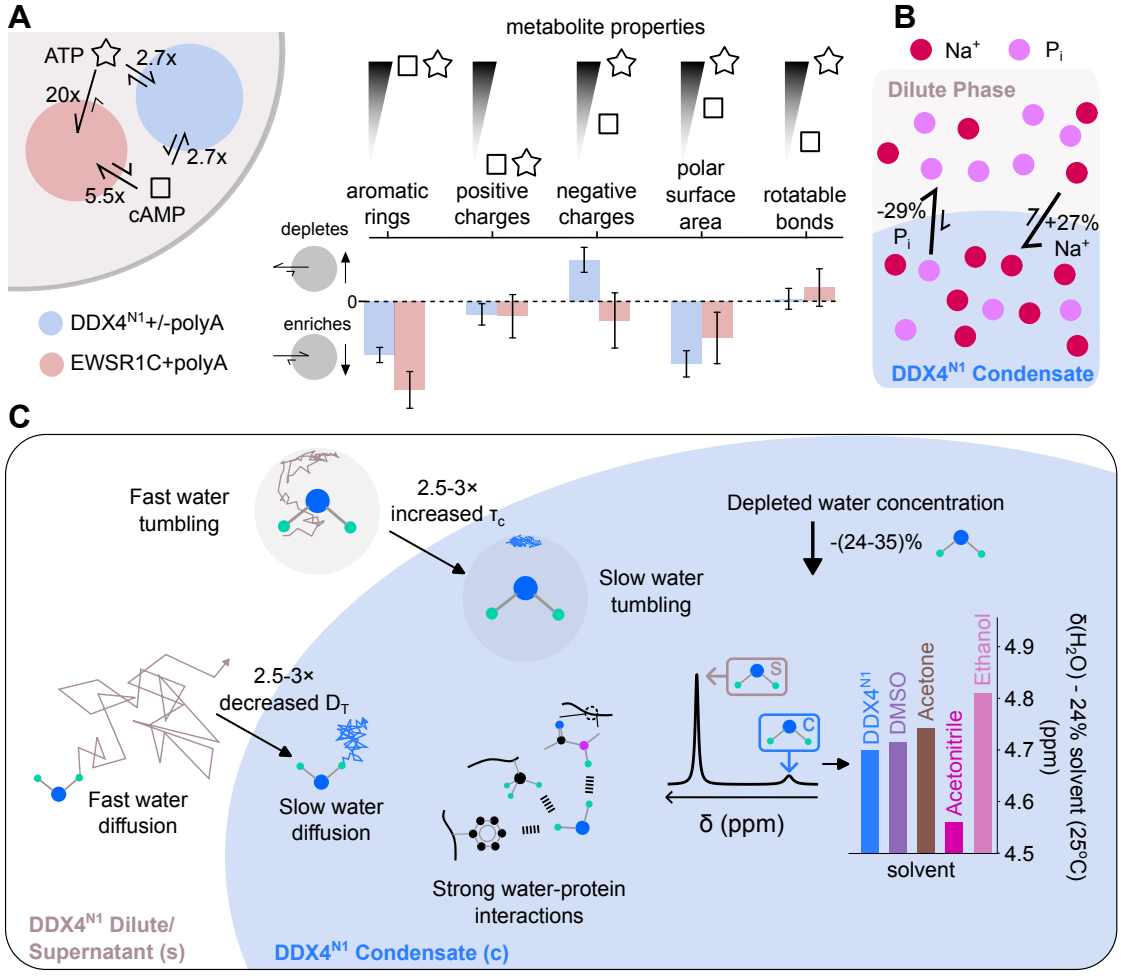


Figure 3.45: **A** A schematic showing how the partitioning of cAMP and ATP into DDX4^{N1}+/-polyA and EWSR1C+polyA condensates can be rationalised based on whether a particular molecular property favours enrichment or depletion. **B** Ion partitioning measurements revealed enrichment of sodium (Na^+) and depletion of phosphate (P_i) into DDX4^{N1} condensates. **C** An illustration of the changes to the solvent dependent properties of water inside DDX4^{N1} condensates. Water was depleted in the condensate phase, with a 2.5-3× reduced translational (D_T) and rotational diffusion (D_r=$1/(6\tau_c)$). Its protons were also more shielded in the condensate phase, driven by enhanced water-protein interactions which led to a large $R_{2,ex}$ rate. The observed increase in shielding of water in DDX4^{N1} condensates was similar (per unit volume protein/organic solvent) to water in DMSO.

into each condensate (Figure 3.45A). These trends provide an indication of how drug candidates with particular molecular properties could be adapted to target a specific condensate selectively. Similar trends governing partitioning could be established for the organic solvents 1-butanol and 1-octanol, highlighting the ability of biomolecular condensates to mimic the organic solvent behaviour.

The fitted coefficients governing the contribution of each molecular property to metabolite partitioning could also be correlated with the properties of the solvent (organic/protein) properties, establishing a set of rational rules similar to the concept of "like dissolves like" used extensively for predicting solubility into organic solvents.

The ability of DDX4^{N1} condensates to partition inorganic ions was then established (Figure 3.45B), with positively charged sodium ions depleted and phosphate ions enriched in the condensate phase. Given the ability of condensates to wet membranes, this could have downstream consequences on the function of ion channels in contact with condensates, such as the post-synaptic density. Furthermore, phosphate depletion could affect reactions coupled to ATP-hydrolysis inside DDX4^{N1} condensates by increasing the energy released for each ATP molecule hydrolysed. However, to completely quantify the effect of phosphate partitioning on ATP-hydrolysis coupled reactions, the partitioning of all substrates and enzymes would have to be deduced. Ion partitioning into condensates also has significant potential consequences linked to the generation of electric potential across a condensate surface, enabling catalytic properties.[174,291]

After establishing the partitioning behaviour caused by altered molecule solvation in biomolecular condensates, further detail on the solvent environment inside the DDX4^{N1} condensate system was explored through the lens of water (Figure 3.45C). A biphasic DDX4^{N1} condensate NMR sample was produced to compare the physicochemical properties of water between the phases. Firstly, imaging experiments revealed that the water concentration inside the condensate phase was reduced by 24-35%. Additionally, the chemical shift of water was decreased inside the DDX4^{N1} condensate, indicating a disruption to the hydrogen bonding network of water, consistent with previous Raman spectroscopy measurements on a different condensate system.[47] The DDX4^{N1} disordered protein chains inside the condensate phase have a similar shielding effect on water per unit

volume as the organic solvent DMSO (Figure 3.45C). This highlights the ability of concentrated disordered protein in the condensate phase to create an alternative solvent environment. Changes to the local water environment brought about by condensate ageing are hypothesised to be linked to the pathological aggregation of proteins.[190] Interestingly, the observed water chemical shift difference between the phases scaled with the protein aliphatic proton-water interactions measured in the condensate phase. In disordered proteins, hydrophobic (aliphatic) residues are more exposed to the surrounding environment than in structured proteins, thereby condensates formed from disordered proteins facilitate a stronger change in solvent environment than similarly concentrated structured proteins.

The dynamical properties of water were also altered inside the DDX4^{N1} condensate phase, with a 2.5 to 3 fold decrease in translational and rotational diffusion rates, indicating an enhanced effective viscosity. The similarities in the fractional change of rotational and translational motion between the phases is evidence of an unaltered hydrodynamic radius of the water between the phases. Viscosity is a key property of a solvent, and an increase in viscosity in biomolecular condensates further demonstrates their ability to provide an alternative solvent environment.

Altogether, the results from this chapter add to the growing evidence[44–49] that biomolecular condensates can provide an alternative solvent environment to cells, changing the physical and chemical properties of water, and enabling the selective partitioning of metabolites and inorganic ions.

3.7 Future work

To further understand the trends governing partitioning into DDX4^{N1}+/-polyA and EWSR1C+polyA biomolecular condensates, it is desirable to increase the pool of metabolites in the dataset from just amino acids and nucleotides to include sugars, vitamins or other biologically relevant molecules/ions. This would

increase the variety of chemical functional groups in the dataset, allowing further trends governing partitioning to be established. Recently, neural networks have been developed to predict the molecular partitioning into certain biomolecular condensates.[48,49] A larger dataset would also enable neural networks to be applied to the condensates studied in this chapter, allowing additional tools to predict the partitioning of new compounds into biomolecular condensates.

Furthermore, experiments need to be performed to discover if the partitioning of a compound depends on factors such as temperature, NaCl concentration, and the concentration of the compound. In-cell experiments have shown the presence of amino acids can alter the size and number of stress granule condensates.[315] Therefore, it would be interesting to determine if the partitioning of a compound into a condensate correlates with compound-dependent condensate stability.

Initial saturation transfer NMR experiments (uSTA[316]) have been performed to detect interactions between EWSR1C and GTP (with a 30× enrichment into EWSR1C+polyA condensates)*. Binding could not be detected in these initial experiments in a non-phase separated sample; however, further NMR experiment optimisation may reveal evidence of interactions. Recently, NMR relaxation experiments have been used to measure weak (K_d>100 μM) small molecule interactions with disordered proteins.[36] Utilising relaxation-based NMR methods to detect binding of the metabolites to the $DDX4^{N1}$ and EWSR1C+polyA condensate constituents may prove more successful in the future. This would enable any correlation between metabolite-protein interactions and metabolite partitioning to be determined.

After establishing the physical and chemical properties of water in $DDX4^{N1}$ condensates, these experiments should be repeated for EWSR1C+polyA condensates to determine if the changes to water properties are condensate specific, and if they correlate with the observed changes to metabolite partitioning.

*These experiments were performed by Ben Naylor and Virgínia Casablancas-Antràs (Baldwin lab).

Furthermore, there is growing interest in how biomolecular condensates can alter enzyme structure, function and kinetics by providing an alternative solvent environment.[46] Therefore, it would be useful to measure if the water and protein chemical shift changes inside condensates, reporting on the local electronic shielding provided by the condensate, are correlated with enzymatic structural changes more generally. In addition, measuring time-dependent changes to the chemical and physical properties of water/protein in biomolecular condensates formed from aggregation-prone proteins (such as FUS or tau) could reveal if the changes to the solvent environment inside the condensates are responsible for the initiation of pathological aggregation.

3.8 Contributions

The NMR-based method to measure amino acid and nucleotide partitioning was conceptualised and optimised in equal partnership with Dr. Virgínia Casablancas-Antràs, with the help of Emily Haasz. DDX4^{N1} and EWSR1C protein produced for research in this chapter was performed by a combination of myself, Dr. Virgínia Casablancas-Antràs, Ben Naylor, Professor Spencer Anthony-Cahill, Emily Haasz, and Eugene Lin. Metabolite and ion partitioning measurements (including NMR data acquisition and analysis) were performed in equal collaboration between myself, Dr. Virgínia Casablancas-Antràs and Ben Naylor. Modification of the HACA(CO) pulse sequence to include Seedless[279] optimal control pulses and C_β decoupling was performed with the help of Dr. Charles Buchanan. The fitting of metabolite partitioning to metabolite properties was conducted with Dr. Virgínia Casablancas-Antràs. Dr. Virgínia Casablancas-Antràs correlated the fitted coefficients for each metabolite property with properties of the condensates/solvents. DDX4^{N1} biphasic sample production was performed in equal collaboration between myself, Emily Haasz and Dr. Charles Buchanan. Dr. Charles Buchanan produced initial slice selective optimal control Seedless[279] pulses, and aided in writing

the water-suppressed slice selective NMR and ^{1}H-^{1}H NOESY pulse sequences. Professor Andrew Baldwin provided project conceptualisation and insight into NMR acquisition, analysis and fitting procedures.

4 Probing the disordered adenovirus capsid protein VI using NMR

4.1 Abstract

Adenoviruses can cause common respiratory illnesses and are also harnessed as viral vectors for vaccination. Protein VI, a disordered adenoviral capsid protein, is multifaceted in the virus life cycle and is essential for endosomal escape into host cells. Other roles of protein VI include nuclear import of the capsid protein hexon, necessary for viral assembly. The predicted N-terminal amphipathic helix is required for endosomal escape. However, little attention has been given to the remainder of the protein. Solution-state NMR was used to determine the properties of N-terminal truncated protein VI (Δ54-pVI). Diffusion NMR measurements revealed Δ54-pVI is highly disordered. However, assignment of the chemical shifts established the presence of secondary structure propensity. Transiently formed amphipathic α-helices were revealed, close to sequence motifs necessary for capsid assembly and microtubule transport of the capsid. Furthermore, Δ54-pVI was shown to interact with membrane-mimicking micelles. Knowledge of the secondary structure propensity in Δ54-pVI provides valuable insights in the context of the wild-type protein VI and its multifaceted roles in the virus life cycle.

4.2 Introduction to adenoviruses and protein VI

Human adenoviruses (HAdV) were first isolated from human adenoids and characterised in the 1950s.[317,318] They are common viruses known to cause mild respiratory illness, particularly in young children.[319] One study found that 8.2% of respiratory infections in children under the age of seven tested positive for adenovirus.[320] However, occasionally the disease can become more severe, leading to pneumonia in young children or immunocompromised patients, which can sometimes be fatal.[321,322] Adenoviruses are also capable of causing infections in locations other than the lungs. For example, approximately 75% of conjunctivitis eye infections are linked to adenoviruses.[323] Therefore, understanding the biological mechanism of infection caused by adenoviruses is essential for the development of treatments.

In 1959, electron microscope images revealed that adenoviruses had an icosahedral shape[324] and are known to be between 70 to 100 nm in diameter.[325] Adenoviruses are non-enveloped, meaning that they are not encapsulated by a lipid membrane, and their outer icosahedral capsid consists of the penton, peripentonal hexon and hexon proteins and contains fiber units emerging from the surface (Figure 4.1A).[325,326] Inside the icosahedral shell lies the genetic material of a human adenovirus in the form of double-stranded DNA (dsDNA) (Figure 4.1B).[327]

There are many species (A-G) of human adenoviruses, with each containing different serotypes. Human adenovirus C serotype 5 (HAdV-C5) is one of the most abundant and well-studied adenoviruses.[328] HAdV-C5 has been used as a viral vector to vaccinate against various diseases, including ebola,[329] chronic hepatitis B,[330] and has potential use for cancer virotherapy.[331] However, the large prevalence of antibodies against HAdV-C5 caused by past human infections leads to natural immunity against these viral vectors. Therefore, the use of the closely related chimpanzee adenoviruses (ChAd) has become common.[332] Notably, the ChAdOx1

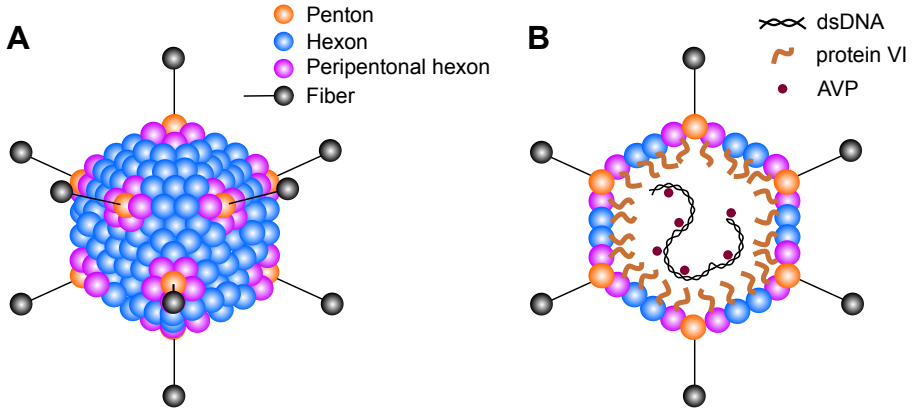


Figure 4.1: Introduction to the human adenovirus. **A** Icosahedral capsid structure of human adenoviruses. **B** Illustration showing the approximate locations of dsDNA, protein VI and adenovirus protease (AVP) within an adenovirus particle. Figure based on Rafie, K. *et al.*[335]

viral vector[333] was used for the ChAdOx1 nCoV-19/AZD1222 vaccine (commonly known as the Oxford-AstraZeneca Covid-19 vaccine) to provide immunity against SARS-Cov-2.[334]

Adenovirus cell entry is first initiated through binding of the fiber unit to a receptor on the host cell membrane, with the coxsackievirus and adenovirus receptor (CAR) being the primary binding partner for HAdV-C5[337] (Figure 4.2). Early electron microscopy images revealed that endocytosis occurs after binding, leading to the adenovirus's encapsulation in an endosome within the cell.[338] During endosome maturation and acidification, the virus capsid proteins, particularly penton and peripentonal hexon proteins at the icosahedral vertices, are removed, leading to the escape of internal proteins such as protein VI.[51] Protein VI has been shown to have significant membrane lytic activity, enabling the endosomal escape of adenovirus into the host cell cytoplasm[51] (Figure 4.2). The adenovirus is then transported to the nucleus, which docks with the nuclear pore complex.[339] The adenoviral dsDNA is injected into the nucleus through the nuclear pore complex,[340] followed by dsDNA replication.[341] The viruses hijack the machinery of the host cell to undertake protein production in the host cell cytoplasm.[342] These proteins are then transported back to the nucleus to partake in adenovirus assembly.[50] It takes approximately 20 hours for

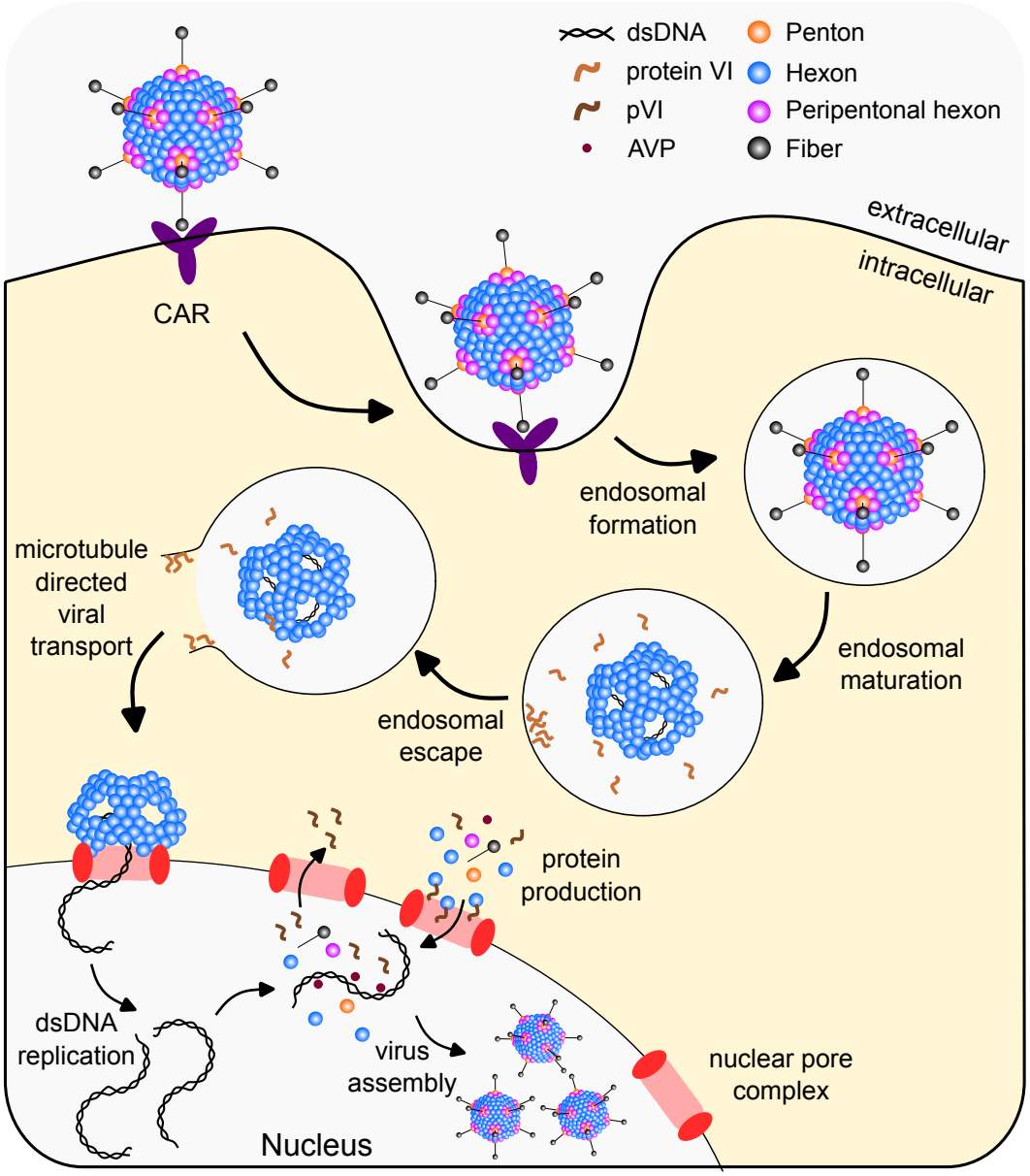


Figure 4.2: The cellular infection pathway of human adenovirus particles. Figure inspired by Nestić, D. *et al.*[336]

new adenovirus particles to be constructed in the host cell nucleus. After 48 hours the host cells begin to lyse and release the newly replicated viruses[343] (Figure 4.2).

During viral replication, precursor protein VI (pVI) is expressed and is utilised to transport hexon expressed in the host cell cytoplasm to the nucleus for virus assembly[50] (Figure 4.2). The pVI sequence contains several proposed nuclear export signals (NES-1/NES-2) and nuclear localisation signals (NLS-1/NLS-2) containing sequence similarities with known NESs and NLSs (Figure 4.3A). Precursor protein

VI has been shown to bind to hexon, and these proposed NES and NLS regions are thought to help shuttle pVI between the nucleus and cytoplasm, enabling the nuclear import of hexon.[50] Furthermore, pVI contains a PPxY (PPSY) motif that recruits Nedd4 (neuronally expressed developmentally downregulated 4) E3 ubiquitin ligases to attach ubiquitin units to pVI, allowing the transport of pVI along host cell microtubules[344] (Figure 4.3A). Adenoviruses containing a mutation at the PPxY motif can escape host cell endosomes but are not transported to the nucleus for replication.[344]

Precursor protein VI (pVI) is cleaved after viral assembly to form mature protein VI as well as the N/C-terminal cleavage products pVIn (residues 1-33) and pVIc (residues 240-250).[349–351] This processing of precursor protein VI occurs during the transition from an immature (non-infectious) to a mature (infectious) adenovirus.[51,351] The cleavage of the C-terminus of precursor protein VI during maturation occurs through the action of the adenovirus protease (AVP) bound to the adenovirus dsDNA, which acts as a cofactor for the reaction (Figure 4.1B). Following this, AVP (with dsDNA and pVIc as cofactors) can cleave the N-terminus of premature protein VI to form pVIn and mature protein VI and process all the other precursor adenovirus proteins necessary to become infectious.[352] The cleaved N-terminus pVIn is found docked into the hexons of mature adenoviruses, leading to the conclusion that the N-terminus (1-33) of pVI is responsible for its localisation to hexon in the virus capsid.[353,354]

A temperature-sensitive human adenovirus mutant (*ts1*), which does not package the AVP during viral assembly at temperatures above the non-permissive temperature (39^{o}C), cannot convert precursor proteins to their mature state.[355,356] Endosomal maturation is coupled to an increase in endosome acidity which has been shown to partially disrupt the capsid structure of mature adenoviruses, leading to the release of penton from the capsid.[357] The immature *ts1* mutant, containing only non-cleaved precursor proteins, fails to escape the endosome.[339]

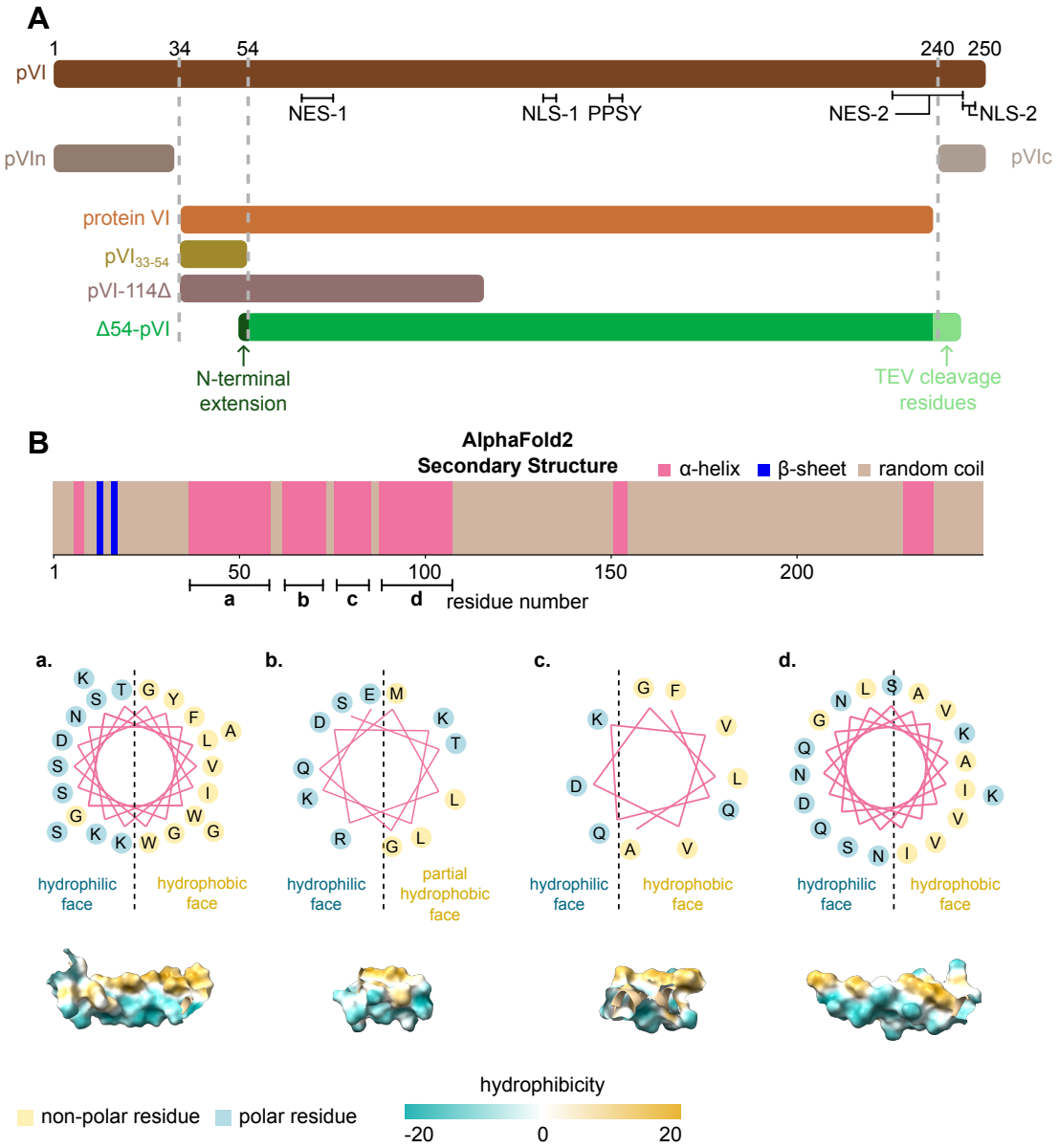


Figure 4.3: An introduction to protein VI constructs, its predicted secondary structure and amphipathic helices. **A** A schematic showing premature protein VI (pVI) which is cleaved by AVP to form pVIc, pVIn and protein VI. The peptide pVI_{33-54} and pVI-114Δ construct are shown, along with the Δ54-pVI construct used in this study, which contains an extra M residue at the N-terminus and the TEV cleavage residues ENLYFQ at the C-terminus. **B** The secondary structure of premature protein VI (pVI), predicted using AlphaFold2.[5,238] The amphipathic nature of the predicted helical regions a-d are represented through helical plots. Helical plots were generated using the pepwheel tool from The European Molecular Biology Open Software Suite (EMBOSS),[345] surface hydrophobicity plots were generated using ChimeraX,[346] based on pyMLP.[347,348]

Immunodepletion of protein VI from the supernatant of temperature-dissociated capsids using antibodies removes the membrane lytic activity of the resultant

solution, highlighting the integral role of protein VI in endosomal adenovirus escape.[51] Furthermore, recombinant protein VI and precursor protein VI possess pH-independent membrane lytic activity.[51] The peptide pVI_{34-54} (Figure 4.3A) has a similar membrane lytic ability as mature protein VI, indicating its high importance in the disruption of endosomes.[52] The peptide has been predicted to have an amphipathic helix structure (Figure 4.3B), an α-helix containing hydrophilic and hydrophobic residues arranged to give hydrophobic and hydrophilic helix faces.[52] Circular dichroism measurements have shown that protein VI contains some α-helical character, predicted to be from the N-terminal amphipathic helix.[52,358] The proposed amphipathic helix lies on the surface of endosomal membranes, inducing positive membrane curvature and causing downstream endosomal escape.[52]

Interestingly, experiments on Δ54-pVI, a construct of mature protein VI missing the proposed amphipathic helix (Figure 4.3), have shown that despite its membrane lysis ability being significantly diminished, it was still able to disrupt membranes.[52] This study uses NMR to establish the presence of transient structure within Δ54-pVI and its potential interaction with lipids. The results indicate that whilst being mostly disordered, Δ54-pVI has a significant presence of residual α-helical structure close to important sequence features that may prove essential to its membrane lytic ability and other viral functions.

Understanding how protein VI can disrupt endosomal membranes is essential to enhance our knowledge of adenovirus cell entry and find future antiviral treatments for adenovirus infections. Furthermore, a greater insight into the multifaceted roles of protein VI in adenovirus host cell entry and replication could improve adenoviral vector vaccine and gene therapy delivery mechanisms.

4.3 Δ54-pVI construct disorder revealed using diffusion NMR

Protein diffusion coefficients are highly sensitive to their relative amount of structure and disorder, with structured proteins being more compact, leading to

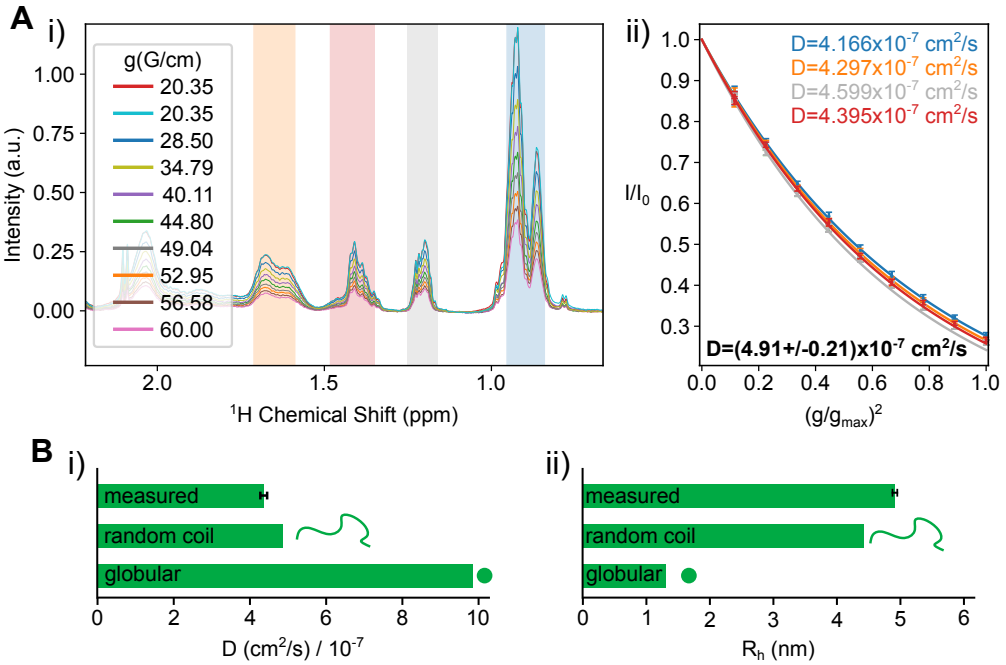


Figure 4.4: Diffusion and hydrodynamic radii measurements of Δ54-pVI. **A** i) Water-suppressed longitudinal encode-decode stimulated echo diffusion NMR zoomed in to the methyl group region. ii) Measured diffusion coefficients for selected methyl group regions obtained from fitting the NMR data to the Stejskal-Tanner equation (equation 2.37).[20] **B** A comparison between the measured and predicted (random coil/globular) diffusion coefficients (i) and hydrodynamic radii (ii) of Δ54-pVI.

faster diffusion through solution.[359] To measure the diffusion coefficient of Δ54-pVI, a 500 μL sample was prepared in a 5 mm NMR tube containing 200 μM Δ54-pVI in a sodium phosphate buffer at pH 7.0 containing 10% D_2O*. Water-suppressed longitudinal encode-decode stimulated echo diffusion NMR measurements (section 2.5) were performed at 20^{o}C (Figure 4.4Ai)†.

The measured data was fitted to the Stejskal-Tanner equation (equation 2.37)[20] for four aliphatic proton regions enabling the average diffusion coefficient of $(4.36\pm0.18)\times10^{-7}$ cm^2/s to be obtained (Figure 4.4Aii). This corresponds to a hydrodynamic radius (R_h) of 4.91±0.21 nm. The Δ54-pVI construct has 192 amino acid residues (N), and the predicted hydrodynamic radii of structured and random coil proteins of this residue number can be calculated using the following

*Δ54-pVI was expressed and purified by Susan Black, Oxford.

†Details on longitudinal encode-decode stimulated echo diffusion NMR are provided in section 2.5

empirically determined formula:[359]

$$R_{h,globular}(nm) = 0.475N^{0.29} \qquad R_{h,random\ coil}(nm) = 0.221N^{0.57} \tag{4.1}$$

This empirical formula predicts hydrodynamic radii of 2.18 nm and 4.42 nm for a 192 residue globular and random coil protein, respectively. Therefore, the NMR determined experimental hydrodynamic radius of Δ54-pVI is more closely aligned with a disordered protein with a random-coil conformation (Figure 4.4B). Size-Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) confirmed that Δ54-pVI was monomeric at the concentration used for the diffusion experiments to rule out aggregation being the cause of the increased hydrodynamic radius*. The diffusion coefficient of Δ54-pVI highly indicated that it was substantially disordered. Nevertheless, it was important to determine if there is any propensity of Δ54-pVI to form any secondary structure potentially important for its viral function. Circular dichroism (CD) measurements on pVI-114Δ (Figure 4.3A) have shown tentative evidence of significant α-helical character at the N-terminus of Δ54-pVI.[52]

4.4 NMR backbone assignment of Δ54-pVI

NMR chemical shifts are highly sensitive to residual secondary structure within a system. The secondary structure of a protein is defined by the dihedral angles Ψ and Φ, representing the angles between adjacent planar amide bond units (Figure 4.5A). Dihedral angles indicating a particular secondary structure type can be visualised using a Ramachandran plot (Figure 4.5B).[360] The C_α and C_β atoms are close to the bonds where Ψ and Φ dihedral angles are defined; therefore, their chemical shifts/electron density are strong reporters on backbone dihedral

*SEC-MALS experiments were performed and analysed by Susan Black, Oxford.

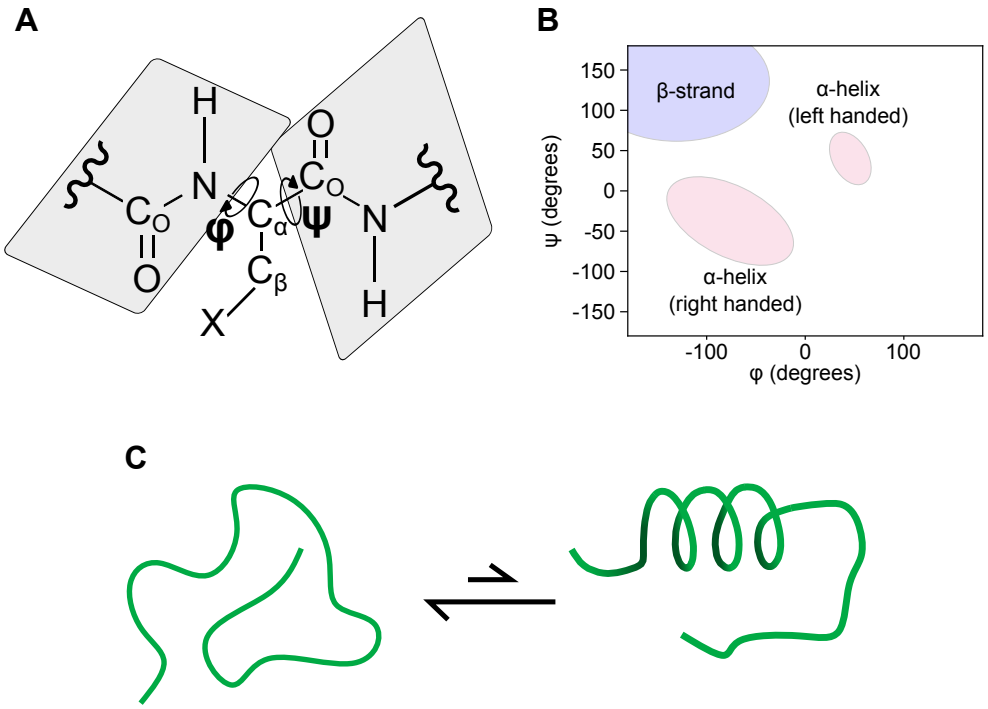


Figure 4.5: An introduction to dihedral angles, Ramachandran plots and transient secondary structure. **A** An illustration of the dihedral angles Ψ and Φ between two neighbouring planar amide units in a protein chain. **B** A plot showing the regions of dihedral angle pairs consistent with right/left-handed α-helical and β-strand secondary structure.[360] **C** A depiction of transiently formed α-helical secondary structure in a primarily disordered protein.

angles. When a protein is in fast conformational exchange, as is the case for highly disordered proteins, solution-state NMR chemical shifts report on the ensemble average environment of each nucleus. However, chemical shifts are highly sensitive to the local environment and can report on the transient secondary structure formed by the protein (Figure 4.5C). When the C_α and C_β chemical shifts are compared to their expected random coil chemical shifts, the presence of transient and residual structure can be quantified.[13,361,362]

Isotopically labelled proteins, enriched in the NMR active ^{15}N and ^{13}C nuclei, are required to assign non-proton chemical shifts in proteins efficiently. Without isotopic enrichment, due to the poor natural abundance of ^{15}N and ^{13}C nuclei (0.37% and 1.1%, respectively[363]), the resultant signal intensity of the spectra would be highly diminished. A 250 μL sample containing 1.7 mM of ^{15}N, ^{13}C-labelled Δ54-pVI was

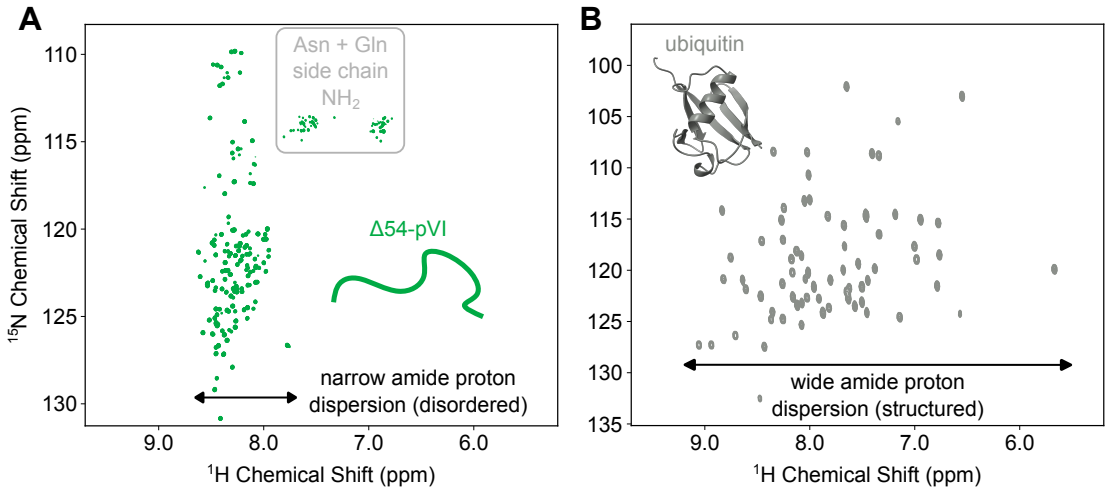


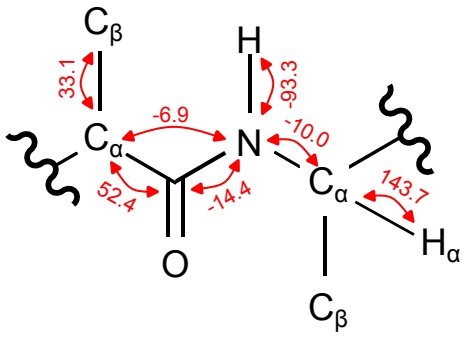
Figure 4.6: Proton chemical shift dispersion in disordered and structured proteins. The ^{1}H, ^{15}N projection of three dimensional HNCO experiments obtained for Δ54-pVI (**A**) and ubiquitin (**B**). The BEST-TROSY version of the HNCO was used to collect data for Δ54-pVI.[364,365] Experimental data for ubiquitin was provided by Professor Andrew Baldwin. The structure of ubiquitin was obtained from the protein data bank (PDB: 1UBQ).[366] Comparison to ubiquitin inspired by Schiavina *et al.*[367]

produced (in 20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O) to enable the backbone ^{1}H, ^{15}N, $^{13}C_O$, $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts to be determined for each amino acid residue.*.

Due to the high degree of accessibility of the backbone amide protons of a disordered protein to solvent water protons, it is desirable to perform experiments on disordered proteins at lower temperatures to reduce the solvent exchange contribution to transverse relaxation of these protons,[367] which would lead to the presence of broad low-intensity peaks that are unresolvable (section 2.4). Furthermore, disordered proteins, such as Δ54-pVI, undertake a rapid conformational exchange in solution, leading to an averaging of their chemical shifts. Therefore, their peaks usually appear in a narrow region of chemical shift space, especially in the proton dimension, leading to a significant degree of peak overlap[367] (Figure 4.6A). This is in contrast to heavily structured proteins, such as ubiquitin, which have high dispersion in their NMR spectra (Figure 4.6B)† and

*Expression and purification of ^{15}N, ^{13}C-labelled Δ54-pVI was performed by Susan Black, Oxford.

†Data for ubiquitin provided by Professor Andrew Baldwin.



Coupling	Coupling Constant (Average, Hz)	Coupling Constant (Range, Hz)
H-N	-93.3	-98.0 to -90.5
N-CO	-14.4	-16.5 to -12
N-$C\alpha_i$	-10.0	-13.1 to -8.2
N-$C\alpha_{i-1}$	-6.9	-11.0 to -3.5
CO-Cα	52.4	49.2 to 55.1
Cα-Cβ	33.1	28.1 to 40.4
Hα-Cα	143.7	126.8 to 153.7

Figure 4.7: A schematic and table showing scalar coupling constants (J) relevant for multidimensional biomolecular NMR experiments. INEPT evolution periods can be set according to $1/|4J|$ in order to selectively transfer magnetisation between two nucleus types. All coupling constants are shown in Hz. Coupling constants from Schmidt, J. M. *et al.*[368]

highlights the importance of multinuclear experiments for disordered proteins.[367] The narrow amide proton dispersion of Δ54-pVI remained in acidic conditions, suggesting its disordered nature remains in an acidic environment (see appendix B.2.5). It is also possible to move beyond two-dimensional NMR experiments to obtain three-dimensional NMR spectra showing ^{1}H-^{15}N-^{13}C correlations between neighbouring ^{1}H-^{15}N-^{13}C nuclei, with added dimensions improving the overall resolution. Coherence transfers between scalar coupled nuclei are performed utilising INEPT pulse sequence units where the evolution delays (τ) are set to $1/(4|J|)$, where J is the scalar coupling constant between the nuclei (see section 2.3). For protein NMR, the relevant coupling constants between coupled nuclei are shown in Figure 4.7. One example of such an experiment is the HNCO experiment,[278] showing correlations between chemically bonded ^{1}H-^{15}N-$^{13}C_O$ connected nuclei (Figure 4.8A/B). This is achieved by adding additional INEPT coherence transfers to a standard ^{1}H-^{15}N HSQC (section 2.3), with delays of the extra INEPT periods set to $1/(4|J_{N-CO}|)$ and incrementation of an evolution period where magnetisation is located on the CO nucleus. Peaks that are heavily overlapped or even indistinguishable in the ^{1}H-^{15}N dimension of Δ54-pVI become discernible when viewing down the $^{13}C_O$ dimension (Figure 4.8C).

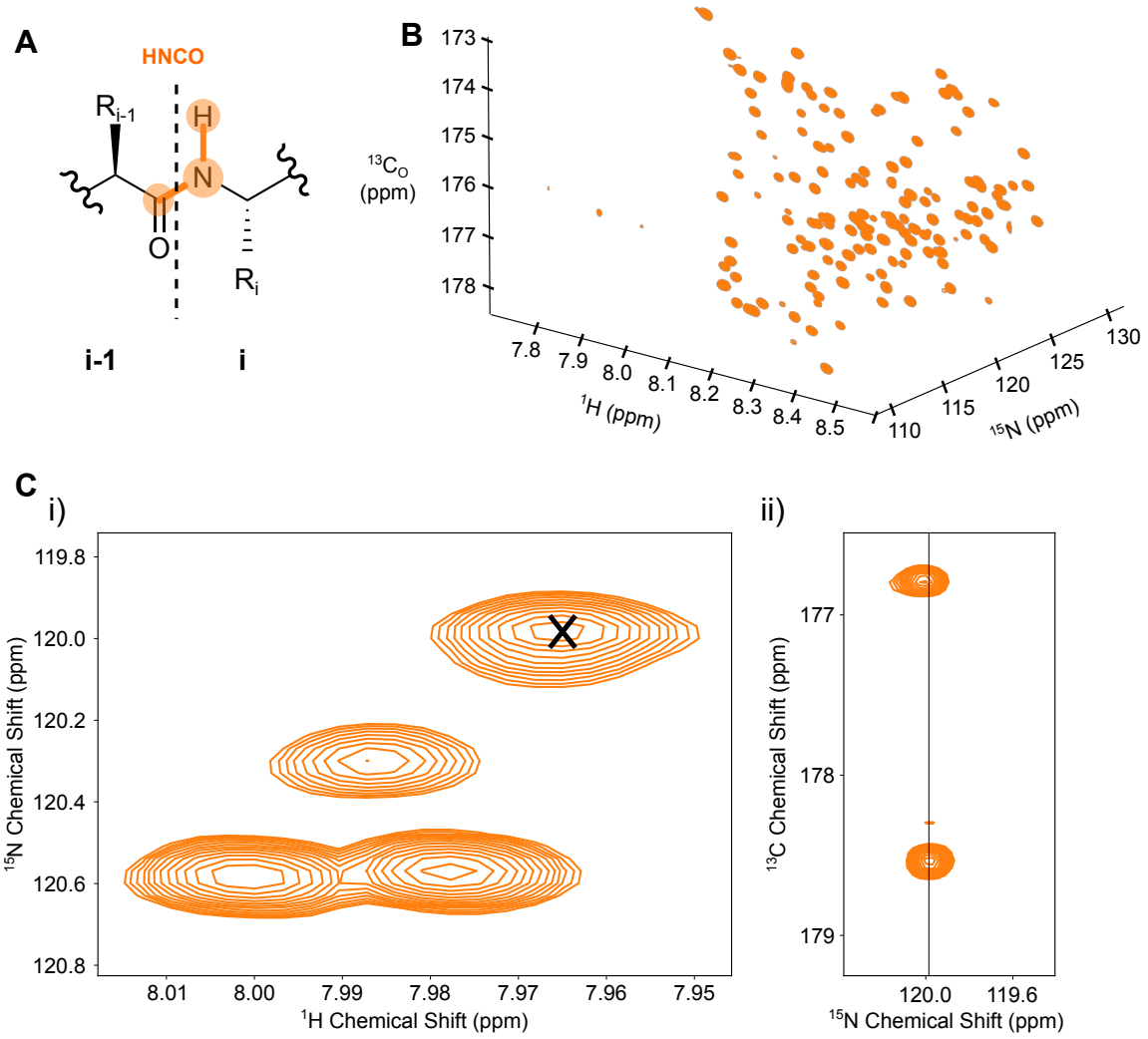


Figure 4.8: An introduction to 3D NMR using the HNCO experiment and its benefits of increasing resolution. **A** A diagram showing the magnetisation transfer pathways in the HNCO experiment leading to each ^{1}H-^{15}N correlation for residue i being connected to the $^{13}C_O$ chemical shift of the $i-1$ residue. **B** A 3D plot of the BEST-TROSY-HNCO spectrum for Δ54-pVI recorded at 5^{o}C. **C** i) A zoom in on an overlapped region of the ^{1}H-^{15}N projection of BEST-TROSY-HNCO for Δ54-pVI. A pair of heavily overlapped/indistinguishable peaks is highlighted by a cross. ii) Viewing the spectrum at the position of the cross down the $^{13}C_O$ axis provides enhanced resolution allowing the two peaks to be effectively resolved.

4.4.1 Band-selective excitation short-transient (BEST) experiments on disordered proteins

Due to the high degree of peak overlap in multidimensional spectra obtained on disordered proteins, it is advantageous to obtain substantial acquisition times in

all dimensions of the spectrum. The maximum achieved digital resolution in each dimension is inversely proportional to acquisition time. Hence, a large acquisition time ensures that the only constraint to spectral resolution is intrinsic relaxation instead of digital resolution. However, recording substantial indirect dimension increments can lead to unfeasible experiment durations. It takes approximately 40 hours to record a 3D NMR experiment with each indirect dimension containing 128 complex points, using an acquisition time of 100 ms, a relaxation delay of 1.0 seconds, and two repeat scans for every indirectly acquired point. This duration is not sustainable, especially if multiple different 3D experiments need to be acquired for a sample. Non-uniform sampling (NUS) was created to combat this problem. For non-uniform sampling, the indirect dimension complex points are sampled sparsely, drastically reducing experimental time.[369] Performing the same experiment as described previously, but instead with 50% sparse sampling in each indirect dimension, reduces the total experiment time to 10 hours. The sparsely sampled data can be reconstructed using a package such as the sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction algorithm[370] or FID-Net.[371]

However, most of this remaining time is taken by relaxation delays necessary to ensure enough magnetisation has returned to equilibrium before starting the following experiment. The following equation represents the expected signal-to-noise ratio (S/N) per unit time as a function of the longitudinal relaxation rate (R_1) and the relaxation delay (d_1):[372]

$$S/N = A\frac{(1 - e^{-\mathrm{R}_1 d_1})}{\sqrt{d_0 + d_1}} \tag{4.2}$$

where d_0 is the duration of the pulse sequence excluding acquisition time, and A is a constant. This equation suggests that the optimum value of d_1 for an NMR experiment is $1.25/\mathrm{R}_1$-d_0.[373] Therefore, by increasing R_1, d_1 can be shortened

without compromising the signal-to-noise ratio per unit time.

It was discovered that band-selective excitation experiments, where the amide protons are selectively excited and aliphatic/water protons remain unperturbed (or are returned to z-magnetisation before acquisition), can substantially increase the amide proton longitudinal relaxation rates.[373] This is known as the Band-selective Excitation Short-Transient (BEST) principle,[374] similar to the band-Selective Optimised Flip-Angle Short-Transient (SOFAST) experiments,[375] and works best in non-background deuterated proteins.[373] Unperturbed protons along the z-axis act as a 'thermal bath' able to decrease the longitudinal relaxation times through cross-relaxation.[373] Therefore, using BEST sequences enables the relaxation delay (d_1) to be shortened to as little as 0.2 seconds without compromising the signal obtained per unit time.[376] This allows a high-resolution BEST-HNCO experiment containing 256 increments along the ^{15}N and ^{13}C indirect dimensions, with 50% non-uniform sampling and two repeat scans, to be recorded in approximately 2.7 hours (acquisition time = 100 ms, relaxation delay = 0.2 s).

Furthermore, it was empirically shown that the use of ^{1}H-^{15}N BEST sequences coupled with Transverse Relaxation Optimised Spectroscopy (TROSY)[53] can increase the intensity of observed peaks when applied to large disordered proteins such as Δ54-pVI.[376] Details on the benefits and theoretical background for TROSY schemes can be found in chapter 5. During non-TROSY biomolecular NMR pulse sequences, each ^{1}H-^{15}N coherence will exist as two peaks in the proton dimension, formed from H_xN_α and H_xN_β magnetisation with frequencies at ω_H+/-πJ_{HN}. Decoupling schemes are usually applied to ensure only one peak exists for each ^{1}H-^{15}N coherence, ensuring enhanced spectral clarity. However, decoupling during acquisition can only be applied for a short amount of time (<100ms) due to hardware restraints. For TROSY pulse sequences, decoupling is not used during acquisition as a single multiplet peak is selected before acquisition. Therefore, the acquisition time can be increased, enhancing the digital resolution in the proton

dimension observed for TROSY sequences. This can help resolve amide coherences for disordered proteins that are highly non-dispersed in the proton dimension (Figure 4.6). Furthermore, BEST-TROSY sequences enable off-pathway proton magnetisation (H_z) formed during relaxation delays in the sequence to be converted to N_z magnetisation.[377] This N_z magnetisation can be utilised in the following experiment scan to enhance signal intensity.[377]

4.4.2 NMR protein backbone assignment experiments

It is necessary to run a suite of assignment experiments to assign resonances observed in NMR spectra to individual residues within a protein. These enable connectivities (edges) between peaks representing connections between neighbouring amino acids to be established. The HNCO[278] and HN(CA)CO[378] are examples of such experiments. The HNCO experiment produces a peak for every non-proline residue (i), containing ^{1}H and ^{15}N chemical shifts associated with residue i (NH_i), and $^{13}C_O$ chemical shifts associated with the $i-1$ residue (CO_{i-1}) (Figure 4.9A/B). In contrast, for the HN(CA)CO experiment, two peaks can be present for each NH_i correlation relating to the $^{13}C_O$ chemical shifts of the $i-1$ (CO_{i-1}) and i (CO_i) residue. The HN(CA)CO experiment involves two additional coherence transfers to a ^{1}H-^{15}N-HSQC with the sequential transfer of magnetisation from N to Cα and then Cα to CO.[†] The presence of two peaks is due to the overlap between the ranges of N-Cα_i and N-Cα_{i-1} coupling constants (Figure 4.7). The most intense CO peak associated with each (NH_i) unit in the HN(CA)CO experiment is the CO(i) peak, with the CO(i-1) peak being less intense.[378] Suppose the $^{13}C_O$ chemical shift associated with an NH coherence in the HNCO is the same as the $^{13}C_O$ chemical shift of the strongest intensity peak associated with another NH peak in the HN(CA)CO experiment. In that case, this indicates a possible connection between these two amino acid residues in the protein. Strip plots highlight the sequential

[†]The brackets around Cα indicate that the magnetisation passes through this nucleus but its chemical shift is not recorded.

connectivity between NH peaks (Figure 4.9C). However, due to the substantial number of peaks within a protein's HNCO/HN(CA)CO spectra, there are multiple amino acid residues with very similar $^{13}C_O$ chemical shifts. Therefore, coherences cannot be definitively connected based on $^{13}C_O$ chemical shifts alone. Instead, further experiments can be used to provide further restraints. Furthermore, $^{13}C_O$ chemical shifts are less sensitive to local secondary structure propensity, and it is essential to obtain residue-specific $^{13}C_\alpha/^{13}C_\beta$ chemical shift information for this purpose.

These additional assignment experiments come in pairs like the HNCO/ HN(CA)CO experiments discussed previously and include the HNCA/ HN(CO)CA, HNCACB/ HN(CO)CACB, and (H)N(CA)NH/ (H)N(CO)(CA)NH spectrum pairs. The HNCA experiment[278] contains two peaks for each NH coherence associated with the Cα of the i and $i-1$ residues due to similarities in the N-Cα_i and N-Cα_{i-1} scalar coupling constants (Figure 4.10A). Like the HN(CA)CO experiment, the most intense peak in the HNCA for each NH coherence (i) is associated with the Cα_i chemical shift. In contrast, the HN(CO)CA experiment[379] gives only one peak (Cα_{i-1}) for NH coherence (i). Therefore, similarly to the HNCO and HN(CA)CO experiments, the HNCA/HN(CO)CA experiments enable peaks associated with neighbouring amino acid residues to be connected based on their Cα chemical shifts. Furthermore, glycine residues are immediately identifiable in these experiments due to their distinctively low Cα chemical shift ($\approx$ 45 ppm)

The HNCACB[380] and HN(CO)CACB[381] are similar to the HNCA and HN(CO)CA experiments, except for the addition of scalar coupling transfer between the Cα and Cβ nuclei. In the HN(CO)CACB experiment, two peaks in the ^{13}C dimension are expected for each NH coherence (i) relating to the Cα and Cβ chemical shifts of the $i-1$ residue. Whereas, the HNCACB experiment contains four peaks for each NH coherence (i) corresponding to the Cα and Cβ chemical shifts of the i and $i-1$ residues, with the Cα_i and Cβ_i being most intense

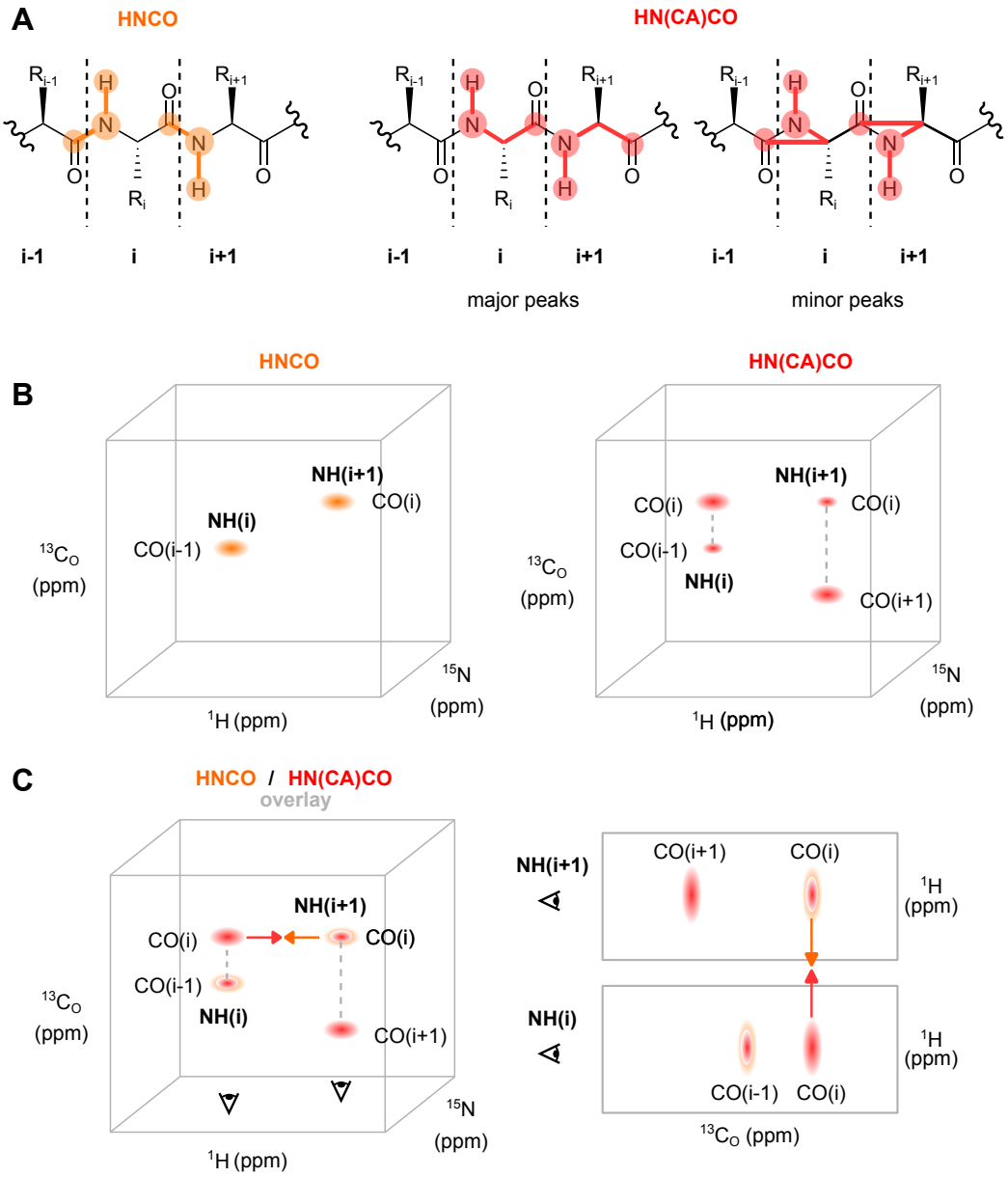


Figure 4.9: NMR backbone connectivities through the HNCO and HN(CA)CO experiments. **A** Magnetisation transfer pathways during the HNCO and HN(CA)CO NMR experiments. Bold lines indicate through-bond magnetisation transfer pathways and shaded circles indicate chemical shifts that are recorded for each experiment. **B** Depiction hypothetical 3D HNCO and HN(CA)CO spectra from two neighbouring NH peaks (NH_i and NH_{i+1}). **C** An overlay of the HNCO and HN(CA)CO spectra and representation as strip plots reveals evidence of the connectivities between NH_i and NH_{i+1}.

(Figure 4.10B). Furthermore, the peaks corresponding to Cα and Cβ for each NH correlation are oppositely signed. However, defining the Cα peaks as positive[380] is standard practice. In addition, serine and threonine residues can be conveniently

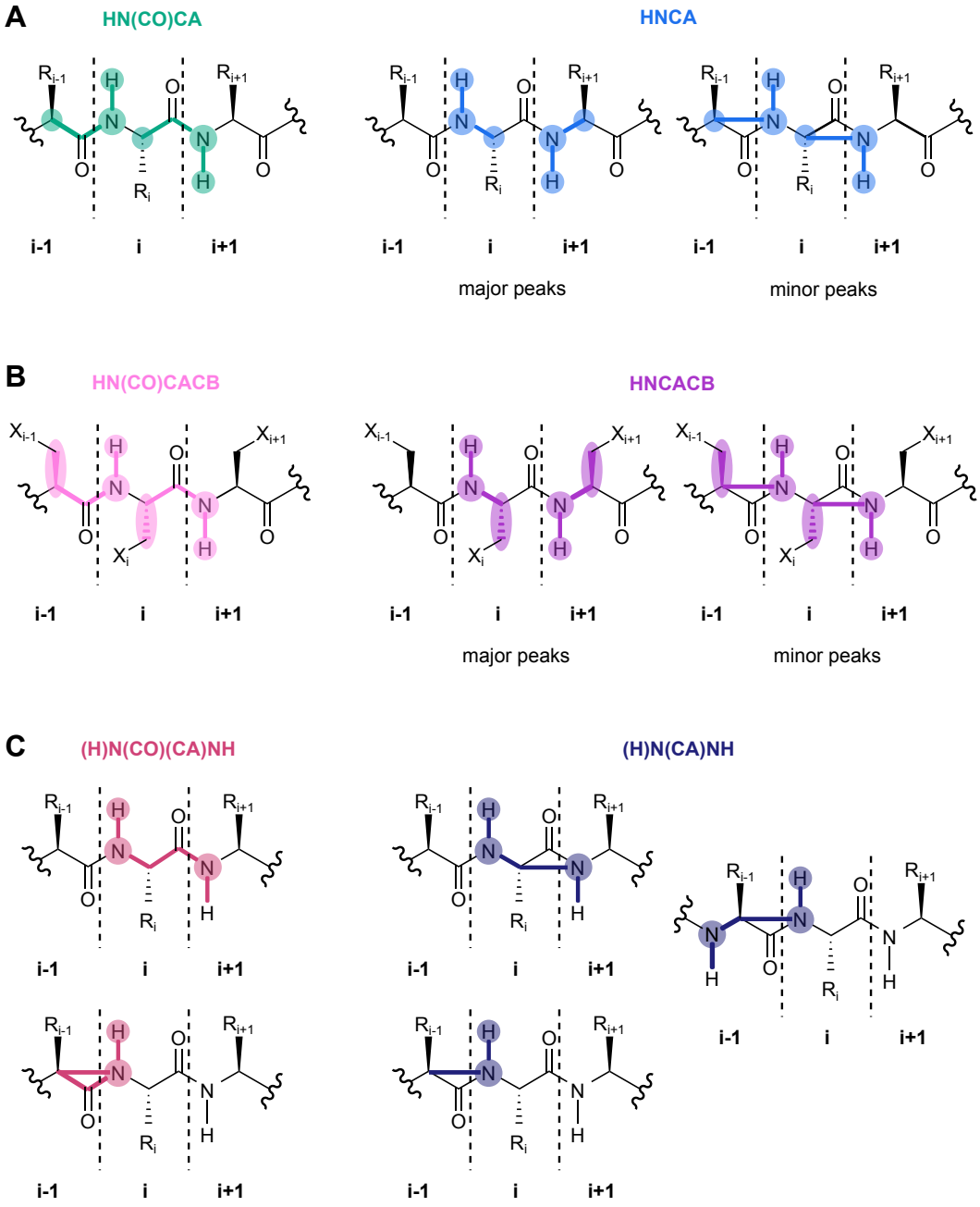


Figure 4.10: Schematics showing the transfer of magnetisation between nuclei along a protein backbone during HN(CO)CA/HNCA experiments (**A**), HN(CO)CACB/HNCACB experiments (**B**), and (H)N(CO)(CA)NH/(H)N(CA)NH experiments (**C**). Bold lines indicate through-bond magnetisation transfer pathways and shaded circles indicate chemical shifts that are recorded for each experiment.

identified through their substantial Cβ chemical shifts (>60 ppm).

Finally, the (H)N(CA)NH[382] and (H)N(CO)(CA)NH[383–385] experiments correlate the ^{15}N chemical shift of neighbouring amide nitrogen nuclei to each NH_i coherence (Figure 4.10C). The (H)N(CO)(CA)NH experiment has two possible coherence

pathways giving a diagonal peak (N_i) and a non-diagonal peak (N_{i+1}) in the additional ^{15}N dimension for each NH_i coherence. The (H)N(CA)NH produces a further peak in the ^{15}N dimension corresponding to N_{i-1}. The diagonal peaks in each experiment are often low in intensity or not visible.

In the application to Δ54-pVI, the experiments described above significantly restrained the connections between NH_i coherences corresponding to neighbouring amino acid residues in the protein sequence. Therefore, utilising such rigorous methodologies provided confidence in the obtained assignment for Δ54-pVI (discussed in section 4.5). This insured that any downstream residue specific information obtained from NMR experiments on Δ54-pVI, such as secondary structure propensity, was accurate.

4.4.3 NMR backbone assignment experimental acquisition and automated peak-picking for Δ54-pVI

BEST-TROSY versions of the HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB were recorded on a 250μL sample of 1.7 mM ^{15}N, ^{13}C-labelled Δ54-pVI (20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O) at 5^{o}C using the IBS NMRlib package.[365] Additionally, (H)N(CA)NH, (H)N(CO)(CA)NH, and (H)CC(CO)NH experiments were performed. In the indirect dimensions of all assignment spectra, 25% sparse sampling was used. Details on the NMR acquisition and processing of the backbone assignment spectra are included in appendix B.2.1.[†]

Initially the UniDecNMR[386] peak-picking algorithm was applied to the HNCO spectrum of Δ54-pVI producing a three-dimensional peak list highlighting ^{1}H,^{15}N, and ^{13}CO correlations. The ^{1}H and ^{15}N chemical shifts were then used as a restraint peak list when applying the UniDecNMR algorithm to peak-pick along the ^{13}C (or ^{15}N) dimension all other obtained spectra. Manual checks were then applied using the UniDecNMR software[386] to ensure the algorithm correctly identified peaks within

[†]Experiments were performed in equal collaboration between myself and Dr. Charles Buchanan.

Experiment	Peak Intensity Information	Classification
HNCO	N/A	CO_{i-1}
HN(CA)CO	Maximum	CO_i
HN(CA)CO	Minimum	CO_{i-1}
HN(CO)CA	N/A	$C\alpha_{i-1}$
HNCA	Maximum	$C\alpha_i$
HNCA	Minimum	$C\alpha_{i-1}$
HN(CO)CACB	Positive	$C\alpha_{i-1}$
HN(CO)CACB	Negative	$C\beta_{i-1}$
HNCACB	Positive Maximum	$C\alpha_i$
HNCACB	Negative Maximum	$C\beta_i$
HNCACB	Positive Minimum	$C\alpha_{i-1}$
HNCACB	Negative Minimum	$C\beta_{i-1}$
(H)N(CO)(CA)NH	Maximum	N_{i+1}
(H)N(CO)(CA)NH	Minimum	N_i
(H)N(CA)NH	Minimum	N_i
(H)N(CA)NH	Max/High intensity	N_{i+1}/N_{i-1}

Table 4.1: A table showing how peaks are classified in the SpinHub (in-house) assignment program for ^{1}H-^{15}N correlations from each residue i (NH_i) based on relative deconvolved peak intensities obtained using UniDecNMR.[386] N/A indicates where only one peak is expected per NH_i correlation. For (H)N(CA)NH peak lists, the peak corresponding to N_{i-1} could be classified by comparing to the (H)N(CO)(CA)NH experiment, where no peak is expected for N_{i-1} magnetisation.

the spectra.

4.5 Semi-automated assignment of Δ54-pVI

Once peak-picking was applied, the SpinHub semi-automated assignment software (in-house) was run on the obtained peak lists. Peaks associated with each NH coherence (NH_i) were initially classified according to table 4.1.

The semi-automated assignment software then found connections (edges) between NH_i coherences (nodes) based on similarities in i, $i-1$ and $i+1$ classification peaks for Δ54-pVI associated with different NH peaks. Strip plots indicating the rigorous justification of an edge (connection) between arbitrary labelled peaks 3H-N and 30H-N are shown in Figure 4.11A. The connectivities (edges) obtained between NH correlation peaks (nodes) are an example of a

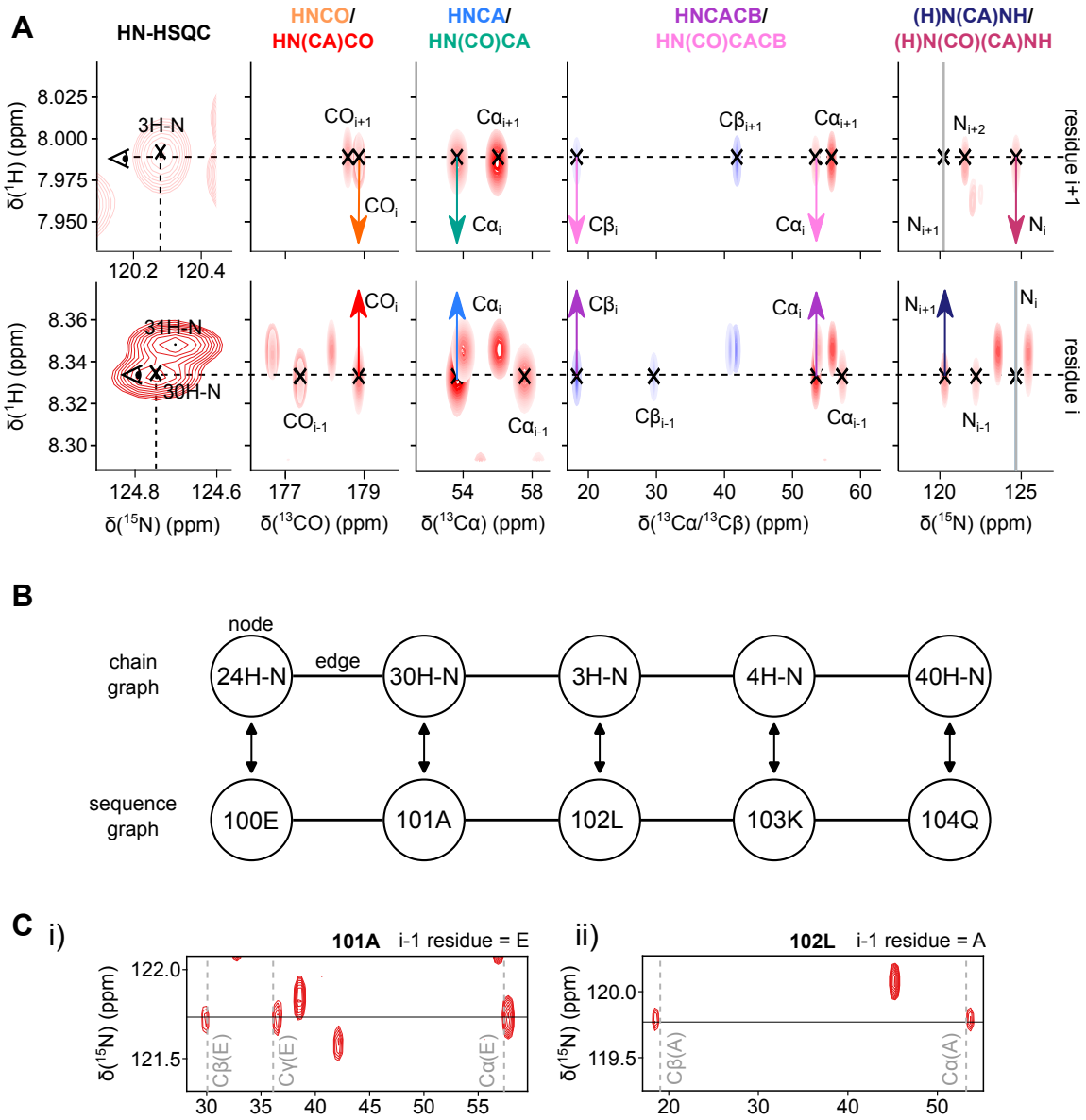


Figure 4.11: An overview of the assignment strategy used within the SpinHub semi-automated backbone assignment software package. **A** Strip plots showing connections between assignment spectra pairs for two arbitrarily named NH peaks (3H-N and 30H-N) for Δ54-pVI, obtained using the SpinHub assignment software. Positive and negative intensity is shown in red and blue, respectively. Peak classifications are shown (relative to residue i, 30H-N) and arrows indicate where the SpinHub assignment software detected proof of connection between the peaks. **B** An illustration of a chain graph determined by the SpinHub assignment software for Δ54-pVI and its mapping onto the protein sequence graph. **C** Strip plots down the ^{13}C dimension for the (H)CC(CO)NH experiment for the assigned 101A (i) and 102L (ii) NH correlation peaks. The mean BMRB[387] ^{13}C chemical shifts for the expected i-1 amino acid type are shown in grey (100E and 101A, respectively).

chain graph (Figure 4.11B). By comparing the obtained backbone carbon Cα/Cβ chemical shifts for each residue i to the Biological Magnetic Resonance Data Bank (BMRB)[387] chemical shift statistics, the software suggested likely residue types for each NH_i correlation peak.

It was necessary to map the chain graph onto the protein sequence graph to complete the assignment of Δ54-pVI (Figure 4.11B). The SpinHub backbone assignment software performs this graph matching based on the Methyl Assignment by Graph Matching (MAGMA) program,[388] which utilises a Maximum Common Edge Subgraph (MCES) approach.[389]

The automated assignment protocol performed well for Δ54-pVI but was incomplete. The disordered nature of the protein caused significant peak overlap in the assignment spectra, occasionally with peaks directly on top of one another. Also, some peaks had low intensity and therefore their intensities were below the threshold for the automatic peak picking routines. Therefore, additional peaks needed to be picked manually. For some of the residues whose peaks were significantly overlapped with others, the carbon chemical shifts had been incorrectly linked to another close-proximity NH residue peak and needed to be manually swapped. Furthermore, some residues had no/minimal peak intensity in the (H)N(CA)NH and (H)N(CO)(CA)NH spectra due to poor signal-to-noise. Therefore, the automatic assignment routine could not find a connection for these residues in these spectra. However, for residues missing peaks in the (H)N(CA)NH and (H)N(CO)(CA)NH spectra, as long as a connection was found to another residue for all other spectral pairs: HNCO/HN(CA)CO, HNCA/HN(CO)CA, HNCACB/HN(CO)CACB, this was added as an edge/connection manually.

The assignment of each residue was then confirmed through the use of the (H)CC(CO)NH experiment, which contains a ^{13}C side-chain fingerprint of the $i-1$ residue for each NH_i correlation peak, specific to each amino acid type (Figure 4.11C). This confirmation strategy provided confidence in the obtained assignments.

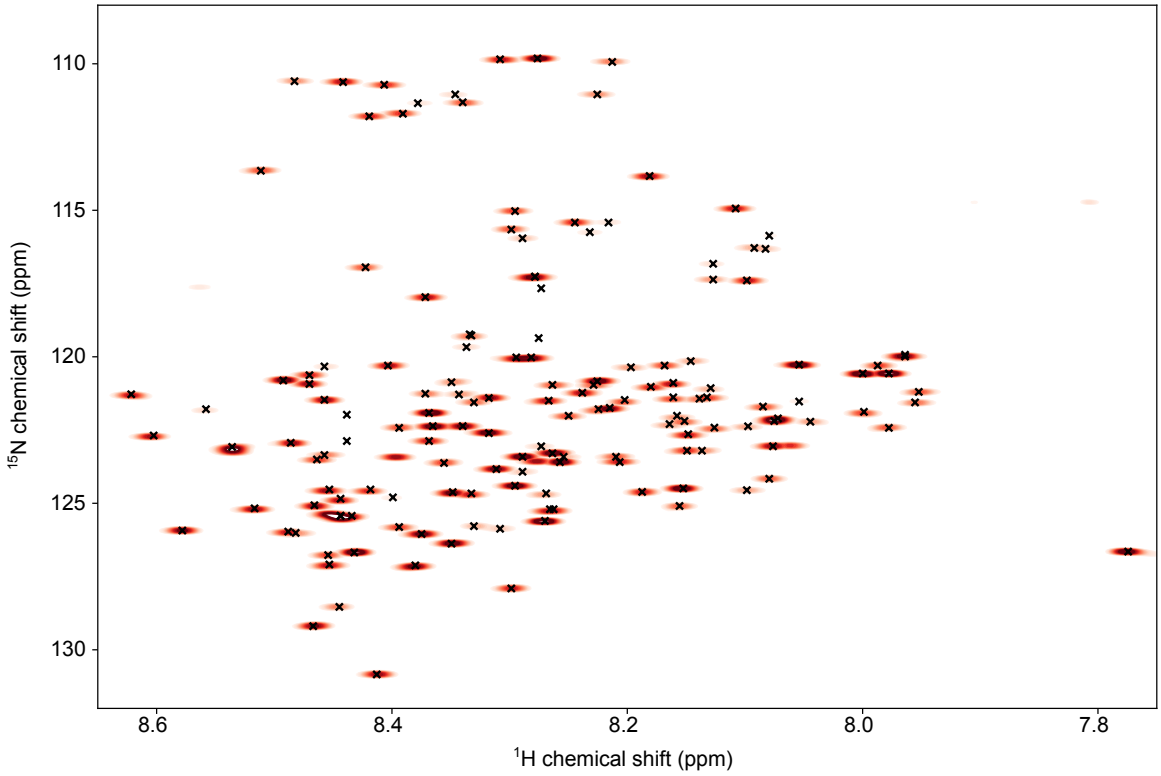


Figure 4.12: A plot showing the assigned ^{1}H-^{15}N backbone amide correlation peaks for Δ54-pVI at 5^{o}C. Crosses indicate peak locations of assigned residues. Information on the residue of each assignment can be found at: `https://james-eaton-1.github.io/d54pVI_Assignment/` (QR code in margin).

Following the semi-automated assignment protocol, 154 NH correlation peaks could be assigned for Δ54-pVI. This corresponds to 92% assignment (154 out of 167 peaks) of the total assignable (non-proline) NH correlation peaks for Δ54-pVI. The assigned NH correlation peaks are indicated in Figure 4.12. In the following sections, the backbone assignment of Δ54-pVI is used to determine the secondary structure propensity of the protein, and assess whether any transiently populated amphipathic helices are present.

4.6 Secondary structure propensity of Δ54-pVI revealed through residue-specific NMR chemical shifts

The residue-specific chemical shift information obtained for Δ54-pVI following the assignment can provide helpful insight into the secondary structure propensity

of the highly disordered protein. Several random coil chemical shift predictors exist, providing the expected chemical shifts of backbone nuclei within a protein, assuming it can access various conformations. Examples of commonly used random coil chemical shift predictors include POTENCI (Prediction Of TEmperature, Neighbour and pH Corrected shifts for Intrinsically disordered proteins),[390] which is a more extensive and rigorous version of ncIDP (Neighbour Corrected IDP Library),[391] and the tools developed from Kjaergaard, M. *et al.*[392,393] (also known as Poulsen IDP chemical shifts).

The presence of secondary structure (or transiently populated secondary structure) caused deviations from the calculated random coil chemical shifts in a predictable manner. Helical conformations will lead to an increase in the Cα chemical shift and a slight decrease in the Cβ chemical shift, whereas β-strand conformations will cause a decrease in the Cα chemical shift and an increase in the Cβ chemical shift.[362,394]

The differences between the observed Cα and Cβ chemical shifts and their POTENCI[390]/Kjaergaard[392,393] random coil values for Δ54-pVI are shown in Figure 4.13. Deviations from the random coil chemical shifts are consistent with the presence of α-helical conformations, especially near the N-terminus (residues 10-30 and 40-51) and a small section near residue 100 (residues 98-104).

TALOS (Torsion Angle Likeliness Obtained from Shift and Sequence Similarity)[395] is commonly used to determine secondary structure propensity from obtained chemical shifts. The TALOS+[396] and TALOS-N[397] programs (extensions of the original TALOS program[395]) were ran for comparison, producing Ramachandran plots for Δ54-pVI (Figure 4.14) representing the dihedral angles within the protein consistent with the observed chemical shifts. There is a significant presence of dihedral angles in the α-helical regions of the Ramachandran plot (Figure 4.5), indicating the presence of α-helical character.

The outputs of residue-specific secondary structure propensity for Δ54-pVI

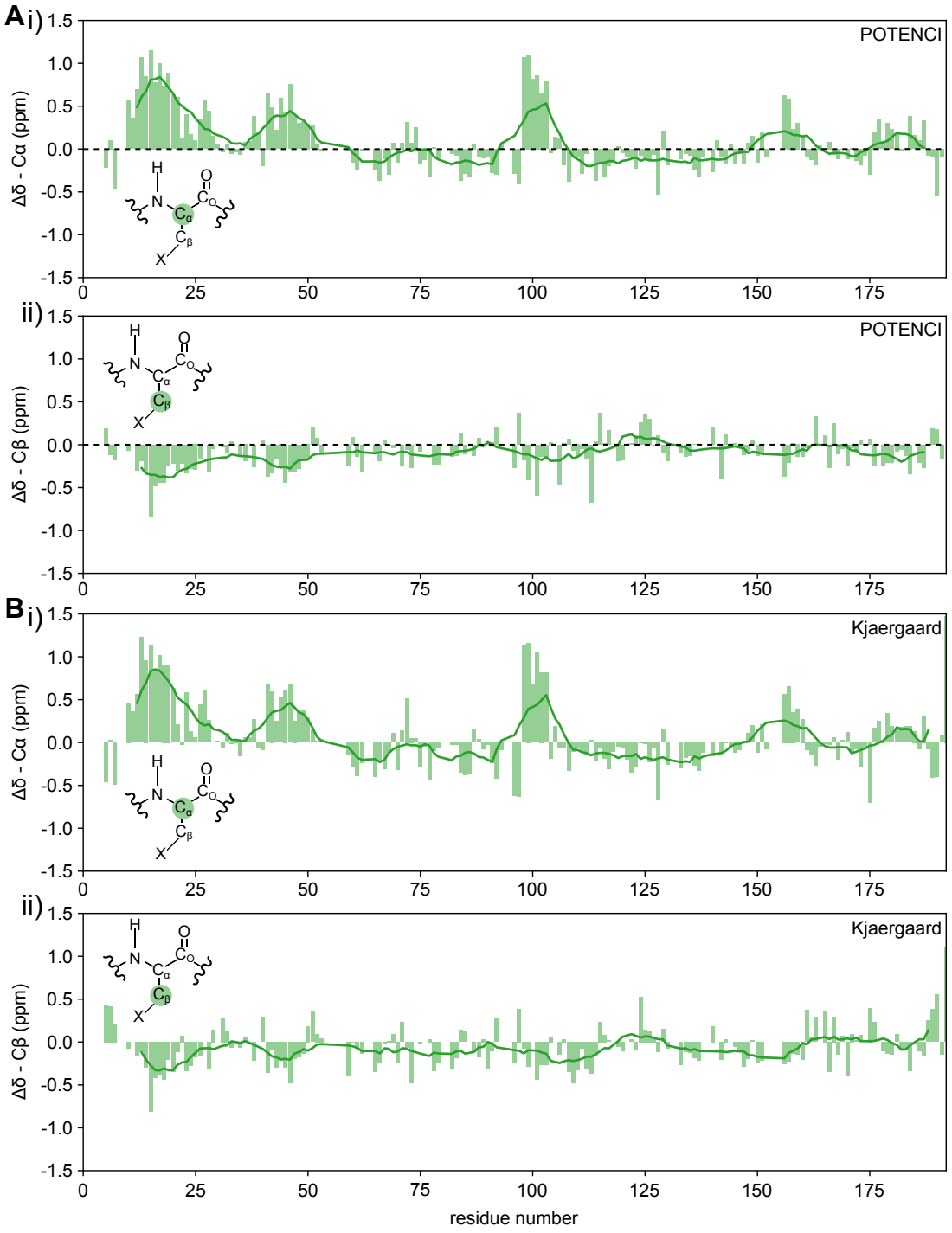


Figure 4.13: Deviations of the Δ54-pVI chemical shifts from random coil values. Random coil values were obtained from POTENCI[390] (**A**) and Kjaergaard[392,393] (**B**) were compared for both Cα (i) and Cβ (ii) chemical shifts from the assignment. Simple moving averages as a function of residue number were calculated using a window size of 10 residues and represented as bold green lines.

consistent with the observed chemical shifts are plotted for TALOS+ and TALOS-N, as well as two other commonly used secondary structure determination

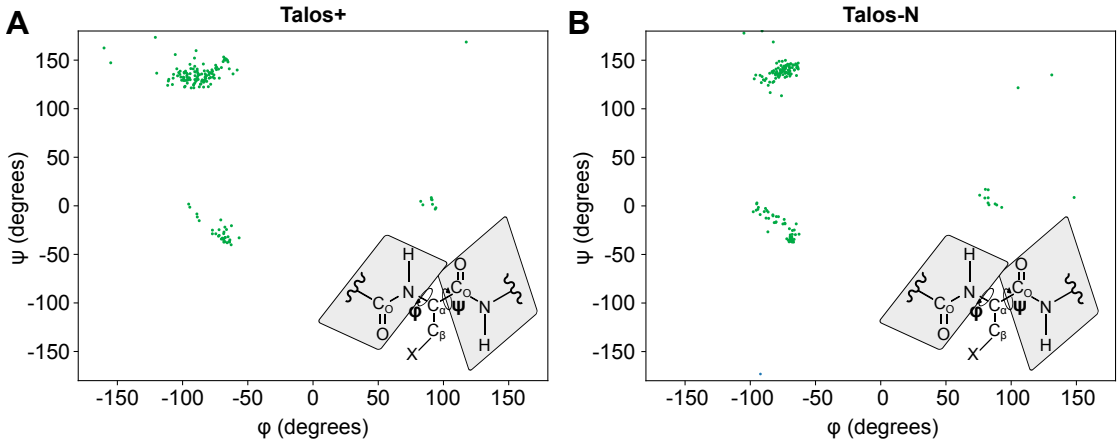


Figure 4.14: Ramachandran plot of dihedral angles in Δ54-pVI obtained using the TALOS+[396] (**A**) and TALOS-N[397] (**B**) programs.

programs: CheSPI (Chemical shift Secondary structure Population Inference)[398] and secondary structure propensity (SSP) score[399] (Figure 4.15). The high degree of agreement between the different programs provides confidence in the obtained residual secondary structure map. The residue-specific NMR peak intensities from a BEST-TROSY HSQC (with their intensities fitted using FuDA[400]) appear to have only minor correlation with the determined secondary structure propensity, indicating that the regions of Δ54-pVI containing secondary structure propensity remain highly dynamic in the solution state.

Previous studies have hypothesised that the amphipathic helix of mature protein VI extends beyond the N-terminus (33-54).[52] Assuming that the secondary structure propensity of the N-terminus of Δ54-pVI determined here is the same as that in the mature protein VI, there is evidence of α-helical secondary structure propensity at particular regions of the N-terminus, which are formed transiently (Figure 4.15). Furthermore, in the N-terminus, the nature of these α-helical regions are amphipathic (Figure 4.16) and likely contribute to the small amount of membrane lytic activity provided by Δ54-pVI.[52] However, it would be interesting to determine how the secondary structure propensity in Δ54-pVI changes as the temperature is increased above 5^{o}C and as the pH is reduced below 7.0. Furthermore, future NMR assignment of mature protein VI using NMR would provide confidence that the

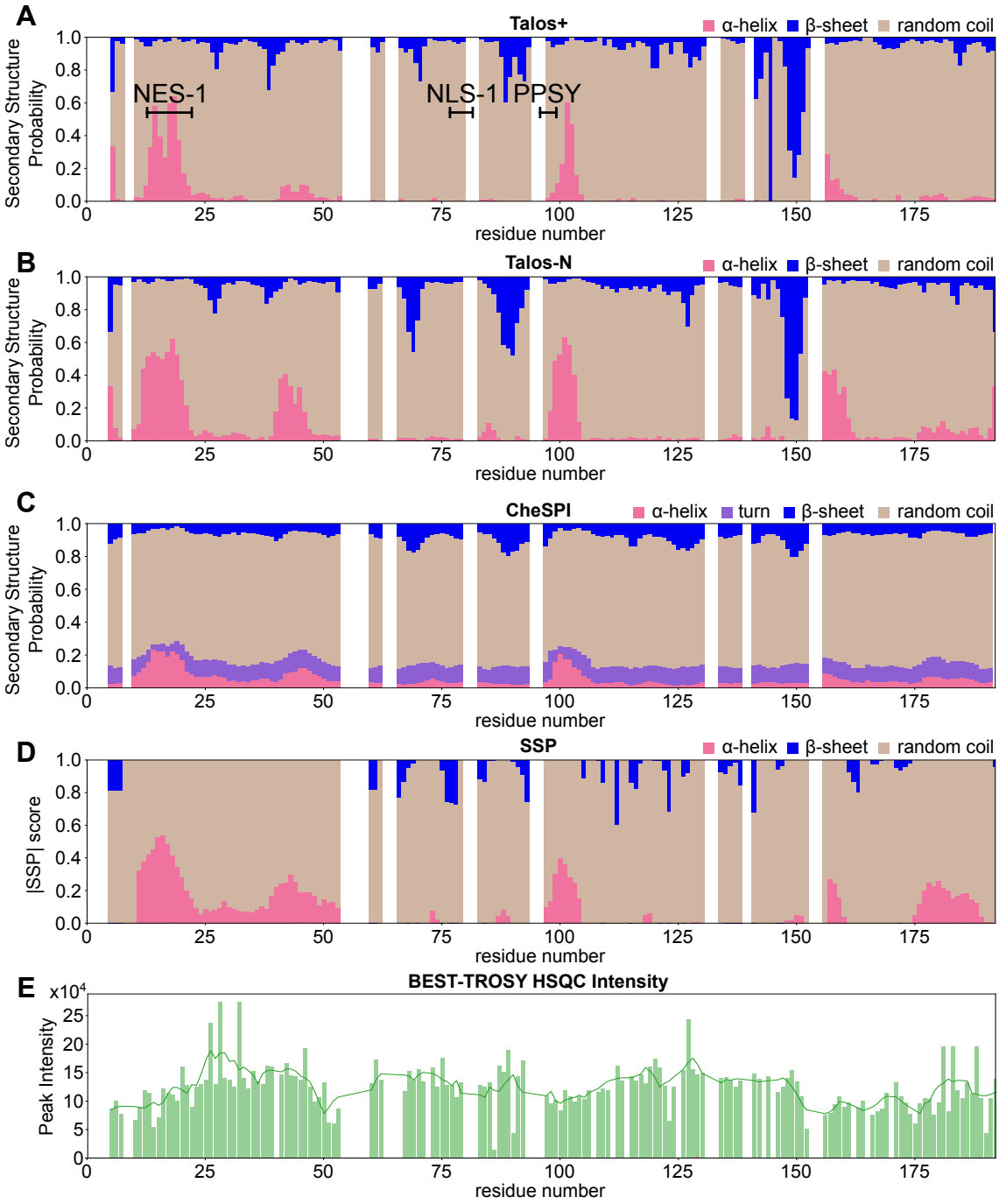


Figure 4.15: NMR determined residue-specific secondary structure propensity of Δ54-pVI obtained using the TALOS+[396] (**A**), TALOS-N[397] (**B**), CheSPI[398] (**C**) and SSP score[399] (**D**) programs are shown. White regions indicate regions where residues could not be assigned using NMR. For the CheSPI predictions in C, the violet band represents a turn conformation, where the backbone direction changes sharply. **E** NMR peak intensities for each residue within the BEST-TROSY HSQC spectrum of Δ54-pVI. Peak intensities (bars) were calculated using FuDA[400] and a smoothed intensities (line) were produced using Savitzky-Golay filtering[401] with a polynomial order of 3 and a window size of 10.

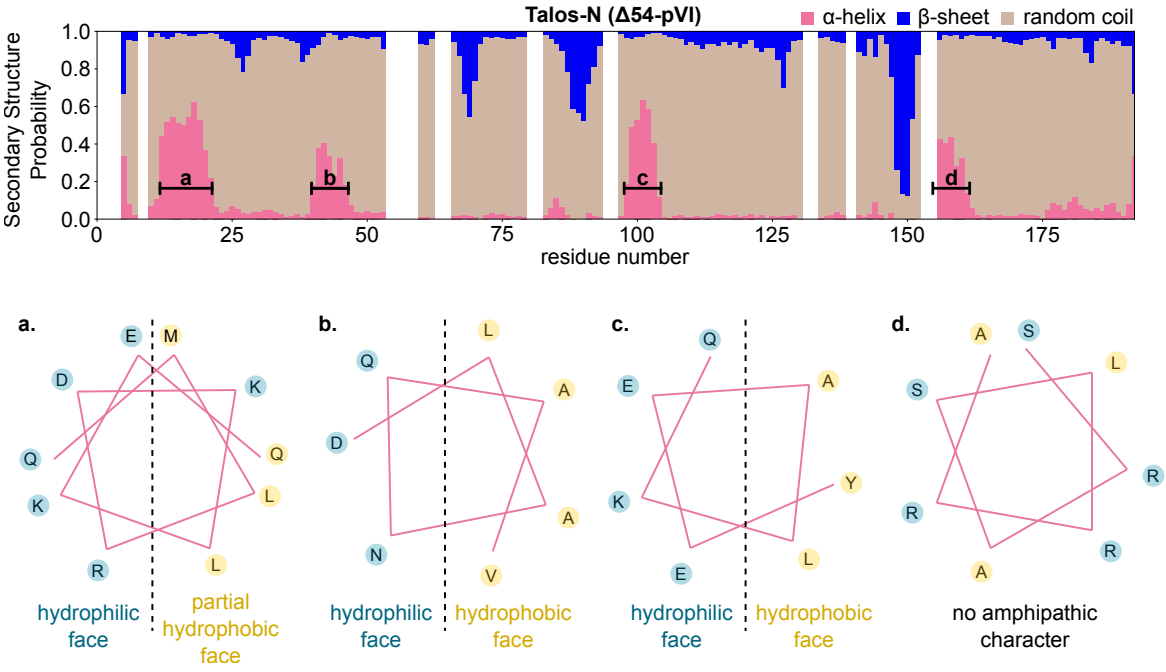


Figure 4.16: The amphipathic nature of experimentally determined α-helical secondary structure propensity (regions a-d) in Δ54-pVI. Helical plots were generated using the pepwheel tool from the European Molecular Biology Open Software Suite (EMBOSS).[345] The secondary structure propensity from TALOS-N is also shown, highlighting the positions of a-d along the sequence.

observed secondary structure found for the Δ54-pVI construct is consistent with that of the active adenoviral protein VI.

Aside from its membrane lytic ability, Δ54-pVI contains three key sequence motifs: the nuclear export signal (NES-1, residues 14-23), the nuclear localisation signal (NLS-1, residues 78-82) and PPxY (PPSY, residues 95-98). Both the NES-1 and PPSY motifs are located in or close to regions of the protein with a substantial degree of α-helical character (Figure 4.15A). NES-1 is a leucine-rich motif and is hypothesised to interact with exportin proteins that enable pVI to be shuttled out of the host cell nucleus and into the cytoplasm.[50,402] This is necessary for pVI to transport hexon produced in the host cell cytoplasm into the nucleus for downstream adenovirus capsid assembly.[50] Furthermore, PPxY motifs are known to recruit Nedd4 E3 ubiquitin ligases, which add ubiquitin units to the protein, so that it can be transported along microtubules. In protein VI, the PPSY has been demonstrated as essential for protein VI ubiquitylation.[344] The experimentally determined α-

helical propensity of Δ54-pVI at the NES-1 and PPSY motifs may prove crucial for its ability to interact with exportin and ubiquitin ligases, respectively, and be transported through a host cell. In the future, quantifying the binding of exportin and ubiquitin-ligases to these motifs and determining how the secondary structure propensity of Δ54-pVI is stabilised or de-stabilised, after the addition of their binding partner, will be essential to establish their full ability to transport pVI through a host cell. These measurements could be performed by NMR, observing chemical shift perturbations when the binding partner is titrated into the Δ54-pVI sample.

4.6.1 Comparison of experimental Δ54-pVI secondary structure to computational predictions

Computational predictors, including AlphaFold, are increasingly being used to predict secondary structure propensity in disordered proteins. Therefore, it was interesting to compare if the experimentally determined secondary structure propensity of Δ54-pVI agreed with computational predictions. Firstly, ColabFold,[5,403–407] a faster version of AlphaFold2,[5] was ran on the Δ54-pVI sequence to predict its structure (Figure 4.17A). The dictionary of protein secondary structure (DSSP) algorithm was performed on the structure to determine the predicted secondary structure along the protein sequence (Figure 4.17B).[414] As expected for a predominantly disordered protein, the confidence scores (pLDDT: predicted local distance difference test) of the AlphaFold2 prediction are relatively low across the sequence (Figure 4.17B). However, the prediction suggested the presence of a significant α-helical region at the N-terminus. A host of alternative secondary structure predictors (ESMFold,[408] JPred4[409]) and disorder predictors (ODiNPred,[410] MetaPredict3.0[411,412] and Disopred3[413]) were run on the Δ54-pVI sequence to compare and contrast the predictions of difference prediction engines (Figure 4.17C-E). All the prediction tools used predicted the presence of α-helical

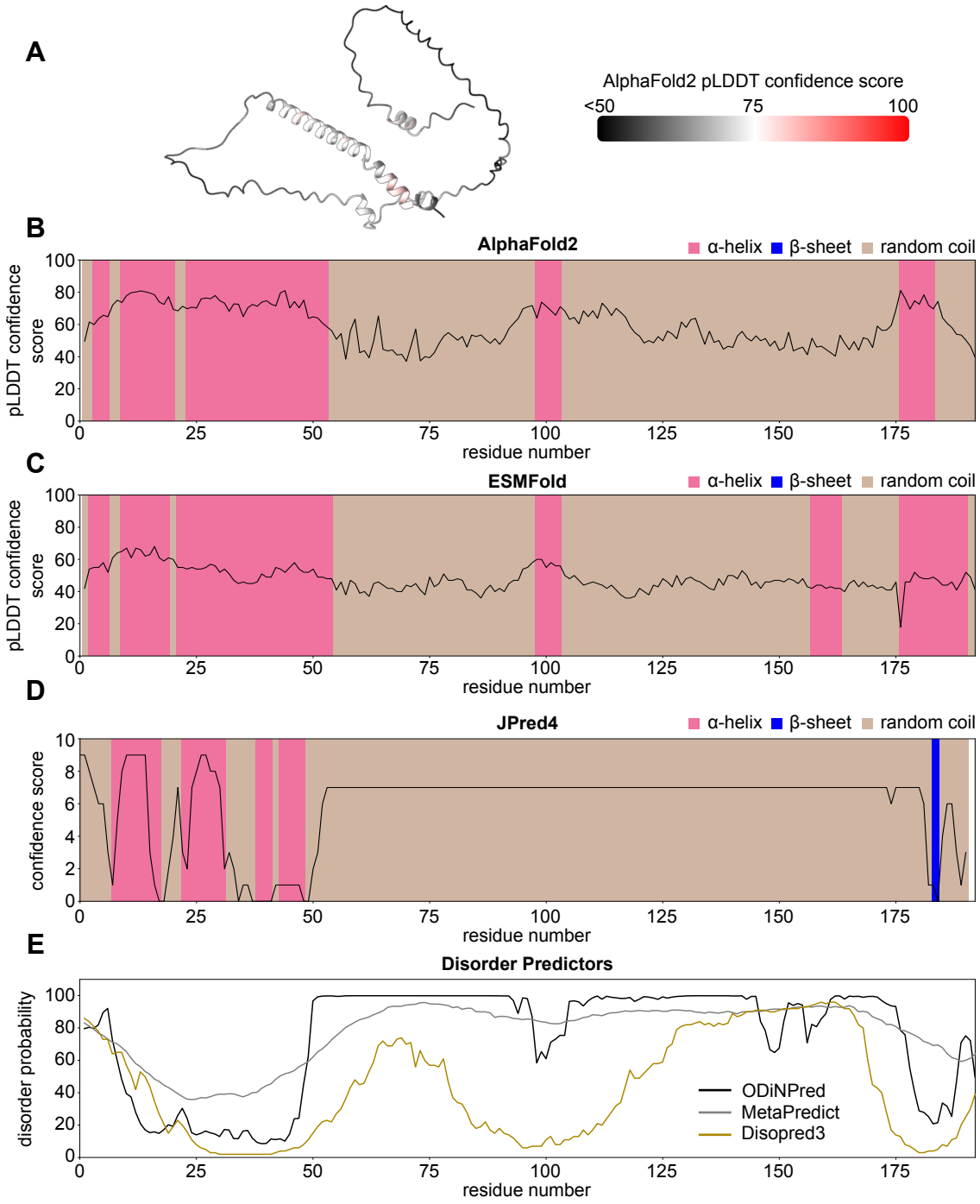


Figure 4.17: Comparison of computational secondary structure/disorder predictions for Δ54-pVI. **A** The predicted structure of Δ54-pVI obtained using ColabFold,[5,403–407] a faster version of AlphaFold2.[5] Comparisons of secondary structure predictions for Δ54-pVI obtained using **B** ColabFold/AlphaFold2,[5,403–407] **C** ESMFold,[408] and **D** JPred4.[409] **E** Disorder probabilities as a function of residue number computed using ODiNPred,[410] MetaPredict3.0[411,412] and Disopred3.[413]

character at the N-terminus of Δ54-pVI,

However, when aligned with the experimentally determined solution state

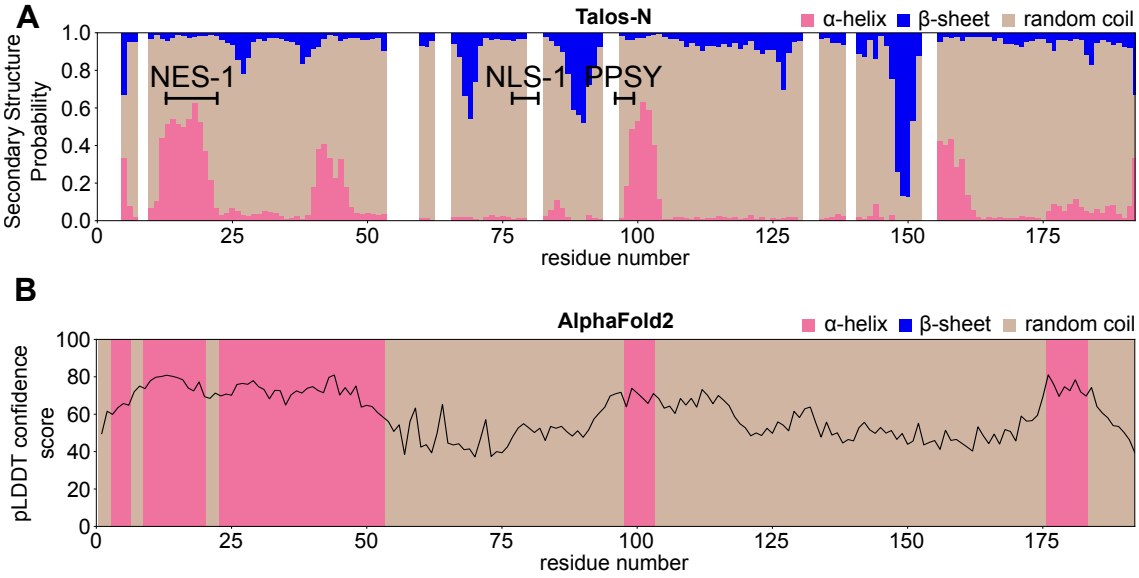


Figure 4.18: Comparison of the experimental secondary structure propensity of the disordered protein Δ54-pVI with the AlphaFold2 prediction. **A** The experimentally determined secondary structure propensity of Δ54-pVI calculated using TALOS-N and alignment of key sequence motifs: the nuclear export signal (NES-1), the nuclear localisation signal (NLS-1) and PPSY, with secondary structure. **B** The AlphaFold2 predicted secondary structure of Δ54-pVI.

secondary structure using TALOS-N, AlphaFold2 (along with the other secondary structure predictors) appeared to have overestimated the α-helical nature of the Δ54-pVI. Similar discrepancies between AlphaFold structures and NMR experimental data of disordered proteins have been observed previously.[415] AlphaFold structures of disordered proteins have been shown to correlate with the structure obtained when bound to a binding partner, termed a conditionally folded state.[415] Therefore, it is possible that when bound to a membrane or another binding partner, the N-terminal domain of Δ54-pVI is much more α-helical than when free in the solution state, and therefore, aligns itself more with the predicted AlphaFold structure. While AlphaFold generates interesting hypotheses for secondary structure propensity in disordered proteins, the results shown here (Figure 4.18) highlight that it is essential to assess secondary structure predictions experimentally and view them in context until models trained on NMR-based data of disordered proteins are produced.

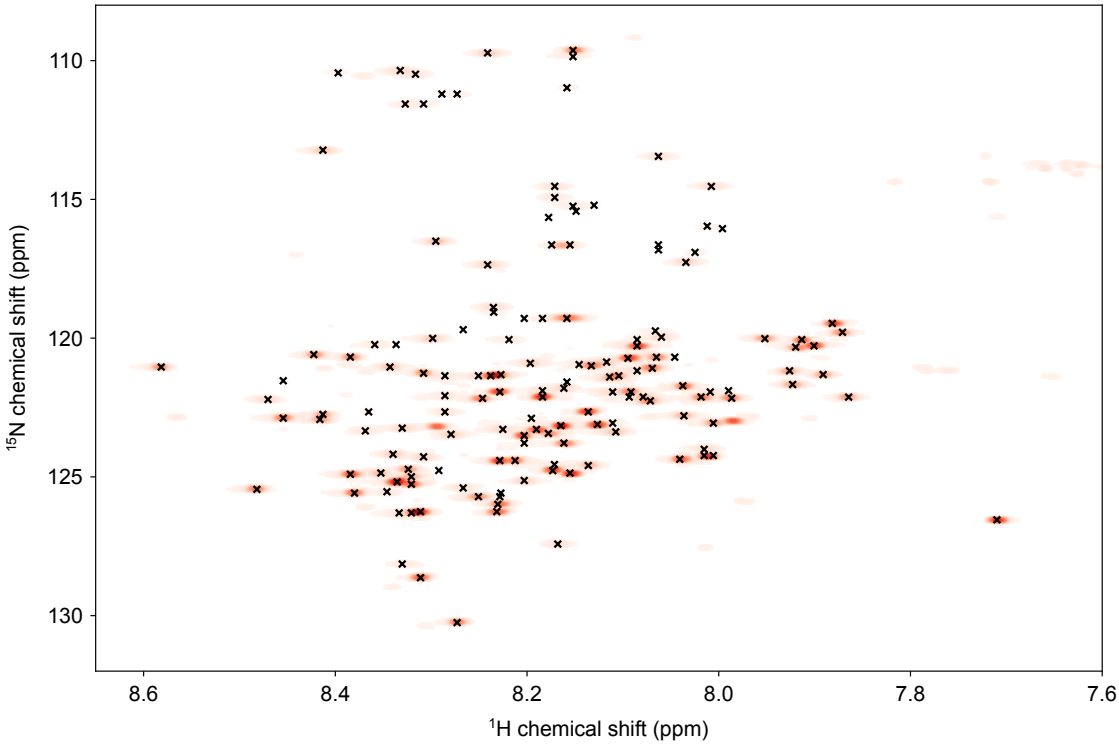


Figure 4.19: A plot showing the assigned ^{1}H-^{15}N backbone amide correlation peaks for Δ54-pVI at 20^{o}C. Crosses indicate peak locations of assigned residues. Information on the residue of each assignment can be found at: `https://james-eaton-1.github.io/d54pVI_Assignment_20C/` (QR code in margin).

4.7 Mapping the ^{1}H-^{15}N backbone amide correlation peaks assignments of Δ54-pVI from 5^{o}C to 20^{o}C

The ^{1}H-^{15}N assignment of Δ54-pVI was mapped from 5^{o}C to 20^{o}C by sequentially running BEST-TROSY HSQC experiments increasing the temperature by 2.5^{o}C at a time (see appendix B.2.2 for details). A total of 152 peaks (out of 154 peaks for 5^{o}C) could be mapped between the temperatures (Figure 4.19); however, a small number of additional peaks appeared that were not present at 5^{o}C. These additional peaks may be from protein degradation products.

The residue-specific chemical shift perturbations of the ^{1}H-^{15}N correlation peaks from 5^{o}C to 20^{o}C ($\Delta\Delta\delta$), relative to the mean chemical shift perturbation of all residues were calculated using the following formula in equation 4.3.[416]

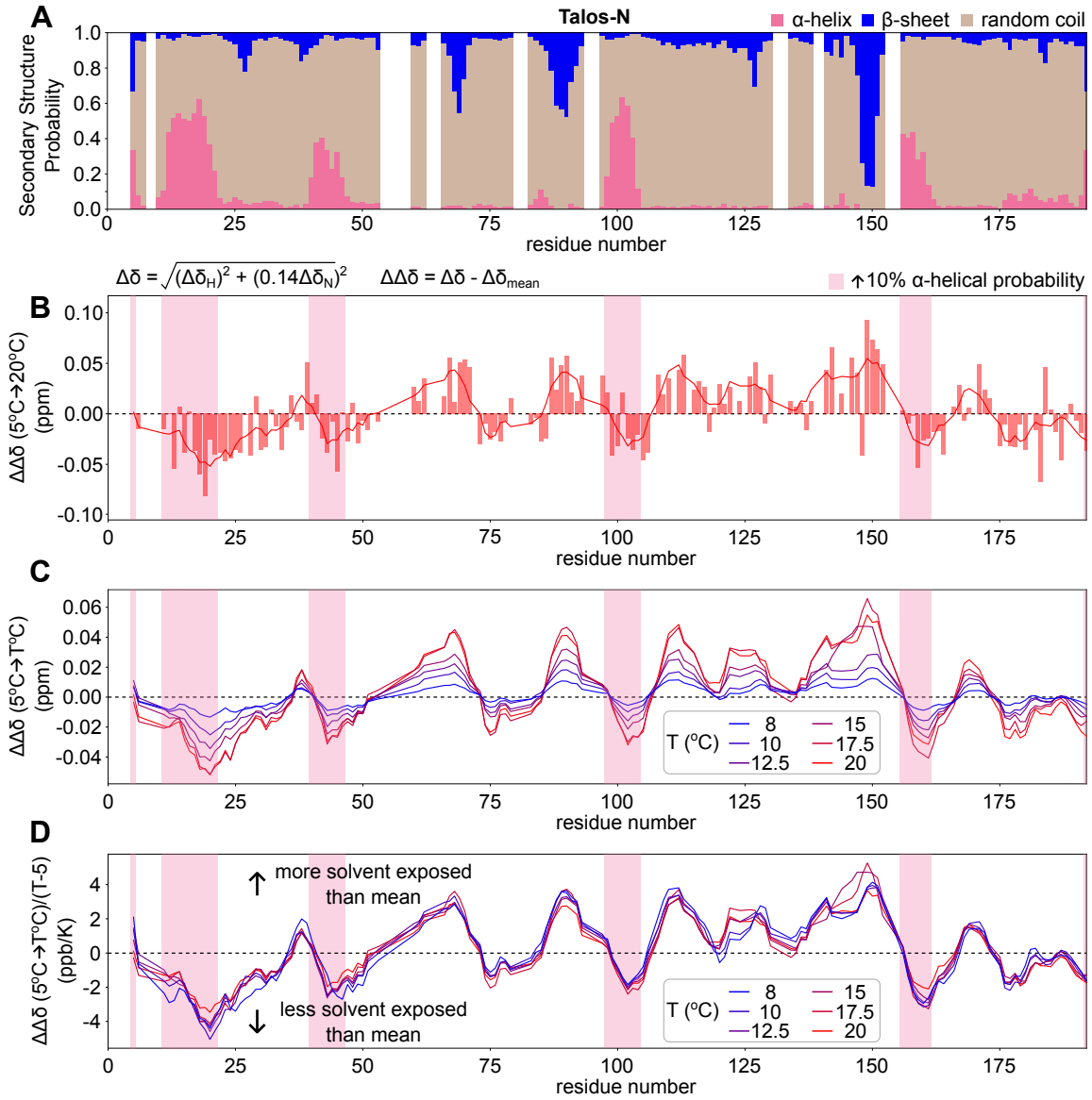


Figure 4.20: Temperature dependent chemical shift perturbations of Δ54-pVI revealing residue-dependent solvent exposure differences. **A** The NMR-based TALOS-N determined secondary structure propensity of Δ54-pVI as a function of residue number. **B** A plot residue-specific $\Delta\Delta\delta(5^o\text{C}\rightarrow20^o\text{C})$ chemical shift perturbation values. **C** Residue-specific $\Delta\Delta\delta(5^o\text{C}\rightarrow\text{T}^o\text{C})$ values as a function of final temperature (T). **D** Residue-specific temperature coefficients $[\Delta\Delta\delta(5^o\text{C}\rightarrow\text{T}^o\text{C})/(\text{T-5})]$ values for different temperatures T. Bars indicate true values and lines represent a smoothed profile calculated using Savitzky-Golay[401] smoothing using a window size of 10 and a polynomial order of 3.

$$\Delta\delta = \sqrt{(\Delta\delta_{\text{H}})^2 + (0.14\Delta\delta_{\text{N}})^2} \qquad \Delta\Delta\delta = \Delta\delta - \Delta\delta_{\text{mean}} \tag{4.3}$$

Subtraction of the mean value ($\Delta\delta_{\text{mean}}$) was performed to remove thermal effects

so that the relative solvent exposure of each residue could be assessed. The degree of solvent exposure is reflected in the temperature coefficients of the chemical shift perturbations observed, which can be calculated using:

$$\text{Temperature coefficient} = \Delta\Delta\delta(T_1 \rightarrow T_2)/(T_2 - T_1) \tag{4.4}$$

Using this definition, residues with a negative temperature coefficient value indicate a lower solvent exposure than the average residue in the protein, indicating enhanced intermolecular hydrogen bonding. The degree of solvent exposure appeared to correlate with α-helical secondary structure propensity within Δ54-pVI (Figure 4.20B). Furthermore, the temperature coefficients of each residue were roughly consistent as a function of final temperature between 8^oC and 20^oC, indicating minimal change in secondary structure propensity as the temperature was increased (Figure 4.20D). This implied that there was no substantial activation-limited conformational rearrangement of the protein that occurred when increasing the temperature to 20^oC. However, further experiments to obtain C_α and C_β chemical shifts of each residue at 20^oC should be obtained to quantify the secondary structure propensity at higher temperatures.

4.8 Chemical exchange experiments

After assessing the secondary structure propensity of Δ54-pVI, CPMG relaxation dispersion and CEST experiments were ran to determine whether there was evidence of chemical exchange in Δ54-pVI between different conformations.[†] CPMG and CEST experiments can assess chemical exchange between excited state and ground state conformations with exchange rates between 100 and 2000 s^{-1},[417] and $<$200 s^{-1},[418] respectively, as long as the excited state population is larger than 1%.

[†]Introductions to the CPMG and CEST experiments are provided in section 2.6

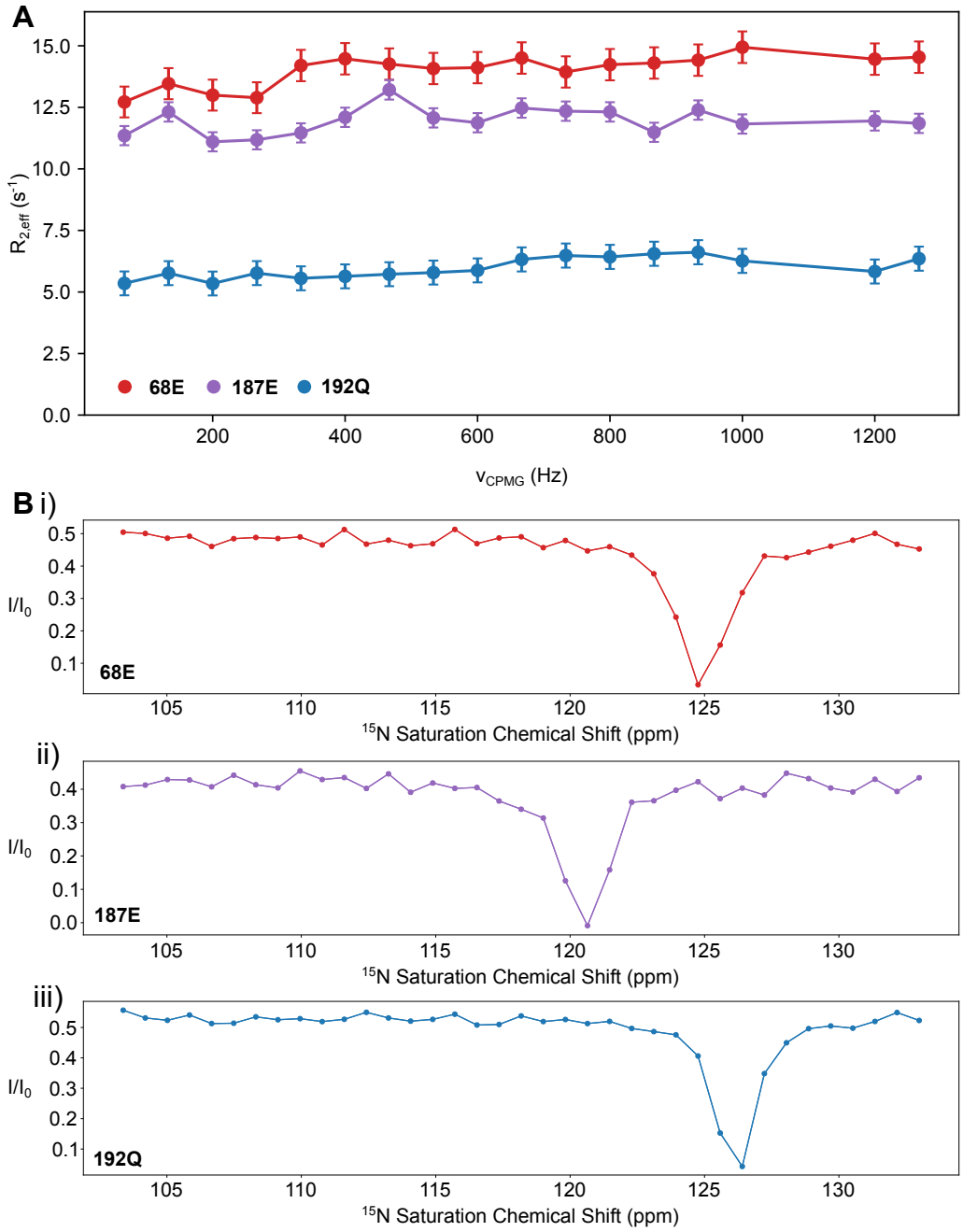


Figure 4.21: Absence of observable chemical exchange with millisecond (or greater) timescale dynamics measured through CPMG relaxation-dispersion and CEST. **A** CPMG relaxation dispersion plots showing $R_{2,eff}$ as a function of ν_{CPMG} for the residues 68E, 187E and 192Q in Δ54-pVI. **B** Z-spectra plots showing the ^{15}N CEST profiles for the residues 68E (i), 187E (ii), and 192Q (iii) in Δ54-pVI. In both experiments, extreme peak overlap prevented residue-specific measurements for most of the whole protein sequence. However, evidence of dispersions/secondary peaks were not observed in any region of the spectrum.

4.8.1 CPMG

A pseudo-3D CPMG experiment, based on a ^{1}H-^{15}N selectivity enhanced HSQC sequence including a relaxation element was performed on Δ54-pVI.* In the relaxation element, transverse ^{15}N magnetisation was created and allowed to relax for a duration of 30 ms (T_{relax}). During this time, a series of τ-180^{o}_{N}-τ spin-echo (CPMG) units were applied with a varying frequency (ν_{CPMG}) up to 1266 Hz giving peak intensities I(ν_{CPMG}). A reference spectrum without any relaxation delay was also recorded giving peak intensities I_0. For each ν_{CPMG} value, a ^{1}H-^{15}N 2D spectrum was obtained and the the effective transverse relaxation rate for each ^{1}H-^{15}N correlation peak ($R_{2,eff}$) was calculated using the following formula:

$$R_{2,eff}(\nu_{CPMG}) = -\frac{1}{T_{relax}} \ln\left(\frac{I(\nu_{CPMG})}{I_0}\right) \tag{4.5}$$

The relaxation dispersions showed no evidence of chemical exchange in Δ54-pVI with exchange rates on the order of 100-2000 s^{-1}, as indicated by a lack of dispersions (Figure 4.21A). This is represented for 3 peaks; however, dispersions could not be found for any peak in Δ54-pVI.

4.8.2 CEST

A pseudo-3D CEST experiment was performed based on a ^{1}H-^{15}N selectivity enhanced HSQC sequence coupled with a continuous wave ^{15}N irradiation element.† The ^{15}N irradiation was applied for a total of 500 ms with a range of chemical shifts between 103 ppm and 133 ppm and had an effective field ($B_{1,eff}$) of 30 Hz. A reference spectrum without any continuous wave ^{15}N irradiation was also recorded

*Experimental details on the CPMG experiment used and subsequent analysis are provided in appendix method B.2.3.

†Experimental details on the CEST experiment used and subsequent analysis are provided in appendix method B.2.3.

(I_0). For each ^{15}N irradiation frequency, a ^{1}H-^{15}N 2D spectrum was obtained and the ratio I/I_0 was calculated enabling the production of Z-spectra for each peak (Figure 4.21B). The Z-spectra showed no secondary peaks indicating there was no evidence of chemical exchange of the Δ54-pVI ground state to excited states (>1% populated) with exchange rates <200 s^{-1} (Figure 4.21B). Figure 4.21B shows the Z-spectrum for 3 peaks; however, evidence of chemical exchange (secondary peaks) could not be found for any peak in Δ54-pVI. The lack of relaxation dispersions in the CPMG experiments or secondary peaks in the CEST does not rule out the presence of chemical exchange. The populations of the excited states within Δ54-pVI may be too low for detection, or the chemical exchange rate may be greater than 2000 s^{-1}.

4.9 Δ54-pVI interaction with micelles

The detergent n-dodecyl-β-D-maltoside (DDM) is known to form micelles (≈70 kDa[420]) in aqueous solutions at concentrations greater than approximately 130 μM,[419] the critical micelle concentration (CMC), and is often used as a lipid membrane mimic and as a medium for purifying hydrophobic membrane proteins[421] (Figure 4.22A). Therefore, DDM was chosen to determine if there was any interaction between Δ54-pVI and lipids responsible for its membrane lytic ability, and possibly relevant to the function of mature protein VI in endosomal escape. At DDM concentrations above the critical micelle concentration, the signal intensity of Δ54-pVI (200 μM) decreased significantly in a residue-independent manner, which indicated binding of Δ54-pVI to the DDM micelles (Figure 4.22B/C).[†] Minimal chemical shift perturbations were observed between the Δ54-pVI only and Δ54-pVI + DDM samples. While interaction of Δ54-pVI with the DDM micelles is likely driven by the amphipathic-helix secondary structure propensity (Figure 4.16), conditionally folding into a more stable helical structure in contact with

[†]For experimental details see appendix B.2.4.

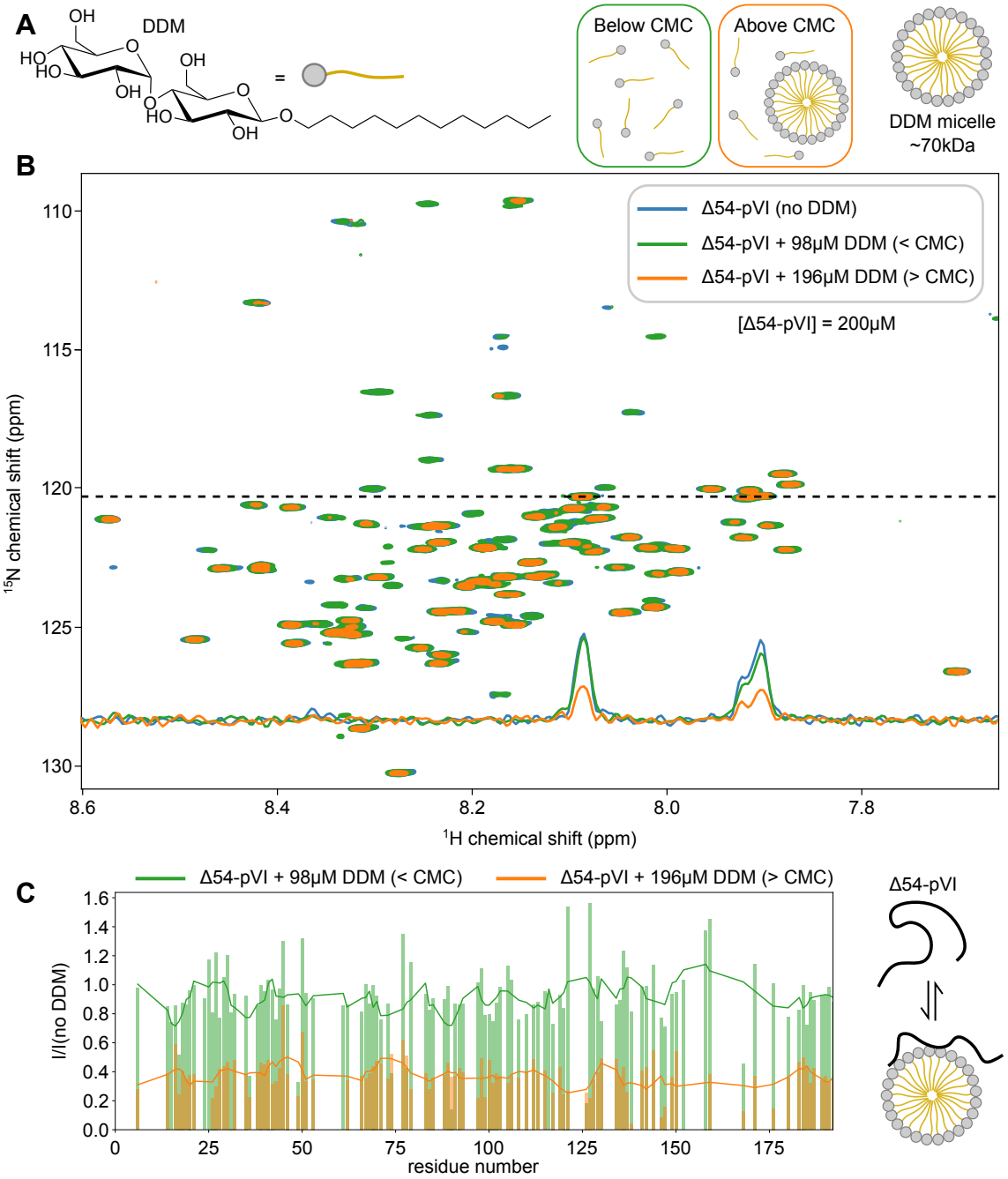


Figure 4.22: Interaction of Δ54-pVI with DDM micelles. **A** The chemical structure of n-dodecyl-β-D-maltoside (DDM) and a depiction of DDM micelle formation when at concentrations above the critical micelle concentration (CMC = 130 μM[419]). **B** BEST-TROSY HSQC spectra of Δ54-pVI containing DDM at concentrations above (196 μM, orange) and below (98 μM, green) the critical micelle concentration. A control with no DDM is also shown (blue). **C** The ratio of fitted peak intensities (I) against those in a DDM-free sample (I(no DDM)), as a function of Δ54-pVI residue number, at DDM concentrations below (98 μM) and above (196 μM) the CMC. Bars indicate raw intensity ratios. A smoothed line was generated using a Savitzky-Golay[401] filter with a window size of 10 and a polynomial order of 3.

the micelle, interactions with more disordered residues cannot be ruled out. The decrease in the global tumbling rate of Δ54-pVI when bound to the large DDM micelles causes an increase in the transverse relaxation rate, thereby decreasing the overall signal intensity observed for each peak (see section 2.1.3.iv). In future work, a detergent with a smaller micelle size could be used so that the extent of residue-specific binding of Δ54-pVI to micelles can be established. The detergent n-dodecyl-β-melibioside (DDMB) forms micelles with a mass of ≈42 kDa,[420] smaller than the ≈70 kDa DDM micelles.[420] Furthermore, DDBM has been shown to give enhanced resolution (smaller linewidths) in NMR experiments on membrane proteins compared to DDM, indicating it as a good alternative for DDM in future work.[420]

4.10 Conclusions

This chapter has provided a detailed overview of the solution state properties of the adenovirus protein Δ54-pVI and its relation to endosomal escape and other functions of protein VI. NMR diffusion measurements were indicative of Δ54-pVI existing in a highly disordered nature as a random coil. However, a ^{15}N-^{13}C isotopically enriched sample of Δ54-pVI was used to assign residues within the protein and determine the residue-specific NMR chemical shifts (H, N, CO, Cα, Cβ). The assignment was performed using an in-house semi-automated assignment pipeline to improve the speed and accuracy of protein NMR assignments. (H)N(CA)NH and (H)N(CO)(CA)NH experiments were added to the assignment package to provide extra confidence in the assignments made, especially in groups of highly overlapped peaks.

The obtained chemical shifts for Δ54-pVI showed deviations from their expected random coil values (Figure 4.23A). These deviations showed that Δ54-pVI had substantial secondary structure propensity, particularly the emergence of α-helical regions (Figure 4.23B). The determined α-helical secondary structure propensity

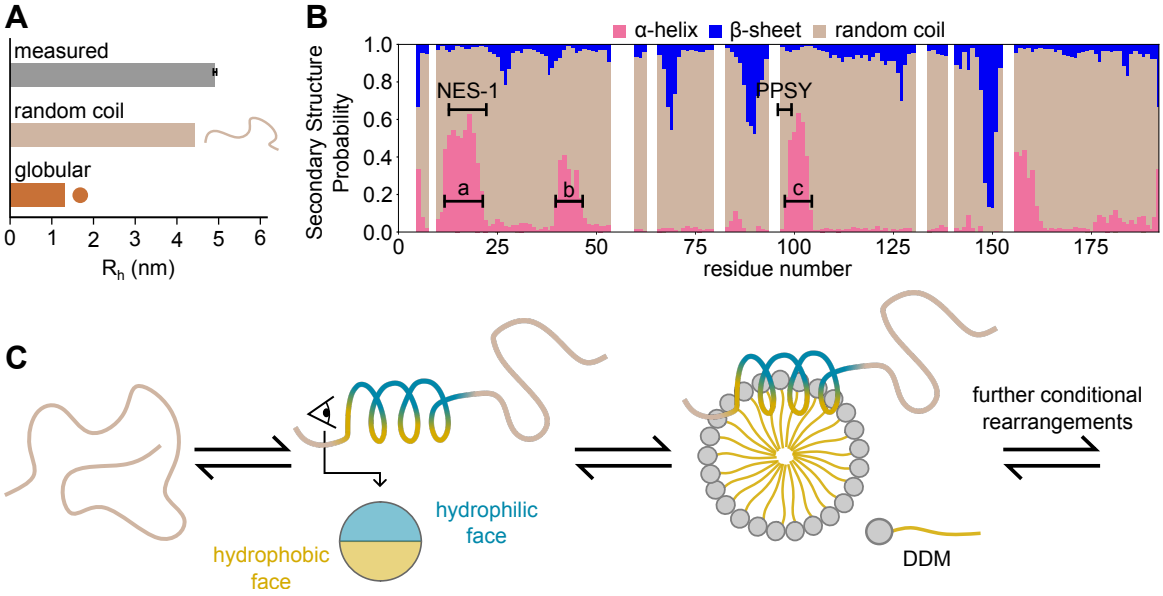


Figure 4.23: An overview of Δ54-pVI secondary structure propensity and its role in membrane binding and endosomal escape. **A** The NMR diffusion determined hydrodynamic radius (R_h) which indicated Δ54-pVI is highly disordered. **B** There is a significant degree of secondary structure propensity for Δ54-pVI, particularly for amphipathic helical regions a-c (secondary structure determined from NMR chemical shifts using TALOS-N). **C** A proposed mechanism by which Δ54-pVI can interact with membranes (or micellar membrane mimics).

correlated highly with the degree of solvent exposure expressed through temperature-dependent chemical shift perturbations (Figure 4.20B). The α-helical secondary structure propensity of Δ54-pVI (regions a,c) aligned with key sequence motifs such as the nuclear export signal (NES-1) and PPxY motifs, proposed to bind to exportin and E3 ubiquitin ligases, respectively. Furthermore, three of these α-helical regions (a-c) are highly amphipathic when formed (Figure 4.23B/C), with a hydrophobic and hydrophilic face, indicating their potential membrane binding properties.

A substantial decrease in NMR signal intensity of Δ54-pVI was observed when placed in a buffer containing membrane-mimicking DDM micelles, implying a significant interaction between Δ54-pVI and the high molecular weight micelles. A proposed mechanism of this interaction is through the amphipathic α-helical propensity (Figure 4.23C). The strong membrane lytic ability of mature adenoviral protein VI lies within the amphipathic N-terminal helix, not present in Δ54-pVI (VI_{33-54});[52] however, the results from this chapter show Δ54-pVI can still

facilitate membrane binding. In conclusion, knowledge of the secondary structure propensity and membrane interaction ability of the Δ54-pVI construct provides crucial information in the context of the wild-type protein VI and its varied roles in endosomal escape, microtubule transport of adenoviruses, and its nuclear export/import role, facilitating the recruitment of hexon for viral capsid assembly.

4.11 Future work

Chemical exchange within μs-ms dynamics timescales were not detected for Δ54-pVI using CEST and CPMG techniques. However, $R_{1\rho}$ experiments could be performed in the future to try to detect chemical exchange with exchange rates up to 25000 s^{-1}.[422] Furthermore, all chemical exchange experiments could be run in the presence of DDM to detect conformational exchanges upon binding to micelles. For example, Dark state Exchange Saturation Transfer (DEST) experiments,[423] suitable for measuring exchange between a fast and slow tumbling species, could be performed for this purpose. Furthermore, provided it could be expressed to a high enough concentration to facilitate NMR experiments, the residue-specific chemical shift of ^{15}N, ^{13}C-labelled full-length mature protein VI could be assigned. This would enable the determination of the secondary structure propensity across the entire virally active form of the protein, and assess whether the secondary structure propensity obtained using Δ54-pVI remains the same.

Aside from its membrane lytic function, the binding of key sequence motifs in protein VI, such as NES-1 and PPSY, to exportin and E3 ubiquitin ligases, respectively, could be determined through chemical shift perturbations. In addition, changes in chemical shifts could be used to assess whether the secondary structure propensity of protein VI is altered upon binding to these binding partners.

Interestingly, mature protein VI, and particularly the construct Δ54-pVI

are predicted to be able to phase separate, forming biomolecular condensates.* Biomolecular condensates are known for their membrane wetting ability.[156,157,294] They could help provide an increased concentration of protein VI at membrane surfaces, enhancing its membrane lytic ability for the endosomal escape of an adenovirus. However, phase separation of Δ54-pVI was not observed using the buffer conditions and concentrations in this chapter. Therefore, future work could be performed to experiment with buffer components and conditions to find an environment where phase separation can occur. If phase separation is induced, the wetting of such condensates to membranes could be deduced. Furthermore, the solvent environment of such condensates (if they can be produced) could be determined to assess how they may aid endosomal membrane lysis.

4.12 Contributions

All Δ54-pVI protein (unlabelled, ^{15}N-labelled, and ^{15}N-^{13}C labelled) was expressed and purified by Dr. Susan Black. The acquisition of all assignment NMR spectra was performed with the help of Dr. Charles Buchanan. All NMR spectrum processing, assignment and downstream analysis was performed by myself, with occasional help from Dr. Charles Buchanan and Professor Andrew Baldwin.

*Phase separation prediction scores (Δ54-pVI, DDX4^{N1}): FuzDrop (0.981, 0.999),[221,424,425] DeePhase (0.53, 0.68),[218] PSPredict (0.629, 0.998).[220] All the programs predicted Δ54-pVI to be a phase-separating protein.

5 Reducing NMR Limitations for Slowly Tumbling Biomolecules

5.1 Abstract

NMR spectra of large biomolecules are limited in signal intensity and resolution because of enhanced transverse relaxation, restricting the routine use of NMR to proteins less than 30 kDa. Transverse relaxation optimised spectroscopy (TROSY) schemes enhance the NMR spectral quality of backbone amide (^{15}N-H) and methyl group ($^{13}CH_3$) spin systems in large proteins. In-house software packages RelCalc and SimSeq were used to evaluate the benefits and limitations of current TROSY techniques. The $^{13}CF_3$ spin system was identified as a potential TROSY target. Unnatural $^{13}CF_3$-labelled methionine was introduced to a large 395 kDa protein to test its relaxation properties. Substantial relaxation interference was observed for the single quantum carbon coherences at 5^oC on a 600 MHz spectrometer. One coherence measured a relaxation rate of just 6.7 s^{-1}. All carbon relaxation rates were significantly reduced compared to their counterpart in a background deuterated $^{13}CH_3$-labelled analogue. Experimental 1D-carbon spectra recorded on the $^{13}CF_3$ labelled system showed the presence of multiple species containing a slow relaxing coherence. Simulations predicted that the slow relaxation rate of this coherence would be largely independent of molecular tumbling. Therefore, $^{13}CF_3$ labelling could be applied to study larger systems such as full virus complexes.

5.2 Introduction

The intensity and full-width half maximum (FWHM) linewidth of an NMR peak is directly dependent on the transverse relaxation rate (R_2) of the signal being observed. Transverse coherences with fast relaxation rates are substantially broader and more diminished than those with slower relaxation rates (Figure 5.1). A brief introduction to relaxation and its effect on NMR signal intensity and resolution is included in section 2.4, which is built upon in this chapter. As a first approximation, the transverse relaxation rate of an observable signal is proportional to $J(0, \tau_c)$, where $J(\omega, \tau_c) = \tau_c/(1 + \omega^2\tau_c^2)$ is the spectral density function (see section 2.1.4) and τ_c is the rotational correlation time (the time taken for the spin system to tumble through 1 radian), respectively.

The rotational correlation time τ_c has been shown to scale linearly with molecular weight (MW) in structured proteins below 30 kDa (τ_c(ns) $\approx$ 0.68MW(kDa)).[426] The function $J(0, \tau_c)$ increases drastically with increased τ_c, thereby causing enhanced relaxation and broader peaks for structured proteins with large molecular weights (Figure 5.1). Furthermore, larger proteins contain more nuclei, which leads to spectral crowding in the resultant NMR spectra. Therefore, using NMR to study structured proteins with molecular weights greater than 30 kDa becomes challenging using conventional protein NMR techniques.[94] This is evident in the number of entries into the BMRB as a function of molecular weight, which peaks at $\approx$10 kDa before falling sharply (Figure 5.2A). The average molecular weight of eukaryotic, bacterial and archeal proteins significantly exceeds 10 kDa,[427] implying that the majority of NMR research is into smaller than average proteins or smaller constructs of larger proteins. The analogous plot for all structural entries in the PDB is also shown for comparison (Figure 5.2B), with proportionally more depositions at higher molecular weights. Therefore, key dynamical information obtainable from NMR studies is missing for most proteins.

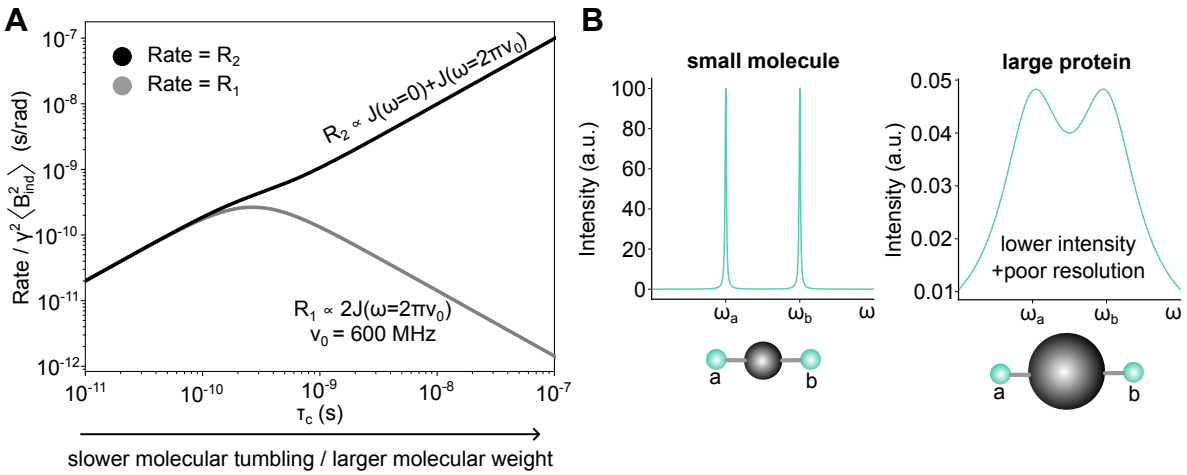


Figure 5.1: Recap of the spectral density function and its relation to relaxation. **A** Plots of the transverse (R_2) and longitudinal (R_1) relaxation rates of a spin with gyromagnetic ratio γ experiencing an induced field with a root-mean-squared magnitude of $\langle B_{ind}^2 \rangle$ as a function of rotational correlation time τ_c for ν_0=600 MHz. **B** Plots demonstrating the NMR signal intensity and resolution problem for slowly tumbling, large molecular weight species.

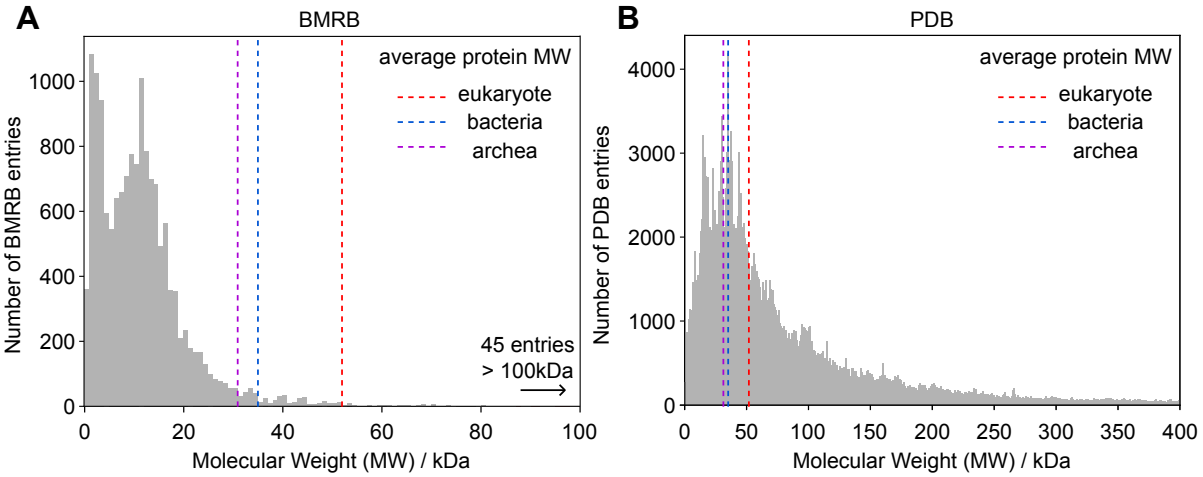


Figure 5.2: BMRB and PDB entries versus molecular weight. **A** The distribution of L-polypeptides with BMRB chemical shift entries as a function of molecular weight using a bin size of 1 kDa. Data obtained from the BMRB[387] on 12/05/2025. The median molecular weight of a BMRB entry was approximately 10 kDa. **B** The distribution of PDB structure entries as a function of molecular weight using a bin size of 20 kDa. Data was obtained from PDB[2] statistics on 05/06/2025. The molecular weights relate to the asymmetric unit for each crystal structure. For both plots, the mean protein molecular weights (MW) of known proteins in eukaryotic, bacterial and archeal cells are also shown.[427] The histogram bin widths were set to 1 kDa.

Section 5.2.2 evaluates the current methods used to combat relaxation and increase NMR signal intensity/resolution in large biomolecules. Following this, the

relaxation properties of the $^{13}CF_3$ spin system are explored in a 395 kDa heat-shock chaperone protein to assess whether it can be utilised as a transverse relaxation optimised NMR probe. Its relaxation properties are compared to those in the current gold standard method of $^{13}CH_3$ labelling and background deuteration. Finally, the most suitable TROSY experiment for the $^{13}CF_3$ probe is assessed.

5.2.1 NMR relaxation and an introduction to RelCalc/SimSeq

Transverse and longitudinal relaxation phenomena (with rates R_2 and R_1, respectively) were introduced in section 2.4. However, in spin systems containing more than one spin, the concept of a single transverse (R_2) and longitudinal (R_1) relaxation rate does not hold. Instead, a more general theory of relaxation is required. The Liouville-von Neumann equation was introduced in section 2.1.5 to explain the evolution of magnetisation ρ in a system. However, including the effects of relaxation leads to the following expression governing the time-dependence of magnetisation ρ:[86]

$$\frac{d\rho}{dt} = -i\hat{\hat{H}}_0\rho + \hat{\hat{R}}(\rho - \rho_{eq}) \tag{5.1}$$

where $\hat{\hat{H}}_0$ is the super-operator representing the static interactions in the system such as the Zeeman interaction, isotropic part of the chemical shift and scalar coupling. The actions of NMR pulses are also included in $\hat{\hat{H}}_0$.[†] The return of the system ρ to equilibrium (ρ_{eq}) is governed by the super-operator $\hat{\hat{R}}$.

The relaxation super-operator contains all the time dependent interaction terms that fluctuate depending on molecular motion but also have an ensemble average of zero. The dipolar interaction and chemical shift anisotropy are examples of these

[†]The super-operator $\hat{\hat{H}}_0$ is shorthand for the term $[\hat{H}_0,]$.

time-dependent interaction terms. The dipolar interaction between two nuclei (I+S) and the chemical shift anisotropy of nucleus I are governed by the constants d_{IS} and c_I, respectively:

$$d_{IS} = -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi r_{IS}^3} \qquad c_I = \gamma_I B_0 \Delta\delta_I / 3 \tag{5.2}$$

where γ_I and γ_S are the gyromagnetic ratios of nuclei I and S, r_{IS} is the internuclear $I-S$ distance, B_0 is the spectrometer magnetic field strength, $\Delta\delta_I$ is the chemical shift anisotropy constant for nucleus I (with the CSA tensors of all nuclei in the remainder of this chapter assumed to be axially symmetric). The constants μ_0 and $\hbar$ represent the vacuum magnetic permeability and Planck's constant, respectively.

Suppose the total magnetisation in a system is represented by an orthonormal basis ($\rho = \sum_i c_i \rho_i$). In that case, a complete Liouvillian evolution matrix (L) can be created to describe the time-evolution of the system ($\frac{d\rho}{dt} = L\rho$, $\rho(t) = e^{Lt}\rho(0)$). Each element of the matrix connecting magnetisation components ρ_a and ρ_b can be expressed as $\langle\rho_a|\,\hat{L}\,|\rho_b\rangle$ which can be split up into static Hamiltonian terms $H^0_{a,b} = -i\,\langle\rho_a|\,\hat{H}_0\,|\rho_b\rangle$, and relaxation terms $R_{a,b} = \langle\rho_a|\,\hat{R}\,|\rho_b\rangle$. This formalism allows the act of relaxation in a general spin system to be evaluated. However, for an isolated proton, the concept of R_2 and R_1 relaxation rates can be recovered with R_{H_+,H_+} representing the transverse relaxation rate R_2, and R_{H_z,H_z} represents the longitudinal relaxation rate R_1.

RelCalc[54] is a Python program that can calculate symbolic expressions for relaxation rates $R_{a,b}$ in a given spin system (AX_n) which can then be evaluated.* Packages producing symbolic relaxation rate expressions using Mathematica[428,429] have been developed previously and perform matrix multiplication to determine

*RelCalc is an in-house piece of software developed by Professor Andrew Baldwin based on Bloch-Wangsness-Redfield (BWR) relaxation theory.[84–86]

the relevant commutator relations. In contrast, RelCalc obtains all required commutators from a pre-determined library for calculations involving a spin system containing only spin-1/2 nuclei. RelCalc can include the effects of local motion (e.g. rotation about the three-fold rotation axis in an AX_3 system) with a rotational correlation time τ_m and global tumbling with a correlation time τ_c. Furthermore, additional external spins (such as protons or deuterons) outside the covalently bound spin system (AX_n) of interest can be added, and their effect can be simulated. A general expression for a relaxation matrix element $R_{a,b}$ and further details are included in appendix C.1.

In addition, SimSeq[55] is a pulse sequence simulator which simulates the full Liouvillian matrix for a spin-system based on the evaluated relaxation rates obtained using RelCalc.* In SimSeq, the matrix elements of the complete Liouvillian relating to pulses and the Zeeman interaction are calculated according to the following:[75]

$$H_{a,b} = -i \langle \rho_a | \hat{H} | \rho_b \rangle = -i \frac{\mathrm{Tr}\left(\rho_a [\hat{H}, \rho_b]\right)}{\mathrm{Tr}(\rho_a \rho_b)} \tag{5.3}$$

RelCalc and SimSeq are used throughout the remainder of this chapter to explain previous methods to reduce the effects of relaxation in NMR experiments on large biomolecules and compare these to the relaxation phenomena observed in $^{13}CF_3$ groups.

5.2.2 Evaluation of the current methods to enhance NMR spectral quality in macromolecules using RelCalc and SimSeq

*SimSeq is an in-house Python-based simulation software developed by Professor Andrew Baldwin to simulate pulse sequences for a given spin system, including the complete Liouvillian relaxation matrix determined through RelCalc.

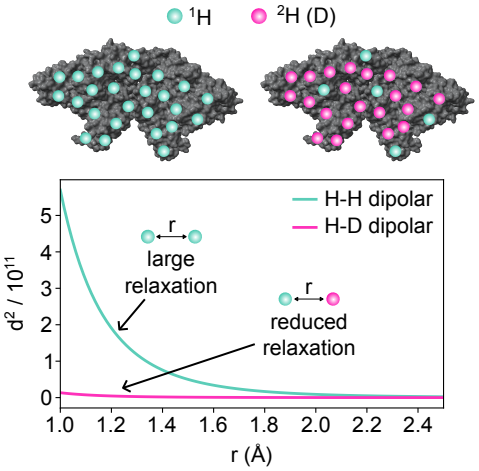


Figure 5.3: A plot showing how background protein deuteration can be used as a technique to reduce the dipolar relaxation experienced by protons, proportional to the dipolar constant squared (d^2). r is the inter-nuclear distance. The protein image shown is BSA, obtained using ChimeraX[346] from PDB[2] entry 4F5S.[431]

5.2.2.i Deuteration

In protonated biomolecular systems, the primary mechanism causing relaxation is the dipolar interaction between nearby protons. This relaxation scales with the square of the dipolar interaction constant (d_{HH}, equation 5.2), which for a proton-proton interaction is dependent on γ_H^4/r_{HH}^6, where γ_H is the proton gyromagnetic ratio (γ_H=267.5×10^6 rad·s^{-1}T^{-1})[430] and r_{HH} is the internuclear distance. Therefore, in highly protonated proteins, substantial proton-proton dipolar interactions cause fast transverse relaxation, thereby increasing peak linewidths. However, by deuterating the whole protein, leaving a small number of protons to be observed, the relaxation of the protons is significantly reduced and the spectral resolution increases. This is caused by the lower gyromagnetic ratio of deuterons (γ_D=41.06×10^6 rad·s^{-1}T^{-1})[430] leading to weaker dipolar interactions with the protons of interest (Figure 5.3).

Background deuteration with selective proton labelling was first shown in 1969 by Crespi *et al.*[432] and is also known as the isotope hybrid method. This method has the benefit of reducing the transverse relaxation rate of the remaining signals but also removing unwanted signals from the NMR spectrum, reducing spectral crowding. Background deuteration is now standard practice in biomolecular NMR; however, it is costly, and the exact degree of deuteration is often challenging to quantify.

5.2.2.ii ^{15}N-^{1}H TROSY

In the coupled HSQC NMR spectrum of a ^{15}N-^{1}H scalar coupled system, such as in the backbone amide group of a protein, the spectrum consists of 4 peaks in the absence of decoupling (Figure 5.4A). During the ^{15}N chemical shift evolution period of a HSQC, nitrogen transverse magnetisation ($(N_x+iN_y)H_z$) can be represented in the shift basis (see equation 2.11) as N_+H_α-N_+H_β. This corresponds to transverse nitrogen magnetisation with the connected proton in the α or β state, respectively. Analogously, the transverse proton magnetisation during the direct detection period (H_x+iH_y) can be represented as $H_+N_\alpha+H_+N_\beta$. The evolution of the components of transverse nitrogen and proton magnetisation under the Zeeman interaction is given by:

$$\frac{d}{dt}\begin{pmatrix} c_{N_+H_\alpha} \\ c_{N_+H_\beta} \end{pmatrix} = \begin{pmatrix} i(\omega_N + \pi J) - R_{N_+H_\alpha} & -R_{N_+H_\alpha,N_+H_\beta} \\ -R_{N_+H_\alpha,N_+H_\beta} & i(\omega_N - \pi J) - R_{N_+H_\beta} \end{pmatrix} \begin{pmatrix} c_{N_+H_\alpha} \\ c_{N_+H_\beta} \end{pmatrix} \quad (5.4)$$

$$\frac{d}{dt}\begin{pmatrix} c_{H_+N_\alpha} \\ c_{H_+N_\beta} \end{pmatrix} = \begin{pmatrix} i(\omega_H + \pi J) - R_{H_+N_\alpha} & -R_{H_+N_\alpha,H_+N_\beta} \\ -R_{H_+N_\alpha,H_+N_\beta} & i(\omega_H - \pi J) - R_{H_+N_\beta} \end{pmatrix} \begin{pmatrix} c_{H_+N_\alpha} \\ c_{H_+N_\beta} \end{pmatrix} \quad (5.5)$$

where c_i is the coefficient of the magnetisation component i, $\omega_{N/H}$ are the Larmor frequencies of the nitrogen and proton nuclei, respectively, and J is the H-N scalar coupling constant. The relaxation rates $R_{N_+H_{\alpha/\beta}}/R_{H_+N_{\alpha/\beta}}$ are known as auto-relaxation rates and represent the transverse relaxation rates of the $N_+H_{\alpha/\beta}$ and $H_+N_{\alpha/\beta}$ coherences, respectively. The rates $R_{H_+N_\alpha,H_+N_\beta}$ and $R_{N_+H_\alpha,N_+H_\beta}$ are known as cross-relaxation rates and represent the interconversion of the H_+N_α/H_+N_β and N_+H_α/N_+H_β states through relaxation, respectively. Note that multiple transverse

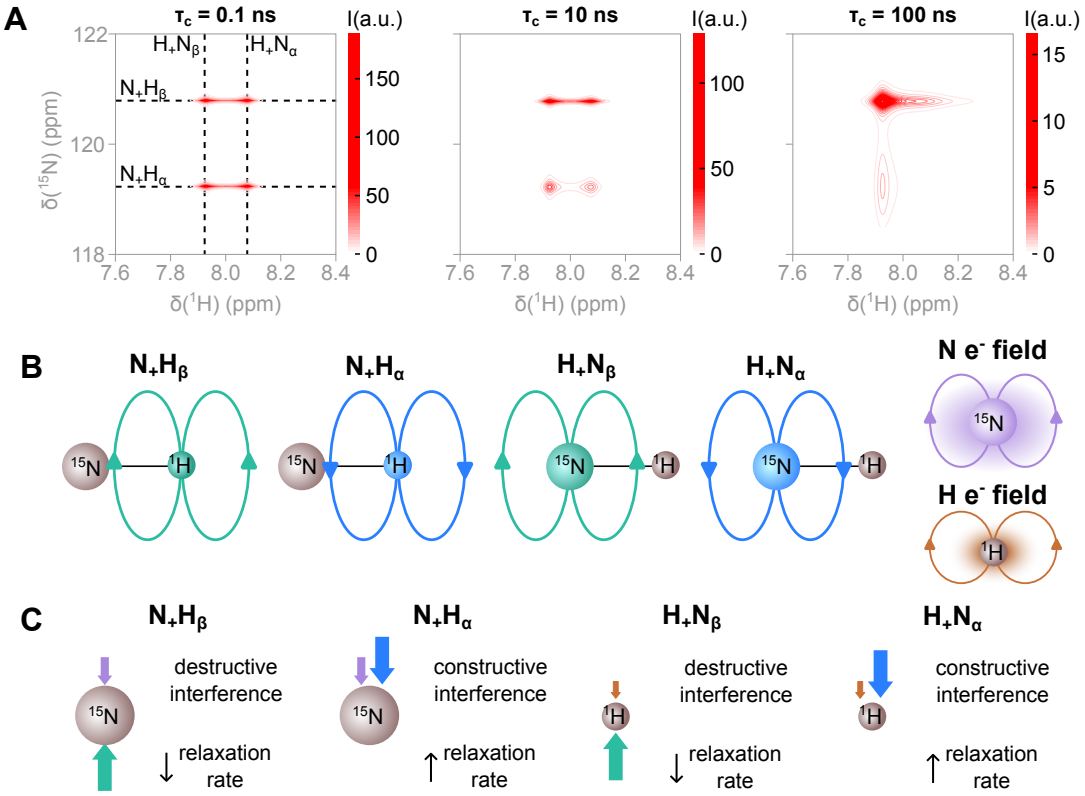


Figure 5.4: An introduction to relaxation interference in ^{15}N-^{1}H spin systems. **A** Coupled HSQC spectra for a ^{15}N-^{1}H spin system (with δ_H and δ_N equal to 8 and 120 ppm, respectively, and the scalar coupling constant J equal to -93.3 Hz) at rotational correlation times (τ_c) of 0.1 ns, 10 ns and 100 ns, simulated using RelCalc and SimSeq. The N-H bond was assumed to be 1 Å, with the N-H dipolar interaction aligned with the N-H bond. Furthermore, the chemical shift anisotropy tensor was assumed to be axially symmetric and colinear with the N-H bond, governed by the values $\Delta\delta_N$=-160 ppm and $\Delta\delta_H$=16 ppm.[53] Further simulation details are included in appendix C.3.1.i. **B** An illustration of the dipolar-CSA relaxation interference in a ^{15}N-^{1}H spin system that gives rise to the differential multiplet linewidths and intensities shown in A. CSA is introduced by orientation dependence of the fields induced by electron (e^-) rotations. **C** Interference between relaxation causing mechanisms in the ^{15}N-^{1}H spin system. The depictions in B are based on those in Levitt, 2008.[95]

relaxation rates are present in this scalar-coupled system. Symbolic expressions describing the relaxation rates in equations 5.4/5.5 were evaluated using RelCalc and are included in appendix C.2.1.

Due to the covalent bond between the ^{15}N and ^{1}H nuclei, the motion of the nuclei is correlated and not independent. This means that cross-correlation, the interference between the induced magnetic fields in the system, is sustained in all

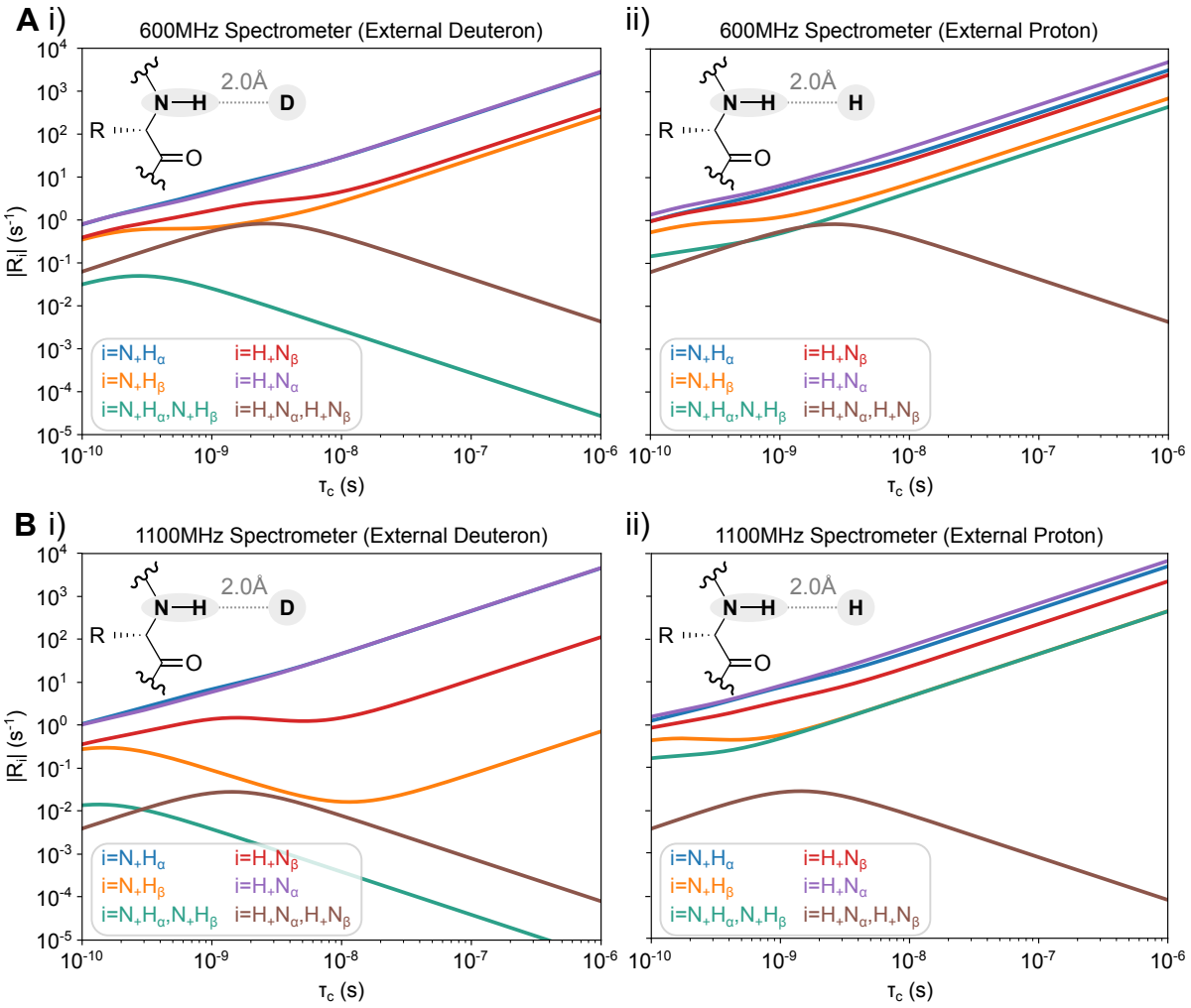


Figure 5.5: Quantification of relaxation interference in ^{15}N-^{1}H spin systems using RelCalc. The auto and cross-relaxation rates in equations 5.4 and 5.5 are evaluated as a function of rotational correlation time for a ^{15}N-^{1}H spin system in **A** a 600 MHz and **B** a 1100 MHz spectrometer, surrounded by an external deuteron (i) or external proton (ii) a distance of 2 Å away. The rates were evaluated using RelCalc, assuming the same spin-system parameters as described in the caption of Figure 5.4.

molecular orientations. In ^{15}N-^{1}H spin system, the dipolar fields in each of the coherences $N_+H_{\alpha/\beta}$ and $H_+N_{\alpha/\beta}$ need to be considered, as well as the chemical shift anisotropy of the ^{15}N and ^{1}H nuclei. For example, the induced dipolar fields of the N_+H_β coherence interact destructively with the induced ^{15}N chemical shift anisotropy in all molecular orientations leading to a lower net induced field (Figure 5.4B). Therefore, the lower magnitude induced fields of the system lower the auto-relaxation rates of these coherences, causing their peaks to be sharper (Figure 5.5). In contrast, the induced dipolar and CSA fields interact constructively for the N_+H_α

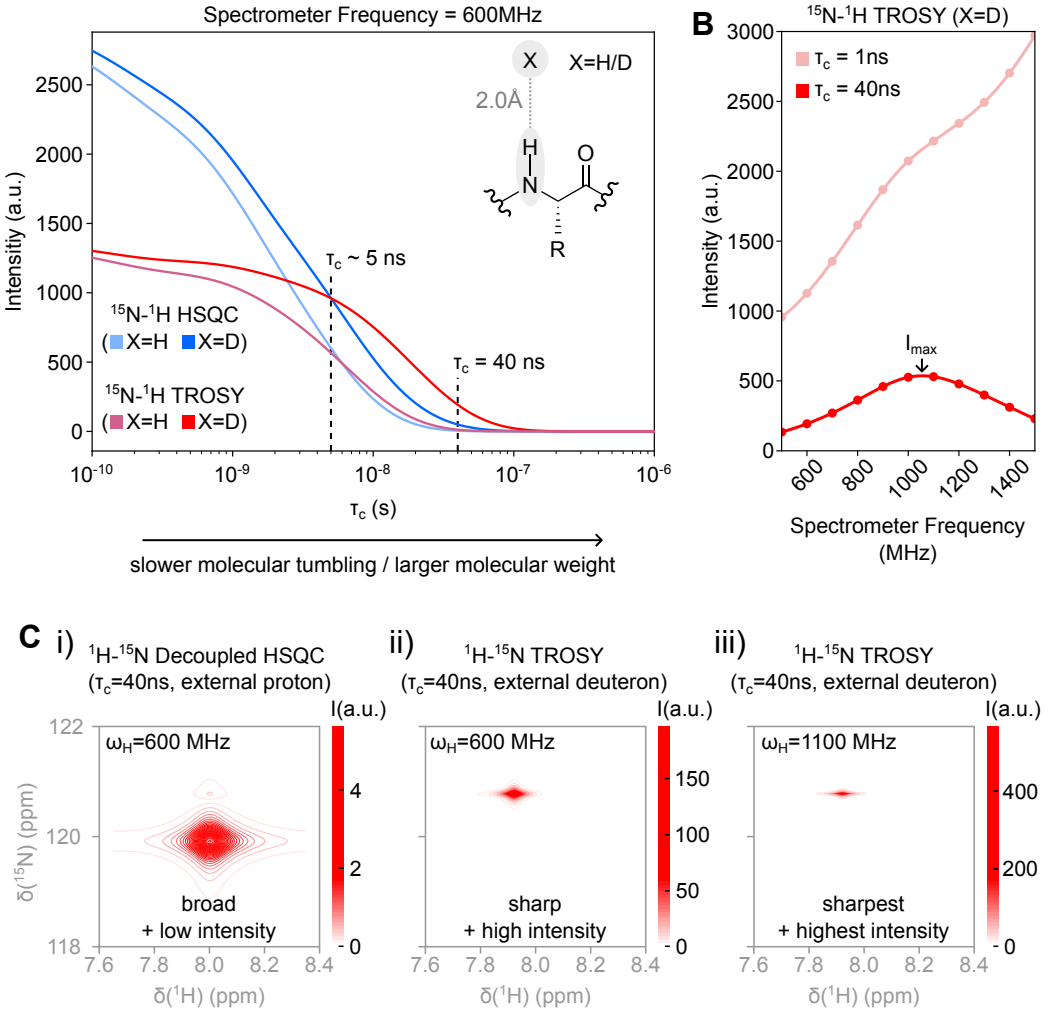


Figure 5.6: Simulations of TROSY and non-TROSY pulse sequences on a ^{15}N-^{1}H spin system. **A** A comparison of the simulated intensities of decoupled sensitivity enhanced HSQC spectra and TROSY spectra obtained for a ^{15}N-^{1}H spin system as a function of rotational correlation time (τ_c). The spin system also contained an external proton or deuteron 2 Å away. **B** Simulated intensities of TROSY spectra for a ^{15}N-^{1}H spin system as a function of spectrometer frequency with τ_c values of 1 ns and 40 ns. **C** A comparison of the simulated 2D NMR spectra for a ^{15}N-^{1}H spin system with a rotational correlation time (τ_c) of 40 ns. (i) A simulated decoupled sensitivity-enhanced HSQC with an external proton at 600 MHz. (ii) A simulated TROSY spectrum with an external deuteron at 600 MHz. (iii) A simulated TROSY spectrum with an external deuteron at 1100 MHz. All simulations were performed using SimSeq, and the relaxation rates of the full Liouvillian matrix were calculated using RelCalc. The same spin-system parameters as described in the caption of Figure 5.4were used. Further details of the simulations can be found in appendix C.3.1.i.

coherence, causing enhanced relaxation rates and a broadening of the resulting peaks at large rotational correlation times (Figure 5.5).

This type of cross-correlated relaxation (relaxation interference) causing differential multiplet peak linewidths has been known since the mid-twentieth century.[85,433–435] However, in 1997, Pervushin *et al.*[53] proposed the utilisation of this phenomenon to provide more intense and sharper coherences in ^{15}N-^{1}H correlation NMR spectra using transverse relaxation optimised spectroscopy (TROSY). In the TROSY sequence, the magnetisation pathway corresponding to N_+H_β during the ^{15}N evolution period, and H_+N_β during the direct acquisition period are selected through phase-cycling, and all other coherences are removed. The TROSY method requires minimal mixing of coherences during the experiment so that the relaxation-optimised coherences are not mixed with coherences containing higher relaxation. Therefore, it is imperative that the cross relaxation rates $R_{H_+N_\alpha,H_+N_\beta}$ and $R_{N_+H_\alpha,N_+H_\beta}$ are negligible. Importantly, this means that background deuteration of non-amide protons is essential for TROSY experiments, as external protons to the ^{15}N-^{1}H spin system substantially enhance the cross-relaxation rates (Figure 5.5). However, in α-helical protein units, ^{15}N-^{1}H nuclei from adjacent amino acids are located nearby, leading to unavoidable external proton effects, even in the presence of background deuteration.[53] At short rotational correlation times, the TROSY experiment produces approximately half of the resultant signal compared to a sensitivity enhanced HSQC[94] (Figure 5.6A). However, in background deuterated systems, simulations reveal the TROSY experiment becomes beneficial intensity-wise over the decoupled sensitivity enhanced HSQC for proteins with τ_c values larger than 5 ns, on a 600 MHz spectrometer (Figure 5.6A). Furthermore, the TROSY sequence benefits from sharper peaks over the decoupled sensitivity-enhanced HSQC as the rotational correlation time is increased.

At enhanced magnetic field strengths, the ^{15}N CSA fields (proportional to magnetic field strength, equation 5.2) can cancel the induced dipolar fields more effectively (Figure 5.5), extending the molecular weight of species which can

be studied using the ^{15}N-^{1}H TROSY method, with the optimum spectrometer frequency for such experiments simulated to be approximately 1100 MHz (Figure 5.6B).

The ^{15}N-^{1}H TROSY method has been instrumental in enabling the study of background deuterated proteins larger than 50 kDa using NMR. It has even been applied to a structured 900 kDa bacteria chaperone complex containing GroEL and GroES proteins.[436] However, even TROSY spectra are limited in their intensities for such substantially sized species due to substantial relaxation during the INEPT delays. Cross relaxation-enhanced polarization transfer (CRINEPT)-TROSY pulse sequences have been used to combat this.[436,437] In addition, delayed decoupling schemes,[58,59] whereby magnetisation is detected during the final evolution period have been used to reduce the effect of this (further details of delayed decoupling schemes are discussed in section 5.2.2.iii).

In chapter 4, the poor dispersion and high signal overlap in the ^{15}N-^{1}H NMR spectra of disordered proteins was discussed. TROSY-based experiments, being able to decrease linewidths and enhance spectral resolution, are also used in the study of large spectrally crowded disordered proteins.[376] In addition, ^{15}N detected TROSY experiments are becoming more common to enhance the resolution in the spectra of disordered proteins.[438]

5.2.2.iii Methyl-TROSY

Methyl groups are highly abundant in proteins, being present in alanine, threonine, leucine, isoleucine, valine and methionine amino acids and report on the side chain interactions and dynamics within a protein.[439] Isotopic ^{13}C labelling strategies of the methyl groups these amino acids have been developed, or $^{13}CH_3$ groups can be added as a post-translational modification to cysteine residues.[439] A $^{13}CH_3$ spin system contains C_{3v} symmetry, with a bond length r_{CH} equal to approximately 1.09 Å and each proton separated by $r_{HH} = \sqrt{3} r_{CH} \sin(\beta)$, where β is 109.5^o (leading to

$r_{HH} = 1.78$Å). In such a system, single quantum carbon magnetisation (C_+) can be separated into the following components in the shift basis:

$$\begin{aligned} C_+ &= C_+H_EH_EH_E \\ &= C_+(H_\alpha + H_\beta)(H_\alpha + H_\beta)(H_\alpha + H_\beta) \\ &= C_+H_{\alpha\alpha\alpha} + 3C_+H_{\alpha\alpha\beta} + 3C_+H_{\alpha\beta\beta} + C_+H_{\beta\beta\beta} \end{aligned} \tag{5.6}$$

The evolution of these components under the Zeeman interaction can be expressed using the following matrix:

$$\frac{d\rho}{dt} = -L\rho \qquad \rho = \begin{pmatrix} c_{C_+H_{\alpha\alpha\alpha}} \\ c_{C_+H_{\alpha\alpha\beta}} \\ c_{C_+H_{\alpha\beta\beta}} \\ c_{C_+H_{\beta\beta\beta}} \end{pmatrix}$$

$$L = \begin{pmatrix} 3i\pi J + R_{C_+H_{\alpha\alpha\alpha}} & R_{C_+H_{\alpha\alpha\alpha},C_+H_{\alpha\alpha\beta}} & R_{C_+H_{\alpha\alpha\alpha},C_+H_{\alpha\beta\beta}} & R_{C_+H_{\alpha\alpha\alpha},C_+H_{\beta\beta\beta}} \\ R_{C_+H_{\alpha\alpha\beta},C_+H_{\alpha\alpha\alpha}} & i\pi J + R_{C_+H_{\alpha\alpha\beta}} & R_{C_+H_{\alpha\alpha\beta},C_+H_{\alpha\beta\beta}} & R_{C_+H_{\alpha\alpha\beta},C_+H_{\beta\beta\beta}} \\ R_{C_+H_{\alpha\beta\beta},C_+H_{\alpha\alpha\alpha}} & R_{C_+H_{\alpha\beta\beta},C_+H_{\alpha\alpha\beta}} & -i\pi J + R_{C_+H_{\alpha\beta\beta}} & R_{C_+H_{\alpha\beta\beta},C_+H_{\beta\beta\beta}} \\ R_{C_+H_{\beta\beta\beta},C_+H_{\alpha\alpha\alpha}} & R_{C_+H_{\beta\beta\beta},C_+H_{\alpha\alpha\beta}} & R_{C_+H_{\beta\beta\beta},C_+H_{\alpha\beta\beta}} & -3i\pi J + R_{C_+H_{\beta\beta\beta}} \end{pmatrix} \tag{5.7}$$

where c_i is the coefficient of magnetisation of the form i, the value of ω_C has been assigned to zero, and J is the C-H scalar coupling constant. The diagonal relaxation rates $R_{C_+H_{\alpha\alpha\alpha}}$, $R_{C_+H_{\alpha\alpha\beta}}$, $R_{C_+H_{\alpha\beta\beta}}$ and $R_{C_+H_{\beta\beta\beta}}$ are auto-relaxation rates, relating to the coherences $C_+H_{\alpha\alpha\alpha}$, $C_+H_{\alpha\alpha\beta}$, $C_+H_{\alpha\beta\beta}$ and $C_+H_{\beta\beta\beta}$, respectively.

Due to the correlated motion of the ^{13}C and ^{1}H nuclei within the $^{13}CH_3$ group, relaxation interference is possible between the dipolar fields created within each of the coherences (Figure 5.7). This leads to differential relaxation of each multiplet component, with $R_{C_+H_{\alpha\beta\beta}}$ exhibiting the smallest relaxation rate (Figure 5.8). This phenomenon in $^{13}CH_3$ groups was first described by Ollerenshaw *et al.* in 2003.[56]

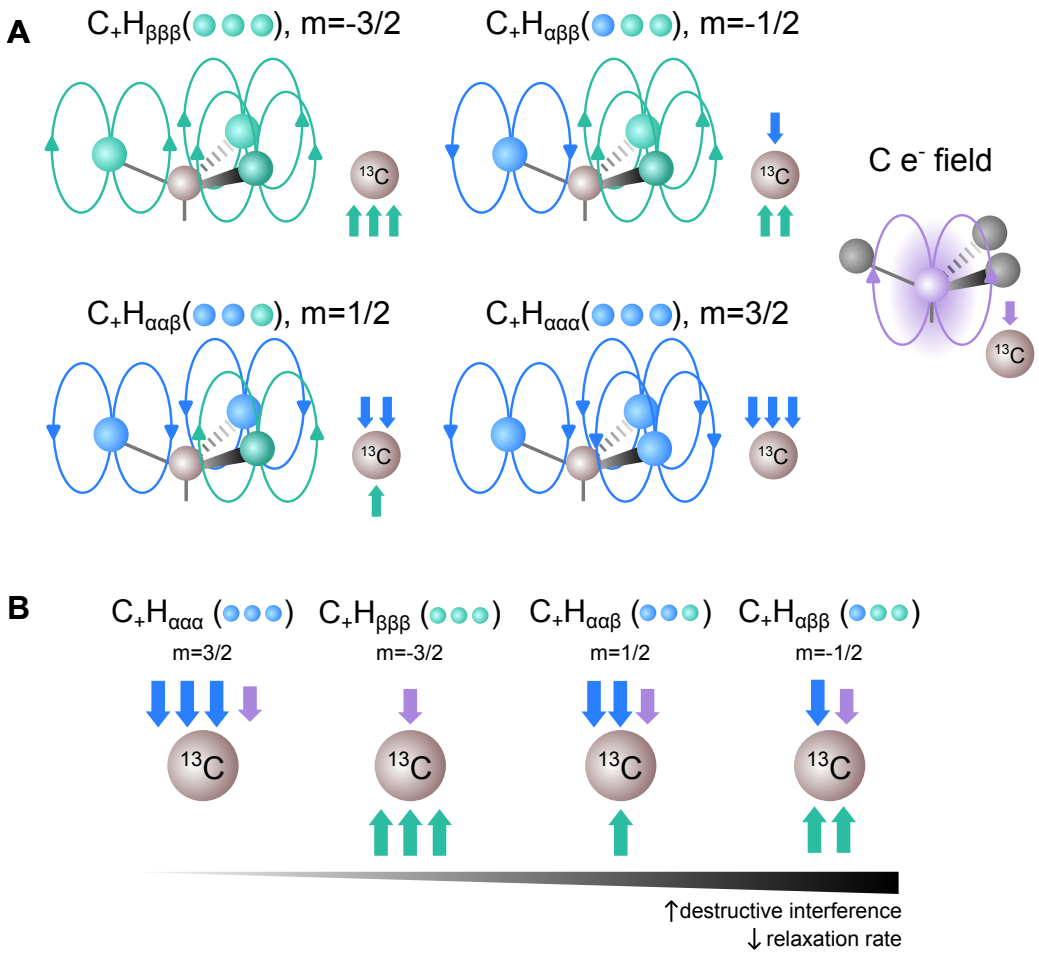


Figure 5.7: The generation of relaxation interference in a $^{13}CH_3$ group. **A** The dipolar fields generated in the single quantum carbon coherences $C_+H_{\alpha\alpha\alpha}$, $C_+H_{\alpha\alpha\beta}$, $C_+H_{\alpha\beta\beta}$, $C_+H_{\beta\beta\beta}$ and the ^{13}C chemical shift field (with anisotropy $\Delta\delta_C$). **B** The constructive and destructive interference between the fields in A.

Similar to the ^{15}N-^{1}H spin system, the mixing of slow and fast relaxing coherences must be minimised during an experiment to isolate the slow relaxation coherences within the system. Mixing of the slow and fast relaxing coherences is mediated by the presence of external protons (Figure 5.8B); therefore, methyl-TROSY techniques require background deuteration to maximise the resultant signal intensity.

Figure 5.9 shows an overlay of the simulated 2D spectral intensities for the HMQC,[440–442] decoupled HSQC[108] and coupled HSQC pulse sequences as a function of rotational correlation time (τ_c). Details of the simulations performed using SimSeq are provided in appendix C.3.1.ii. The results conform with those observed in the literature[56] that the HMQC[440–442] is the best TROSY experiment for $^{13}CH_3$ groups

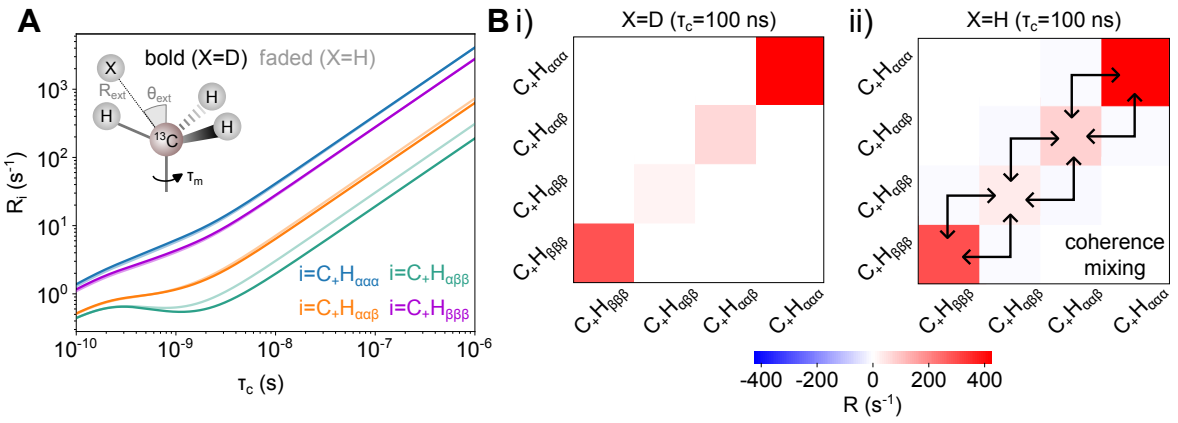


Figure 5.8: Relaxation interference in single quantum carbon auto-relaxation rates for a $^{13}CH_3$ group. **A** The auto-relaxation rates of the single quantum carbon coherences in a $^{13}CH_3$ group as a function of rotational correlation time (τ_c). The effects of an external nucleus (X) a distance R_{ext} of 2 Å away at an angle θ_{ext} of 60^o to the CH_3 rotation axis was included. Bold and faded lines represent where X was set to a deuteron or proton. **B** The values of the auto and cross-relaxation rates in the evolution matrix of equation 5.7 at a τ_c value of 100 ns where X was set to (i) a deuteron and (ii) a proton. All relaxation rates were calculated for a 600 MHz spectrometer using RelCalc assuming a local 3-site hopping motion of the $^{13}CH_3$ group about its rotational symmetry axis with rate constant $1/\tau_m$ (where τ_m was set to 10 ps). The internuclear dipolar interactions were assumed to be aligned with the internuclear vectors. The chemical shift anisotropy tensors were assumed to be axially symmetric and colinear with the C-H bonds, governed by the constants $\Delta\delta_{C/H}$, with values of 20 ppm and 1 ppm for ^{13}C and ^{1}H nuclei, respectively (consistent with fitted values in Figure 5.21). Rate expressions are included in appendix C.2.3.

in background deuterated proteins, enabling species with τ_c larger than 10^{-7} s to be observed (Figure 5.9). Furthermore, due to the minimal chemical shift anisotropy contribution to relaxation within the system (in contrast to the ^{15}N-^{1}H spin system), the observed signal intensity will always be enhanced in spectrometers with increased magnetic fields. HMQC spectra have been used to study the structure and dynamics of a background deuterated $^{13}CH_3$-labelled 670 kDa proteosome protein.[443]

The plots in Figure 5.9 emphasise the high importance of background deuteration of proteins used for methyl-TROSY studies on large proteins. However, recently developed neural networks such as FID-NET have been used to enable similar methyl-TROSY spectral quality in non-deuterated proteins as in background deuterated systems up to 360 kDa in molecular weight.[444] This leads to a reduction

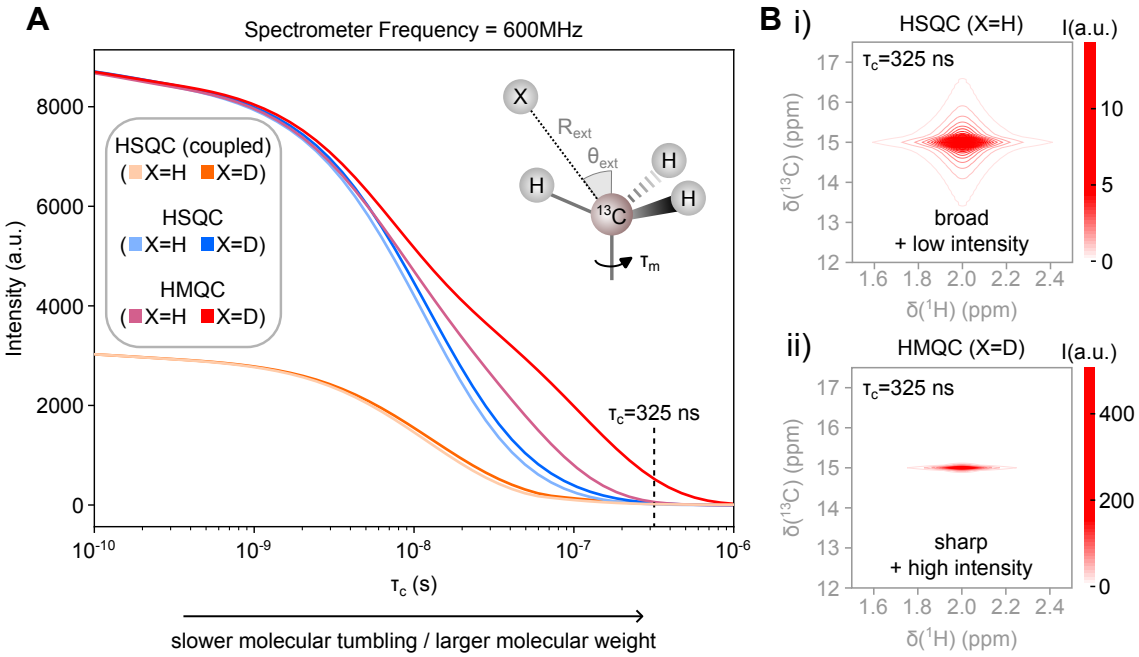


Figure 5.9: A comparison of simulated spectra for a $^{13}CH_3$ group as a function of rotational correlation time. **A** Simulated HSQC[108] (coupled), HSQC[108] and HMQC[440–442] 2D spectra intensities as a function of τ_c. The effects of an external nucleus (X=H/D) a distance R_{ext} of 2 Å away at an angle θ_{ext} of 60^o to the CH_3 rotation axis was included. **B** The simulated (i) HSQC spectrum in a non-deuterated background and (ii) HMQC spectrum with background deuteration at a τ_c values of 325 ns. The complete set of Liovillian relaxation rates was evaluated using RelCalc, and simulations were performed using SimSeq (details on the simulations are included in appendix C.3.1.ii). The spin system was set up with the same parameters as described in Figure 5.8.

in the cost and an increase in the yield of the labelled proteins produced.

To understand why the HMQC experiment is superior to the HSQC, it is best to enter the irreducible basis for the system. The total spin angular momentum of three spin-1/2 protons can be calculated using the tensor product $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2}$. This can be decomposed into its irreducible states with total spin angular momentum (I) terms determined using Clebsch Gordon coefficients leading to the direct sum $\frac{3}{2} \oplus \frac{1}{2} \oplus \frac{1}{2}$. The I=3/2 states have A_1 symmetry in a C_{3v} group, and the I=1/2 states represent doubly degenerate E symmetry states. For the I=3/2 state, the projected angular momentum (m_I) can have values of $+\frac{3}{2}$, $+\frac{1}{2}$, $-\frac{1}{2}$ and $-\frac{3}{2}$. These represent the wavefunctions ($|1\rangle$-$|4\rangle$). The two I=1/2 states can have projected angular momentum values of $\pm\frac{1}{2}$ and represent the proton spin wavefunctions ($|5\rangle$-$|8\rangle$).[56]

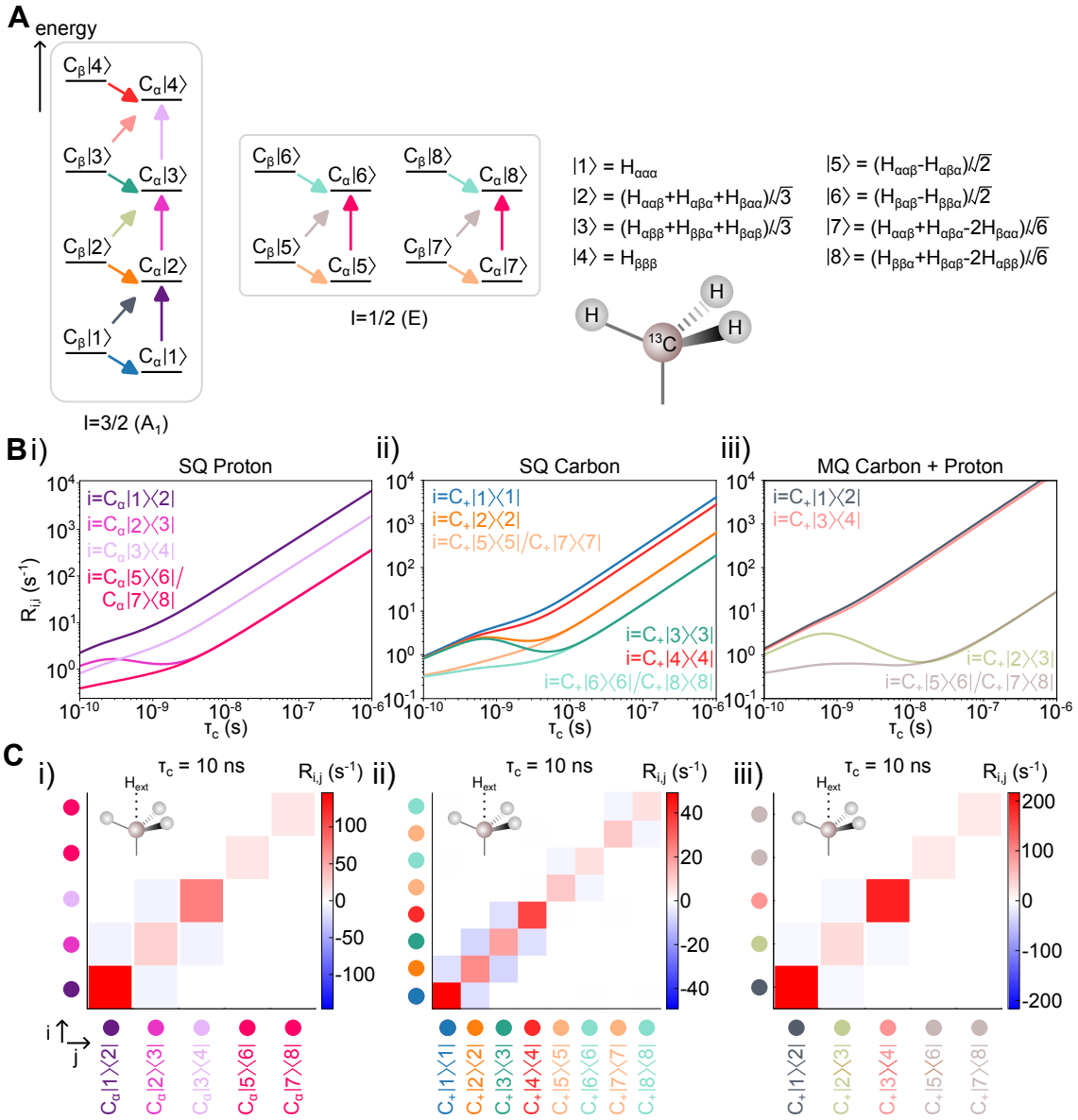


Figure 5.10: Irreducible basis single and multiple quantum relaxation rates for a $^{13}CH_3$ group. **A** The complete spin-state energy level diagram for a $^{13}CH_3$ group. The eigenstates of the protons are represented by the states $|1\rangle$-$|8\rangle$. Single quantum (SQ) carbon and proton transitions as well as multiple quantum (MQ) transitions are shown. **B** Auto-relaxation rates of (i) the SQ proton transitions, (ii) the SQ carbon transitions and (iii) MQ transitions shown in A as a function of τ_c on a 600 MHz spectrometer. **C** Relaxation rate matrices of (i) the SQ proton transitions, (ii) the SQ carbon transitions and (iii) MQ transitions shown in A for τ_c=10 ns on a 600 MHz spectrometer. Relaxation rates were evaluated using RelCalc with the spin system set up with the parameters described in Figure 5.8. The rates in B were calculated assuming no external nuclei and the relaxation matrix in C was calculated assuming an external proton was placed 2Å away. Rate expressions for the auto-relaxation rates in B are included in appendix C.2.3.

The form of the proton spin wavefunctions $|1\rangle$-$|8\rangle$ are shown in (Figure 5.10A).

Single quantum proton transitions where carbon is in the C_α state are shown by the vertical arrows representing transitions between the proton eigenstates (Figure 5.10A). These coherences are present during the τ delays of the HMQC and HSQC sequences,[56,439] and their auto-relaxation rates exhibit differences with some having lower rates than others (Figure 5.10Bi). Single quantum (SQ) carbon (C_+) transitions present during the indirect dimension of a HSQC are represented by those from the C_β to the C_α states (with a constant proton state $|i\rangle$), with their relaxation rates depicted in Figure 5.10Bii/Cii for a 600 MHz spectrometer.

In contrast to the HSQC, during the indirect dimension of a HMQC, multiple quantum (MQ) coherences are present, with their relaxation rates shown in Figure 5.10Biii/Ciii. Notably, the $C_+ |2\rangle \langle 3|$ and $C_+ |5\rangle \langle 6|/C_+ |7\rangle \langle 8|$ MQ transitions have remarkably small auto-relaxation rates at large τ_c values in the absence of external protons.

Proton 90^o pulses mix the slow and fast relaxing coherences, leading to signal losses. In the HMQC sequence, the absence of 90^o proton pulses after the first pulse prevents the mixing of slow and fast relaxing coherences, leading to enhanced intensity at large τ_c values (Figure 5.9A). During the sequence, slow relaxing SQ proton magnetisation is converted to slow relaxing MQ coherences before returning to slow relaxing SQ proton magnetisation before acquisition.[56,439] In contrast, the HSQC contains many 90^o proton pulses leading to the mixing of the slow and fast relaxing transitions, reducing the overall spectrum intensity at large τ_c values.[56,439] Furthermore, during the HMQC, all coherences in the indirect dimension appear at the same chemical shift, meaning the filtering out of unwanted coherences like in the ^{15}N-^{1}H TROSY sequence is not required.[56,439]

However, in proteins with substantial rotational correlation times ($\tau_c \gg 100$ ns), intensity in the HMQC experiment is overcome by relaxation during the coherence transfer delays (Figure 5.11Ai). Recently, delayed decoupling HMQC sequences

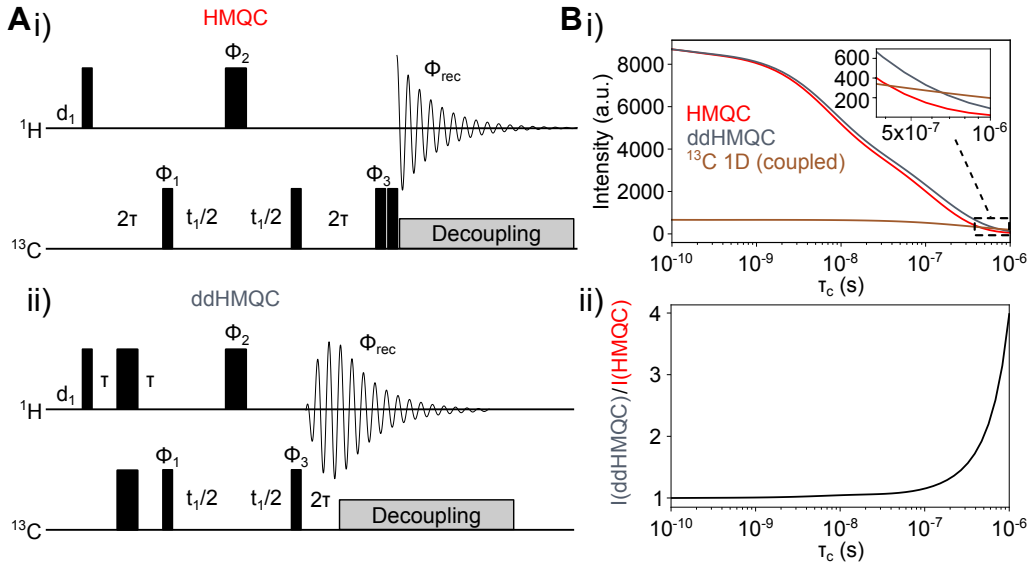


Figure 5.11: Delayed decoupling sequences for the study of $^{13}CH_3$ group in slow tumbling systems. **A** The pulse sequences for the ^{13}C-^{1}H (i) HMQC[56] and (ii) ddHMQC.[59] The values of τ are set to $1/(4J)$, and t_1 is the indirect dimension evolution time. The phase cycles for each experiment are included in appendix C.3.1.ii. **B** (i) Simulated signal intensities of HMQC, ddHMQC and coupled ^{13}C 1D spectra applied to $^{13}CH_3$ group as a function of τ_c. Simulations were performed in SimSeq at a spectrometer frequency of 600 MHz, where the total duration of each experiment was the same (see appendix C.3.1.ii for simulation details). Relaxation rates were evaluated using RelCalc with the spin system set up with the parameters described in Figure 5.8 with an external deuteron.

(ddHMQC) have been proven to give significant enhancements in the studies of substantially sized proteins.[59,60] During a delayed decoupling HMQC sequence (Figure 5.11Aii), magnetisation is recorded during the final 2τ delay period (in the absence of decoupling) in which it evolves from C_zH_x to H_y through scalar coupling (τ=$1/(4J)$). Decoupling is then applied for the remainder of the acquisition period.[59] The distorted FID can be modified by division of $\sin(\pi Jt)$ during $0 < t < 1/(2J)$. By early detection of the FID, the signal is enhanced at large τ_c values due to a reduction in the time available for relaxation before the acquisition begins (Figure 5.11B). Delayed-decoupling HMQC pulse sequences have been used to study $^{13}CH_3$-labelled 1 MDa virus complexes[59] and slowly tumbling folded proteins/RNAs in highly viscous biomolecular condensates.[60] In these examples, using a delayed decoupling sequence is the difference between observing sufficient

signal from the desired protein of interest or not.

Interestingly, despite its low gyromagnetic ratio, the simulated coupled carbon 1D spectrum* of a $^{13}CH_3$ label in a background deuterated protein is predicted to give significant signal intensity and resolution at large τ_c values ($\approx 10^{-6}$ s) when compared to the two-dimensional techniques (Figure 5.11B). However, for most practical applications at these τ_c ranges, the 2D ddHMQC spectrum would still be preferable.

5.2.2.iv TROSY in ^{13}C-F aromatic spin systems

Similar to the ^{15}N-H spin system, the ^{13}C-F spin system is capable of relaxation interference between the C-F dipolar and fluorine CSA mechanisms (Figure 5.12).[57] The benefit of this system over the ^{15}N-H spin system is that background protein deuteration is not essential to preserve the relaxation interference within the system (Figure 5.12). The cross-relaxation rate expression between the slow relaxing (C_+F_α) and fast relaxing (C_+F_β) coherences is dependent on the term p^2J(ω_H-ω_F), where p is the dipolar constant between the carbon and fluorine nuclei and $J(\omega)$ is the spectral density function (see appendix C.2.4 for complete rate expressions). This contrasts with the ^{15}N-H spin system, where the analogous cross-relaxation rate is dependent on the larger term $f^2\tau_c$, with f being the proton-proton dipolar interaction constant (see appendix C.2.1 for complete rate expressions). Furthermore, the proton-fluorine dipolar constant (p), dependent on $\gamma_H\gamma_F$ is reduced compared to the proton-proton dipolar constant (f), which is dependent on γ_H^2.† The ratio $[p^2J(\omega_H - \omega_F)] \,/\, [f^2\tau_c]$ is approximately 0.15 on a 600 MHz spectrometer if τ_c is equal to 10^{-8} seconds and the external proton is placed the same distance away from the C-F or N-H/C-H spin systems.

The problem associated with the ^{13}C-F spin system arises at enhanced magnetic

*The number of scans in the carbon 1D spectrum was chosen to give the same total experiment time as the 2D experiments.

†γ represents the gyromagnetic ratio for each nucleus.

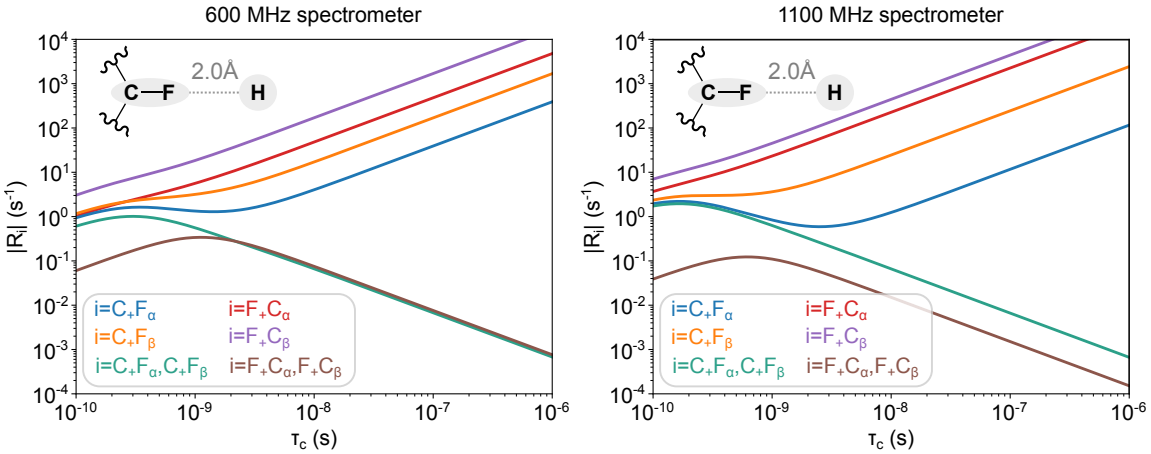


Figure 5.12: Quantification of the single quantum auto and cross-relaxation rates as a function of rotational correlation time for a ^{13}C-^{19}F spin system in a **A** 600 MHz and **B** 1100 MHz spectrometer, surrounded by an external proton a distance of 2 Å away. The rates were evaluated using RelCalc, assuming a C-F bond length of 1.38 Å and axially symmetric chemical shift anisotropies $\Delta\delta_C$=40 ppm and $\Delta\delta_F$=100 ppm.

field strengths, with the fluorine relaxation rates dependent on the substantial ^{19}F CSA terms, which increase with magnetic field strength. However, two-field spectrometer systems utilising shuttle systems have been proposed to enable the ^{19}F magnetisation to be encoded at lower magnetic fields with a reduced relaxation rate, and ^{13}C magnetisation detected at high magnetic field.[445] Another recently developed method to combat the ^{19}F relaxation issue with the ^{13}C-F spin system is to avoid transferring magnetisation to the ^{19}F nucleus entirely and instead transfer the magnetisation from nearby protons to the carbon atom in the ^{13}C-F spin system through multiple bond scalar couplings.[57]

5.3 Proposed relaxation interference in $^{13}CF_3$ groups

The discovery of additional spin-labelling schemes and experiments is required to increase the size of proteins amenable to NMR beyond that made possible through ^{15}N^{1}H-TROSY, ^{13}CF-TROSY or methyl-TROSY techniques.

Fluorine NMR is increasingly popular in biological studies.[446–450]. Only a few natural ^{19}F-containing compounds have been found to exist.[451] Therefore, employing

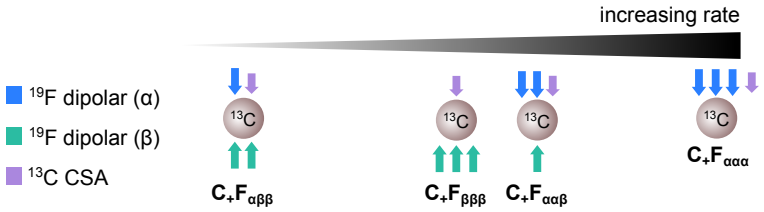


Figure 5.13: An illustration of expected interference between dipolar and chemical shift anisotropy interactions in $^{13}CF_3$ groups causing an expected differential relaxation rate of each single quantum carbon (C_+) coherence.

^{19}F labels in biological NMR creates spectra devoid of unwanted background peaks due to the absence of ^{19}F nuclei within natural protein structures.[446] This has allowed fluorine NMR to excel in studies of cellular systems.[452] The broad chemical shift range ($\approx$ 200 ppm) inherent to ^{19}F nuclei renders them highly responsive to the local chemical environment.[453] Furthermore, ^{19}F nuclei are 100% abundant spin 1/2 with a high gyromagnetic ratio (251.8×10^6 rad s^{-1} T^{-1}),[454] similar to that of protons (γ_H=267.5×10^6 rad·$s^{-1}T^{-1}$),[430] enabling fluorine NMR to be 83% as sensitive as proton NMR.[448,455] Furthermore, the progress of synthetic techniques has increased the scope of ^{19}F labelling in biology using unnatural amino acid incorporation and post-translational modification.[456–458] Fluorine and hydrogen atoms have been shown to have similar Van der Waals radii (F=1.35 Å, H=1.2 Å),[459] enabling ^{19}F labelled systems to mimic the natural protonated system.

Destructive interference between relaxation mechanisms is required to produce slow relaxing coherences.[53,56] As discussed previously (section 5.2.2.iv), the aromatic ^{13}C-F spin system contains significant relaxation interference, enabling TROSY schemes to be developed.[57] However, the $^{13}CF_3$ group has been proposed as another potential label for this purpose. The larger C-F bond length is predicted to give a smaller dipolar interaction, reducing the relaxation of all C_+ coherences compared to their counterparts in the $^{13}CH_3$ system (Figure 5.13). Furthermore, the addition of a ^{19}F nucleus with a large electron density produces a substantial ^{19}F CSA,[460] capable of opposing the strong interactions within the system. Earlier attempts to measure relaxation interference within $^{13}CF_3$ groups were inconclusive;[461] however,

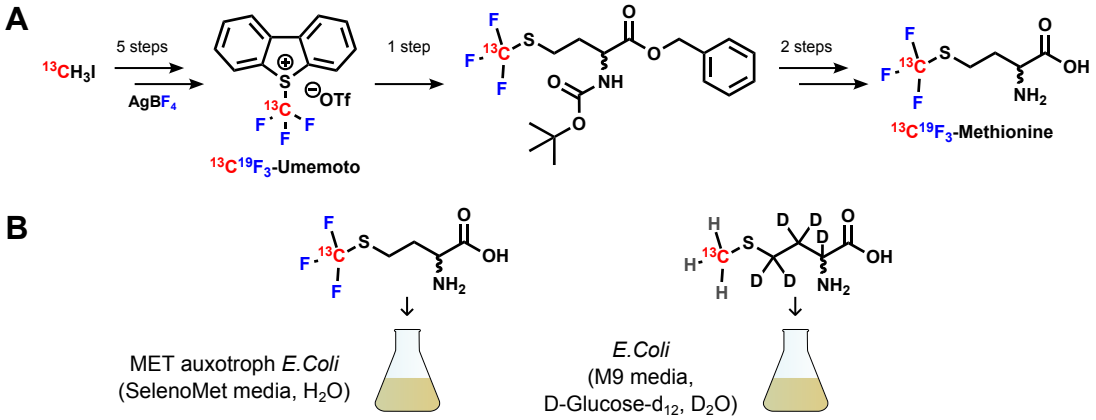


Figure 5.14: **A** An overview of a synthetic labelling scheme to produce $_{DL}$-$^{13}CF_3$-Methionine. **B** An illustration of the expression strategies used to produce $^{13}CF_3$-labelled HSP16.5 and $^{13}CH_3$-labelled background deuterated HSP16.5.

this analysis was conducted on a relatively small 11 kDa CF_3 labelled protein (small τ_c), where significant relaxation interference was not necessarily expected.

5.4 $^{13}CF_3$ labelling of a 395 kDa protein

Producing a $^{13}CF_3$-labelled molecule that could be integrated into proteins was crucial to test the hypothesis of relaxation interference in the spin system. Previous labelling strategies have introduced $^{13}CF_3$ into isoleucine, through the synthesis of perdeutero-5-^{13}C,5,5,5-trifluoroisoleucine VI.[462] However, methionine was chosen as a target due to its relative scarcity in proteins (2.33%)[†], aiming to minimise the disruption to protein structure and function. $_{DL}$-$^{13}CF_3$-Methionine was produced using the synthesis outlined in Figure 5.14A, where commercially available $^{13}CH_3I$ and $AgBF_4$ were used as the ^{13}C and ^{19}F sources, respectively.[‡] During the synthesis, the synthetic intermediate of $^{13}CF_3$ Umemoto's reagent[464] was produced, which could be utilised for side chain post-translational modification. However, incorporation of the unnatural amino acid (UAA) $_{DL}$-$^{13}CF_3$-Methionine was chosen as the labelling scheme.

[†]Obtained from UniProt[463] statistics on 29/05/2025.

[‡]Synthesis of $_{DL}$-$^{13}CF_3$-Methionine was performed by former Baldwin lab member Dr. Satoshi Kishigami.

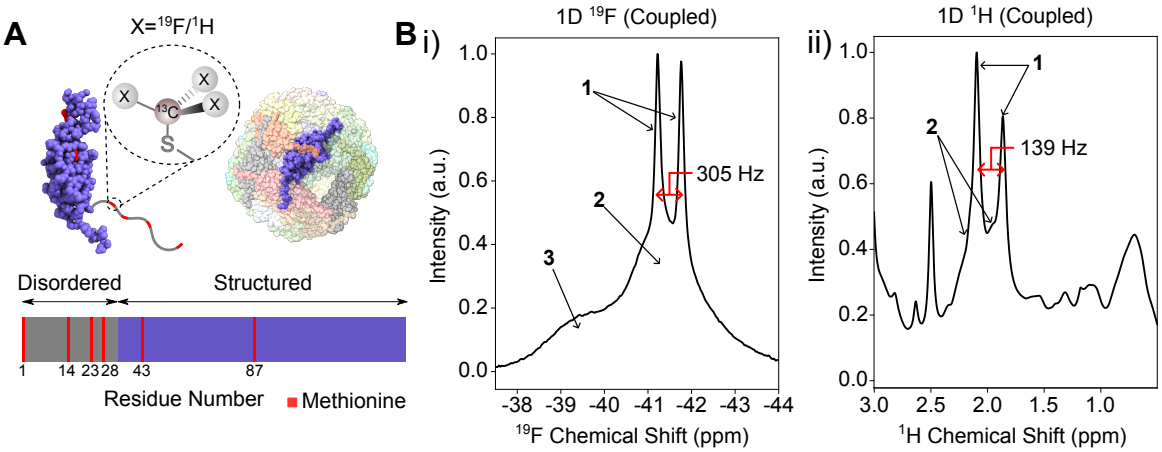


Figure 5.15: **A** The structure of the 24-mer HSP16.5 with the sequence locations of methionine residues highlighted. The protein image was generated in ChimeraX,[346] using PDB[2] entry 1SHS.[465] **B** Fluorine (i) and proton (ii) 1D NMR spectra obtained for $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5, respectively. The labels 1-3 symbolise the observed chemical environments in each experiment. NMR experimental details are included in appendix C.3.2.

The 395 kDa heat shock protein HSP16.5 was chosen to compare the relaxation properties between $^{13}CF_3$ and $^{13}CH_3$ coherences in proteins. The protein is monodispersed and contains 24 identical subunits 16.5 kDa in size, with each subunit containing six methionine residues (Figure 5.15A).[465] Four of the methionine residues are in the disordered N-terminus of the protein chain (Met1, Met14, Met23, Met28), with Met43 in a highly structured location and Met87 located within a loop domain in the subunit interface that stabilises the high-order macromolecular structure.[465] The protein originates in the hyperthermophilic Archaea *Methanococcus jannaschii*, and its high stability and molecular weight made it an ideal target to test the benefits of the $^{13}CF_3$ probe.

Both $_{DL}$-$^{13}CF_3$-methionine labelled and $_{DL}$-$^{13}CH_3$-methionine labelled variations of HSP16.5 were expressed and purified (Figure 5.14B).* The $^{13}CH_3$-labelled variant was additionally background deuterated. Both samples were exchanged into a buffer containing 50 mM NaPi, 100 mM NaCl, 1 mM EDTA, 2 mM NaN_3 at pH 7.0, with the $^{13}CF_3$ and $^{13}CH_3$ labelled-systems in 10% and 100% D_2O, respectively.

*Expression and purification were performed primarily by formed Baldwin lab member Dr Satoshi Kishigami, with help from myself and Dr Virgínia Casablancas-Antràs.

Additionally, 100μM 1,1,1-trifluoroacetic acid (TFA) was added to the $^{13}CF_3$ labelled sample to perform calibrations. Approximately 200 μL of each sample was placed into a Bruker Shigemi tube matched to D_2O. The final concentrations of HSP16.5 monomer in the $^{13}CF_3$ and $^{13}CH_3$ samples was 4 mM (170 μM oligomer) and 1 mM (41 μM oligomer), respectively. However, the labelling efficacy in each sample was not measured.

The ^{19}F 1D NMR spectrum of $^{13}CF_3$-labelled HSP16.5, recorded at 5^{o}C, showed evidence of three fluorine chemical environments (Figure 5.15Bi). The spectrum contained two broad resonances centred at -41.90 and -39.95 ppm, whose coupling was not resolved. These are predicted to originate from $^{13}CF_3$ groups in the structured methionine residues (Met43, Met87). The sharp doublet (J=-305 Hz), centred at -42.2 ppm, is expected to originate from the methionine residues in the disordered N-terminus.* In comparison, the proton 1D spectrum of deuterated $^{13}CH_3$-labelled HSP16.5 recorded at 5^{o}C contained one sharp doublet at 2.0 ppm (J=139 Hz) which was shouldered by a second environment containing slightly broader peaks (Figure 5.15Bii).

Fluorine and proton NMR diffusion measurements were performed to confirm the $^{13}CF_3$/$^{13}CH_3$ labelling of HSP16.5 had not disrupted the integrity of the 24-mer complex. Stimulated-echo diffusion experiments utilising bipolar gradients were performed at 20^{o}C and fitted to the Stejskal-Tanner[19,20] equation (equation 2.37) to obtain diffusion coefficients of $(2.61\pm0.2)\times10^{-7}$ cm^{2}s^{-1} and $(1.85\pm0.12)\times10^{-7}$ cm^{2}s^{-1} for $^{13}CF_3$ and $^{13}CH_3$-labelled HSP16.5, respectively (Figure 5.16). Using the Stokes-Einstein-Sutherland equation (equation 3.8),[304–306] the hydrodynamic radii were calculated to be 8.22 nm and 9.29 nm. The viscosities of H_2O ($^{13}CF_3$-HSP16.5) and D_2O ($^{13}CH_3$-HSP16.5) were taken to be 1.005 and 1.250 mPa s for the hydrodynamic radii calculations. The differences in observed hydrodynamic radii between $^{13}CF_3$ and $^{13}CH_3$-labelled HSP16.5 are likely caused

*Note that the sign of the one bond ^{13}C-^{19}F coupling constant is negative.[57,466,467]

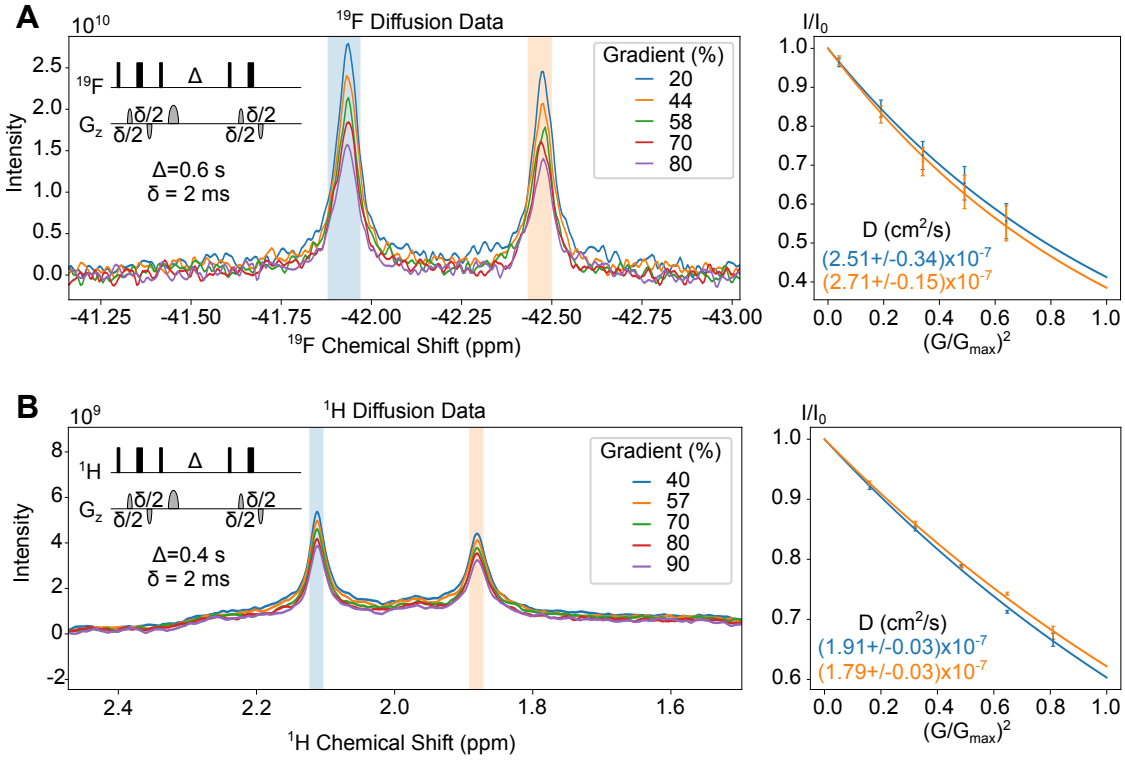


Figure 5.16: Stimulated-echo diffusion NMR spectra obtained for **A** $^{13}CF_3$ labelled and **B** background deuterated $^{13}CH_3$ labelled HSP16.5 using ^{19}F and ^{1}H NMR, respectively. The peak intensities were fitted as a function of gradient strength in each case using the Stejskal-Tanner[19,20] equation (equation 2.37) to obtain the diffusion coefficient. Experimental NMR details are included in appendix C.3.2.

by structural differences between each construct, either through the additional ^{19}F nuclei ($^{13}CF_3$-labelled HSP16.5) or the background deuteration ($^{13}CH_3$-labelled HSP16.5). Despite arguments suggesting fluorine atoms are a suitable mimic for hydrogen atoms in biology,[459] there are numerous cases where fluorine labelling can affect protein intermolecular and intramolecular interactions.[44,468–470] Heavily fluorinated amino acids are suggested to lead to a stabilisation of folded protein states, and this stabilisation could be contributing to the observed compaction compared to the non-fluorinated analogue.[468] Native mass spectrometry or X-ray crystallography techniques could be performed in the future to obtain structural information on $^{13}CF_3$-labelled HSP16.5.

5.5 NMR relaxation rates for the $^{13}CF_3$ and $^{13}CH_3$ systems

Relaxation studies on the coherences in HSP16.5 were performed at 5^{o}C to reduce molecular tumbling as much as possible and to emulate an even larger protein system. Assuming linearity between molecular weight and rotational correlation time,[426] the tumbling of the 395 kDa HSP16.5 system at 5^{o}C is predicted to be similar to that of a protein with a molecular weight of 700 kDa at 25^{o}C. Single quantum ^{13}C and ^{19}F/^{1}H relaxation rates were measured on the $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 systems using the schemes shown in Figure 5.17 on a 600 MHz spectrometer. The relevant coherences were produced, followed by a CPMG relaxation period of T_{relax}. Due to the significant transverse relaxation durations involved in the relaxation experiments, only the sharp peaks could be observed in the resulting spectra (Figures 5.18 and 5.19).

At small T_{relax} durations, the ^{13}C multiplet components in the $^{13}CF_3$ labelled system appeared to roughly conform to a 3:1:1:3 ratio (Figure 5.18Ai/ii). An exact 3:1:1:3 ratio is expected for single quantum carbon magnetisation in the limit of minimal relaxation interference (see appendix C.6). However, differential coherence relaxation becomes apparent at large T_{relax} durations (Figure 5.18Aiii). In contrast, even at short relaxation times, there is significant evidence of differential multiplet peak relaxation in the $^{13}CH_3$ labelled system, which substantially deviates from the 3:1:1:3 ratio expected in the absence of relaxation interference (Figure 5.18B).

For each coherence (single quantum carbon or fluorine/proton), the intensities of each coherence were fitted to an exponential decay as a function of relaxation duration (T_{relax}), enabling the relaxation rate to be determined (Figure 5.20A). Notably, the coherence $C_+F_{\alpha\beta\beta}$ had a slow relaxation rate of 6.7 s^{-1} for $^{13}CF_3$ labelled HSP16.5 at 5^{o}C (Figure 5.20B). In addition, all single quantum carbon (C_+) relaxation rates in $^{13}CF_3$ labelled HSP16.5 were reduced compared to their counterparts in background deuterated $^{13}CH_3$ labelled HSP16.5. This can be

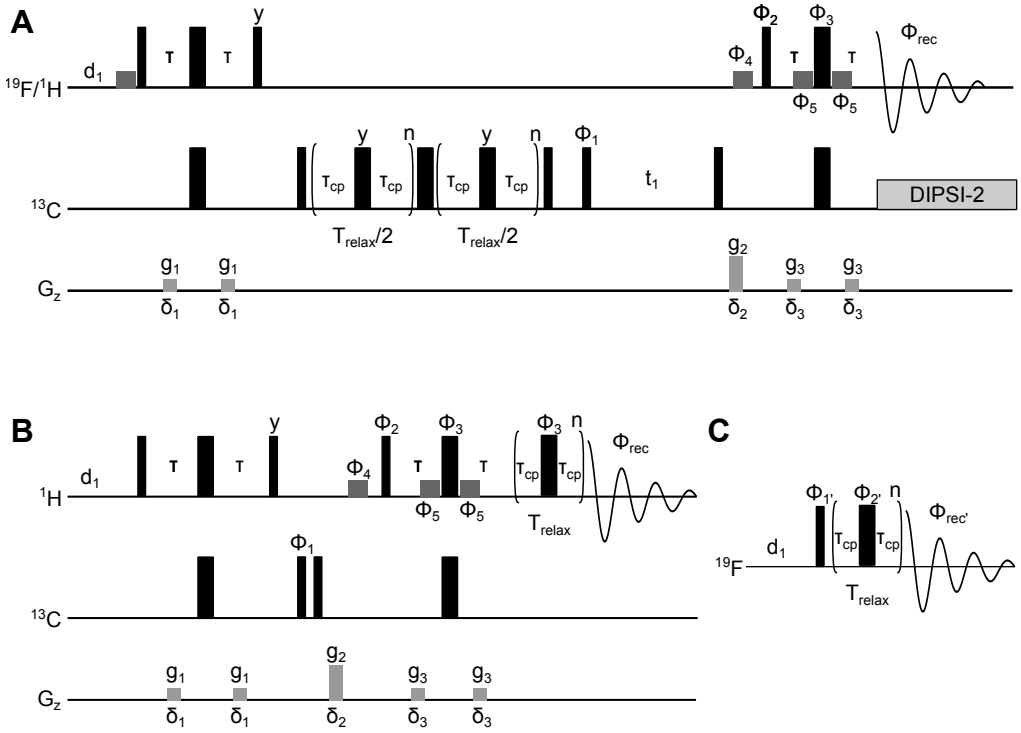


Figure 5.17: Sequences used to measured single quantum relaxation rates in $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5. **A** Pulse sequence used to measure single quantum carbon relaxation rates in $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5. For the $^{13}CH_3$ system, water-selective proton pulses (dark grey) were applied for water suppression. The total relaxation duration (T_{relax}) was divided into two CPMG blocks with n set to either 1 or 2 and τ_{cp} varied. The coherence transfer durations (τ) were set to $1/|4J|$ for each system, equal to 1.8 ms and 0.82 ms for $^{13}CH_3$ and $^{13}CF_3$, respectively. DIPSI-2 carbon decoupling was provided during acquisition. **B** Pulse sequence used to measure single quantum proton relaxation rates in $^{13}CH_3$ labelled HSP16.5. Water-selective proton pulses (dark grey) were applied to suppress water signals. The coherence transfer durations (τ) were set to $1/|4J|$ for each system, equal to 1.8 ms. The total relaxation duration (T_{relax}) was varied by increasing n and keeping τ_{cp} fixed at 200 μs. **C** The 1D CPMG sequence used to measure single quantum fluorine relaxation rates. The total relaxation duration (T_{relax}) was varied by increasing n and keeping τ_{cp} fixed at 50 μs. In all sequences, thin and wide black rectangles symbolise non-selective 90^o and 180^o pulses, respectively. The phase cycles used were as follows: Φ_1=(4(x),4(-x)), Φ_2=(x,-x), Φ_3=(2(x),2(-x)), Φ_4=(2(-x),2(x)), Φ_5=(2(-x),2(x)), Φ_{rec}=(2(x,-x),2(-x,x)), $\Phi_{1'}$=(2(x),2(-x),2(y),2(-y)), $\Phi_{2'}$=(2(y,-y),2(x,-x)), $\Phi_{rec'}$=(2(x),2(y),2(-x),2(-y)). Quadrature detection in sequence A was applied in a States manner, with 90^o incrementation of Φ_1. Further experimental details are provided in appendix C.3.2.

rationalised due to the longer C-F bond length of 1.38 Å (compared to 1.09 Å for the C-H bond length) leading to reduced C-F and F-F dipolar relaxation. The detection of significant relaxation interference within the $^{13}CF_3$-labelled protein

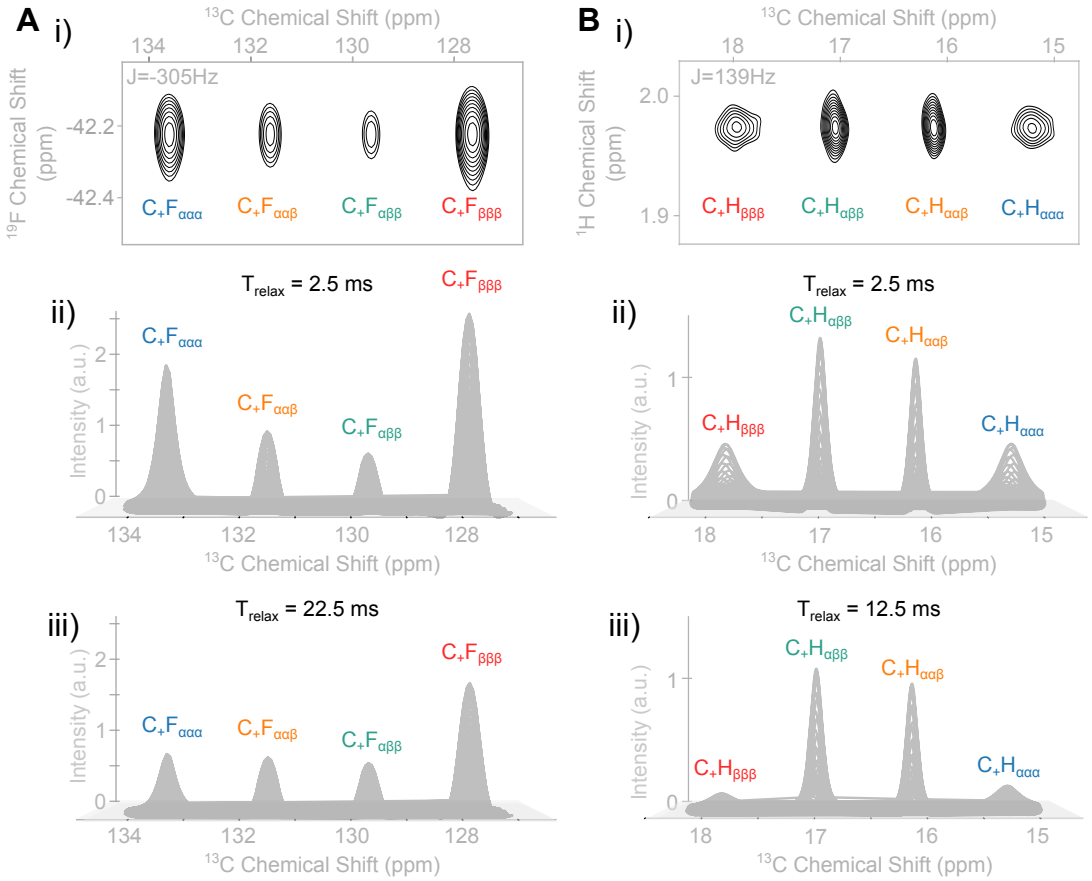


Figure 5.18: Spectra from the single quantum carbon relaxation experiment on $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^{o}C. **A** (i) A ^{13}C-^{19}F 2D correlation plot for $^{13}CF_3$-labelled HSP16.5, obtained using the sequence in Figure 5.17A at 5^{o}C with T_{relax}=2.5 ms. The peak corresponding to each single quantum carbon coherence is also defined. (ii) The intensity profile of the peaks in i. (iii) The intensity profile of the single quantum carbon coherences at a T_{relax} duration of 22.5 ms. **B** (i) A ^{13}C-^{1}H 2D correlation plot for $^{13}CH_3$-labelled HSP16.5 obtained using the sequence in Figure 5.17A at 5^{o}C with T_{relax}=2.5 ms. The peak corresponding to each single quantum carbon coherence is also defined. (ii) The intensity profile of the peaks in i. (iii) The intensity profile of the single quantum carbon coherences at a T_{relax} duration of 12.5 ms.

without deuterium labelling is excepted (like for the ^{13}C-F spin system), rationalised through the fact that the cross-relaxation rate, mixing single quantum carbon coherences in the presence of an external proton, is dependent on $f^2C(\omega_H - \omega_F)$ for $^{13}CF_3$ groups. This is much smaller than the corresponding term $f^2C(0)$ for $^{13}CH_3$ groups (see appendix C.2.3). Typically, background deuterium labelling is required for techniques such as ^{15}N-^{1}H and methyl TROSY (section 5.2.2).[53,56] Interestingly, the relaxation rate ordering of the $C_+H_{\alpha\alpha\alpha}$ (low δ_C) and $C_+H_{\beta\beta\beta}$

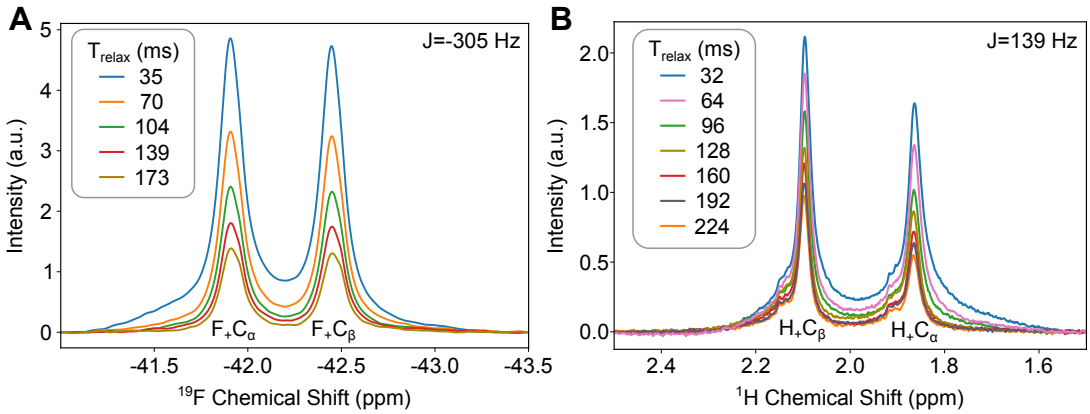


Figure 5.19: Spectra from the single quantum fluorine/proton relaxation experiments on $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^{o}C. **A** Fluorine NMR spectra obtained on $^{13}CF_3$-labelled HSP16.5 at 5^{o}C using the CPMG sequence shown in Figure 5.17C for different T_{relax} durations. **B** Proton NMR spectra obtained on $^{13}CH_3$-labelled HSP16.5 at 5^{o}C using the sequence shown in Figure 5.17B for different T_{relax} durations.

(high δ_C) for $^{13}CH_3$ labelled HSP16.5 appears opposite from what is expected from the literature.[471,472] The current rationale for this behaviour is that the interactions of the methyl group within HSP16.5 are potentially inverting the ^{13}C CSA; however, this is not confirmed.

In contrast to the SQ carbon coherences, the relaxation rates of SQ fluorine coherences (F_+) in $^{13}CF_3$ labelled HSP16.5 were heightened compared to the corresponding H_+ rates. This can be rationalised due to the large expected ^{19}F CSA component to these relaxation rates.

In addition to the SQ relaxation rates obtained on labelled HSP16.5, relaxation rates on $_{DL}$-$^{13}CF_3$-Methionine and $_{L}$-$^{13}CH_3$-Methionine in 95% glycerol solutions were obtained at numerous temperatures between 10^{o}C and 40^{o}C on a 600 MHz spectrometer to mimic a small protein and sample a larger region of τ_c space.* The relaxation rates were also measured at 20^{o}C for $^{13}CF_3$ labelled HSP16.5.

All obtained relaxation rates were subsequently fitted to the relevant symbolic rate expressions (appendix C.2.3) to determine estimates for the CSA values ($\Delta\sigma$), methyl tumbling time (τ_m) and local rotational correlation time (τ_c) (Figure 5.21).

*The results on the 95% glycerol samples were originally deposited in my masters thesis;[473] however, the spectra have been reprocessed and analysed since.

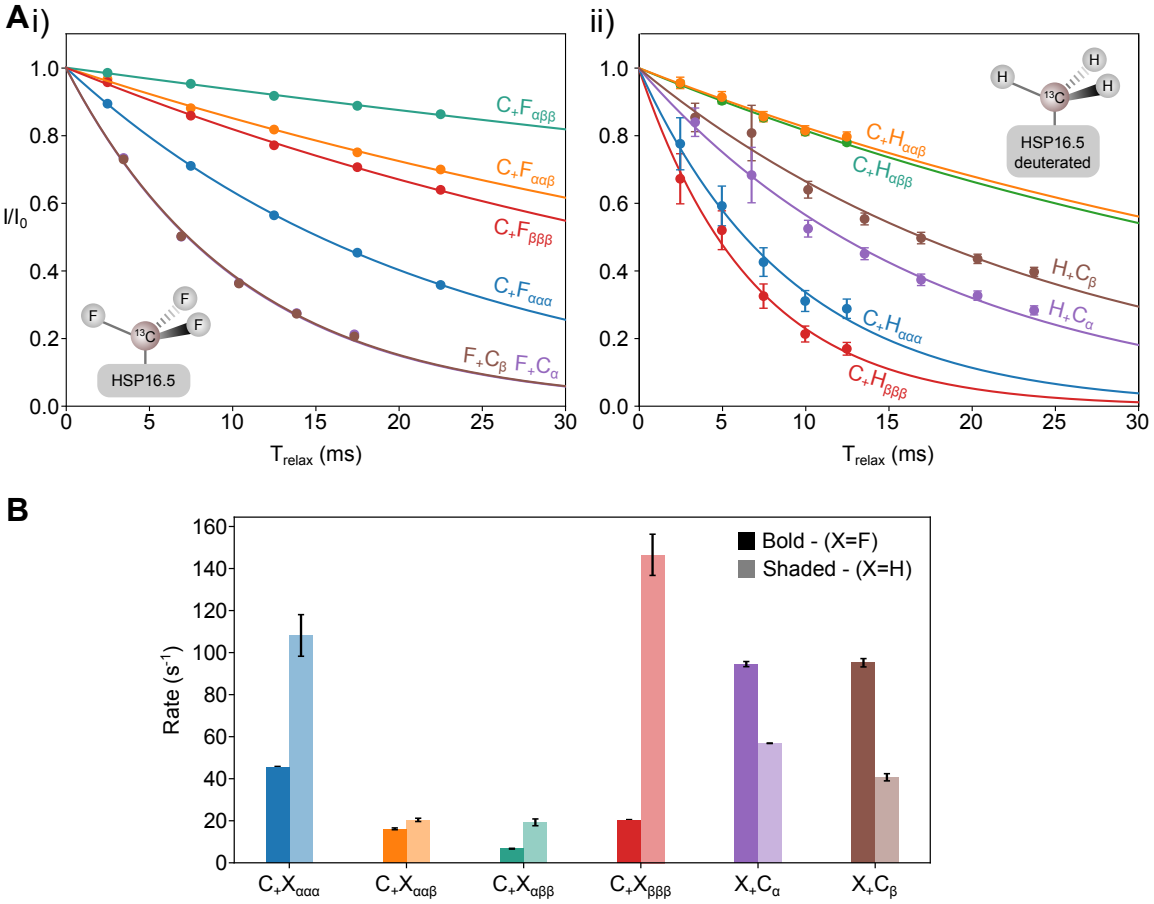


Figure 5.20: Experimental single quantum relaxation rates for $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 at 5^{o}C. **A** Plots of normalised peak intensity against relaxation delay (T_{relax}) for single quantum carbon and fluorine coherences in (i) $^{13}CF_3$ and (ii) $^{13}CH_3$-labelled HSP16.5 at 5^{o}C. **B** A single exponential curve was fitted to each profile to obtain a relaxation rate for each coherence, with the errors determined using a bootstrap procedure.

The least squares fit was performed using the Scipy Levenberg–Marquardt algorithm minimising χ^{2*}. The fitting procedure was run using multiple starting values for each variable to ensure the global optimisation minimum was found. For the $^{13}CH_3$ data, it was necessary to include the effect of external protons (a distance R_{ext} away) to obtain a good fit (see appendix C.4).

Even though the fits of the relaxation rates for the $^{13}CF_3$ and $^{13}CH_3$ labels were performed independently, the fitted τ_c values are similar, providing confidence in the fitting procedure. The errors for the fitted parameters were calculated using

*$\chi^2 = \sum_i \left(\frac{y_{\text{exp,i}} - y_{\text{calc,i}}}{\sigma_i}\right)^2$, where y_{exp} and y_{calc} are the experimental and calculated relaxation rates, with σ_i being the experimental error.

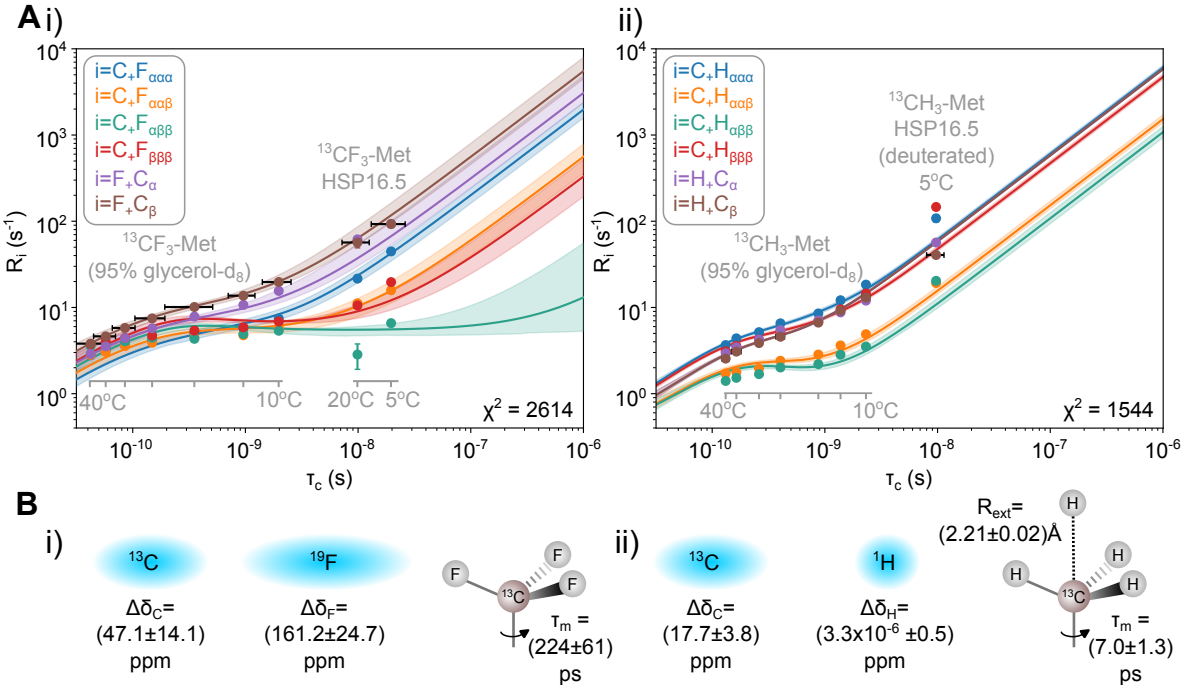


Figure 5.21: Global fitting of single quantum relaxation rates for $^{13}CF_3$ and $^{13}CH_3$ labelled systems. Fits (**A**) of the experimentally determined relaxation rates and fitted parameters (**B**) for (i) $^{13}CF_3$ and (ii) $^{13}CH_3$ single quantum coherences to symbolic rate expressions determined using RelCalc (equations in appendix C.2.3). Unlike for the $^{13}CF_3$ label, the fit of the $^{13}CH_3$ data was significantly improved through the incorporation of external proton effects (see appendix C.4). The C-H/C-F bond lengths were 1.09 and 1.38 Å, respectively. From the fit, estimates of τ_c for each set of relaxation rates were determined, along with the fitted values of $\Delta\sigma_C$ for ^{13}C, $\Delta\sigma_{F/H}$ for $^{1}H/^{19}F$, as well the methyl tumbling rate τ_m. For the $^{13}CH_3$ fit, the distance of an external proton to the carbon atom (R_{ext}) was also obtained. Errors of each parameter were calculated using a bootstrapping procedure. The error in fitted τ_c for each set of relaxation rates is shown only for the F_+C_β/H_+C_β rates. Relaxation rate expressions were then evaluated as a function of τ_c using the best-fit parameter values. The errors in fitted parameters were then propagated to errors in the predicted relaxation rates through Monte Carlo simulations. The 90% confidence intervals of these simulated values for each coherence are shown as a shaded region.

a bootstrapping procedure (see appendix C.5). The fitted parameters for each system were then used to evaluate the symbolic relaxation rates as a function of global rotational correlation time (τ_c), where τ_m was assumed to be constant (Figure 5.21). Using Monte-Carlo simulations,[474] whereby parameters were sampled from the bootstrap fitted parameter distributions, the errors of the fitted parameters were propagated to give an error in the predicted rates at each τ_c value. The expected

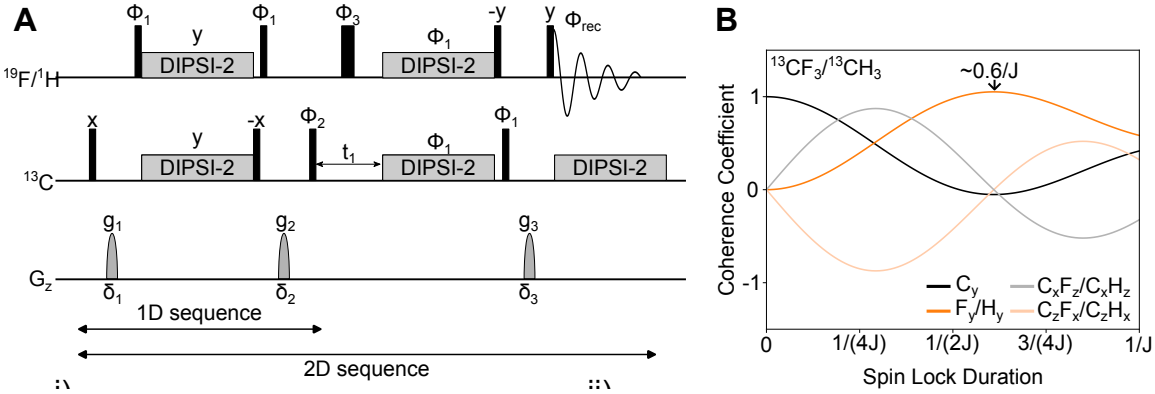


Figure 5.22: **A** A pulse sequence for one and two-dimensional [^{19}F/^{1}H,^{13}C]-HEHAHA experiments, based on previous heteronuclear Hartmann-Hahn schemes.[475,476] Rectangular 90^o and 180^o pulses are shown using thin and thick black rectangles, respectively. A pair of 4.16 kHz ^{19}F/^{1}H and ^{13}C DIPSI-2x elements were applied for durations of 1.96 ms or 4.3 ms (0.6/J) for the ^{1}H/^{19}F labelled systems to achieve cross polarisation. A 4.16 kHz ^{13}C DIPSI-2x decoupling element was used when necessary. The phase cycle employed is as follows: Φ_1=(x,-x,x,-x), Φ_2=(y,-x), Φ_3=(x,-x), Φ_4=(x,-x,x,-x) where Φ_2 is responsible for quadrature detection in two-dimensional experiments. **B** A plot showing the coherence transfer from in phase C_x magnetisation during a heteronuclear spin-lock as a function of spin-lock duration for $^{13}CF_3$ and $^{13}CH_3$ spin systems.

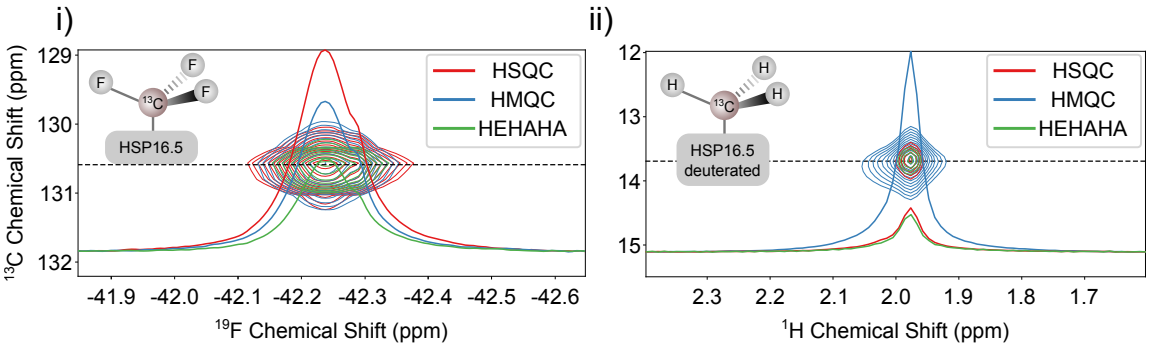


Figure 5.23: A comparison of 2D HSQC, HMQC and HEHAHA experiments recorded on (i) $^{13}CF_3$ and (ii) deuterated $^{13}CH_3$ labelled HSP16.5 at 5^{o}C. Pulse sequences for the HSQC and HMQC sequences can be found in appendix C.3.2, with the HEHAHA sequence shown in B.

rates suggested that the relaxation rate for the $C_+F_{\alpha\beta\beta}$ coherence was minimal, even at large rotational correlation times (τ_c), when in a 600 MHz spectrometer.

5.6 Determination of the optimal 2D NMR experiment for $^{13}CF_3$ labelled HSP16.5

Decoupled HSQC, HMQC and heteronuclear Hartmann-Hahn (HEHAHA) experiments were compared to determine their suitability when applied to the $^{13}CF_3$ labelled sample. Standard HSQC and HMQC pulse sequences were used (appendix C.3.2),[108,372] with the pulse sequence used for the HEHAHA experiment based on previously designed schemes (Figure 5.22A).[475,476†] During the HEHAHA experiment, conversion of magnetisation between the ^{19}F and ^{13}C nuclei occurs using a spin-lock instead of an INEPT transfer. The spin-lock was applied at the Hartmann-Hahn condition[102] using a carefully calibrated DIPSI-2x sequence.[477] For an AX_3 spin system, the predicted optimum spin-lock duration for maximum transfer of in-phase magnetisation between the A-X spins is approximately 0.6/J, where J is the A-X scalar coupling constant (Figure 5.22B).[106] This was shown empirically to give maximum signal intensity in the HEHAHA spectra on $^{13}CF_3$ and $^{13}CH_3$-labelled HSP16.5. However, imperfect pulses caused incomplete magnetisation transfer during the spin-lock element. Therefore, z-filters were added to the HEHAHA experiment to remove undesirable magnetisation terms present at the end of the spin-lock scheme (Figure 5.22A).[475]

For the HSQC and HMQC, standard magnetisation transfer times of 0.5/J were used. Therefore, due to the extended transfer duration of 0.6/J for the HEHAHA, during which fast transverse (F_+) relaxation was present, reduced signal intensity was observed when compared to the HSQC and HMQC sequences (Figure 5.23i).

For $^{13}CH_3$-labelled HSP16.5, the HMQC expectedly gave highest intensity spectrum when compared to the HSQC (Figure 5.23ii, section 5.2.2.iii). However, the opposite was observed for $^{13}CF_3$-labelled HSP16.5, with the HSQC giving higher intensity spectra (Figure 5.23i). Pulse sequence simulations performed using

†The HEHAHA sequence was developed in equal collaboration between myself and Dr. Charles Buchanan.

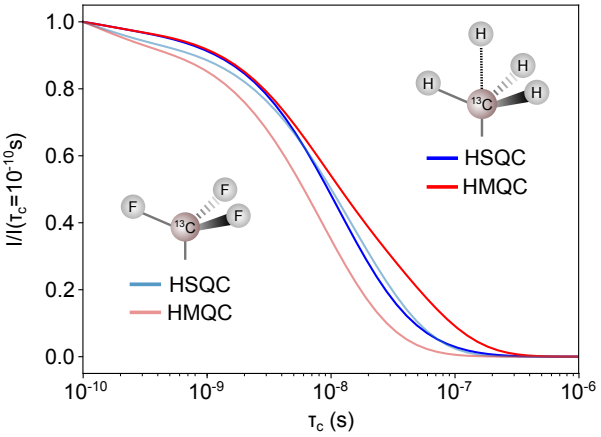


Figure 5.24: A plot showing the simulated intensities of HMQC/HSQC 2D peak intensities as a function of τ_c. Spectra were simulated in SimSeq using a complete relaxation model provided by RelCalc. The fitted constants in Figure 5.21 were used to evaluate the relaxation rates, with the spectrometer frequency being 600 MHz. Intensities were normalised against their values at τ_c=10^{-10} s. Further details can be found in appendix C.3.1.ii/C.3.1.iii.

SimSeq containing the relevant obtained physical parameters ($\Delta\delta_C$, $\Delta\delta_F$ and τ_m) for $^{13}CF_3$ groups agreed with this behaviour (Figure 5.24). To understand the reasons for this, it is again useful to view the $^{13}CF_3$ relaxation rates in the irreducible basis (Figure 5.25A). In contrast to $^{13}CH_3$ labelled systems, the lowest relaxation single quantum fluorine coherence is for $|1\rangle\langle 2|$ transitions (Figure 5.25Bi). This is opposed to the more common transitions ($|2\rangle\langle 3|$, $|5\rangle\langle 6|$, $|7\rangle\langle 8|$) which have slower relaxation rates in $^{13}CH_3$ systems (Figure 5.10). In addition, the multiple quantum transitions present during the indirect detection period of a HMQC are significantly enhanced compared to those for a $^{13}CH_3$ system (Figures 5.25Biii, 5.10Biii). Furthermore, there is only one minimally fast relaxing MQ coherence for $^{13}CF_3$, less common than the 3 slow relaxing MQ coherences for $^{13}CH_3$ groups ($C_+|2\rangle\langle 3|$, $C_+|5\rangle\langle 6|$, $C_+|7\rangle\langle 8|$). These differences in relaxation rates are largely attributed to the enhanced ^{19}F CSA. Furthermore, due to the relatively small difference in SQ carbon relaxation rates for $^{13}CF_3$-labelled HSP16.5, decoupling during the indirect dimension of a HSQC does not significantly increase signal loss, in contrast to $^{13}CH_3$-labelled HSP16.5. Altogether, this explains why the HMQC has reduced intensity when compared to the HSQC for $^{13}CF_3$-labelled HSP16.5. Nevertheless, simulations suggest that the HMQC (or ddHMQC) on $^{13}CH_3$ spin systems remains more beneficial in terms of signal intensity over the HSQC of the corresponding $^{13}CF_3$ labelled counterpart (Figure 5.24).

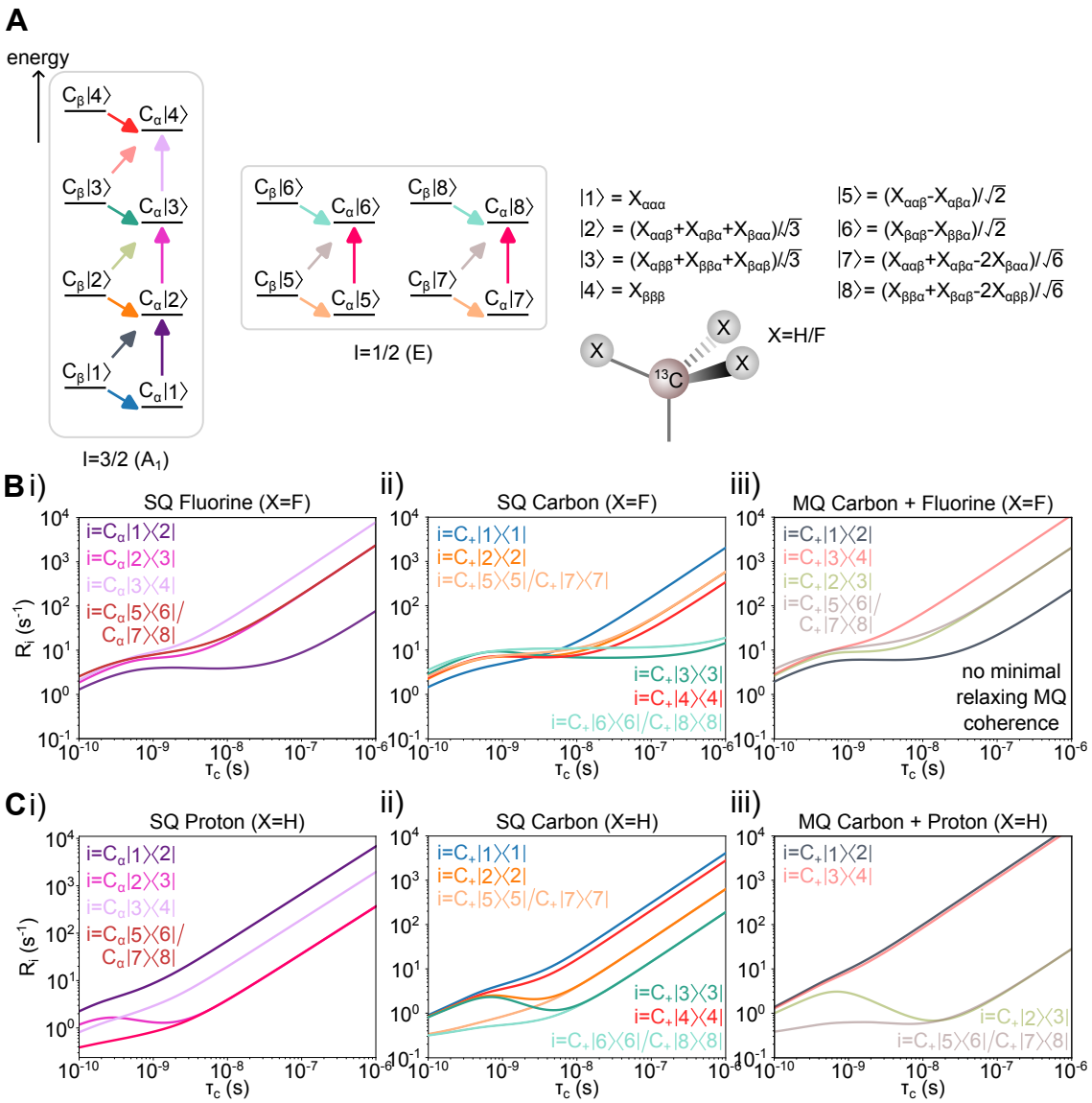


Figure 5.25: An evaluation and comparison of relaxation rates in the irreducible basis for $^{13}CF_3$ and $^{13}CH_3$ spin systems. **A** The complete spin-state energy level diagram for a $^{13}CX_3$ group (X=H/F). The eigenstates of the X nuclei are represented by the states $|1\rangle$-$|8\rangle$. Single quantum (SQ) carbon and fluorine/proton transitions, as well as multiple quantum (MQ) transitions, are shown. **B** Auto-relaxation rates of (i) the SQ fluorine transitions, (ii) SQ carbon transitions (iii) MQ transitions shown in A for a $^{13}CF_3$ spin system as a function of τ_c on a 600 MHz spectrometer. Relaxation rates were evaluated using RelCalc with the spin system set up with the same parameters determined in Figure 5.21. **C** Auto-relaxation rates of (i) the SQ proton transitions, (ii) SQ carbon transitions (iii) MQ transitions shown in A for a $^{13}CH_3$ spin system as a function of τ_c on a 600 MHz spectrometer. The same parameters as introduced in Figure 5.8 were used, in the absence of any external nuclei. Rate expressions are included in appendix C.2.3.

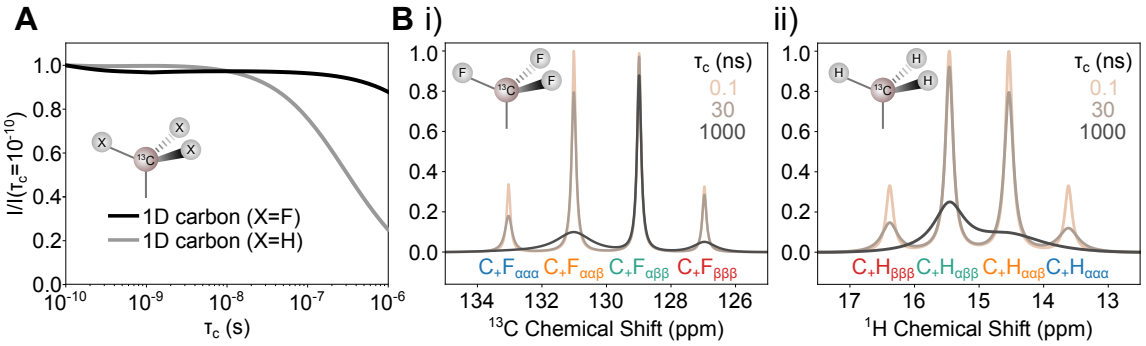


Figure 5.26: Simulated carbon 1D spectra for $^{13}CF_3$ and $^{13}CH_3$ spin systems as a function of τ_c. **A** Simulated maximum intensity in the non-decoupled carbon 1D pulse-acquire spectrum for $^{13}CF_3$ and $^{13}CF_3$ spin systems as a function of rotational correlation time. Intensities were normalised by the intensity at τ_c=10^{-10} s. **B** A plot of the simulated 1D carbon spectra for (i) $^{13}CF_3$ and (ii) $^{13}CH_3$ spin systems at τ_c values of 0.1, 30 and 1000 ns. All simulations were performed using SimSeq for a 600 MHz spectrometer, including a full relaxation model provided by RelCalc. Relaxation rates were evaluated using RelCalc with the spin system set up with the same parameters as determined from the global fit in Figure 5.21. External protons were removed from the relaxation model in the simulation for the $^{13}CH_3$ system to simulate the 1D carbon spectrum in the case of 100% background deuteration. Further details on the simulations are provided in appendix C.3.1.ii/C.3.1.iii.

5.7 1D Carbon Spectra

Despite $^{13}CF_3$ labelling being predicted to produce lower intensity 2D spectra when compared to $^{13}CH_3$ labelling (Figure 5.24), the relaxation of the $C_+F_{\alpha\beta\beta}$ coherence is predicted to be minimal even at enhanced τ_c values (Figure 5.21A). Therefore, a 1D carbon pulse-acquire spectrum, avoiding the production of fast relaxing fluorine transverse magnetisation, is predicted to retain magnetisation over an extended period and produce spectra with sharp peaks, even at large τ_c values (Figure 5.26). In contrast, at substantial τ_c values, the intensity and resolution of the carbon-pulse acquire spectrum is expected to be significantly lower for $^{13}CH_3$ labelled systems, even with the complete removal of external protons.

To test this hypothesis, 1D ^{13}C pulse acquire spectra of $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 were recorded at 25^{o}C on a 600 MHz spectrometer (Figure 5.27). Carbon direct-detect experiments were performed on a probe optimised for $^1H/^{19}F$ direct detection, yet still gave sufficient signal intensity for analysis. The $^{13}CF_3$

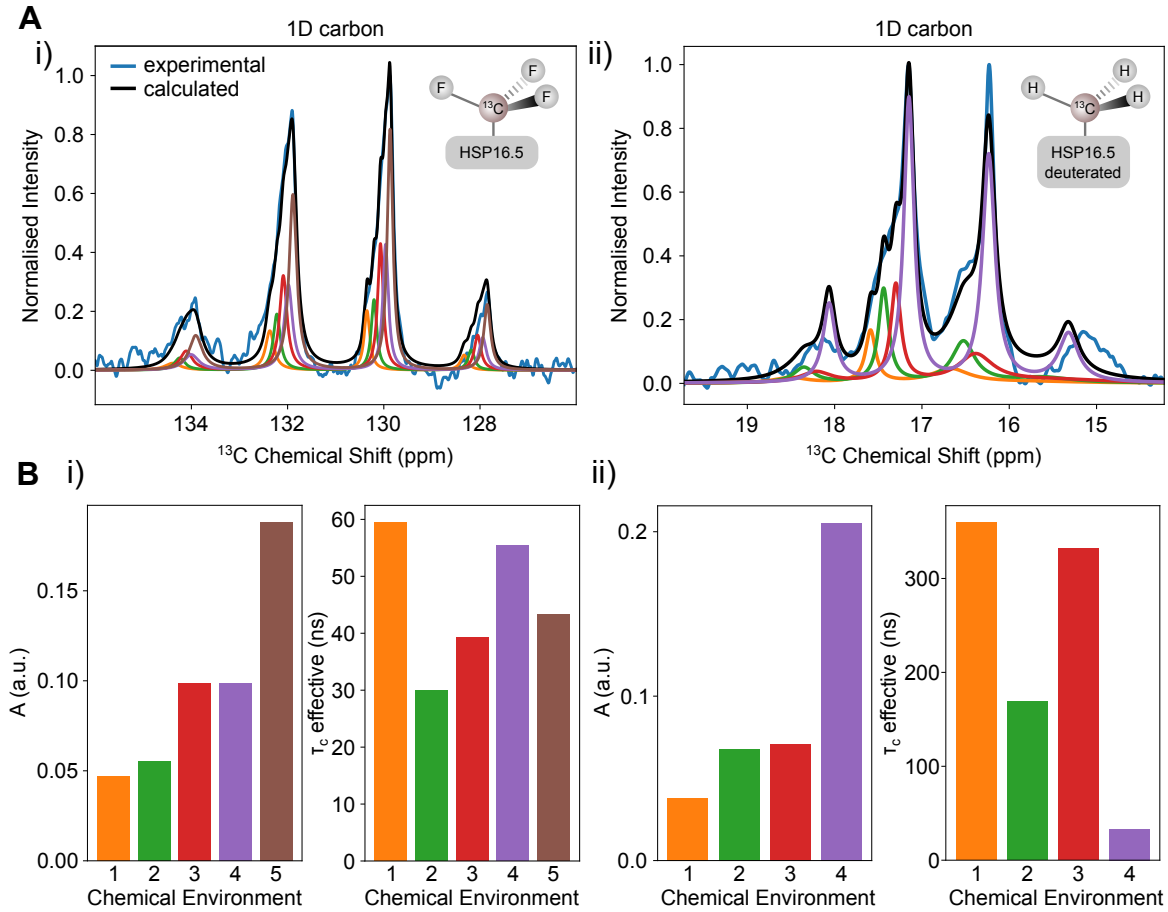


Figure 5.27: Experimental carbon 1D pulse-acquire spectra for $^{13}CF_3$ and $^{13}CH_3$ HSP16.5. **A** The coupled 1D-carbon spectra obtained on (i) $^{13}CF_3$ and (ii) $^{13}CH_3$ Met labelled HSP16.5 overlaid with fitted simulated spectra in each case. The fitted multiplets from each chemical environment are shown in different colours with their sum, the total simulated spectrum, shown in black. The 1D carbon spectra were obtained on a non-broadband probe optimised for proton detection. Further experimental details are included in appendix C.3.2.**B** The fitted relative peak amplitudes and τ_c values for the detected chemical environments in (i) $^{13}CF_3$ and (ii) $^{13}CH_3$ labelled HSP16.5.

spectrum contained sharper multiplet peaks than its $^{13}CH_3$ counterpart. In both systems, multiple chemical environments could be detected, in contrast to the two dimensional experiments. Peaks were picked on the slowest relaxing $C_+X_{\alpha\beta\beta}$ lines (X=^{19}F/^{1}H) from the spectra using UnidecNMR[386] followed by minor manual adjustment, with five peaks obtained for $^{13}CF_3$ and four for $^{13}CH_3$. Identifying multiple chemical environments in the spectra is consistent with HSP16.5, which contains six methionine residues. From the $C_+X_{\alpha\beta\beta}$ peak lists, the spectra could be

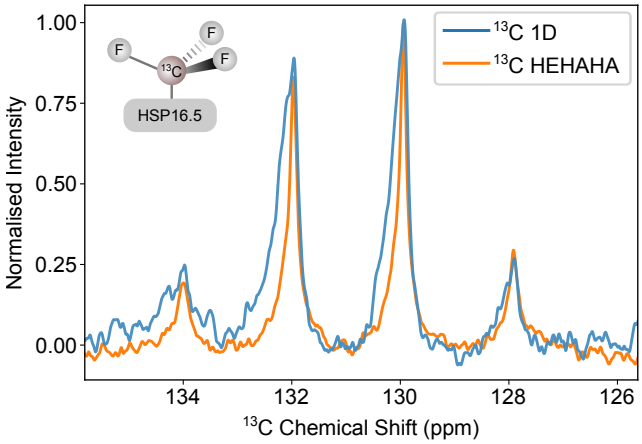


Figure 5.28: A plot comparing the normalised spectra from the 1D carbon HEHAHA experiment (described in Figure 5.22) and the carbon pulse-acquire experiment. The shoulder peaks observed for the carbon pulse-acquire experiment were not visible in the HEHAHA spectrum.

simulated using relaxation rate expressions from RelCalc to match the experimental data, whereby the values of $\Delta\delta_C$, $\Delta\delta_{H/F}$, τ_m, and R_{ext} in the case of $^{13}CH_3$ groups, were those obtained from the global fitting routine (Figure 5.21). Least squares minimisation was applied to determine the values of τ_c and relative proportions (A) of each that produced simulated spectra which agreed most strongly with the data (Figure 5.27). Interestingly, the relative proportions of each chemical environment (denoted by A) were not consistent, indicating that the labelling efficacy was not identical for each residue.

For $^{13}CF_3$-HSP16.5 the fitted τ_c values range between 30 ns and 55 ns (Figure 5.27Aii), slightly elevated above to approximate 10 ns value obtained from the global fit of relaxation rates at 20^{o}C. The most prominent peak in $^{13}CH_3$-labelled HSP16.5 also has a similar fitted τ_c value of 30 ns (5.27Bii). However, the additional peaks were significantly broader than expected. Therefore, the fitted effective τ_c values are substantially elevated above that expected $^{13}CH_3$-labelled HSP16.5. However, the elevated effective τ_c fitting parameters are likely an artefact of the fitting, manifesting from the fact that the effect of residual external protons is assumed constant for all residues but likely has a more pronounced effect in residues from the more structured regions of the protein.

One-dimensional HEHAHA experiments were performed on $^{13}CF_3$-labelled HSP16.5 in an attempt to increase the sensitivity from the carbon pulse-acquire experiment. The refocussed-INEPT experiment could also have been performed

for this purpose.[478] While the sensitivity of the experiment was increased, the number of observed chemical environments was reduced due to the presence of fast relaxation of transverse fluorine magnetisation during the coherence transfer period (Figure 5.28).

All together, the experimental line shapes in carbon pulse-acquire spectra on $^{13}CF_3$-labelled HSP16.5 appeared consistent with the predicted spectra in Figure 5.26. Therefore, it is hypothesised that the 1D carbon pulse-acquire sequence can be applied to study very high molecular weight, $^{13}CF_3$ labelled systems, without the requirement of deuteration. Furthermore, recent developments in ^{13}C optimised NMR probe production (such as the Bruker TXO CryoProbe) are also predicted to enhance the observed signal-to-noise ratio of ^{13}C spectra on $^{13}CF_3$ labelled systems.[479]

5.8 Conclusion

The $^{13}CF_3$ system was proposed as a probe to be used in large (slowly tumbling) proteins to leverage relaxation interference and increase the resolution in its NMR spectra. Earlier attempts to detect relaxation interference in proteins involving this spin system proved unsuccessful, albeit on a small 11 kDa protein.[461] The $^{13}CF_3$ group was introduced into the 395 kDa heat shock protein complex HSP16.5 without background deuteration. For comparison, background deuterated $^{13}CH_3$-labelled HSP16.5 was produced. Analysis of the $^{19}F/^{1}H$ NMR spectra at 5^{o}C revealed three distinct fluorine chemical environments within the $^{13}CF_3$ labelled system (Figure 5.15), with two proton chemical environments detected in the $^{13}CH_3$ labelled counterpart.

The ^{13}C and $^{19}F/^{1}H$ single quantum relaxation rates for the most prominent chemical environment in the $^{13}CF_3$ and $^{13}CH_3$ labelled systems were assessed at 5^{o}C on a 600 MHz spectrometer (Figures 5.18, 5.19). The ^{13}C relaxation rates in both labelled systems exhibited clear indications of relaxation interference. However, the

relaxation rate of $C_+F_{\alpha\beta\beta}$ in the $^{13}CF_3$ labelled system was significantly smaller than the $^{13}CH_3$ analogue, even in the absence of background protein deuteration (Figure 5.20). The single quantum relaxation rates of each coherence in these systems, together with the relaxation rates obtained for $^{13}CF_3$/$^{13}CH_3$-Methionine in 95% glycerol solutions, were fitted to symbolic rate expressions evaluated using RelCalc (Figure 5.21). The fits predicted that the $C_+F_{\alpha\beta\beta}$ relaxation rate would remain extremely small, even at enhanced molecular weights and τ_c values.

A comparative analysis of two-dimensional experiments, namely the HSQC, HMQC, and HEHAHA, indicated that the HSQC yielded the highest signal intensity for $^{13}CF_3$-labelled HSP16.5 at 5^oC (Figure 5.23i). This contrasts with $^{13}CH_3$ labelled HSP16.5 where the HMQC is more effective (Figure 5.23ii).[56]

Simulations using SimSeq predicted that the intensity and resolution of the slow relaxing $C_+F_{\alpha\beta\beta}$ coherence in the pulse acquire 1D carbon spectra of the $^{13}CF_3$ spin system would be largely independent of rotational correlation time (Figure 5.26). An experimental 1D-carbon spectrum was obtained for $^{13}CF_3$ labelled HSP16.5 and was consistent with the prediction of a slowly relaxing coherence across a range of τ_c values (Figure 5.27). Owing to the diminished relaxation rate of $C_+F_{\alpha\beta\beta}$, in the case of larger molecular weight systems, incorporating $^{13}CF_3$ labels and conducting 1D ^{13}C spectra using may facilitate the study of larger biomolecules using NMR. The simplicity of 1D carbon experiments makes them accessible to non-experts, and advancements in carbon-13 optimised probes[479] could enhance the potential of $^{13}CF_3$ labels in NMR. Nevertheless, the use of $^{13}CF_3$ labelling schemes to perform NMR on high molecular weight biomolecules is expected to remain highly specialised.

5.9 Future work

To verify that the relaxation rate of $C_+F_{\alpha\beta\beta}$ coherence is indeed small at extreme rotational correlation times (τ_c>100 ns), labelling of larger molecular weight species such as a virus capsid with the $^{13}CF_3$ probe is required. Carbon pulse-acquire spectra

can then be obtained to see if there is evidence of a sharp $C_+F_{\alpha\beta\beta}$ coherence.

However, as discussed previously, the main drawback of the $^{13}CF_3$ probe is the vast ^{19}F relaxation rates at large τ_c values due to the substantial ^{19}F CSA contributions. These rates become even more enhanced at larger spectrometer frequencies. The previously proposed two-field spectrometer systems for ^{13}C-F spin systems utilising shuttle systems could be used to enhance the resolution in the resultant 2D C-F correlation spectra on $^{13}CF_3$ labelled systems.[445] This would enable ^{19}F magnetisation to be encoded at lower magnetic fields with a reduced relaxation rate, and ^{13}C magnetisation detected at high magnetic field with enhanced sensitivity.

Recently, schemes have been developed by Boeszoermenyi *et al.*[480] for ^{1}H-^{13}C-F aromatic scalar coupled spin systems using heteronuclear multiple bond couplings to transfer the magnetisation from ^{1}H protons on neighbouring carbon atoms to the ^{13}CF carbon atom. SQ ^{13}C magnetisation can be transferred with SQ proton magnetisation (opposed to SQ fluorine), enabling the production of higher resolution 2D spectra. Similar schemes could be developed for $^{13}CF_3$ labelled amino acids such as alanine, threonine, leucine, isoleucine and valine. In these systems, the two-bond coupling constants (J) between the methyl group carbon nuclei and the hydrogen nucleus on the preceding carbon atom are suggested to be on the order of 3-6 Hz.[481] Therefore, coherence transfers between trifluoromethyl ^{13}C nucleus and the ^{1}H nucleus would take on the order of 100 ms ($1/2J$). These long transfer times would no doubt limit the intensity of the produced spectra by allowing more time for transverse relaxation. However, the resultant peaks will likely be sharp due to the reduced overall relaxation rates. To test this hypothesis, synthetic procedures to produce $^{13}CF_3$ labelled variants of alanine, threonine, leucine, isoleucine, or valine will have to be developed. Sequences similar to that employed in Boeszoermenyi *et al.*[480] could be re-purposed to apply to ^{1}H-$^{13}CF_3$ spin systems.

5.10 Contributions

I performed all RelCalc and SimSeq simulations, NMR experiments and analysis as part of this project. Professor Andrew Baldwin provided the initial conceptualisation for this project and wrote the RelCalc and SimSeq software packages. Dr. Satoshi Kishigami synthesised the $_{DL}$-$^{13}CF_3$-Methionine molecule and performed protein expression and purification of the $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 constructs. The HEHAHA pulse sequence development was performed in equal collaboration between myself and Dr. Charles Buchanan.

6 SpinExplorer: A graphical interface for NMR processing and analysis

6.1 Introduction

Accurate processing and analysis is essential for biomolecular NMR. Since its development, nmrPipe[482] has become the go-to package for all biomolecular NMR processing. However, the graphical user interface of nmrPipe (nmrDraw) is no longer available natively on MacOS operating systems (as of June 2025). Furthermore, nmrPipe itself does not have a native build for Windows operating systems. Several workarounds using Docker (`https://compbiocore.github.io/nmrdock/`), Windows-Subsystem for Linux (WSL), or virtual machines are available. However, these inherently limit the accessibility of the package. Since the development of the Python package nmrglue,[483] most of this functionality can be recreated; however, it often requires the production of single-use Python scripts for processing and analysis. SpinExplorer is a graphical user interface (GUI) Python package for nmrPipe and nmrglue available through pip (installation instructions at `https://pypi.org/project/SpinExplorer/`). Full documentation and test data is provided at `https://github.com/james-eaton-1/SpinExplorer`. This package is designed to allow users to quickly process and analyse their data and perform routine analysis tasks such as diffusion or relaxation analysis, with the hope of adding further functionality in the future.

6.2 What SpinExplorer can do

SpinExplorer is divided into three sub-packages (SpinConverter, SpinProcess and SpinView) responsible for converting raw NMR data to nmrPipe format, processing NMR data and viewing/analysing spectra. A large quantity of spectra and analysis in this thesis were performed using SpinExplorer.

6.2.1 SpinConverter

SpinConverter works in a similar manner to the bruker/varian commands of nmrPipe. SpinConverter aims to find the correct spectral parameters from the spectrometer files ("acqus" or "procpar" for Bruker/Varian spectrometer types). However all conversion script inputs are fully customisable. SpinConverter can then provide conversion of the FID to nmrPipe format using nmrPipe (if installed) or using nmrglue.

6.2.2 Processing

SpinProcess allows the customisable processing of nmrPipe FID files (.fid) to spectrum files (.ft,.ft2,.ft3). If nmrPipe is not installed, all processing will be performed using nmrglue. Note that SMILE NUS reconstruction is not available without an nmrPipe installation and 3D NMR processing using nmrPipe is in beta-mode. Additionally, digital filter removal for Bruker data using nmrglue is not perfect and can sometimes lead to baseline distortions which can usually be corrected for using baseline subtraction.

6.2.3 Viewing/overlaying spectra

SpinView enables processed spectrum files in nmrPipe format (.ft, .ft2, .ft3) or TopSpin format to be viewed and subsequently analysed. Spectrum files (nmrPipe) or folders (TopSpin) can be dragged and dropped on top of other

spectra provided they are of the same dimension. Modules for diffusion and relaxation (R_1, R_2) analysis are included for data in pseudo-2D format. Additional analysis modules could be added in the future to improve the impact of the project. Full documentation and a discussion of available functionality is included in Documentation.pdf at `https://github.com/james-eaton-1/SpinExplorer`.

6.3 Contributions

I would like to thank the members of the Baldwin lab for their help in making this package easier to install/use and their input on what functionality they would like it to have. Professor Andrew Baldwin provided insight on how to best layout the diffusion and relaxation analysis modules. Eugene Lin provided useful comments on how to improve the code accessibility.

7 Conclusion

This thesis used NMR to probe biomolecular condensates and a disordered virus capsid protein. Additionally, an evaluation of methods to combat molecular size limitations in NMR were discussed.

As highlighted in Chapter 1, solution-state NMR is an invaluable technique to extract atom-specific structural and dynamical properties in molecules. Chapter 2 introduced the NMR theory and key experiments used throughout this thesis.

Chapter 3 explored the solvent environment within biomolecular condensates. *In-vitro* NMR-based methods were developed to quantify the partitioning of metabolites (nucleotides and amino acids) and ions into biomolecular condensates derived from the proteins $DDX4^{N1}$ and EWSR1C. The metabolite partitioning measurements in the condensates were then compared to their values in the organic solvents 1-butanol and 1-octanol. A model to assess the contributions of metabolite properties to partitioning was established for both the condensates and organic solvents. The solubility trends showed similarities and differences in the factors affecting metabolite partitioning between the condensates and organic solvents. The factors affecting metabolite partitioning were correlated to the properties of the solvent. This is similar to the concept of "like-dissolves-like", which is widely used to predict partitioning into organic solvents. The trends provide a potential method for improved rational drug design to target or use condensates as therapeutic delivery mechanisms.

The physical and chemical properties of water inside $DDX4^{N1}$ condensates were probed using NMR on a biphasic sample. MRI experiments revealed an

approximately 30% reduction in the water concentration inside the condensate. Additionally, water inside the condensate was shielded to a similar extent as water in DMSO solutions. The rotational and translational effective viscosities experienced by water in the condensate phase were 2.5 to 3 times higher than in the dilute phase. In summary, NMR was used to quantify the partitioning and physicochemical properties in biomolecular condensates, which adds to the growing evidence that condensates act as distinct solvents.[44–49]

In Chapter 4, NMR was used to explore the properties of the N-terminal truncated adenoviral capsid protein VI (Δ54-pVI). Adenoviruses can cause common respiratory illnesses and are also used as viral vectors for vaccination.[320,329–334] The predicted N-terminal amphipathic helix of protein VI is essential for adenoviral infectivity, by disrupting endosomal membranes and allowing the virus to enter a host cell.[50–52] In comparison to the N-terminus, less attention has been given to the remainder of the protein sequence and its functions. Diffusion NMR measurements showed that Δ54-pVI was highly disordered. However, the semi-automated backbone assignment of Δ54-pVI revealed significant amphipathic α-helical propensity near key sequence features. Micelles were used to mimic a phospholipid membrane. Δ54-pVI interacted substantially with the micelles. The Δ54-pVI-micelle interaction is predicted to be driven by the amphipathic α-helical propensity. Knowledge of the disordered nature and secondary structure propensity within Δ54-pVI provides a valuable understanding in the context of full-length protein VI and its intricate roles in the virus life cycle.

Chapter 5 evaluated methods to combat reduced signal intensity and resolution of NMR spectra for slowly tumbling species. The simulation packages RelCalc[54] and SimSeq[55] were used to evaluate the benefits and limitations of current TROSY methods.[53,56–60] Simulations identified the $^{13}CF_3$ spin system as a potential TROSY target. Relaxation measurements on $^{13}CF_3$-labelled HSP16.5 (395 kDa) at 5^{o}C on a 600 MHz spectrometer showed a slow relaxing carbon coherence. Experimental

1D-carbon spectra recorded on the $^{13}CF_3$ labelled system showed the presence of multiple species containing a slow relaxing coherence. Simulations predicted that the slow relaxation rate of this coherence would be largely independent of molecular tumbling. Further optimisation of TROSY sequences applied to $^{13}CF_3$-labelled samples could enable enhanced NMR spectral quality of larger systems such as complete virus complexes.

Finally, SpinExplorer was introduced; a Python-based graphical user interface for NMR processing and analysis. Many of the spectra presented in this thesis were processed and analysed using this package. While the package is not currently released for public use, it will be made open-source for release on PyPI shortly.

NMR is a powerful tool to probe biology, enabling the characterisation of biomolecules across phases, for both disordered and structured systems. This work contributed to the understanding of: biological solvent environments, the mechanism of protein VI in the adenovirus life cycle, and methods to combat the limitations of NMR for slow tumbling systems.

Bibliography

(1) Alderson, T. R.; Kay, L. E. *Cell* **2021**, *184*, 577–595, DOI: 10.1016/j.cell.2020.12.034.

(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235–242, DOI: 10.1093/nar/28.1.235.

(3) Baldwin, A. J.; Kay, L. E. *Nature Chemical Biology* **2009**, *5*, 808–814, DOI: 10.1038/nchembio.238.

(4) Punjani, A.; Fleet, D. J. *Journal of Structural Biology* **2021**, *213*, 107702, DOI: https://doi.org/10.1016/j.jsb.2021.107702.

(5) Jumper, J. et al. *Nature* **2021**, *596*, 583–589, DOI: 10.1038/s41586-021-03819-2.

(6) Del Alamo, D.; Sala, D.; McHaourab, H. S.; Meiler, J. *eLife* **2022**, *11*, e75751, DOI: 10.7554/eLife.75751.

(7) Guo, H.-B.; Perminov, A.; Bekele, S.; Kedziora, G.; Farajollahi, S.; Varaljay, V.; Hinkle, K.; Molinero, V.; Meister, K.; Hung, C.; Dennis, P.; Kelley-Loughnane, N.; Berry, R. *Scientific Reports* **2022**, *12*, 10696, DOI: 10.1038/s41598-022-14382-9.

(8) Gimenez, D.; Phelan, A.; Murphy, C. D.; Cobb, S. L. *Beilstein Journal of Organic Chemistry* **2021**, *17*, 293–318, DOI: 10.3762/bjoc.17.28.

(9) Williamson, M. P.; Havel, T. F.; Wüthrich, K. *Journal of molecular biology* **1985**, *182*, 295–315, DOI: 10.1016/0022-2836(85)90347-X.

(10) Kay, L. E.; Marion, D.; Bax, A. *Journal of Magnetic Resonance (1969)* **1989**, *84*, 72–84, DOI: 10.1016/0022-2364(89)90006-1.

(11) Ikura, M.; Kay, L. E.; Bax, A. *Biochemistry* **1990**, *29*, 4659–4667, DOI: 10.1021/bi00471a022.

(12) Clore, G. M.; Gronenborn, A. M. *Annu Rev Biophys Biophys Chem* **1991**, *20*, 29–63, DOI: 10.1146/annurev.bb.20.060191.000333.

(13) Spera, S.; Bax, A. *Journal of the American Chemical Society* **1991**, *113*, 5490–5492, DOI: 10.1021/ja00014a071.

(14) Clore, G. M.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Journal of Molecular Biology* **1990**, *214*, 811–817, DOI: 10.1016/0022-2836(90)90336-K.

(15) Clore, G. M.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1991**, *30*, 2315–2323, DOI: 10.1021/bi00223a005.

(16) Forman-Kay, J. D.; Clore, G. M.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1991**, *30*, 2685–2698, DOI: 10.1021/bi00224a017.

(17) Carr, H. Y.; Purcell, E. M. *Physical Review* **1954**, *94*, 630–638, DOI: 10.1103/PhysRev.94.630.

(18) Meiboom, S.; Gill, D. *Review of Scientific Instruments* **1958**, *29*, 688–691, DOI: 10.1063/1.1716296.

(19) Stejskal, E. O.; Tanner, J. E. *The Journal of Chemical Physics* **1965**, *42*, 288–292, DOI: 10.1063/1.1695690.

(20) Tanner, J. E. *The Journal of Chemical Physics* **1970**, *52*, 2523–2526, DOI: 10.1063/1.1673336.

(21) Palmer III, A. G.; Williams, J.; McDermott, A. *Journal of Physical Chemistry* **1996**, *100*, 13293–13310, DOI: 10.1021/jp9606117.

(22) Dobson, C. M.; Hore, P. J. *Nature Structural Biology* **1998**, *5*, 504–507, DOI: 10.1038/744.

(23) Mok, K. H.; Hore, P. J. *Methods* **2004**, *34*, 75–87, DOI: 10.1016/j.ymeth.2004.03.006.

(24) Tsang, B.; Pritišanac, I.; Scherer, S. W.; Moses, A. M.; Forman-Kay, J. D. *Cell* **2020**, *183*, 1742–1756, DOI: 10.1016/j.cell.2020.11.050.

(25) Karplus, M.; Kuriyan, J. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6679–6685, DOI: 10.1073/pnas.0408930102.

(26) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964–972, DOI: 10.1038/nature06522.

(27) Wüthrich, K.; Wagner, G. *Trends in Biochemical Sciences* **1978**, *3*, 227–230, DOI: 10.1016/S0968-0004(78)94607-8.

(28) Palmer III, A. G. *Current Opinion in Biotechnology* **1993**, *4*, 385–391, DOI: 10.1016/0958-1669(93)90002-E.

(29) Smith, L. J.; Dobson, C. M. *International Journal of Quantum Chemistry* **1996**, *59*, 315–332, DOI: 10.1002/(SICI)1097-461X(1996)59:4<315::AID-QUA6>3.0.CO;2-0.

(30) Kay, L. E. *Journal of Magnetic Resonance* **2005**, *173*, 193–207, DOI: 10.1016/j.jmr.2004.11.021.

(31) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *Journal of the American Chemical Society* **2012**, *134*, 8148–8161, DOI: 10.1021/ja3001419.

(32) Heller, G. T.; Yu, L.; Hansen, D. F. In *NMR Spectroscopy for Probing Functional Dynamics at Biological Interfaces*, Bhunia, A., Atreya, H. S., Sinha, N., Eds.; The Royal Society of Chemistry: 2022.

(33) Fischer, E. *Berichte der deutschen chemischen Gesellschaft* **1894**, *27*, 2985–2993, DOI: 10.1002/cber.18940270364.

(34) Koshland, D. E. *Proceedings of the National Academy of Sciences* **1958**, *44*, 98–104, DOI: 10.1073/pnas.44.2.98.

(35) Heller, G. T.; Aprile, F. A.; Bonomi, M.; Camilloni, C.; De Simone, A.; Vendruscolo, M. *Journal of Molecular Biology* **2017**, *429*, 2772–2779, DOI: 10.1016/j.jmb.2017.07.016.

(36) Heller, G. T.; Shukla, V. K.; Figueiredo, A. M.; Hansen, D. F. *Journal of the American Chemical Society* **2024**, *146*, DOI: 10.1021/jacs.3c11614.

(37) Xie, T.; Saleh, T.; Rossi, P.; Kalodimos, C. G. *Science* **2020**, *370*, eabc2754, DOI: 10.1126/science.abc2754.

(38) Lazar, T. et al. *Nucleic Acids Research* **2020**, *49*, D404–D411, DOI: 10.1093/nar/gkaa1021.

(39) Shukla, V. K.; Heller, G. T.; Hansen, D. F. *Structure* **2023**, *31*, 1360–1374, DOI: 10.1016/j.str.2023.09.011.

(40) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936–962, DOI: 10.1002/cbic.200300700.

(41) Lauterbur, P. C. *Nature* **1973**, *242*, 190–191, DOI: 10.1038/242190a0.

(42) Emwas, A.-H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. *Metabolites* **2019**, *9*, 123, DOI: 10.3390/metabo9070123.

(43) Hyman, A. A.; Weber, C. A.; Jülicher, F. *Annual Review of Cell and Developmental Biology* **2014**, *30*, 39–58, DOI: 10.1146/annurev-cellbio-100913-013325.

(44) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. *Molecular Cell* **2015**, *57*, 936–947, DOI: 10.1016/j.molcel.2015.01.013.

(45) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E. *Proceedings of the National Academy of Sciences* **2017**, *114*, E8194–E8203, DOI: 10.1073/pnas.1706197114.

(46) Ahmed, R.; Liang, M.; Hudson, R. P.; Rangadurai, A. K.; Huang, S. K.; Forman-Kay, J. D.; Kay, L. E. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2408554121, DOI: 10.1073/pnas.2408554121.

(47) Joshi, A.; Avni, A.; Walimbe, A.; Rai, S. K.; Sarkar, S.; Mukhopadhyay, S. *The Journal of Physical Chemistry Letters* **2024**, *15*, 7724–7734, DOI: 10.1021/acs.jpclett.4c01153.

(48) Ambadi Thody, S.; Clements, H. D.; Baniasadi, H.; Lyon, A. S.; Sigman, M. S.; Rosen, M. K. *Nature Chemistry* **2024**, *16*, 1794–1802, DOI: 10.1038/s41557-024-01630-w.

(49) Kilgore, H. R.; Mikhael, P. G.; Overholt, K. J.; Boija, A.; Hannett, N. M.; Van Dongen, C.; Lee, T. I.; Chang, Y.-T.; Barzilay, R.; Young, R. A. *Nature Chemical Biology* **2024**, *20*, 291–301, DOI: 10.1038/s41589-023-01432-0.

(50) Wodrich, H.; Guan, T.; Cingolani, G.; Von Seggern, D.; Nemerow, G.; Gerace, L. *The EMBO Journal* **2003**, *22*, 6245–6255, DOI: 10.1093/emboj/cdg614.

(51) Wiethoff, C. M.; Wodrich, H.; Gerace, L.; Nemerow, G. R. *Journal of Virology* **2005**, *79*, 1992–2000, DOI: 10.1128/jvi.79.4.1992-2000.2005.

(52) Maier, O.; Galan, D. L.; Wodrich, H.; Wiethoff, C. M. *Virology* **2010**, *402*, 11–19, DOI: 10.1016/j.virol.2010.03.043.

(53) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proceedings of the National Academy of Sciences* **1997**, *94*, 12366–12371, DOI: 10.1073/pnas.94.23.12366.

(54) Eaton, J. T.; Stedman, D.; Baldwin, A. J. RelCalc: symbolic evaluation of BWR theory relaxation rates, applications to TROSY effects in AXn spin systems (in preparation).

(55) Baldwin, A. J. SimSeq - A fully functional virtual NMR spectrometer (in preparation).

(56) Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. *Magnetic Resonance in Chemistry* **2003**, *41*, 843–852, DOI: 10.1002/mrc.1256.

(57) Boeszoermenyi, A.; Chhabra, S.; Dubey, A.; Radeva, D. L.; Burdzhiev, N. T.; Chanev, C. D.; Petrov, O. I.; Gelev, V. M.; Zhang, M.; Anklin, C.; Kovacs, H.; Wagner, G.; Kuprov, I.; Takeuchi, K.; Arthanari, H. *Nat Methods* **2019**, *16*, 333–340, DOI: 10.1038/s41592-019-0334-x.

(58) Rößler, P.; Mathieu, D.; Gossert, A. D. *Angewandte Chemie International Edition* **2020**, *59*, 19329–19337, DOI: 10.1002/anie.202007715.

(59) Bolik-Coulon, N.; Sever, A. I. M.; Harkness, R. W.; Aramini, J. M.; Toyama, Y.; Hansen, D. F.; Kay, L. E. *Journal of Magnetic Resonance* **2023**, *346*, 107326, DOI: 10.1016/j.jmr.2022.107326.

(60) Ahmed, R.; Rangadurai, A. K.; Ruetz, L.; Tollinger, M.; Kreutz, C.; Kay, L. E. *Journal of Magnetic Resonance* **2024**, *362*, 107667, DOI: 10.1016/j.jmr.2024.107667.

(61) Hore, P. J., *Nuclear magnetic resonance*, Second edition.; Nuclear magnetic resonance; Oxford University Press: Oxford, 2023.

(62) Zeeman, P. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1897**, *43*, 226–239, DOI: 10.1080/14786449708620985.

(63) Frisch, R.; Stern, O. *Zeitschrift für Physik* **1933**, *85*, 4–16, DOI: 10.1007/BF01330773.

(64) Estermann, I.; Stern, O. *Zeitschrift für Physik* **1933**, *85*, 17–24, DOI: 10.1007/BF01330774.

(65) Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. *Physical Review* **1938**, *53*, 318–318, DOI: 10.1103/PhysRev.53.318.

(66) Bloch, F.; Hansen, W. W.; Packard, M. *Physical Review* **1946**, *70*, 474–485, DOI: 10.1103/PhysRev.70.474.

(67) Purcell, E. M.; Torrey, H. C.; Pound, R. V. *Physical Review* **1946**, *69*, 37–38, DOI: 10.1103/PhysRev.69.37.

(68) Proctor, W. G.; Yu, F. C. *Physical Review* **1950**, *77*, 717–717, DOI: 10.1103/PhysRev.77.717.

(69) Dickinson, W. C. *Physical Review* **1950**, *77*, 736–737, DOI: 10.1103/PhysRev.77.736.2.

(70) Lamb, W. E. *Physical Review* **1941**, *60*, 817–819, DOI: 10.1103/PhysRev.60.817.

(71) Ramsey, N. F. *Physical Review* **1950**, *77*, 567–567, DOI: 10.1103/PhysRev.77.567.

(72) Ramsey, N. F. *Physical Review* **1950**, *78*, 699–703, DOI: 10.1103/PhysRev.78.699.

(73) Lenz, E. *Annalen der Physik* **1834**, *107*, 483–494, DOI: 10.1002/andp.18341073103.

(74) Larmor, J. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1897**, *44*, 503–512, DOI: 10.1080/14786449708621095.

(75) Hore, P. J.; Jones, J. A.; Wimperis, S., *NMR : the toolkit : how pulse sequences work*, Second edition.; NMR : the toolkit : how pulse sequences work; Oxford University Press: Oxford, 2023.

(76) Heisenberg, W. *Zeitschrift für Physik* **1927**, *43*, 172–198, DOI: 10.1007/BF01397280.

(77) Mandelstam, L. *J. Phys.(USSR)* **1945**, *9*, 249.

(78) Hahn, E. L. *Physical Review* **1950**, *77*, 297–298, DOI: 10.1103/PhysRev.77.297.2.

(79) Bloch, F. *Physical Review* **1946**, *70*, 460–474, DOI: 10.1103/PhysRev.70.460.

(80) Ernst, R. R.; Anderson, W. A. *Review of Scientific Instruments* **1966**, *37*, 93–102, DOI: 10.1063/1.1719961.

(81) Bloembergen, N.; Purcell, E. M.; Pound, R. V. *Physical Review* **1948**, *73*, 679–712, DOI: 10.1103/PhysRev.73.679.

(82) Solomon, I. *Physical Review* **1955**, *99*, 559–565, DOI: 10.1103/PhysRev.99.559.

(83) Bloembergen, N.; Morgan, L. O. *The Journal of Chemical Physics* **1961**, *34*, 842–850, DOI: 10.1063/1.1731684.

(84) Wangsness, R. K.; Bloch, F. *Physical Review* **1953**, *89*, 728–739, DOI: 10.1103/PhysRev.89.728.

(85) Redfield, A. G. *IBM Journal of Research and Development* **1957**, *1*, 19–31, DOI: 10.1147/rd.11.0019.

(86) Redfield, A. G. In *Advances in Magnetic and Optical Resonance*, Waugh, J. S., Ed.; Academic Press: 1965; Vol. 1, pp 1–32.

(87) Keeler, J., *Understanding NMR spectroscopy*, 2nd; Understanding nuclear magnetic resonance spectroscopy; John Wiley and Sons: Chichester, U.K, 2010.

(88) Einstein, A. *Phys Zeit* **1917**, *18*, 121.

(89) Dirac, P. A. M.; Bohr, N. H. D. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1927**, *114*, 243–265, DOI: 10.1098/rspa.1927.0039.

(90) Woessner, D. E. *The Journal of chemical physics* **1962**, *37*, 647–654, DOI: 10.1063/1.1701390.

(91) Fano, U. *Reviews of Modern Physics* **1957**, *29*, 74–93, DOI: 10.1103/RevModPhys.29.74.

(92) Schrödinger, E. *Physical Review* **1926**, *28*, 1049–1070, DOI: 10.1103/PhysRev.28.1049.

(93) Pauli, W. *Zeitschrift für Physik* **1927**, *43*, 601–623, DOI: 10.1007/BF01397326.

(94) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skeleton, N. J., *Protein NMR Spectroscopy*; Elsevier: Cambridge, Massachusetts, 2007; Vol. Second Edition.

(95) Levitt, M. H., *Spin dynamics : basics of nuclear magnetic resonance*, Second edition; Spin dynamics : basics of nuclear magnetic resonance; John Wiley & Sons: Chichester, England, 2008.

(96) Haase, A.; Frahm, J.; Matthaei, D.; Hanicke, W.; Merboldt, K. D. *Journal of Magnetic Resonance (1969)* **1986**, *67*, 258–266, DOI: 10.1016/0022-2364(86)90433-6.

(97) Stott, K.; Stonehouse, J.; Keeler, J.; Hwang, T.-L.; Shaka, A. J. *Journal of the American Chemical Society* **1995**, *117*, 4199–4200, DOI: 10.1021/ja00119a048.

(98) Glunde, K.; Artemov, D.; Penet, M.-F.; Jacobs, M. A.; Bhujwalla, Z. M. *Chemical Reviews* **2010**, *110*, 3043–3059, DOI: 10.1021/cr9004007.

(99) Hahn, E. L.; Maxwell, D. E. *Physical Review* **1951**, *84*, 1246–1247, DOI: 10.1103/PhysRev.84.1246.

(100) Atkins, P. W.; De Paula, J.; Keeler, J., *Atkins' physical chemistry*, Twelfth edition / Peter Atkins, Julio de Paula, James Keeler.; Physical chemistry; Oxford University Press: Oxford, 2023.

(101) Morris, G. A.; Freeman, R. *Journal of the American Chemical Society* **1979**, *101*, 760–762, DOI: 10.1021/ja00497a058.

(102) Hartmann, S. R.; Hahn, E. L. *Physical Review* **1962**, *128*, 2042–2053, DOI: 10.1103/PhysRev.128.2042.

(103) Müller, L.; Ernst, R. R. *Molecular Physics* **1979**, *38*, 963–992, DOI: 10.1080/00268977900102161.

(104) Braunschweiler, L.; Ernst, R. *Journal of Magnetic Resonance (1969)* **1983**, *53*, 521–528, DOI: 10.1016/0022-2364(83)90226-3.

(105) Glaser, S. J.; Quant, J. J. In Advances in Magnetic and Optical Resonance, 1996, 59–254e.

(106) Ernst, M.; C., G.; R., E. R.; Bermel, W. *Molecular Physics* **1991**, *74*, 219–252, DOI: 10.1080/00268979100102191.

(107) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *The Journal of Chemical Physics* **1976**, *64*, 2229–2246, DOI: 10.1063/1.432450.

(108) Bodenhausen, G.; Ruben, D. J. *Chemical Physics Letters* **1980**, *69*, 185–189, DOI: 10.1016/0009-2614(80)80041-8.

(109) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *Journal of Magnetic Resonance (1969)* **1982**, *48*, 286–292, DOI: 10.1016/0022-2364(82)90279-7.

(110) Vold, R. L.; Waugh, J. S.; Klein, M. P.; Phelps, D. E. *The Journal of Chemical Physics* **1968**, *48*, 3831–3832, DOI: 10.1063/1.1669699.

(111) Hahn, E. L. *Physical Review* **1950**, *80*, 580–594, DOI: 10.1103/PhysRev.80.580.

(112) Sauerwein, A. C.; Hansen, D. F. In *Protein NMR: Modern Techniques and Biomedical Applications*; Springer: 2015, pp 75–132.

(113) Kingsley, P. B. *Concepts in Magnetic Resonance* **1995**, *7*, 115–136, DOI: 10.1002/cmr.1820070203.

(114) Forsén, S.; Hoffman, R. A. *The Journal of Chemical Physics* **1963**, *39*, 2892–2901, DOI: 10.1063/1.1734121.

(115) Overhauser, A. W. *Physical Review* **1953**, *92*, 411–415, DOI: 10.1103/PhysRev.92.411.

(116) Bracewell, R. N., *The Fourier transform and its applications*, 3rd; The Fourier transform and its applications; McGraw Hill: Boston ; 2000.

(117) Ward, K. M.; Aletras, A. H.; Balaban, R. S. *Journal of Magnetic Resonance* **2000**, *143*, 79–87, DOI: 10.1006/jmre.1999.1956.

(118) Vinogradov, E.; Sherry, A. D.; Lenkinski, R. E. *Journal of Magnetic Resonance* **2013**, *229*, 155–172, DOI: 10.1016/j.jmr.2012.11.024.

(119) Luz, Z.; Meiboom, S. *The Journal of Chemical Physics* **1963**, *39*, 366–370, DOI: 10.1063/1.1734254.

(120) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochemical and Biophysical Research Communications* **1980**, *95*, 1–6, DOI: 10.1016/0006-291X(80)90695-6.

(121) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 285–298, DOI: 10.1038/nrm.2017.7.

(122) Forman-Kay, J. D.; Ditlev, J. A.; Nosella, M. L.; Lee, H. O. *RNA* **2022**, *28*, 36–47, DOI: 10.1261/rna.079026.121.

(123) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. *Nature* **2012**, *483*, 336–340, DOI: 10.1038/nature10879.

(124) Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. *Nature Reviews Molecular Cell Biology* **2023**, *24*, 288–304, DOI: 10.1038/s41580-022-00558-8.

(125) Pessina, F.; Giavazzi, F.; Yin, Y.; Gioia, U.; Vitelli, V.; Galbiati, A.; Barozzi, S.; Garre, M.; Oldani, A.; Flaus, A.; Cerbino, R.; Parazzoli, D.; Rothenberg, E.; d'Adda di Fagagna, F. *Nature Cell Biology* **2019**, *21*, 1286–1299, DOI: 10.1038/s41556-019-0392-4.

(126) Oshidari, R.; Huang, R.; Medghalchi, M.; Tse, E. Y. W.; Ashgriz, N.; Lee, H. O.; Wyatt, H.; Mekhail, K. *Nature Communications* **2020**, *11*, 695, DOI: 10.1038/s41467-020-14546-z.

(127) Charman, M.; Grams, N.; Kumar, N.; Halko, E.; Dybas, J. M.; Abbott, A.; Lum, K. K.; Blumenthal, D.; Tsopurashvili, E.; Weitzman, M. D. *Nature* **2023**, *616*, 332–338, DOI: 10.1038/s41586-023-05887-y.

(128) Du, M.; Chen, Z. J. *Science* **2018**, *361*, 704–709, DOI: 10.1126/science.aat1022.

(129) Wen, Y.; Ma, J. *Front Immunol* **2022**, *13*, 986589, DOI: 10.3389/fimmu.2022.986589.

(130) Fontana, F., *Traité sur le vénin de la vipere, sur les poisons americains, sur le laurier-cerise et sur quelques autres poisons vegetaux*; Nyon l'Ainé: 1781; Vol. 2.

(131) Hughes, A., *The mitotic cycle. The cytoplasm and nucleus during interphase and mitosis*; Butterworths Scientific Publications: London, 1952.

(132) Wagner, R. *Müller's Archiv Anat Physiol Wissenschaft Med* **1835**, *268*, 373–7.

(133) Valentin, G., *Repertorium für anatomie und physiologie*; Verlag von Veit und Comp: Berlin, 1836; Vol. 1.

(134) Valentin, G., *Repertorium für anatomie und physiologie*; Verlag von Veit und Comp: Berlin, 1839; Vol. 4.

(135) Kölliker, A. v. *Zeitschr f wissensch Zool* **1857**, *8*, 311–325.

(136) Golgi, C. *Bollettino della Società Medico-Chirurgica di Pavia* **1898**, *13:1–14*.

(137) Montgomery Jr, T. H., *Comparative cytological studies, with especial regard to the morphology of the nucleolus*; Journal of morphology, Vol. 15; Ginn & Company: Boston, 1898.

(138) Bütschli, O.; Minchin, E. A., *Investigations on microscopic foams and on protoplasm. Experiments & observations directed towards a solution of the question of the physical conditions of the phenomena of life*; A. and C. Black: London, 1894.

(139) Wilson, E. B. *Science* **1899**, *10*, 33–45, DOI: 10.1126/science.10.237.33.

(140) Hardy, W. B. *The Journal of physiology* **1899**, *24*, 158–210, DOI: 10.1113/jphysiol.1899.sp000755.

(141) Cajal, R. *Trabajos del lab. de investig. biolog. de la Univ. de Madrid* **1903**, *2*, 129–129.

(142) Oparin, A. I., *The Origin of Life*; The Macmillan Company: New York, 1938.

(143) Oparin, A. I., *The Origin of Life*, 2nd; Dover Publications, Inc: New York, 1953.

(144) Darling, A. L.; Uversky, V. N. In *Droplets of Life*, Uversky, V. N., Ed.; Academic Press: 2023, pp 271–335.

(145) Platani, M.; Goldberg, I.; Swedlow, J. R.; Lamond, A. I. *Journal of Cell Biology* **2000**, *151*, 1561–1574, DOI: 10.1083/jcb.151.7.1561.

(146) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. *Science* **2009**, *324*, 1729–1732, DOI: 10.1126/science.1172046.

(147) Keating, C. D. *Accounts of Chemical Research* **2012**, *45*, 2114–2124, DOI: 10.1021/ar200294y.

(148) Weidtkamp-Peters, S.; Lenser, T.; Negorev, D.; Gerstner, N.; Hofmann, T. G.; Schwanitz, G.; Hoischen, C.; Maul, G.; Dittrich, P.; Hemmerich, P. *Journal of Cell Science* **2008**, *121*, 2731–2743, DOI: 10.1242/jcs.031922.

(149) Spector, D. L.; Lamond, A. I. *Cold Spring Harbor perspectives in biology* **2011**, *3*, a000646, DOI: 10.1101/cshperspect.a000646.

(150) Rostam, N.; Ghosh, S.; Chow, C. F. W.; Hadarovich, A.; Landerer, C.; Ghosh, R.; Moon, H.; Hersemann, L.; Mitrea, D. M.; Klein, I. A.; Hyman, A. A.; Toth-Petroczy, A. *Nature Methods* **2023**, *20*, 673–676, DOI: 10.1038/s41592-023-01831-0.

(151) Ladouceur, A.-M.; Parmar, B. S.; Biedzinski, S.; Wall, J.; Tope, S. G.; Cohn, D.; Kim, A.; Soubry, N.; Reyes-Lamothe, R.; Weber, S. C. *Proceedings of the National Academy of Sciences* **2020**, *117*, 18540–18549, DOI: 10.1073/pnas.2005019117.

(152) Vernon, R. M.; Forman-Kay, J. D. *Current Opinion in Structural Biology* **2019**, *58*, 88–96, DOI: 10.1016/j.sbi.2019.05.016.

(153) De Jong, H. G. B. *Protoplasma* **1932**, *15*, 110–173, DOI: 10.1007/BF01610198.

(154) Dubois, M.-L.; Boisvert, F.-M. In *The Functional Nucleus*, Bazett-Jones, D. P., Dellaire, G., Eds.; Springer International Publishing: Cham, 2016, pp 29–49.

(155) King, M. R.; Ruff, K. M.; Lin, A. Z.; Pant, A.; Farag, M.; Lalmansingh, J. M.; Wu, T.; Fossat, M. J.; Ouyang, W.; Lew, M. D.; Lundberg, E.; Vahey, M. D.; Pappu, R. V. *Cell* **2024**, *187*, 1889–1906.e24, DOI: 10.1016/j.cell.2024.02.029.

(156) Beutel, O.; Maraspini, R.; Pombo-García, K.; Martin-Lemaitre, C.; Honigmann, A. *Cell* **2019**, *179*, 923–936.e11, DOI: 10.1016/j.cell.2019.10.011.

(157) Pombo-García, K.; Adame-Arana, O.; Martin-Lemaitre, C.; Jülicher, F.; Honigmann, A. *Nature* **2024**, *632*, 647–655, DOI: 10.1038/s41586-024-07726-0.

(158) Petralia, R. S.; Sans, N.; Wang, Y.-X.; Wenthold, R. J. *Molecular and Cellular Neuroscience* **2005**, *29*, 436–452, DOI: 10.1016/j.mcn.2005.03.013.

(159) Böckers, T. M. *Cell and tissue research* **2006**, *326*, 409–422, DOI: 10.1007/s00441-006-0274-5.

(160) Zeng, M.; Shang, Y.; Araki, Y.; Guo, T.; Huganir, R. L.; Zhang, M. *Cell* **2016**, *166*, 1163–1175. e12, DOI: 10.1016/j.cell.2016.07.008.

(161) Śliwińska, M. A.; Cały, A.; Borczyk, M.; Ziółkowska, M.; Skonieczna, E.; Chilimoniuk, M.; Bernaś, T.; Giese, K. P.; Radwanska, K. *Cerebral Cortex* **2020**, *30*, 2573–2585, DOI: 10.1093/cercor/bhz261.

(162) Patel, A. et al. *Cell* **2015**, *162*, 1066–1077, DOI: 10.1016/j.cell.2015.07.047.

(163) Shen, Y.; Ruggeri, F. S.; Vigolo, D.; Kamada, A.; Qamar, S.; Levin, A.; Iserman, C.; Alberti, S.; George-Hyslop, P. S.; Knowles, T. P. J. *Nature Nanotechnology* **2020**, *15*, 841–847, DOI: 10.1038/s41565-020-0731-4.

(164) Ray, S. et al. *Nature Chemistry* **2020**, *12*, 705–716, DOI: 10.1038/s41557-020-0465-9.

(165) Wegmann, S. et al. *The EMBO Journal* **2018**, *37*, e98049, DOI: 10.15252/embj.201798049.

(166) Kanaan, N. M.; Hamel, C.; Grabinski, T.; Combs, B. *Nature Communications* **2020**, *11*, 2809, DOI: 10.1038/s41467-020-16580-3.

(167) Garcia-Cabau, C. et al. *bioRxiv* **2024**, 2023.03.19.532587, DOI: 10.1101/2023.03.19.532587.

(168) Boija, A.; Klein, I. A.; Young, R. A. *Cancer Cell* **2021**, *39*, 174–192, DOI: 10.1016/j.ccell.2020.12.003.

(169) Tong, X.; Tang, R.; Xu, J.; Wang, W.; Zhao, Y.; Yu, X.; Shi, S. *Signal Transduction and Targeted Therapy* **2022**, *7*, 221, DOI: 10.1038/s41392-022-01076-x.

(170) Benedetti, R.; Altucci, L. *Nature Cancer* **2024**, *5*, 1602–1604, DOI: 10.1038/s43018-024-00843-9.

(171) Klein, I. A. et al. *Science* **2020**, *368*, 1386–1392, DOI: 10.1126/science.aaz4427.

(172) Leppert, A.; Feng, J.; Railaite, V.; Bohn Pessatti, T.; Cerrato, C. P.; Mörman, C.; Osterholz, H.; Lane, D. P.; Maia, F. R. N. C.; Linder, M. B.; Rising, A.; Landreh, M. *Journal of the American Chemical Society* **2024**, *146*, DOI: 10.1021/JACS.4C06688.

(173) Lin, Y.-H.; Kim, T. H.; Das, S.; Pal, T.; Wessén, J.; Rangadurai, A. K.; Kay, L. E.; Forman-Kay, J. D.; Chan, H. S. Electrostatics of Salt-Dependent Reentrant Phase Behaviors Highlights Diverse Roles of ATP in Biomolecular Condensates, 2024.

(174) Posey, A. E.; Bremer, A.; Erkamp, N. A.; Pant, A.; Knowles, T. P. J.; Dai, Y.; Mittag, T.; Pappu, R. V. *Journal of the American Chemical Society* **2024**, *146*, 28268–28281, DOI: 10.1021/jacs.4c08946.

(175) Dai, Y.; Zhou, Z.; Yu, W.; Ma, Y.; Kim, K.; Rivera, N.; Mohammed, J.; Lantelme, E.; Hsu-Kim, H.; Chilkoti, A.; You, L. *Cell* **2024**, *187*, 5951–5966.e18, DOI: 10.1016/j.cell.2024.08.018.

(176) Schaffert, N.; Hossbach, M.; Heintzmann, R.; Achsel, T.; Lührmann, R. *EMBO Journal* **2004**, *23*, 3000–3009, DOI: 10.1038/SJ.EMBOJ.7600296/ASSET/B14D490E-609C-466C-8A7B-592073E3B4B8/ASSETS/GRAPHIC/EMBJ7600296-FIG-0005-M.JPG.

(177) Nott, T. J.; Craggs, T. D.; Baldwin, A. J. *Nature Chemistry* **2016**, *8*, 569–575, DOI: 10.1038/nchem.2519.

(178) Jain, S.; Wheeler, J. R.; Walters, R. W.; Agrawal, A.; Barsic, A.; Parker, R. *Cell* **2016**, *164*, 487–498, DOI: 10.1016/J.CELL.2015.12.038.

(179) Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. *Science* **2016**, *352*, 595–599, DOI: 10.1126/SCIENCE.AAD9964/SUPPL_FILE/SU-SM.PDF.

(180) Woodruff, J. B.; Ferreira Gomes, B.; Widlund, O.; Mahamid, J.; Honigmann, A.; Hyman, A. A. *Cell* **2017**, *169*, DOI: 10.1016/j.cell.2017.05.028.

(181) Lu, Y.; Wu, T.; Gutman, O.; Lu, H.; Zhou, Q.; Henis, Y. I.; Luo, K. *Nature Cell Biology 2020 22:4* **2020**, *22*, 453–464, DOI: 10.1038/s41556-020-0485-0.

(182) Xing, W.; Muhlrad, D.; Parker, R.; Rosen, M. K. *eLife* **2020**, *9*, 1–63, DOI: 10.7554/ELIFE.56525.

(183) Langdon, E. M.; Qiu, Y.; Niaki, A. G.; McLaughlin, G. A.; Weidmann, C. A.; Gerbich, T. M.; Smith, J. A.; Crutchley, J. M.; Termini, C. M.; Weeks, K. M.; Myong, S.; Gladfelter, A. S. *Science* **2018**, *360*, 922–927, DOI: 10.1126/SCIENCE.AAR7432/SUPPL_FILE/PAPV2.PDF.

(184) Hickling, M. The selective partitioning of mRNA in biomolecular condensates, Ph.D. Thesis, 2022.

(185) Dumelie, J. G.; Chen, Q.; Miller, D.; Attarwala, N.; Gross, S. S.; Jaffrey, S. R. *Nature Chemical Biology 2023 20:3* **2023**, *20*, 302–313, DOI: 10.1038/s41589-023-01474-4.

(186) Testa, A.; Dindo, M.; Rebane, A. A.; Nasouri, B.; Style, R. W.; Golestanian, R.; Dufresne, E. R.; Laurino, P. *Nature Communications* **2021**, *12*, 6293, DOI: 10.1038/s41467-021-26532-0.

(187) Peeples, W.; Rosen, M. K. *Nature Chemical Biology* **2021**, *17*, 693–702, DOI: 10.1038/s41589-021-00801-x.

(188) Gil-Garcia, M.; Benítez-Mateos, A. I.; Papp, M.; Stoffel, F.; Morelli, C.; Normak, K.; Makasewicz, K.; Faltova, L.; Paradisi, F.; Arosio, P. *Nature Communications* **2024**, *15*, 3322, DOI: 10.1038/s41467-024-47435-w.

(189) Jackson, M.; Mantsch, H. H. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1991**, *1078*, 231–235, DOI: 10.1016/0167-4838(91)90563-F.

(190) Camino, J. D.; Gracia, P.; Cremades, N. *Biophysical Chemistry* **2021**, *269*, 106520–106520, DOI: 10.1016/J.BPC.2020.106520.

(191) Ausserwöger, H. et al. *bioRxiv* **2024**, 2024.05.23.595321, DOI: 10.1101/2024.05.23.595321.

(192) Shapiro, D. M.; Ney, M.; Eghtesadi, S. A.; Chilkoti, A. *The Journal of Physical Chemistry B* **2021**, *125*, 6740–6759, DOI: 10.1021/acs.jpcb.1c01146.

(193) Jenkins, I. C.; Milligan, J. J.; Chilkoti, A. *Advanced Healthcare Materials* **2021**, *10*, 2100209, DOI: 10.1002/adhm.202100209.

(194) Gimenez, G.; Marin, E.; Zanon, A.; Lapeyre, V.; Douliez, J.-P.; Ravaine, V.; Perro, A. *Langmuir* **2024**, *40*, 24854–24862, DOI: 10.1021/acs.langmuir.4c02738.

(195) Graham, T. *Philosophical Transactions of the Royal Society of London* **1861**, *151*, 183–224, DOI: 10.1098/rstl.1861.0011.

(196) Kossel, A. *Biological Chemistry* **1897**, *22*, 176–187, DOI: 10.1515/bchm2.1897.22.2.176.

(197) Tiebackx, F. W. *Zeitschrift für Chemie und Industrie der Kolloide* **1911**, *8*, 198–201, DOI: 10.1007/BF01503532.

(198) De Jong, H. G. B.; Kruyt, H. R. *Kolloid-Zeitschrift* **1930**, *50*, 39–48, DOI: 10.1007/BF01422833.

(199) Ehrenberg, L. *Hereditas* **1946**, *32*, 407–418, DOI: 10.1111/j.1601-5223.1946.tb02783.x.

(200) Frey-Wyssling, A., *Submicroscopic morphology of protoplasm*; Elsevier Pub. Co: Amsterdam, 1953.

(201) Flory, P. J. *The Journal of Chemical Physics* **1942**, *10*, 51–61, DOI: 10.1063/1.1723621.

(202) Huggins, M. L. *The Journal of Physical Chemistry* **1942**, *46*, 151–158, DOI: 10.1021/j150415a018.

(203) Qian, D.; Michaels, T. C. T.; Knowles, T. P. J. *The Journal of Physical Chemistry Letters* **2022**, *13*, 7853–7860, DOI: 10.1021/acs.jpclett.2c01986.

(204) Fritsch, A. W.; Diaz-Delgadillo, A. F.; Adame-Arana, O.; Hoege, C.; Mittasch, M.; Kreysing, M.; Leaver, M.; Hyman, A. A.; Jülicher, F.; Weber, C. A. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2102772118, DOI: 10.1073/pnas.2102772118.

(205) Li, S.; Yoshizawa, T.; Yamazaki, R.; Fujiwara, A.; Kameda, T.; Kitahara, R. *The Journal of Physical Chemistry B* **2021**, *125*, 6821–6829, DOI: 10.1021/acs.jpcb.1c01451.

(206) Qian, D.; Ausserwoger, H.; Sneideris, T.; Farag, M.; Pappu, R. V.; Knowles, T. P. J. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2407453121, DOI: 10.1073/pnas.2407453121.

(207) Arter, W. E. et al. *Nature Communications* **2022**, *13*, 7845, DOI: 10.1038/s41467-022-35265-7.

(208) Mao, S.; Kuldinow, D.; Haataja, M. P.; Košmrlj, A. *Soft Matter* **2019**, *15*, 1297–1311, DOI: 10.1039/C8SM02045K.

(209) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. *Nature Physics* **2015**, *11*, 899–904, DOI: 10.1038/nphys3532.

(210) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. *Cell* **2018**, *174*, 688–699.e16, DOI: 10.1016/j.cell.2018.06.006.

(211) Posey, A. E.; Holehouse, A. S.; Pappu, R. V. In *Methods in Enzymology*, Rhoades, E., Ed.; Academic Press: 2018; Vol. 611, pp 1–30.

(212) Holehouse, A. S. In *Intrinsically Disordered Proteins*, Salvi, N., Ed.; Academic Press: 2019, pp 209–255.

(213) Martin, E. W.; Holehouse, A. S. *Emerg Top Life Sci* **2020**, *4*, 307–329, DOI: 10.1042/etls20190164.

(214) Yamaguchi, A.; Kitajo, K. *PLOS ONE* **2012**, *7*, e49267, DOI: 10.1371/journal.pone.0049267.

(215) Hearst, S. M.; Gilder, A. S.; Negi, S. S.; Davis, M. D.; George, E. M.; Whittom, A. A.; Toyota, C. G.; Husedzinovic, A.; Gruss, O. J.; Hebert, M. D. *Journal of Cell Science* **2009**, *122*, 1872–1881, DOI: 10.1242/jcs.044040.

(216) Li, X.; Pinou, l.; Du, Y.; Chen, X.; Liu, C. *Current Opinion in Chemical Biology* **2023**, *75*, 102314, DOI: 10.1016/j.cbpa.2023.102314.

(217) Orlando, G.; Raimondi, D.; Tabaro, F.; Codicè, F.; Moreau, Y.; Vranken, W. F. *Bioinformatics* **2019**, *35*, 4617–4623, DOI: 10.1093/bioinformatics/btz274.

(218) Saar, K. L.; Morgunov, A. S.; Qi, R.; Arter, W. E.; Krainer, G.; Lee, A. A.; Knowles, T. P. J. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2019053118, DOI: 10.1073/pnas.2019053118.

(219) Cai, H.; Vernon, R. M.; Forman-Kay, J. D. *Biomolecules* **2022**, *12*, 1131, DOI: 10.3390/biom12081131.

(220) Chu, X.; Sun, T.; Li, Q.; Xu, Y.; Zhang, Z.; Lai, L.; Pei, J. *BMC Bioinformatics* **2022**, *23*, 72, DOI: 10.1186/s12859-022-04599-w.

(221) Hatos, A.; Tosatto, S. C. E.; Vendruscolo, M.; Fuxreiter, M. *Nucleic Acids Research* **2022**, *50*, W337–W344, DOI: 10.1093/nar/gkac386.

(222) Liang, Q.; Peng, N.; Xie, Y.; Kumar, N.; Gao, W.; Miao, Y. *The EMBO Journal* **2024**, *43*, 1898-1918–1918, DOI: 10.1038/s44318-024-00090-9.

(223) Hadarovich, A.; Singh, H. R.; Ghosh, S.; Scheremetjew, M.; Rostam, N.; Hyman, A. A.; Toth-Petroczy, A. *Nature Communications* **2024**, *15*, 10668, DOI: 10.1038/s41467-024-55089-x.

(224) Pantoja, C. F.; Ibáñez de Opakua, A.; Cima-Omori, M.-S.; Zweckstetter, M. *Angewandte Chemie International Edition* **2023**, *62*, e202218078, DOI: 10.1002/anie.202218078.

(225) André, A. A. M.; Yewdall, N. A.; Spruijt, E. *Biophysical Journal* **2023**, *122*, 397–407, DOI: 10.1016/j.bpj.2022.12.001.

(226) Xu, C.; Cao, Y.; Bao, J. *Cellular and Molecular Life Sciences* **2021**, *79*, 4, DOI: 10.1007/s00018-021-04069-1.

(227) Dobrynin, M. A.; Bashendjieva, E. O.; Enukashvily, N. I. *Journal of Developmental Biology* **2022**, *10*, 43, DOI: 10.3390/jdb10040043.

(228) Thomas, L.; Putnam, A.; Folkmann, A. *Development* **2023**, *150*, DOI: 10.1242/dev.201037.

(229) Kotaja, N.; Sassone-Corsi, P. *Nature Reviews Molecular Cell Biology* **2007**, *8*, 85–90, DOI: 10.1038/nrm2081.

(230) Meikar, O.; Da Ros, M.; Korhonen, H.; Kotaja, N. *REPRODUCTION* **2011**, *142*, 195–209, DOI: 10.1530/rep-11-0057.

(231) Updike, D. L.; Knutson, A. K.; Egelhofer, T. A.; Campbell, A. C.; Strome, S. *Current Biology* **2014**, *24*, 970–975, DOI: 10.1016/j.cub.2014.03.015.

(232) Kotaja, N.; Bhattacharyya, S. N.; Jaskiewicz, L.; Kimmins, S.; Parvinen, M.; Filipowicz, W.; Sassone-Corsi, P. *Proceedings of the National Academy of Sciences* **2006**, *103*, 2647–2652, DOI: 10.1073/pnas.0509333103.

(233) Liang, L.; Diehl-Jones, W.; Lasko, P. *Development* **1994**, *120*, 1201–1211, DOI: 10.1242/dev.120.5.1201.

(234) Tanaka, S. S.; Toyooka, Y.; Akasu, R.; Katoh-Fukui, Y.; Nakahara, Y.; Suzuki, R.; Yokoyama, M.; Noce, T. *Genes Dev* **2000**, *14*, 841–53, DOI: 10.1101/gad.14.7.841.

(235) Schudrowitz, N.; Takagi, S.; Wessel, G. M.; Yajima, M. *Cancer Science* **2017**, *108*, 1612–1619, DOI: 10.1111/cas.13299.

(236) Olotu, O. et al. *Cell Reports* **2024**, *43*, 114430, DOI: 10.1016/j.celrep.2024.114430.

(237) Noyes, C. et al. *Communications Biology* **2023**, *6*, 65, DOI: 10.1038/s42003-023-04444-7.

(238) Varadi, M. et al. *Nucleic Acids Research* **2023**, *52*, D368–D375, DOI: 10.1093/nar/gkad1011.

(239) The Scripps Research, I. Protein Calculator.

(240) Crabtree, M. D.; Holland, J.; Pillai, A. S.; Kompella, P. S.; Babl, L.; Turner, N. N.; Eaton, J. T.; Hochberg, G. K. A.; Aarts, D. G. A. L.; Redfield, C.; Baldwin, A. J.; Nott, T. J. *Cell Reports* **2023**, *42*, 113375, DOI: 10.1016/j.celrep.2023.113375.

(241) Bielskutė, S.; Garcia-Cabau, C.; Frigolé-Vivas, M.; Szulc, E.; De Mol, E.; Pesarrodona, M.; García, J.; Salvatella, X. *Protein Science* **2021**, *30*, 1427–1437, DOI: 10.1002/pro.4110.

(242) Scheiner, S.; Čuma, M. *Journal of the American Chemical Society* **1996**, *118*, 1511–1521, DOI: 10.1021/ja9530376.

(243) O'Neil, M. J., *The Merck index: an encyclopedia of chemicals, drugs, and biologicals*; RSC Publishing: 2013.

(244) Chase, M. W. *Journal of Physical and Chemical Reference Data* **1998**, *9*, 1–1951.

(245) Hoell, J. I.; Larsson, E.; Runge, S.; Nusbaum, J. D.; Duggimpudi, S.; Farazi, T. A.; Hafner, M.; Borkhardt, A.; Sander, C.; Tuschl, T. *Nature Structural & Molecular Biology* **2011**, *18*, 1428–1431, DOI: 10.1038/nsmb.2163.

(246) Schwartz, J. C.; Cech, T. R.; Parker, R. R. *Annu Rev Biochem* **2015**, *84*, 355–79, DOI: 10.1146/annurev-biochem-060614-034325.

(247) Lee, J.; Nguyen, P. T.; Shim, H. S.; Hyeon, S. J.; Im, H.; Choi, M.-H.; Chung, S.; Kowall, N. W.; Lee, S. B.; Ryu, H. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2019**, *1865*, 1938–1945, DOI: 10.1016/j.bbadis.2018.10.042.

(248) Luo, Y.; Blechingberg, J.; Fernandes, A. M.; Li, S.; Fryland, T.; Børglum, A. D.; Bolund, L.; Nielsen, A. L. *BMC Genomics* **2015**, *16*, 929, DOI: 10.1186/s12864-015-2125-9.

(249) Park, J. H.; Kang, H.-J.; Lee, Y. K.; Kang, H.; Kim, J.; Chung, J. H.; Chang, J. S.; McPherron, A. C.; Lee, S. B. *Proceedings of the National Academy of Sciences* **2015**, *112*, 6074–6079, DOI: 10.1073/pnas.1504391112.

(250) Yoon, Y.; Park, H.; Kim, S.; Nguyen, P. T.; Hyeon, S. J.; Chung, S.; Im, H.; Lee, J.; Lee, S. B.; Ryu, H. *Experimental Neurobiology* **2018**, *27*, 103–111, DOI: 10.5607/en.2018.27.2.103.

(251) Azuma, M.; Embree, L. J.; Sabaawy, H.; Hickstein, D. D. *PLOS ONE* **2007**, *2*, e979, DOI: 10.1371/journal.pone.0000979.

(252) Li, H.; Watford, W.; Li, C.; Parmelee, A.; Bryant, M. A.; Deng, C.; O'Shea, J.; Lee, S. B. *J Clin Invest* **2007**, *117*, 1314–23, DOI: 10.1172/jci31222.

(253) Ewing, J. *Proc New York Path* **1921**, *21*, 17–24.

(254) Embree, L. J.; Azuma, M.; Hickstein, D. D. *Cancer Research* **2009**, *69*, 4363–4371, DOI: 10.1158/0008-5472.Can-08-3229.

(255) Flucke, U.; van Noesel, M. M.; Siozopoulou, V.; Creytens, D.; Tops, B. B. J.; van Gorp, J. M.; Hiemcke-Jiwa, L. S. *Diagnostics* **2021**, *11*, 1093, DOI: 10.3390/diagnostics11061093.

(256) Wei, H.; Yu, X. *Genomics, Proteomics & Bioinformatics* **2016**, *14*, 131–139, DOI: 10.1016/j.gpb.2016.05.001.

(257) Altmeyer, M.; Neelsen, K. J.; Teloni, F.; Pozdnyakova, I.; Pellegrino, S.; Grøfte, M.; Rask, M.-B. D.; Streicher, W.; Jungmichel, S.; Nielsen, M. L.; Lukas, J. *Nature Communications* **2015**, *6*, 8088, DOI: 10 . 1038 / ncomms9088.

(258) Couthouis, J. et al. *Human Molecular Genetics* **2012**, *21*, 2899–2911, DOI: 10.1093/hmg/dds116.

(259) Blechingberg, J.; Luo, Y.; Bolund, L.; Damgaard, C. K.; Nielsen, A. L. *PLOS ONE* **2012**, *7*, e46251, DOI: 10.1371/journal.pone.0046251.

(260) Alberts, B.; Heald, R.; Johnson, A.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P.; Wilson, J.; Hunt, T., *Molecular Biology of the Cell (Seventh Edition)*, 7th; Molecular Biology of the Cell (Seventh Edition); W. W. Norton & Company: 2022.

(261) Passmore, L. A.; Coller, J. *Nature Reviews Molecular Cell Biology* **2022**, *23*, 93–106, DOI: 10.1038/s41580-021-00417-y.

(262) Houseley, J.; Tollervey, D. *Cell* **2009**, *136*, 763–776, DOI: 10.1016/j.cell.2009.01.019.

(263) Eaton, J. T.; Casablancas-Antràs, V.; Naylor, B.; Buchanan, C. J.; Haasz, E.; Bercovici, J.; Spencer, A.-C.; Baldwin, A. J. Metabolite partitioning into biomolecular condensates mimics organic solvent behavior (in preparation).

(264) Chemat, F.; Vian, M. A.; Cravotto, G. *International Journal of Molecular Sciences* **2012**, *13*, 8615–8627, DOI: 10.3390/ijms13078615.

(265) Rice, N. M.; Irving, H. M. N. H.; Leonard, M. A. *Pure and Applied Chemistry* **1993**, *65*, 2373–2396, DOI: 10 . 1351 / PAC199365112373 / MACHINEREADABLECITATION/RIS.

(266) Berthelot, M.; Jungfleisch, E. *Ann Chim Phys.* **1872**, *26*, 396–407.

(267) Nernst, W. *Zeitschrift für Physikalische Chemie* **1891**, *8U*, 110–139, DOI: 10.1515/ZPCH-1891-0806.

(268) Leo, A.; Hansch, C.; Elkins, D. *Chemical Reviews* **1971**, *71*, 525–616, DOI: 10.1021/cr60274a001.

(269) Sangster, J. *GDF Databanks Bulletin* **1997**, *1*.

(270) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced Drug Delivery Reviews* **1997**, *23*, 3–25, DOI: 10.1016/S0169-409X(96)00423-1.

(271) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. *Journal of Computer-Aided Molecular Design* **2005**, *19*, 453–463, DOI: 10.1007/s10822-005-8694-y.

(272) Albertsson, P. Å. *Biochemical Pharmacology* **1961**, *5*, 351–358, DOI: 10 . 1016/0006-2952(61)90028-4.

(273) Madeira, P. P.; Reis, C. A.; Rodrigues, A. E.; Mikheeva, L. M.; Zaslavsky, B. Y. *The Journal of Physical Chemistry B* **2010**, *114*, 457–462, DOI: 10.1021/jp907346s.

(274) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. *Nature Chemical Biology* **2009**, *5*, 593–599, DOI: 10.1038/nchembio.186.

(275) Co-operation, O. f. E.; Development, *Test No. 107: Partition Coefficient (n-octanol/water): Shake Flask Method*; OECD Publishing: 1995.

(276) Dearden, J. C.; Bresnen, G. M. *Quantitative Structure-Activity Relationships* **1988**, *7*, 133–144, DOI: 10.1002/qsar.19880070304.

(277) Smokers, I. B. A.; Lavagna, E.; Freire, R. V. M.; Paloni, M.; Voets, I. K.; Barducci, A.; White, P. B.; Khajehpour, M.; Spruijt, E. *bioRxiv* **2024**, 2024.12.24.630169, DOI: 10.1101/2024.12.24.630169.

(278) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *Journal of Magnetic Resonance (1969)* **1990**, *89*, 496–514, DOI: 10.1016/0022-2364(90)90333-5.

(279) Buchanan, C.; Bhole, G.; Karunanithy, G.; Casablancas-Antràs, V.; Poh, A.; Davis, B. G.; Jones, J. A.; Baldwin, A. J. *bioRxiv* **2024**, 2024.01.31.578133, DOI: 10.1101/2024.01.31.578133.

(280) Lane, D.; Soong, R.; Bermel, W.; Ning, P.; Dutta Majumdar, R.; Tabatabaei-Anaraki, M.; Heumann, H.; Gundy, M.; Bönisch, H.; Liaghati Mobarhan, Y.; Simpson, M. J.; Simpson, A. J. *ACS Omega* **2019**, *4*, 9017–9028, DOI: 10.1021/acsomega.9b00931.

(281) Kay, L. E.; Xu, G. Y.; Singer, A. U.; Muhandiram, D. R.; Formankay, J. D. *Journal of Magnetic Resonance, Series B* **1993**, *101*, 333–337, DOI: 10.1006/JMRB.1993.1053.

(282) Wishart, D. S. et al. *Nucleic Acids Research* **2021**, *50*, D622–D631, DOI: 10.1093/nar/gkab1062.

(283) Kier, L. B.; Hall, L. H., *Molecular structure description : the electrotopological state*; Molecular structure description : the electrotopological state; Academic Press: San Diego ; 1999.

(284) Huuskonen, J. J.; Villa, A. E. P.; Tetko, I. V. *Journal of Pharmaceutical Sciences* **1999**, *88*, 229–233, DOI: 10.1021/js980266s.

(285) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *Journal of Medicinal Chemistry* **2002**, *45*, 2615–2623, DOI: 10.1021/jm020017n.

(286) Press, W. H., *Numerical recipes 3rd edition: The art of scientific computing*; Cambridge university press: 2007.

(287) Chatterjee, S.; Simonoff, J. S., *Handbook of Regression Analysis*, 1. Aufl.; Handbook of Regression Analysis, Vol. 5; Wiley: Newark, 2013.

(288) Li, L.; Srivastava, S.; Andreev, M.; Marciel, A. B.; de Pablo, J. J.; Tirrell, M. V. *Macromolecules* **2018**, *51*, 2988–2995, DOI: 10.1021/acs.macromol.8b00238.

(289) Mintis, D. G.; Mavrantzas, V. G. *Macromolecules* **2020**, *53*, 7618–7634, DOI: 10.1021/acs.macromol.0c00728.

(290) Wagner, C. A. *Nephrology Dialysis Transplantation* **2023**, *39*, 190–201, DOI: 10.1093/ndt/gfad188.

(291) Yu, W.; Guo, X.; Xia, Y.; Ma, Y.; Tong, Z.; Yang, L.; Song, X.; Zare, R. N.; Hong, G.; Dai, Y. *Nature Chemistry* **2025**, DOI: 10.1038/s41557-025-01762-7.

(292) Kim, N.; Yun, H.; Lee, H.; Yoo, J.-Y. *Experimental & Molecular Medicine* **2024**, *56*, 2357–2364, DOI: 10.1038/s12276-024-01337-5.

(293) Naylor, B. Investigating the partitioning of ions within membraneless organelles, and the phase separation theory underlying their behaviour, Ph.D. Thesis, 2024.

(294) Holland, J.; Nott, T. J.; Aarts, D. G. A. L. *The Journal of Chemical Physics* **2025**, *162*, DOI: 10.1063/5.0253522.

(295) Rosing, J.; Slater, E. C. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1972**, *267*, 275–290, DOI: 10.1016/0005-2728(72)90116-8.

(296) Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. *Nature* **2023**, *619*, 876–883, DOI: 10.1038/s41586-023-06329-5.

(297) Guo, X.; Farag, M.; Qian, N.; Yu, X.; Ni, A.; Ma, Y.; Yu, W.; King, M. R.; Liu, V.; Lee, J.; Zare, R. N.; Min, W.; Pappu, R. V.; Dai, Y. *bioRxiv* **2024**, 2024.07.06.602359, DOI: 10.1101/2024.07.06.602359.

(298) Yan, X. et al. *bioRxiv* **2024**, 2024.01.23.576837, DOI: 10.1101/2024.01.23.576837.

(299) Khaneja, N.; Reiss, T.; Kehlet, C.; Schulte-Herbrüggen, T.; Glaser, S. J. *Journal of Magnetic Resonance* **2005**, *172*, 296–305, DOI: 10.1016/j.jmr.2004.11.004.

(300) Webb, A. *Concepts in Magnetic Resonance Part A* **2011**, *38A*, 148–184, DOI: 10.1002/cmr.a.20219.

(301) Gavish, N.; Promislow, K. *Physical Review E* **2016**, *94*, 012611, DOI: 10.1103/PhysRevE.94.012611.

(302) Piotto, M.; Saudek, V.; Sklenář, V. *Journal of Biomolecular NMR* **1992**, *2*, 661–665, DOI: 10.1007/BF02192855.

(303) Jin, T.; Wang, P.; Zong, X.; Kim, S.-G. *Magnetic Resonance in Medicine* **2013**, *69*, 760–770, DOI: 10.1002/mrm.24315.

(304) Stokes, G. G. **1851**.

(305) Einstein, A. *Annalen der Physik* **1905**, *322*, 549–560, DOI: 10.1002/andp.19053220806.

(306) Sutherland, W. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1905**, *9*, 781–785, DOI: 10.1080/14786440509463331.

(307) Michalski, J.; Kalwarczyk, T.; Kwapiszewska, K.; Enderlein, J.; Poniewierski, A.; Karpińska, A.; Kucharska, K.; Hołyst, R. *Soft Matter* **2024**, *20*, 5810–5821, DOI: 10.1039/D4SM00422A.

(308) Axpe, E.; Chan, D.; Offeddu, G. S.; Chang, Y.; Merida, D.; Hernandez, H. L.; Appel, E. A. *Macromolecules* **2019**, *52*, 6889–6897, DOI: 10.1021/acs.macromol.9b00753.

(309) Kausik, R.; Srivastava, A.; Korevaar, P. A.; Stucky, G.; Waite, J. H.; Han, S. *Macromolecules* **2009**, *42*, 7404–7412, DOI: 10.1021/ma901137g.

(310) Arrhenius, S. *Zeitschrift für Physikalische Chemie* **1889**, *4U*, 226–248, DOI: 10.1515/zpch-1889-0416.

(311) Karunanithy, G.; Wheeler, R. J.; Tear, L. R.; Farrer, N. J.; Faulkner, S.; Baldwin, A. J. *Journal of Magnetic Resonance* **2019**, *302*, 1–13, DOI: 10.1016/j.jmr.2018.12.001.

(312) Tritt-Goc, J.; Piślewski, N.; Haeberlen, U. *Chemical Physics* **1986**, *102*, 133–140, DOI: 10.1016/0301-0104(86)85124-2.

(313) Ropp, J.; Lawrence, C.; Farrar, T. C.; Skinner, J. L. *Journal of the American Chemical Society* **2001**, *123*, 8047–8052, DOI: 10.1021/ja010312h.

(314) Sweet, W. L.; Blanchard, J. S. *Archives of Biochemistry and Biophysics* **1990**, *277*, 196–202, DOI: 10.1016/0003-9861(90)90569-K.

(315) Xu, X.; Rebane, A. A.; Roset Julia, L.; Rosowski, K. A.; Dufresne, E. R.; Stellacci, F. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2407633121, DOI: 10.1073/pnas.2407633121.

(316) Buchanan, C. J. et al. *Science* **2022**, *377*, eabm3125, DOI: 10.1126/science.abm3125.

(317) Rowe, W. P.; Huebner, R. J.; Gilmore, L. K.; Parrott, R. H.; Ward, T. G. *Proceedings of the Society for Experimental Biology and Medicine* **1953**, *84*, 570–573, DOI: 10.3181/00379727-84-20714.

(318) Hilleman, M. R.; Werner, J. H. *Proceedings of the Society for Experimental Biology and Medicine* **1954**, *85*, 183–8, DOI: 10.3181/00379727-85-20825.

(319) Shieh, W.-J. *Biomedical Journal* **2022**, *45*, 38–49, DOI: 10.1016/j.bj.2021.08.009.

(320) Edwards, K. M.; Thompson, J.; Paolini, J.; Wright, P. F. *Pediatrics* **1985**, *76*, 420–424, DOI: 10.1542/peds.76.3.420.

(321) Chany, C.; Lépine, P.; Lelong, M.; Le-Tan-Vinh; Satagé, P.; Virat, J. *American Journal of Epidemiology* **1958**, *67*, 367–378, DOI: 10.1093/oxfordjournals.aje.a119941.

(322) Hierholzer, J. C. *Clinical Microbiology Reviews* **1992**, *5*, 262–274, DOI: 10.1128/cmr.5.3.262.

(323) Woodland, R. M.; Darougar, S.; Thaker, U.; Cornell, L.; Siddique, M.; Wania, J.; Shah, M. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **1992**, *86*, 317–320, DOI: 10.1016/0035-9203(92)90328-A.

(324) Horne, R. W.; Brenner, S.; Waterson, A. P.; Wildy, P. *Journal of molecular biology* **1959**, *1*, 84, IN13–86, IN15, DOI: 10.1016/S0022-2836(59)80011-5.

(325) Liu, H.; Jin, L.; Koh, S. B. S.; Atanasov, I.; Schein, S.; Wu, L.; Zhou, Z. H. *Science* **2010**, *329*, 1038–1043, DOI: 10.1126/science.1187433.

(326) Kennedy, M. A.; Parks, R. J. *Molecular Therapy* **2009**, *17*, 1664–1666, DOI: 10.1038/mt.2009.202.

(327) Atchison, R. W.; Casto, B. C.; Hammon, W. M. *Science* **1965**, *149*, 754–756, DOI: 10.1126/science.149.3685.754.

(328) Dhingra, A.; Hage, E.; Ganzenmueller, T.; Böttcher, S.; Hofmann, J.; Hamprecht, K.; Obermeier, P.; Rath, B.; Hausmann, F.; Dobner, T.; Heim, A. *Scientific Reports* **2019**, *9*, 1039, DOI: 10.1038/s41598-018-37249-4.

(329) Hunegnaw, R.; Vicari, A. S.; Sullivan, N. J. In *Plotkin's Vaccines (Eighth Edition)*, Orenstein, W., Offit, P., Edwards, K. M., Plotkin, S., Eds.; Elsevier: Philadelphia, 2023, 311–329.e6.

(330) Martin, P. et al. *Gut* **2015**, *64*, 1961–1971, DOI: 10.1136/gutjnl-2014-308041.

(331) Baker, A. T.; Davies, J. A.; Bates, E. A.; Moses, E.; Mundy, R. M.; Marlow, G.; Cole, D. K.; Bliss, C. M.; Rizkallah, P. J.; Parker, A. L. *Journal of Virology* **2021**, *95*, 10.1128/jvi.01849–20, DOI: 10.1128/jvi.01849-20.

(332) Zhang, H.; Wang, H.; An, Y.; Chen, Z. *Molecular Therapy Nucleic Acids* **2023**, *34*, 102027, DOI: 10.1016/j.omtn.2023.09.004.

(333) Antrobus, R. D.; Coughlan, L.; Berthoud, T. K.; Dicks, M. D.; Hill, A. V. S.; Lambe, T.; Gilbert, S. C. *Molecular Therapy* **2014**, *22*, 668–674, DOI: 10.1038/mt.2013.284.

(334) Voysey, M. et al. *The Lancet* **2021**, *397*, 99–111, DOI: 10.1016/S0140-6736(20)32661-1.

(335) Rafie, K.; Lenman, A.; Fuchs, J.; Rajan, A.; Arnberg, N.; Carlson, L.-A. *Science Advances* **2021**, *7*, eabe0974, DOI: 10.1126/sciadv.abe0974.

(336) Nestić, D.; Božinović, K.; Pehar, I.; Wallace, R.; Parker, A. L.; Majhen, D. *Pharmaceutics* **2021**, *13*, 1585, DOI: 10.3390/pharmaceutics13101585.

(337) Tomko, R. P.; Xu, R.; Philipson, L. *Proceedings of the National Academy of Sciences* **1997**, *94*, 3352–3356, DOI: 10.1073/pnas.94.7.3352.

(338) Morgan, C.; Rosenkranz, H. S.; Mednis, B. *Journal of Virology* **1969**, *4*, 777–796, DOI: 10.1128/jvi.4.5.777-796.1969.

(339) Greber, U. F.; Webster, P.; Weber, J.; Helenius, A. *The EMBO Journal* **1996**, *15*, 1766–1777, DOI: 10.1002/j.1460-2075.1996.tb00525.x.

(340) Greber, U. F.; Suomalainen, M.; Stidwill, R. P.; Boucke, K.; Ebersold, M. W.; Helenius, A. *The EMBO Journal* **1997**, *16*, 5998–6007, DOI: 10.1093/emboj/16.19.5998.

(341) Hoeben, R. C.; Uil, T. G. *Cold Spring Harb Perspect Biol* **2013**, *5*, a013003, DOI: 10.1101/cshperspect.a013003.

(342) Velicer, L. F.; Ginsberg, H. S. *Proceedings of the National Academy of Sciences* **1968**, *61*, 1264–1271, DOI: 10.1073/pnas.61.4.1264.

(343) William, S. M. W.; Karoly, T. *Current Gene Therapy* **2013**, *13*, 421–433, DOI: 10.2174/15665232133666131125095046.

(344) Wodrich, H.; Henaff, D.; Jammart, B.; Segura-Morales, C.; Seelmeir, S.; Coux, O.; Ruzsics, Z.; Wiethoff, C. M.; Kremer, E. J. *PLOS Pathogens* **2010**, *6*, e1000808, DOI: 10.1371/journal.ppat.1000808.

(345) Rice, P.; Longden, I.; Bleasby, A. *Trends in Genetics* **2000**, *16*, 276–277, DOI: 10.1016/S0168-9525(00)02024-2.

(346) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. *Protein Sci* **2021**, *30*, 70–82, DOI: 10.1002/pro.3943.

(347) Broto, P.; Moreau, G.; Vandycke, C. *European journal of medicinal chemistry* **1984**, *19*, 71–78.

(348) Laguerre, M.; Saux, M.; Dubost, J.; Carpy, A. *Pharmacy and Pharmacology Communications* **1997**, *3*, 217–222.

(349) Greber, U. F.; Willetts, M.; Webster, P.; Helenius, A. *Cell* **1993**, *75*, 477–486, DOI: 10.1016/0092-8674(93)90382-Z.

(350) Mangel, W. F.; San Martín, C. *Viruses* **2014**, *6*, 4536–4570, DOI: 10.3390/v6114536.

(351) Wiethoff, C. M.; Nemerow, G. R. *Virology* **2015**, *479-480*, 591–599, DOI: 10.1016/j.virol.2015.03.006.

(352) Graziano, V.; Luo, G.; Blainey, P. C.; Pérez-Berná, A. J.; McGrath, W. J.; Flint, S. J.; San Martín, C.; Xie, X. S.; Mangel, W. F. *Journal of Biological Chemistry* **2013**, *288*, 2068–2080, DOI: 10.1074/jbc.M112.407312.

(353) Snijder, J.; Benevento, M.; Moyer, C. L.; Reddy, V.; Nemerow, G. R.; Heck, A. J. R. *Journal of Molecular Biology* **2014**, *426*, 1971–1979, DOI: 10.1016/j.jmb.2014.02.022.

(354) Dai, X.; Wu, L.; Sun, R.; Zhou, Z. H. *J Virol* **2017**, *91*, DOI: 10.1128/jvi.00850-17.

(355) Weber, J. *Journal of Virology* **1976**, *17*, 462–471, DOI: 10.1128/jvi.17.2.462-471.1976.

(356) Rancourt, C.; Keyvani-Amineh, H.; Sircar, S.; Labrecque, P.; Weber, J. M. *Virology* **1995**, *209*, 167–173, DOI: 10.1006/viro.1995.1240.

(357) Martinez, R.; Schellenberger, P.; Vasishtan, D.; Aknin, C.; Austin, S.; Dacheux, D.; Rayne, F.; Siebert, A.; Ruzsics, Z.; Gruenewald, K.; Wodrich, H. *J Virol* **2015**, *89*, 2121–35, DOI: 10.1128/jvi.02257-14.

(358) Moyer, C. L.; Wiethoff, C. M.; Maier, O.; Smith, J. G.; Nemerow, G. R. *J Virol* **2011**, *85*, 2631–41, DOI: 10.1128/jvi.02321-10.

(359) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424–16431, DOI: 10.1021/bi991765q.

(360) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *Journal of Molecular Biology* **1963**, *7*, 95–99, DOI: 10.1016/S0022-2836(63)80023-6.

(361) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647–1651, DOI: 10.1021/bi00121a010.

(362) Wishart, D. S.; Sykes, B. D. *Journal of Biomolecular NMR* **1994**, *4*, 171–180, DOI: 10.1007/BF00175245.

(363) Rosman, K. J. R.; Taylor, P. D. P. *Pure and Applied Chemistry* **1998**, *70*, 217–235, DOI: 10.1351/pac199870010217.

(364) Lescop, E.; Schanda, P.; Brutscher, B. *Journal of Magnetic Resonance* **2007**, *187*, 163–169, DOI: 10.1016/j.jmr.2007.04.002.

(365) Favier, A.; Brutscher, B. *Journal of Biomolecular NMR* **2019**, *73*, 199–211, DOI: 10.1007/s10858-019-00249-1.

(366) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *Journal of Molecular Biology* **1987**, *194*, 531–544, DOI: 10.1016/0022-2836(87)90679-6.

(367) Schiavina, M.; Bracaglia, L.; Bolognesi, T.; Rodella, M. A.; Tagliaferro, G.; Tino, A. S.; Pierattelli, R.; Felli, I. C. *Journal of Magnetic Resonance Open* **2024**, *18*, 100143, DOI: 10.1016/j.jmro.2023.100143.

(368) Schmidt, J. M.; Zhou, S.; Rowe, M. L.; Howard, M. J.; Williamson, R. A.; Löhr, F. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 428–443, DOI: 10.1002/prot.22893.

(369) Barna, J. C. J.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J. P. *Journal of Magnetic Resonance (1969)* **1987**, *73*, 69–77, DOI: 10.1016/0022-2364(87)90225-3.

(370) Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A. *Journal of Biomolecular NMR* **2017**, *68*, 101–118, DOI: 10.1007/s10858-016-0072-7.

(371) Karunanithy, G.; Hansen, D. F. *Journal of Biomolecular NMR* **2021**, *75*, 179–191, DOI: 10.1007/s10858-021-00366-w.

(372) Ernst, R. R.; Bodenhausen, G.; Wokaun, A.; Redfield, A. G., *Principles of nuclear magnetic resonance in one and two dimensions*; American Institute of Physics: 1989.

(373) Pervushin, K.; Vögeli, B.; Eletsky, A. *Journal of the American Chemical Society* **2002**, *124*, 12898–12902, DOI: 10.1021/ja027149q.

(374) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *Journal of the American Chemical Society* **2006**, *128*, 9042–9043, DOI: 10.1021/ja062025p.

(375) Schanda, P.; Brutscher, B. *Journal of the American Chemical Society* **2005**, *127*, 8014–8015, DOI: 10.1021/ja051306e.

(376) Solyom, Z.; Schwarten, M.; Geist, L.; Konrat, R.; Willbold, D.; Brutscher, B. *Journal of Biomolecular NMR* **2013**, *55*, 311–321, DOI: 10.1007/s10858-013-9715-0.

(377) Favier, A.; Brutscher, B. *Journal of Biomolecular NMR* **2011**, *49*, 9–15, DOI: 10.1007/s10858-010-9461-5.

(378) Clubb, R. T.; Thanabal, V.; Wagner, G. *Journal of Magnetic Resonance (1969)* **1992**, *97*, 213–217, DOI: 10.1016/0022-2364(92)90252-3.

(379) Bax, A.; Ikura, M. *Journal of Biomolecular NMR* **1991**, *1*, 99–104, DOI: 10.1007/BF01874573.

(380) Grzesiek, S.; Bax, A. *Journal of Magnetic Resonance (1969)* **1992**, *99*, 201–207, DOI: 10.1016/0022-2364(92)90169-8.

(381) Grzesiek, S.; Bax, A. *Journal of the American Chemical Society* **1992**, *114*, 6291–6293, DOI: 10.1021/ja00042a003.

(382) Weisemann, R.; Rüterjans, H.; Bermel, W. *Journal of Biomolecular NMR* **1993**, *3*, 113–120, DOI: 10.1007/BF00242479.

(383) Bracken, C.; Palmer III, A. G.; Cavanagh, J. *Journal of Biomolecular NMR* **1997**, *9*, 94–100, DOI: 10.1023/A:1018679819693.

(384) Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V. *Journal of Biomolecular NMR* **2001**, *20*, 135–147, DOI: 10.1023/A:1011239023422.

(385) Sun, Z.-Y. J.; Frueh, D. P.; Selenko, P.; Hoch, J. C.; Wagner, G. *Journal of Biomolecular NMR* **2005**, *33*, 43–50, DOI: 10.1007/s10858-005-1284-4.

(386) Buchanan, C.; Karunanithy, G.; Tkachenko, O.; Barber, M.; Marty, M. T.; Nott, T. J.; Redfield, C.; Baldwin, A. J. *Nature Communications* **2025**, *16*, 449, DOI: 10.1038/s41467-024-54899-3.

(387) Hoch, J. C. et al. *Nucleic Acids Research* **2022**, *51*, D368–D376, DOI: 10.1093/nar/gkac1050.

(388) Pritišanac, I.; Degiacomi, M. T.; Alderson, T. R.; Carneiro, M. G.; Ab, E.; Siegal, G.; Baldwin, A. J. *Journal of the American Chemical Society* **2017**, *139*, 9523–9533, DOI: 10.1021/jacs.6b11358.

(389) Raymond, J. W.; Willett, P. *Journal of Computer-Aided Molecular Design* **2002**, *16*, 521–533, DOI: 10.1023/A:1021271615909.

(390) Nielsen, J. T.; Mulder, F. A. A. *Journal of Biomolecular NMR* **2018**, *70*, 141–165, DOI: 10.1007/s10858-018-0166-5.

(391) Tamiola, K.; Acar, B.; Mulder, F. A. A. *Journal of the American Chemical Society* **2010**, *132*, 18000–18003, DOI: 10.1021/ja105656t.

(392) Kjaergaard, M.; Poulsen, F. M. *Journal of Biomolecular NMR* **2011**, *50*, 157–165, DOI: 10.1007/s10858-011-9508-2.

(393) Kjaergaard, M.; Brander, S.; Poulsen, F. M. *Journal of Biomolecular NMR* **2011**, *49*, 139–149, DOI: 10.1007/s10858-011-9472-x.

(394) Avbelj, F.; Kocjan, D.; Baldwin, R. L. *Proceedings of the National Academy of Sciences* **2004**, *101*, 17394–17397, DOI: 10.1073/pnas.0407969101.

(395) Cornilescu, G.; Delaglio, F.; Bax, A. *Journal of Biomolecular NMR* **1999**, *13*, 289–302, DOI: 10.1023/A:1008392405740.

(396) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *Journal of Biomolecular NMR* **2009**, *44*, 213–223, DOI: 10.1007/s10858-009-9333-z.

(397) Shen, Y.; Bax, A. In *Artificial Neural Networks*, Cartwright, H., Ed.; Springer New York: New York, NY, 2015, pp 17–32.

(398) Nielsen, J. T.; Mulder, F. A. A. *Journal of Biomolecular NMR* **2021**, *75*, 273–291, DOI: 10.1007/s10858-021-00374-w.

(399) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Science* **2006**, *15*, 2795–2804, DOI: 10.1110/ps.062465306.

(400) Hansen, D. F.; Yang, D.; Feng, H.; Zhou, Z.; Wiesner, S.; Bai, Y.; Kay, L. E. *Journal of the American Chemical Society* **2007**, *129*, 11468–11479, DOI: 10.1021/ja072717t.

(401) Savitzky, A.; Golay, M. J. E. *Analytical Chemistry* **1964**, *36*, 1627–1639, DOI: 10.1021/ac60214a047.

(402) Stade, K.; Ford, C. S.; Guthrie, C.; Weis, K. *Cell* **1997**, *90*, 1041–1050, DOI: 10.1016/S0092-8674(00)80370-0.

(403) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nature Methods* **2022**, *19*, 679–682, DOI: 10.1038/s41592-022-01488-1.

(404) Mirdita, M.; Steinegger, M.; Söding, J. *Bioinformatics* **2019**, *35*, 2856–2858, DOI: 10.1093/bioinformatics/bty1057.

(405) Mirdita, M.; von den Driesch, L.; Galiez, C.; Martin, M. J.; Söding, J.; Steinegger, M. *Nucleic Acids Research* **2016**, *45*, D170–D176, DOI: 10.1093/nar/gkw1081.

(406) Mitchell, A. L. et al. *Nucleic Acids Research* **2019**, *48*, D570–D578, DOI: 10.1093/nar/gkz1035.

(407) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLOS Computational Biology* **2017**, *13*, e1005659, DOI: 10.1371/journal.pcbi.1005659.

(408) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. *Science* **2023**, *379*, 1123–1130, DOI: 10.1126/science.ade2574.

(409) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. *Nucleic Acids Research* **2015**, *43*, W389–W394, DOI: 10.1093/nar/gkv332.

(410) Dass, R.; Mulder, F. A. A.; Nielsen, J. T. *Scientific Reports* **2020**, *10*, 14780, DOI: 10.1038/s41598-020-71716-1.

(411) Emenecker, R. J.; Griffith, D.; Holehouse, A. S. *Biophysical Journal* **2021**, *120*, 4312–4319, DOI: 10.1016/j.bpj.2021.08.039.

(412) Lotthammer, J. M.; Hernández-García, J.; Griffith, D.; Weijers, D.; Holehouse, A. S.; Emenecker, R. J. *bioRxiv* **2024**, 2024.11.05.622168, DOI: 10.1101/2024.11.05.622168.

(413) Jones, D. T.; Cozzetto, D. *Bioinformatics* **2014**, *31*, 857–863, DOI: 10.1093/bioinformatics/btu744.

(414) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637, DOI: 10.1002/bip.360221211.

(415) Alderson, T. R.; Pritišanac, I.; Kolarić, Đ.; Moses, A. M.; Forman-Kay, J. D. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2304302120, DOI: 10.1073/pnas.2304302120.

(416) Williamson, M. P. In *Modern Magnetic Resonance*, Webb, G. A., Ed.; Springer International Publishing: Cham, 2018, pp 995–1012.

(417) Palmer III, A. G.; Kroenke, C. D.; Patrick Loria, J. In *Methods in Enzymology*, James, T. L., Dötsch, V., Schmitz, U., Eds.; Academic Press: 2001; Vol. 339, pp 204–238.

(418) Vallurupalli, P.; Tiwari, V. P.; Ghosh, S. *The Journal of Physical Chemistry Letters* **2019**, *10*, 3051–3056, DOI: 10.1021/acs.jpclett.9b00985.

(419) Chung, K. Y.; Kim, T. H.; Manglik, A.; Alvares, R.; Kobilka, B. K.; Prosser, R. S. *Journal of Biological Chemistry* **2012**, *287*, 36305–36311, DOI: 10.1074/jbc.M112.406371.

(420) Hutchison, J. M.; Lu, Z.; Li, G. C.; Travis, B.; Mittal, R.; Deatherage, C. L.; Sanders, C. R. *Biochemistry* **2017**, *56*, 5481–5484, DOI: 10.1021/acs.biochem.7b00810.

(421) Woubshete, M.; Cioccolo, S.; Byrne, B. *ChemPlusChem* **2024**, *89*, e202300678, DOI: 10.1002/cplu.202300678.

(422) Palmer III, A. G.; Massi, F. *Chemical Reviews* **2006**, *106*, 1700–1719, DOI: 10.1021/cr0404287.

(423) Fawzi, N. L.; Ying, J.; Torchia, D. A.; Clore, G. M. *Nature Protocols* **2012**, *7*, 1523–1533, DOI: 10.1038/nprot.2012.077.

(424) Hardenberg, M.; Horvath, A.; Ambrus, V.; Fuxreiter, M.; Vendruscolo, M. *Proceedings of the National Academy of Sciences* **2020**, *117*, 33254–33262, DOI: 10.1073/pnas.2007670117.

(425) Vendruscolo, M.; Fuxreiter, M. *Journal of Molecular Biology* **2022**, *434*, 167201, DOI: 10.1016/j.jmb.2021.167201.

(426) Taylor, G. M.; Ma, L.; Vogt, V. M.; Post, C. B. *Biochemistry* **2010**, *49*, 4006–4017, DOI: 10.1021/bi902196e.

(427) Tiessen, A.; Pérez-Rodríguez, P.; Delaye-Arredondo, L. J. *BMC Research Notes* **2012**, *5*, 85, DOI: 10.1186/1756-0500-5-85.

(428) Kuprov, I.; Wagner-Rundell, N.; Hore, P. J. *Journal of Magnetic Resonance* **2007**, *184*, 196–206, DOI: 10.1016/j.jmr.2006.09.023.

(429) Bengs, C.; Levitt, M. H. *Magnetic Resonance in Chemistry* **2018**, *56*, 374–414, DOI: 10.1002/mrc.4642.

(430) Frait, Z.; Gemperle, R. *Czechoslovak Journal of Physics B* **1972**, *22*, 1259–1268, DOI: 10.1007/BF01689904.

(431) Bujacz, A. *Biological Crystallography* **2012**, *68*, 1278–1289, DOI: 10.1107/S0907444912027047.

(432) Crespi, H.; Katz, J. *Nature* **1969**, *224*, 560–562, DOI: 10.1038/224560a0.

(433) McConnell, H. M. *The Journal of chemical physics* **1956**, *25*, 709–711, DOI: 10.1063/1.1743033.

(434) Shimizu, H. *The Journal of chemical physics* **1964**, *40*, 3357–3364, DOI: 10.1063/1.1725007.

(435) Werbelow, L. G.; Grant, D. M. *The Journal of Chemical Physics* **1975**, *63*, 544–556, DOI: 10.1063/1.431085.

(436) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. *nature* **2002**, *418*, 207–211, DOI: 10.1038/nature00860.

(437) Riek, R.; Wider, G.; Pervushin, K.; Wüthrich, K. *Proceedings of the National Academy of Sciences* **1999**, *96*, 4918–4923, DOI: 10.1073/pnas.96.9.4918.

(438) Rodella, M. A.; Schneider, R.; Kümmerle, R.; Felli, I. C.; Pierattelli, R. *Journal of Biomolecular NMR* **2025**, *79*, 15–24, DOI: 10.1007/s10858-024-00453-8.

(439) Schütz, S.; Sprangers, R. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2020**, *116*, 56–84, DOI: 10.1016/j.pnmrs.2019.09.004.

(440) Mueller, L. *Journal of the American Chemical Society* **1979**, *101*, 4481–4484, DOI: 10.1021/ja00510a007.

(441) Bax, A.; Griffey, R. H.; Hawkins, B. L. *Journal of Magnetic Resonance (1969)* **1983**, *55*, 301–315, DOI: 10.1016/0022-2364(83)90241-X.

(442) Bax, A.; Griffey, R. H.; Hawkins, B. L. *Journal of the American Chemical Society* **1983**, *105*, 7188–7190, DOI: 10.1021/ja00362a038.

(443) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445*, 618–622, DOI: 10.1038/nature05512.

(444) Karunanithy, G.; Shukla, V. K.; Hansen, D. F. *Nature Communications* **2024**, *15*, 5073, DOI: 10.1038/s41467-024-49378-8.

(445) Bolik-Coulon, N.; Pelupessy, P.; Bouvignies, G.; Ferrage, F. *Journal of Magnetic Resonance Open* **2020**, *4-5*, 100007, DOI: 10.1016/j.jmro.2020.100007.

(446) Cobb, S. L.; Murphy, C. D. *Journal of Fluorine Chemistry* **2009**, *130*, 132–143, DOI: 10.1016/j.jfluchem.2008.11.003.

(447) Spotswood, T. M.; Evans, J. M.; Richards, J. H. *Journal of the American Chemical Society* **1967**, *89*, 5052–5054, DOI: 10.1021/ja00995a047.

(448) Sykes, B. D.; Hull, W. E. In *Methods in Enzymology*; Academic Press: Cambridge, Massachusetts, 1978; Vol. 49, pp 270–295.

(449) Lau, E. Y.; Gerig, J. *Biophysical journal* **1997**, *73*, 1579–1592, DOI: 10.1016/S0006-3495(97)78190-6.

(450) Frieden, C.; Hoeltzli, S. D.; Bann, J. G. In *Methods in enzymology*; Academic Press: Cambridge, Massachusetts, 2004; Vol. 380, pp 400–415.

(451) O'Hagan, D.; B. Harper, D. *Journal of Fluorine Chemistry* **1999**, *100*, 127–133, DOI: 10.1016/S0022-1139(99)00201-8.

(452) Ye, Y.; Li, C. In *In-cell NMR Spectroscopy: From Molecular Sciences to Cell Biology*, Ito, Y., Dötsch, V., Shirakawa, M., Eds.; The Royal Society of Chemistry: 2019.

(453) Divakaran, A.; Kirberger, S. E.; Pomerantz, W. C. K. *Accounts of Chemical Research* **2019**, *52*, 3407–3418, DOI: 10.1021/acs.accounts.9b00377.

(454) Yu, J.-X.; Hallac, R. R.; Chiguru, S.; Mason, R. P. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2013**, *70*, 25–49, DOI: 10.1016/j.pnmrs.2012.10.001.

(455) Claridge, T. D., *High-resolution NMR techniques in organic chemistry*; Elsevier: 2016; Vol. 27.

(456) Imiołek, M.; Karunanithy, G.; Ng, W.-L.; Baldwin, A. J.; Gouverneur, V.; Davis, B. G. *Journal of the American Chemical Society* **2018**, *140*, 1568–1571, DOI: 10.1021/jacs.7b10230.

(457) Komarov, I. V.; Afonin, S.; Ulrich, A. S. In *Fluorine in Life Sciences: Pharmaceuticals, Medicinal Diagnostics, and Agrochemicals*; Elsevier: Cambridge, Massachusetts, 2019, pp 349–395.

(458) Isenegger, P. G.; Josephson, B.; Gaunt, B.; Davy, M. J.; Gouverneur, V.; Baldwin, A. J.; Davis, B. G. *Nature Protocols* **2023**, *18*, 1543–1562, DOI: 10.1038/s41596-022-00800-9.

(459) Marsh, E. N. G.; Suzuki, Y. *ACS Chemical Biology* **2014**, *9*, 1242–1250, DOI: 10.1021/cb500111u.

(460) Hull, W. E.; Sykes, B. D. *Journal of molecular biology* **1975**, *98*, 121–153, DOI: 10.1021/bi00714a002.

(461) Klein, B. A.; Sykes, B. D. *J Biomol NMR* **2019**, *73*, 519–523, DOI: 10.1007/s10858-019-00266-0.

(462) Naugler, D.; Prosser, R. S. *BioRxiv* **2017**, 140681, DOI: 10.1101/140681.

(463) Consortium, T. U. *Nucleic Acids Research* **2022**, *51*, D523–D531, DOI: 10.1093/nar/gkac1052.

(464) Umemoto, T.; Ishihara, S. *Journal of Fluorine Chemistry* **1998**, *92*, 181–187, DOI: 10.1016/S0022-1139(98)00276-0.

(465) Kim, K. K.; Kim, R.; Kim, S.-H. *Nature* **1998**, *394*, 595–599, DOI: 10.1038/29106.

(466) Emsley, J. W.; Phillips, L.; Wray, V. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1976**, *10*, 83–752, DOI: 10.1016/S0079-6565(76)80005-2.

(467) Skeel, B. A.; Boreen, M. A.; Lohrey, T. D.; Arnold, J. *Inorganic Chemistry* **2020**, *59*, 17259–17267, DOI: 10.1021/acs.inorgchem.0c02516.

(468) Chiu, H.-P.; Kokona, B.; Fairman, R.; Cheng, R. P. *Journal of the American Chemical Society* **2009**, *131*, 13192–13193, DOI: 10.1021/ja903631h.

(469) Salwiczek, M.; Nyakatura, E. K.; Gerling, U. I. M.; Ye, S.; Koksch, B. *Chemical Society Reviews* **2012**, *41*, 2135–2171, DOI: 10.1039/C1CS15241F.

(470) Mukhopadhyay, A.; Li, Y.; Cliff, M. J.; Golovanov, A. P.; Dalby, P. A. *Scientific Reports* **2024**, *14*, 28080, DOI: 10.1038/s41598-024-79711-6.

(471) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *Journal of the American Chemical Society* **2009**, *131*, 12745–12754, DOI: 10.1021/ja903897e.

(472) Waudby, C. A.; Christodoulou, J. *Magn. Reson.* **2021**, *2*, 777–793, DOI: 10.5194/mr-2-777-2021.

(473) Eaton, J. T. Investigating the TROSY Effect in 13CH3 and 13CF3 Nuclear Spin Systems: Increasing the Size Limit of NMR, Ph.D. Thesis, University of Oxford, 2021.

(474) Schwartz, L. M. *Analytical Chemistry* **1975**, *47*, 963–964, DOI: 10.1021/ac60356a027.

(475) Gibbs, A.; Morris, G. A. *Magnetic resonance in chemistry* **1992**, *30*, 662–665, DOI: 10.1002/mrc.1260300716.

(476) Majumdar, A.; Wang, H.; Morshauser, R. C.; Zuiderweg, E. R. *Journal of Biomolecular NMR* **1993**, *3*, 387–397, DOI: 10.1007/BF00176006.

(477) Shaka, A.; Lee, C.; Pines, A. *Journal of Magnetic Resonance (1969)* **1988**, *77*, 274–293, DOI: 10.1016/0022-2364(88)90178-3.

(478) Burum, D. P.; Ernst, R. R. *Journal of Magnetic Resonance (1969)* **1980**, *39*, 163–168, DOI: 10.1016/0022-2364(80)90168-7.

(479) Felli, I. C.; Pierattelli, R. *Chemical Reviews* **2022**, *122*, 9468–9496, DOI: 10.1021/acs.chemrev.1c00871.

(480) Boeszoermenyi, A. et al. *Nature Chemistry* **2025**, DOI: 10.1038/s41557-025-01818-8.

(481) Vuister, G. W.; Yamazaki, T.; Torchia, D. A.; Bax, A. *Journal of Biomolecular NMR* **1993**, *3*, 297–306, DOI: 10.1007/BF00212516.

(482) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *Journal of biomolecular NMR* **1995**, *6*, 277–293, DOI: 10.1007/BF00197809.

(483) Helmus, J. J.; Jaroniec, C. P. *J Biomol NMR* **2013**, *55*, 355–67, DOI: 10.1007/s10858-013-9718-x.

(484) Cumming, H.; Rücker, C. *ACS Omega* **2017**, *2*, 6244–6249, DOI: 10.1021/acsomega.7b01102.

(485) Kristensen, S. M.; Hansen, D. F. FuDA: a function and data fitting and analysis package, 2006.

(486) Wider, G.; Dreier, L. *Journal of the American Chemical Society* **2006**, *128*, 2571–2576, DOI: 10.1021/ja055336t.

(487) Flory, P. J., *Principles of polymer chemistry*; Cornell University Press: 1953.

(488) Cowie, J. M. G.; Arrighi, V., *Polymers: chemistry and physics of modern materials*; CRC press: 2007.

(489) Loria, J. P.; Rance, M.; Palmer III, A. G. *Journal of the American Chemical Society* **1999**, *121*, 2331–2332, DOI: 10.1021/ja983961a.

(490) Palmer III, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *Journal of Magnetic Resonance (1969)* **1991**, *93*, 151–170, DOI: 10.1016/0022-2364(91)90036-S.

(491) Kay, L. E.; Keifer, P.; Saarinen, T. *Journal of the American Chemical Society* **1992**, *114*, 10663–10665, DOI: 10.1021/ja00052a088.

(492) Kontaxis, G.; Bax, A. *Journal of Biomolecular NMR* **2001**, *20*, 77–82, DOI: 10.1023/A:1011280529850.

A Appendices for chapter 3: Exploration of the solvent environment in biomolecular condensates

A.1 Protein sequences

A.1.1 DDX4^{N1}

GDEDWEAEINPHMSSYVPIFEKDRYSGENGDNFNRTPASSSEMDDGP
SRRDHFMKSGFASGRNFGNRDAGECNKRDNTSTMGGFGVGKSFGNRGF
SNSRFEDGDSSGFWRESSNDCEDNPTRNRGFSKRGGYRDGNNSEASGP
YRRGGRGSFRGCRGGFGLGSPNNDLDPDECMQRTGGLFGSRRPVLSGT
GNGDTSQSRSGSGSERGGYKGLNEEVITGSGKNSWKSEAEGGES

A.1.2 EWSR1C

MNSMRGGLPPREGRGMPPPLRGGPGGPGGPGGPMGRMGGRGGDRGGFPP
RGPRGSRGNPSGGGNVQHRAGDWQCPNPGCGNQNFAWRTECNQCKAPKP
EGFLPPPFPPPGGDRGRGGPGGMRGGRGGLMDRGGPGGMFRGGRGGDRG
GFRGGRGMDRGGFGGGRRGGPGGPPGPLMEQMGGRRGGRGGPGKMDKGE
HRQERRDRPY

A.2 Protein expression and purification (DDX4^{N1}/EWSR1C)

A.2.1 DDX4^{N1} protein production and purification

Competent *Escherichia coli* cells (BL21-CodonPlus (DE3)-RIL from Agilent) were thawed on ice. The plasmid vector pET-His-SUMO containing the encoding sequence for DDX4^{N1}, a gift from Timothy Nott (Department of Chemistry, King's College London), was transformed into the competent cells using heat shock. Pre-warmed Luria broth (1 mL, LB) was added to the cells and then incubated at 37^{o}C for 1 hour. The cells were then plated on an LB Agar plate containing 1% glucose, 50 μg/mL Kanamycin and 25 μg/mL Chloramphenicol and incubated for 12 hours at 37^{o}C. Transformed colonies were placed into 3mL LB pre-cultures (containing 1% glucose, 50 μg/mL Kanamycin, 25μg/mL Chloramphenicol) and shaken at 37^{o}C for 8 hours. Pre-warmed flasks containing 50 mL LB (+ 1% glucose, 50 μg/mL Kanamycin and 25 μg/mL Chloramphenicol) were inoculated with 1:1000 of the pre-cultures and shaken at 37^{o}C for 12 hours. Pre-warmed flasks containing 500 mL of Terrific Broth (TB) (+ 10g/L glycerol, 50 μg/mL Kanamycin) were inoculated to an initial OD_{600} of 0.1. The cells were shaken at 37^{o}C until an OD_{600} of 0.8-1.0 was achieved. Expression was induced by adding 1 mM IPTG followed by shaking at 22^{o}C overnight. Cells were pelleted through centrifugation (4000×g, 4^{o}C, 10 minutes, Beckman Avanti JLA 8.1000) followed by resuspension in lysis buffer (6M Guanidine Hydrochloride, 500 mM NaCl, 20 mM sodium phosphate, 10 mM imidazole, 1 mM TCEP, pH 8.0) containing protease inhibitor (cOmplete, EDTA-free, Roche). Cells were lysed on ice through sonication (15 minutes, 40% power, 1 sec on/2 sec off, Sonics Vibra-Cell). The lysate was centrifuged (20000×g, 20 minutes, Beckman Avanti J-26 XP, 4^{o}C) to obtain the aqueous supernatant containing the expressed His-SUMO-DDX4^{N1}, and filtered using 22 μm syringe filters (Millex® syringe filters, Durapore® PVDF membrane). The impure protein fractions were mixed with 10mL Ni-NTA agarose resin bed

volume (Agarose Bead Technologies) and rolled for 1 hour. The mixture was then flowed through a gravity flow chromatography column (Econo-Pac®, Bio-Rad), with the flow through discarded. The resin was washed with 3 × 10 mL of lysis buffer, followed by 2 × Ni-wash buffer 1 (3M Guanidine Hydrochloride, 500 mM NaCl, 20 mM sodium phosphate, 10 mM imidazole, 1 mM TCEP, pH 8.0) and 5 × 10 mL Ni-wash buffer 2 (500 mM NaCl, 20 mM sodium phosphate, 10 mM imidazole, 1 mM TCEP, pH 8.0). ULP1 was added followed by rolling for 12 hours at room temperature to cleave off the His-SUMO tag leaving DDX4^{N1} in solution. The flow through was collected and the column was washed with 5 × 10 mL Ni-wash buffer 2. Fractions containing DDX4^{N1} were collected and concentrated to 1 mL in Vivaspin®-20 (Sartorius) centrifugal concentrator units with a 3 kDa MWCO in a benchtop centrifuge (UNIVERSAL 320R, 6000×g, 37^{o}C). A size exclusion column (HiLoad™ 16/600 Superdex™ 75pg, 120 mL) was then attached to the ÅKTA Pure™ 25M system and equilibrated in sodium phosphate buffer (20 mM sodium phosphate, 300 mM NaCl, 5mM TCEP, pH=8.0). The concentrated protein (1mL) was injected into a 1 mL loop attached to the ÅKTA Pure™ 25M system using a 1 mL syringe (Hamilton syringe, 1000 series GASTIGHT®). The injected protein was run through the column at a speed of 0.5mL/min and collected in 1 mL fractions until a total of 1.5 column volumes of buffer had been ran over the column. All fractions containing a non-zero intensity at 280 nm on the UV chromatogram were analysed by SDS-PAGE (NuPAGE™ , 10-well, 4-12% Bis-Tris Gel from Invitrogen by Thermo Fisher Scientific, MES running buffer, 150V, 50 min) using InstantBlue® Coomassie protein stain. All fractions containing pure DDX4^{N1} were concentrated in Vivaspin®-20 (Sartorius) centrifugal concentrator units with a 3 kDa MWCO using a benchtop centrifuge (UNIVERSAL 320R, 6000×g, 37^{o}C) until a final concentration of 430 μM.*

*Expression and purification of DDX4^{N1} used for research shown in this thesis was performed by myself and other members of the Baldwin lab (Virgínia Casablancas-Antràs, Ben Naylor, Emily Haasz). Eugene Lin (Baldwin lab) also has performed extensive optimisation to the protocol to maximise yield of protein.

A.2.2 EWSR1C protein production and purification

Competent *Escherichia coli* cells (BL21-CodonPlus (DE3)-RIL from Agilent) were thawed on ice. The plasmid vector pETM-30 containing the encoding sequence for EWSR1C, a gift from Timothy Nott (Department of Chemistry, King's College London), was transformed into the competent cells using heat shock. Pre-warmed Luria broth (1 mL, LB) was added to the cells and then incubated at 37 ^{o}C for 1 hour. The cells were then plated on an LB Agar plate containing 1% glucose, 50 μg/mL Kanamycin and 25 μg/mL Chloramphenicol and incubated for 12 hours at 37^{o}C. Transformed colonies were placed into 3mL LB pre-cultures (containing 1% glucose, 50 μg/mL Kanamycin, 25 μg/mL Chloramphenicol) and shaken at 37^{o}C for 8 hours. Pre-warmed flasks containing 50 mL LB (+ 1% glucose, 50 μg/mL Kanamycin and 25 μg/mL Chloramphenicol) were inoculated with 1:1000 of the pre-cultures and shaken at 37^{o}C for 12 hours. Pre-warmed flasks containing 500mL of Terrific Broth (TB) (+ 10g/L glycerol, 50 μg/mL Kanamycin) were inoculated to an initial OD_{600} of 0.1. The cells were shaken at 37^{o}C until an OD_{600} of 0.8-1.0 was reached. Expression was induced by adding 1 mM IPTG followed by shaking at 22^{o}C overnight. Cells were pelleted through centrifugation (4000×g, 4^{o}C, 10 minutes, Beckman Avanti JLA 8.1000). Cell pellets were resuspended in lysis buffer (50 mM TRIS, 500 mM NaCl, 20 mM imidazole, 2 mM TCEP, pH 8.0) followed by the addition of a protease inhibitor tablet (cOmplete, EDTA-free, Roche) and DNAse. The cells were lysed using a microfluidiser an then centrifuged (16500×g, 60 minutes, Beckman Avanti J-26 XP, 4^{o}C) to obtain the aqueous supernatant containing the expressed His-GST-EWSR1C, and filtered using 22 μm syringe filters (Millex® syringe filters, Durapore® PVDF membrane). EWSR1C was purified using Pierce Glutathione Spin Columns, following the manufacturer's instructions. The columns were pre-equilibrated with Equilibration/Wash buffer (50 mM Tris, 500 mM NaCl, 2 mM TCEP, 20 mM imidazole, pH 8). The impure aqueous supernatant containing His-GST-EWSR1C was added to the columns and allowed to bind for 1 hour on a

rocking platform. The flow through was eluted by centrifugation and the columns were washed twice with two resin bed volumes of Glutathione-free buffer. The protein was eluted using Elution buffer (50 mM Tris, 500 mM NaCl, 10 mM reduced glutathione, 2 mM TCEP, 20 mM imidazole pH 8) until there was no absorbance detected in the eluted fractions, which were analysed by SDS-PAGE to locate those corresponding to GST-EWSR1C (51 kDa). After concentrating the fractions to around 10 ml, the GST was cleaved using Tobacco Etch Virus (TEV) protease at 1:1000 ratio for 1 hour at room temperature, on a rocking bed. EWSR1C was separated from His-TEV and His-GST using reverse Nickel affinity chromatography. The flow through, containing EWSR1C, was concentrated down to 2 ml and injected onto a HiLoad™ 26/60060 Superdex™ 200 pg size exclusion chromatography column connected to an ÅKTA Pure™ 25M system, pre-equilibrated with the final storage buffer (20 mM sodium phosphate, 300 mM NaCl, 5 mM TCEP). After elution, fractions containing pure EWSR1C (20.9 kDa) were pooled and concentrated to 225 μM, and flash frozen using liquid nitrogen for storage at -80^{o}C.

*

A.3 Phase separation tests on DDX4^{N1} and EWSR1C

A.3.1 DDX4^{N1} phase separation dependence on D_2O

A solution of DDX4^{N1} at a concentration of 335 μM was prepared in an aqueous buffer containing 20 mM NaPi, 300 mM NaCl, 5 mM TCEP at pH 8.0. Two no salt buffers containing 20 mM NaPi, 5 mM TCEP at a pH of 8.0 were prepared, one containing no D_2O, with the other containing 90% D_2O. Phase separation of DDX4^{N1} was induced by mixing the high concentration protein solution with no salt buffer in a 1:2 ratio, thereby decreasing the salt concentration to 100 mM. To

*Expression and purification of EWSR1C used for the partitioning research (section 3.4) was performed by Virgínia Casablancas-Antràs (Baldwin lab). EWSR1C produce to probe EWSR1C phase separation dependence on NaCl and D_2O was prepared by myself (Figure 3.9).

achieve the desired D_2O percentages of 0%, 25%, 50% and 60% in the final buffer, the samples were prepared in 0.5 mL Eppendorf tubes according to table A.1.

% D_2O	Volume of protein stock (μL)	Volume of no salt buffer 0% D_2O (μL)	Volume of no salt buffer 90% D_2O (μL)
0	10.00	20.00	0.00
25	10.00	11.67	8.33
50	10.00	16.67	3.33
60	10.00	20.00	0.00

Table A.1: A table showing how the $DDX4^{N1}$ solutions were prepared at different percentages of D_2O.

The solutions were stored at 4 ^{o}C for 1.5 hours and then centrifuged at 10000 $\times$g for 5 minutes at 4 ^{o}C to separate the condensate phase and dilute phase. A 10 μL sample of the supernatant (dilute phase) of each sample was removed from each sample. The concentrations of $DDX4^{N1}$ in the dilute phase (c_{sat}) were measured using a Thermo Scientific NanoDrop 1000 UV-Vis spectrophotometer using an extinction coefficient of 23950 $M^{-1}cm^{-1}$ at 280 nm. The results of the saturation concentrations (c_{sat}) of each solution used in Figure 3.6Bi are shown in table A.2.

% D_2O	$DDX4^{N1}$ Saturation Concentration (c_{sat}) (μM)				
	Repeat 1	Repeat 2	Repeat 3	Mean	Standard Deviation
0	51.18	52.76	52.36	52.09	0.79
25	37.00	38.19	34.25	36.48	1.97
50	27.95	25.98	27.56	27.17	1.02
60	24.80	22.44	24.02	23.75	1.21

Table A.2: The results of the saturation concentrations (c_{sat}) of each solution at different percentages of D_2O. Three separate repeats were performed simultaneously to obtain a mean and standard deviation for each D_2O percentage.

The change in free energy of phase separation by the addition of D_2O could be established using equation A.1.

$$\Delta\Delta G = RT \ln\left(\frac{c_{sat}}{c_{sat,0\%\ D_2O}}\right) \tag{A.1}$$

Applying equation A.1 to the results in table A.2 leads to the values in table A.3 which are plotted in Figure 3.6Bii.

% D_2O	Mean ΔΔG (kJ/mol)	Standard Deviation ΔΔG (kJ/mol)
0	0.000	0.049
25	-0.794	0.127
50	-1.472	0.092
60	-1.789	0.121

Table A.3: The differences in free energy upon phase separation (ΔΔG) for different D_2O percentages, obtained by inserting the c_{sat} values in table A.2 into equation A.1.

A.3.2 EWSR1C phase separation dependence on NaCl

A solution of EWSR1C at a concentration of 215 μM was prepared in a buffer containing 20 mM NaPi, 100 mM NaCl, and 2 mM TCEP at a pH of 8.0. Additional buffers were prepared containing 20 mM NaPi, 2 mM TCEP, pH 8.0. One of the additional buffers also contained 4 M NaCl. Samples of EWSR1C at different NaCl concentrations were prepared according to table A.4.

NaCl Concentration (mM)	Volume of protein stock (μL)	Volume of 0 *M* NaCl buffer (μL)	Volume of 4 *M* NaCl buffer (μL)
77	10.00	3.00	0.00
538	10.00	1.50	1.50
1000	10.00	0.00	3.00

Table A.4: A table showing how the EWSR1C solutions were prepared at different NaCl concentrations.

The solutions were stored at 4 ^{o}C for 30 minutes and then centrifuged at 10000 $\times$g for 10 minutes at 4 ^{o}C to separate the condensate phase and dilute phase. A 5 μL sample of the supernatant (dilute phase) of each sample was removed from each sample. The concentrations of EWSR1C in the dilute phase (c_{sat}) were measured using a Thermo Scientific NanoDrop 1000 UV-Vis spectrophotometer using an

extinction coefficient of 12490 $M^{-1}cm^{-1}$ at 280 nm. The results of the saturation concentrations (c_{sat}) of each solution used in Figure 3.6A are shown in table A.5.

NaCl Concentration (mM)	EWSR1C Saturation Concentration (c_{sat}) (μM)
77	162.38
538	166.19
1000	133.33

Table A.5: The results of the saturation concentrations (c_{sat}) of each solution at different NaCl concentrations.

A.3.3 EWSR1C phase separation dependence on D_2O

A solution of EWSR1C at a concentration of 240 μM was prepared in an aqueous buffer containing 20 mM NaPi, 100 mM NaCl, 2 mM TCEP at pH 8.0. Two high salt buffers containing 4 M NaCl, 20 mM NaPi, 2 mM TCEP, pH 8.0 were prepared, one containing no D_2O, with the other containing 100% D_2O. Phase separation of EWSR1C was induced by mixing 10 μL of the high concentration protein solution with 5.6 μL high salt buffer, thereby increasing the salt concentration to 1.5 mM. To achieve the desired D_2O percentages of 0%, 18% and 36% in the final buffer, the samples were prepared in 0.5 mL Eppendorf tubes according to table A.6.

% D_2O	Volume of protein stock (μL)	Volume of 4 *M* NaCl buffer 0% D_2O (μL)	Volume of 4 *M* NaCl buffer 100% D_2O (μL)
0	10.00	5.60	0.00
18	10.00	2.80	2.80
36	10.00	0.00	5.60

Table A.6: A table showing how the EWSR1C solutions were prepared at different percentages of D_2O.

The solutions were stored at 4 ^{o}C for 1 hour and then centrifuged at 10000 $\times$g for 5 minutes at 4 ^{o}C to separate the condensate phase and dilute phase. A 5 μL sample of the supernatant (dilute phase) of each sample was removed from each sample. The

concentrations of EWSR1C in the dilute phase (c_{sat}) were measured using a Thermo Scientific NanoDrop 1000 UV-Vis spectrophotometer using an extinction coefficient of 12490 $M^{-1}cm^{-1}$ at 280 nm. The results of the saturation concentrations (c_{sat}) of each solution used in Figure 3.6Bi are shown in table A.7.

% D_2O	EWSR1C Saturation Concentration (c_{sat}) (μM)				
	Repeat 1	Repeat 2	Repeat 3	Mean	Standard Deviation
0	91.90	90.48	90.00	90.79	2.24
18	99.52	86.19	N/A	92.86	7.67
36	81.90	82.14	N/A	82.02	1.91

Table A.7: The results of the saturation concentrations (c_{sat}) of each solution at different percentages of D_2O. Three (0% D_2O) or two (18%,36% D_2O) separate repeats were performed simultaneously to obtain a mean and standard deviation for each D_2O percentage.

The change in free energy of phase separation by the addition of D_2O could be established using equation A.1. Applying equation A.1 to the results in table A.7 leads to the values in table A.8 which are plotted in Figure 3.9Bii.

% D_2O	Mean $\Delta\Delta G$ (kJ/mol)	Standard Deviation $\Delta\Delta G$ (kJ/mol)
0	0.000	0.079
18	-0.051	0.198
36	-0.231	0.077

Table A.8: The differences in free energy upon phase separation ($\Delta\Delta G$) for different D_2O percentages, obtained by inserting the c_{sat} values in table A.7 into equation A.1.

A.4 Methods (partitioning experiments)

A.4.1 Condensate partitioning method

Solutions of $DDX4^{N1}$ (466 μM) and EWSR1C (225 μM) were prepared in high-salt buffer (20 mM sodium phosphate, 300 mM NaCl, 5 mM TCEP, pH 8.0, H_2O). A second solution was prepared, containing a mixture of the nucleotides (dA, dG,

cGMP, cAMP, GTP, ATP) and ^{13}C,^{15}N-labelled amino acids in no-salt buffer (14 mM sodium phosphate, 3.5 mM TCEP, pH 8.0, 70% D_2O), and in some cases (DDX4^{N1}+polyA, EWSR1C+polyA datasets) 90 μM of 28mer polyadenosine (polyA) purchased from IDT™. The nucleotides and amino acids ((5-100 mM Cell Free Amino Acid Mixture – ^{13}C,^{15}N) were purchased from Sigma-Aldrich®. The solution contained the following nucleotides at a concentration of 1.0 mM: deoxyguanosine (dG), deoxyadenosine (dA), guanosine 3',5'-cyclic monophosphate (cGMP), adenosine 3',5'-cyclic monophosphate (cAMP), guanosine 5'-triphosphate (GTP), adenosine 5'-triphosphate (ATP). The amino acids had concentrations between 1.5-30 mM, diluted 70% from their supplied concentration. The no-salt buffer was filtered using a 0.22 μm Corning 3161 filter to ensure no undissolved solid particulates of nucleotides or amino acids remained that might affect the partitioning measurements. No-salt buffer (with and without polyA) was added to the DDX4^{N1} solution in a packed cell volume, PCV, tube (TPP®, Product No. 87005) in a 2:1 ratio to induce phase separation followed by thorough mixing. The final buffer solution consisted of 16 mM sodium phosphate, 100 mM NaCl, 4 mM TCEP, 46.6% D2O. The same procedure was applied to EWSR1C with only no-salt buffer containing the polyA (EWSR1C could not phase separate under the final buffer conditions without the presence of polyA). Furthermore, D_2O was observed to enhance the phase separation for the DDX4^{N1} systems, but not for EWSR1C (Figure 3.6/ Figure 3.9). The samples were left on ice for approximately 2 hours. The PCV tubes containing the phase separated solutions were placed inside a 15 mL Falcon™ tube, equilibrated on ice, followed by centrifugation to form a pellet (1500×g, 4 ^{o}C, 10 minutes, Sigma-Aldrich® 3-18KS, swing out rotor 11180 + bucket 13190). Transfer to a 15 mL tube followed by centrifugation in a spin-out rotor was used to ensure the resulting pellet had a level surface. The pellet volumes (V_p) were then measured using the PCV "easy read" measuring device (TPP®, Product No. 87010) at 4 ^{o}C. 100 μL Supernatant was removed ($V_{i,s}$) and stored,

followed by the rapid removal of as much supernatant as possible from the samples without disturbing the pellet. For this task, thin 20 μL pipette tips (epT.I.P.S. GELoader™, Eppendorf™, Product No. 10411193) were used. The samples were then centrifuged gently using a manual centrifuge, followed by further removal of the supernatant. This step was necessary to remove leftover liquid from the walls of the tube. This was followed by measurement of the leftover supernatant (V_{LOS}) in the tubes using the "easy-read" measuring device. The PCV tubes containing the remaining liquid (V_p + V_{LOS}) were then placed into 15 mL Falcon™ tubes in an inverted fashion and centrifuged (1500×g, 4^{o}C, 10 minutes, Sigma-Aldrich® 3-18KS, swing out rotor 11180 + bucket 13190). This step was necessary to ensure that all the liquid had been adequately removed from the PCV tubes. To the pellet samples, now in the 15 mL tube, 200 μL of 100 mM salt buffer was added (16 mM sodium phosphate, 100 mM NaCl, 4 mM TCEP, 46.6% D_2O), followed by the addition of 50 μL of 3 M salt buffer (100% D_2O, 20 mM NaPi, 3 M NaCl, 5 mM TCEP, pH 8.0) to reach a final salt concentration of 600 mM (57.3% D_2O, 16.8 mM sodium phosphate, 600 mM NaCl, 4.2 mM TCEP, pH 8.0). To the stored supernatant samples ($V_{i,s}$), 43.6 μL of 3 M salt buffer (100% D_2O, 20 mM NaPi, 3 M NaCl, 5 mM TCEP, pH 8.0) and 6.7 μL of no salt buffer (100% D_2O, 20 mM sodium phosphate, 5 mM TCEP, pH 8.0), without dissolved nucleotides/amino acids, was added. The pH of each supernatant and resuspended condensate sample was then carefully adjusted to exactly 8.0 with any additional volume added to the samples measured. The final supernatant and resuspended condensate samples were then placed into 3 mm Bruker® SampleJet® NMR tubes.

Note: repeats were performed on the same day with the same stocks of protein and small molecule mix (nucleotides and amino acids), except for DDX4^{N1} where repeat 1 was done on a different day, using a small molecule mixture prepared on that day. *

*Condensate partitioning experiments were performed in equal collaboration between myself and Virgínia Casablancas-Antràs (Baldwin lab).

A.4.2 Butanol partitioning method

The following method is based on the method introduced by Cumming *et al.*[484]. The partitioning of molecules in a 1-butanol / buffer mixture was measured by measuring the signals of the aqueous phase initially and then again after addition of 1-butanol and mixing. A mixture containing 1.5-30 mM ^{13}C,^{15}N amino acids and 1 mM of ATP, cAMP, dA, GTP, cGMP, dG was prepared in no Salt D_2O buffer, and mixed in a 2:1 ratio with H_2O High salt buffer without protein (20 mM sodium phosphate, 300 mM NaCl, 5 mM TCEP, pH 8), in order to achieve the same conditions of the phase separated samples described above (16 mM sodium phosphate, 100 mM NaCl, 4 mM TCEP, 46.6% D_2O). 550 μL of sample were loaded on a 5 mm NMR tube and the length of tube made up of the sample was measured with a ruler. Then NMR spectra of the initial water-phase were recorded. After, 500 μL of 1-butanol were added to the sample and after sealing the top with parafilm the mixture was equilibrated by inverting 50 times and allowing 30 minutes to equilibrate. The length of tube taken by each phase was measured with a ruler. Then the same NMR spectra were acquired on the water-phase. The measurements were performed in duplicate.*

A.4.3 NMR data acquisition for metabolite partitioning measurements

All NMR spectra on supernatant and resuspended condensate NMR samples were performed on a 950 MHz Oxford Instruments Company NMR magnet equipped with a Bruker® Avance III HD console and a 5 mm TCI (^1H, ^{13}C, ^{15}N) CryoProbe at a temperature of 25^{o}C. Nucleotide signals for each sample were obtained using one-dimensional water suppressed proton experiments, which were acquired using a Bruker-standard water suppressed pulse sequence with excitation sculpting (zgesgp). Amino acid signals were obtained using a HACA(CO) pulse sequence, adapted

*Partitioning experiments into butanol were performed in equal collaboration between myself and Virgínia Casablancas-Antràs (Baldwin lab).

from the standard HACACO pulse sequence developed by L.E. Kay *et al*[278], by the addition of seedless optimal control selective pulses for carbon, as well as C_β decoupling.[279] The pulse sequence (Figure 3.14A) and example output reports of pulse optimisations (appendix A.9) are included. The phase cycle used was: The seedless pulses provided high selectivity, and the seedless C_β decoupling enabled high-resolution spectra to be obtained even when amino acid H_α-C_α peaks were heavily overlapped (Figure 3.15). Peaks in each spectrum were picked using UniDecNMR[386] and mapped to H_α-C_α amino acid assignments (appendix A.10) performed using a standard HCCH TOCSY sequence (hcchdigp3d).[281] The combination of many increments in the indirect dimension and the C_β decoupling enabled the GLN, ARG, LYS and GLU peaks to be resolved and isolated. However, the C_β decoupling destroyed all GLY magnetisation. Therefore, a separate experiment without C_β decoupling was used to measure the partitioning coefficient of GLY throughout this paper. Furthermore, C_β decoupling was not possible for THR and SER whose C_β chemical shifts are significantly overlapped with the C_α shifts and outside the range of the seedless C_β selective pulse. Unfortunately, the SER peak in the HACA(CO) spectra was too low in intensity and could not be quantified for all supernatant/resuspended condensate pairs. Therefore, it has been omitted from all results and downstream quantification.

A.4.4 Condensate metabolite partitioning NMR data analysis

All NMR spectra were processed using SpinExplorer, a graphical user interface (GUI) for processing and analysing NMR spectra with a backend utilising nmrPipe[482] and nmrglue.[483] The one-dimensional proton spectra were processed using exponential apodization with a line broadening factor of 1.0 Hz. The spectra were extracted between 5.2 ppm and 9.0 ppm followed by the application of a linear baseline. For one-dimensional NMR spectra, SpinView was used calculate

the signal factor between peaks in the supernatant and resuspended condensate spectra. SpinView allowed manual baseline corrections to each spectrum in a peak/nucleotide dependent manner to ensure the accuracy of determined signal factors for each molecule. Furthermore, for nucleotide peaks significantly overlapped with the polyA, a reference polyA only background spectrum was subtracted from the supernatant and resuspended condensate spectra to enable the nucleotide signal factor to be calculated. For the two-dimensional HACA(CO) NMR spectra, Lorentz to Gauss apodization was applied so that quantification of peak intensities could performed using FuDA[485] with a gaussian peak shape.

A.4.5 Condensate ion partitioning method

Solutions of DDX4^{N1} (400 μM) were prepared in high-salt buffer (20 mM sodium phosphate, 300 mM NaCl, 5 mM TCEP, pH 8.0, H_2O). A second buffer was prepared containing no-salt (20 mM sodium phosphate, 5 mM TCEP, pH 8.0, D_2O). The no-salt buffer (360 μL) was added to the DDX4^{N1} solution (180 μL) in a packed cell volume, PCV, tube (TPP®, Product No. 87005) to induce phase separation followed by thorough mixing. The final buffer solution consisted of 20 mM sodium phosphate, 100 mM NaCl, 5 mM TCEP, 66% D_2O. The samples were left at 4^{o}C for 1 hour. The PCV tubes containing the phase separated solutions were placed inside a 15 mL Falcon™ tube, equilibrated on ice, followed by centrifugation to form a pellet (1500×g, 4 ^{o}C, 10 minutes, Sigma-Aldrich® 3-18KS, swing out rotor 11180 + bucket 13190). Transfer to a 15 mL tube followed by centrifugation in a spin-out rotor was used to ensure the resulting pellet had a level surface. The pellet volumes (V_p) were then measured using the PCV "easy read" measuring device (TPP®, Product No. 87010) at 4 ^{o}C. 100 μL Supernatant was removed ($V_{i,s}$) and stored (note: all tubes used for storing samples from this point onwards were cleaned using UHPLC-grade water before use). This was followed by the rapid removal of as much supernatant as possible from the samples without disturbing the pellet. For

this task, thin 20 μL pipette tips (epT.I.P.S. GELoader™, Eppendorf™, Product No. 10411193) were used. The samples were then centrifuged gently using a manual centrifuge, followed by further removal of the supernatant. This was followed by measurement of the leftover supernatant (V_{LOS}) in the tubes using the "easy-read" measuring device. The PCV tubes containing the remaining liquid (V_p + V_{LOS}) were then placed into 15 mL Falcon™ tubes in an inverted fashion and centrifuged (1500×g, 4^{o}C, 10 minutes, Sigma-Aldrich® 3-18KS, swing out rotor 11180 + bucket 13190). This step was necessary to ensure that all the liquid had been adequately removed from the PCV tubes. To the pellet samples, now in the 15 mL tube, 600 μL of UHPLC-grade water containing 10% D_2O was added. To the stored supernatant samples (V_i), 500 μL of UHPLC-grade water containing 10% D_2O was added. A 600 μL control sample of UHPLC-grade water containing 10% D_2O was first placed into a 15 mL Falcon™ tube. The final supernatant, resuspended condensate and control samples were then placed into 5 mm NMR tubes.

Note: repeats were performed on the same day. *

A.4.6 NMR data acquisition for the condensate ion partitioning measurements

NMR measurements were recorded on a 600 MHz Oxford Instruments spectrometer in a room-temperature 10 mm BB600-10K Z-SPEC® Varian broadband probe tuned to either ^{23}Na or ^{31}P. All measurements were performed at 25^{o}C. One-dimensional pulse acquire spectra were performed with relaxation delays of 1 s and 9 s for ^{23}Na or ^{31}P experiments, respectively. † These relaxation delays were measured to give spectra linear with number of scans. All NMR spectra were processed using SpinExplorer, a graphical user interface (GUI) for processing and analysing NMR

*Condensate ion partitioning experiments were performed in equal collaboration between myself and Ben Naylor (Baldwin lab).

†Condensate ion partitioning NMR experiments were performed in equal collaboration between myself and Ben Naylor (Baldwin lab).

spectra with a backend utilising nmrPipe[482] and nmrglue.[483] The one-dimensional proton spectra were processed using exponential apodization with a line broadening factor of 5 Hz and 50 Hz, for ^{23}Na or ^{31}P, respectively. The substantial apodization allowed an increase in signal to noise of the resulting spectra. SpinView was used calculate signal intensities in the supernatant, resuspended condensate and control spectra. Given the low concentration of ^{23}Na/^{31}P in the resuspended condensate samples, calibration curves were performed to confirm linearity of signal intensity with concentration.*

A.5 Methods (biphasic DDX4^{N1} NMR experiments)

NMR measurements on the biphasic DDX4^{N1} sample were recorded on a 600 MHz Oxford Instruments spectrometer in a room-temperature 5 mm ^{1}H,^{15}N,^{13}C-triple resonance probe. All NMR spectra were processed using SpinExplorer, a graphical user interface (GUI) for processing and analysing NMR spectra with a backend utilising nmrPipe[482] and nmrglue.[483] Details of acquisition, processing and analysis of each set of experiments are included below. All spectra were referenced to H_2O in the supernatant phase.

A.5.1 Creation of a DDX4^{N1} biphasic NMR sample

The following method outlines how a biphasic DDX4^{N1} sample was created (Figure 3.28). A 400 μM DDX4^{N1} solution (in 20 mM NaPi, 300 mM NaCl, 5 mM Tris (2-carboxyethyl) phosphine (TCEP), pH 8.0, H_2O) was mixed in a 1:2 ratio with no salt D_2O buffer (20 mM NaPi, 5 mM TCEP, pH 8.0, D_2O) in a packed cell volume (PCV) tube. Therefore, the final sample buffer conditions were 20 mM NaPi, 100 mM NaCl, 5 mM TCEP, pH 8.0, 66% D_2O. The solution was left at 4^{o}C for one hour to allow equilibrium to be reached before being centrifuged at 1500×g at 4^{o}C for 10 minutes to separate the condensate and supernatant phases.

*The calibration curves were performed by Ben Naylor (Baldwin lab).

Excess supernatant was removed through pipetting before transferring the mixture carefully to a 1 mm NMR tube. The 1 mm tube was then centrifuged using a hand centrifuge to produce the biphasic $DDX4^{N1}$ sample. The final NMR sample was then created using a "Russian doll" approach of placing the tube in increasing sizes of NMR tube (1.7 mm, 3 mm, 5 mm) to be used in a 5 mm NMR probe.

A.5.2 MRI experiments

The gradient echo (GRE) experiment was ran to obtained MRI spectra using a pulse sequence outlined in Figure 2.10. The initial gradient of -10000 DAC was applied for a total of $(at)/2 + d4$, where at (acquisition time) was equal to 2.5 ms and the delay $d4$ was equal to 1 ms. A total of 8 steady-state scans and 64 transients were used separated with a relaxation delay of 5 s. The initial 90^o pulse (Φ_1) and receiver (Φ_{rec}) phase cycles used were: $\Phi_1 = \Phi_{rec}$=(x,-x,x,-x,x,-x,x,-x). Spectra were processed using sine-bell (SP) apodization with a power/exponent of three. After Fourier transform, the magnitude of each spectrum was taken, giving the MRI spectra.

A.5.3 Slice selective NMR experiments

The DPFGSE experiment was ran to obtain slice selective NMR spectra as outlined in Figure 2.11. Detailed descriptions of the pulse sequences are included in Figure A.1

Non-water suppressed For non-water suppressed slice-selective experiments (Figure A.1A), the experimental parameters used were as follows.[†] The gradients G_3, G_4 and G_5 had magnitudes of 6000, 15000 and 11450 DAC, respectively. The selective 180^o pulses were applied for 2 ms with simultaneous application of gradient G_3. The gradient durations δ_4 and δ_5 were 1 ms and 500 μs, respectively. The phase

[†]This sequence was written in a collaboration between myself and Charles Buchanan (Baldwin lab)

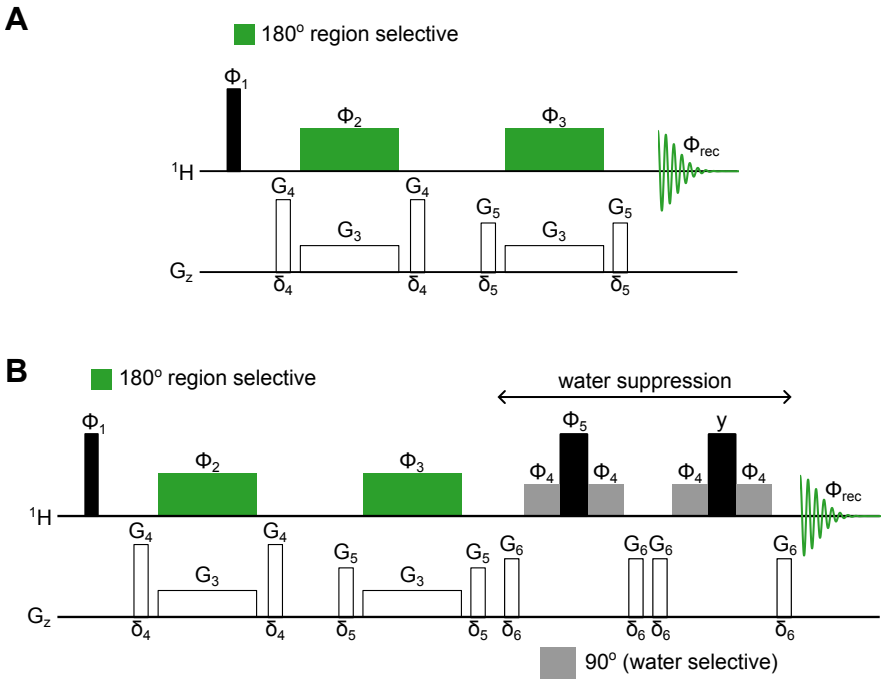


Figure A.1: Detailed pulse sequences of the slice selective (**A**) and water-suppressed slice selective (**B**) sequences employed.

cycle employed was as follows: Φ_1=(x,-x,x,-x,x,-x,x,-x), Φ_2=(x,-y,-x,y,x,y,-x,-y), Φ_3=(x), Φ_{rec}=(x). The acquisition time was set to 2 s and a relaxation delay of 3 s was used. A total of 2 steady-state scans were used and 4 transients were recorded. Spectra were processed using exponential apodization with a line broadening factor of 10 Hz. Slice-selective MRI spectra were recorded by adding a gradient-echo unit to the end of the pulse sequence and reducing the acquisition time to below 40 ms.

Water suppressed For non-water suppressed slice-selective experiments (Figure A.1B), the experimental parameters used were as follows.* The gradients G_3, G_4, G_5 and G_6 had magnitudes of 6000, 15000, 11450 and 26425 DAC, respectively. The selective 180^o pulses were applied for 2 ms with simultaneous application of gradient G_3. The gradient durations δ_4, δ_5, and δ_6 were 1 ms, 500 μs and 2 ms, respectively. The phase cycle employed was as follows: Φ_1=(x,-x,x,-x,x,-x,x,-x), Φ_2=(x,-y,-x,y,x,y,-x,-y), Φ_3=(x), Φ_4=(-x,-y,y,x,-x,-y,y,x),

*This sequence was written in a collaboration between myself and Charles Buchanan (Baldwin lab)

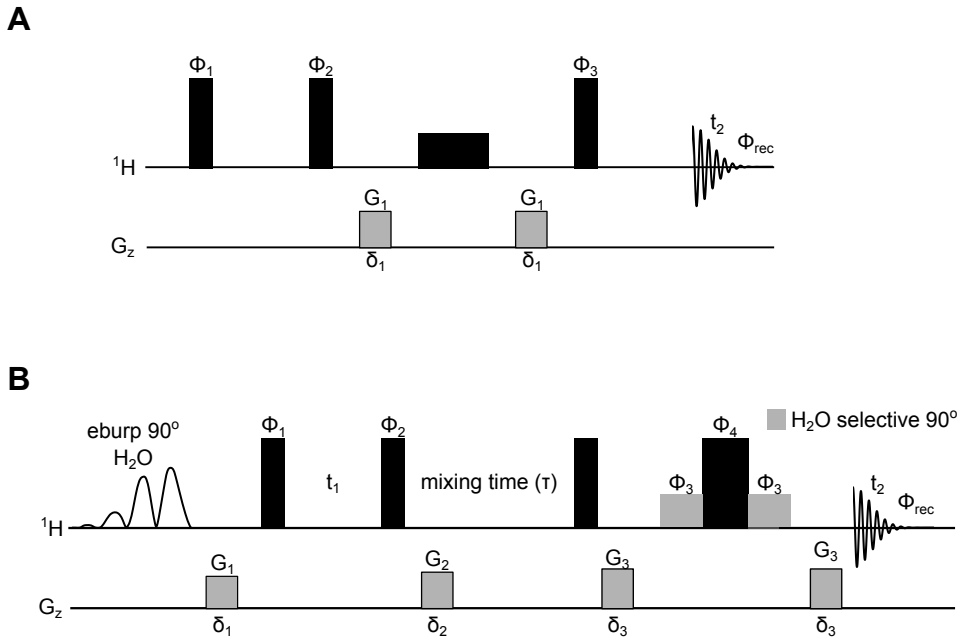


Figure A.2: Detailed pulse sequences of the CEST (**A**) and NOESY (**B**) sequences employed.

Φ_5=(x,y,-y,-x,x,y,-y,-x), Φ_{rec}=(x,-x,-x,x,x,-x,-x,x). The water-selective 90^o pulses were centred on the supernatant water peak. The acquisition time was set to 1 s and a relaxation delay of 2 s was used. A total of 2 steady-state scans were used and 3000 transients were recorded. Digital solvent suppression was applied and spectra were processed using exponential apodization with a line broadening factor of 5 Hz.

A.5.4 CEST experiments

The CEST experiment (discussed in section 2.6.1) used for this experiment was based on a longitudinal encode-decode diffusion CEST experiment, with the encode-decode gradients set to zero (Figure A.2A). The gradient strength (G_1) and duration (δ_1) used were 25000 DAC and 1 ms, respectively. The CEST pulse was applied at low power on the proton channel at 40 chemical shifts evenly spaced between -0.45 ppm and 11.25 ppm for 1 s. Each application of a saturation pulse was interleaved with an experiment containing only a delay of 1 s without a saturation pulse, enabling a reference spectrum to be obtained to account for fluctuations in the sample over

time. The acquisition time was set to 1.5 s, and a relaxation delay of 1.5 s was used. A total of 4 steady-state scans were applied, and 4 transients were recorded for each saturation pulse application. Spectra were processed using exponential apodization where 1.0 Hz of line broadening was applied. CEST-profiles were generated using the CEST analysis module of SpinExplorer.

A.5.5 NOESY experiment

The ^{1}H-^{1}H NOESY experiment is shown in (Figure A.2B).[†] The eburp water selective pulse had phase x. Hypercomplex quadrature in the indirect dimension (t_1) was obtained through a 90^o phase shift on Φ_2, with the total number of complex points along (t_1) set to 512. A total of 64 scans were performed per time increment t_1, with the relaxation delay (d_1) set to 1.2 s. The phase cycle employed was as follows: Φ_1 = (x,-x,x,-x,x,-x,x,-x), Φ_2 = x, Φ_3 = (-x,-y,y,x,-x,-y,y,x),Φ_4 = (x,y,-y,-x,x,y,-y,-x), Φ_{rec} = (x,x,-x-x,x,x,x,-x,-x). The gradient strengths of G_1, G_2 and G_3 used were 15000, 10000 and 26475 DAC, respectively. The gradient durations δ_1, δ_2 and δ_3 were all set to 1 ms. The mixing time τ was set to 200 ms. Water selective pulses were centred on the water-supernatant peak. Spectra were processed using exponential apodization in both dimensions with a line broadening factor of 5.0 Hz applied. The imaginary points were negated in the indirect dimension for the Fourier transform.

[†]This sequence was written in a collaboration between myself and Charles Buchanan (Baldwin lab)

A.6 Condensate volume measurements

A.6.1 Metabolite condensate partitioning volume measurements

Sample	$DDX4^{N1}$			$DDX4^{N1}$ + polyA		EWSR1C + polyA	
Repeat	1	2	3	1	2	1	2
Condensate Volume V_p (μL)	3.9	4.4	4.1	1.95	1.875	0.775	0.9
Leftover supernatant V_{LOS} (μL)	0.525	0.4	0.35	0.075	0.075	0.825	0.85
Total volume of sample (μL)	600	600	495	270	270	249	249
Condensate Volume Fraction (%)	0.65	0.73	0.82	0.72	0.69	0.31	0.36
Average volume fraction (%) ± standard deviation	0.73 ± 0.069			0.705 ± 0.015		0.335 ± 0.025	

Table A.9: Table of volumes of pellet (V_p) and leftover supernatant (V_{LOS}) for the condensate partitioning experiments. Experiments were performed in an equal collaboration between myself and Virgínia Casablancas-Antràs (Baldwin lab).

A.6.2 DDX4^{N1} condensate ion partitioning volume measurements

Sample	DDX4^{N1}		
Repeat	1	2	3
Condensate Volume V_p (μL)	2.775	3.000	2.95
Leftover supernatant V_{LOS} (μL)	0.575	0.425	0.300
Total volume of sample (μL)	540	540	540
Condensate Volume Fraction (%)	0.51	0.56	0.55
Average volume fraction (%) $\pm$ standard deviation	0.54 $\pm$ 0.022		

Table A.10: DDX4^{N1} condensate ion partitioning volume measurements

A.7 Partitioning coefficient formulae derivations

A.7.1 Derivation of partitioning coefficient formula used for the biomolecular condensate partitioning experiments

This section outlines the derivation of the formula to determine the partitioning coefficient of small molecules into biomolecular condensates. For a molecule A, the concentration inside the pellet is given by $[A]_p$, with the concentration in the supernatant given by $[A]_s$. Therefore the partitioning coefficient $PC = [A]_p/[A]_s$. However, the ratio of pellet and supernatant sample intensities (concentrations) per unit scan is not the same as the raw partitioning coefficient $[A]_p/[A]_s$. This is because the pellet and supernatant samples are modified before recording the NMR experiments. We therefore define the ratio of the pellet and supernatant sample intensities per unit scan as the signal factor $S = [A]_{s,m}/[A]_{p,m}$, where the subscript

m stands for modified. The supernatant removed after centrifugation (100μL, V_i) at the concentration $[A]_s$ is diluted by adding $V_{add,s}$ of high salt buffer to form the NMR sample at a concentration of $[A]_{s,m}$. Therefore:

$$[A]_{s,m} = \frac{[A]_s V_i}{V_i + V_{add,s}} \tag{A.2}$$

The pellet remaining has a volume V_p and, after trying to remove some as much supernatant as possible, still contains a small amount of supernatant with a volume V_{LOS}, where the LOS stands for leftover supernatant. In addition, the pellet sample is diluted to form the NMR sample by adding $V_{add,p}$ of high salt buffer. Therefore:

$$[A]_{p,m} = \frac{[A]_p V_p + [A]_s V_{LOS}}{V_p + V_{LOS} + V_{add,p}} \tag{A.3}$$

An equation for the signal factor (S) can be obtained by dividing equation 1 by equation 2. Therefore:

$$S = \frac{[A]_s V_i \left(V_p + V_{LOS} + V_{add,p}\right)}{\left([A]_p V_p + [A]_s V_{LOS}\right)\left(V_i + V_{add,s}\right)} \tag{A.4}$$

$$\frac{1}{S} = \frac{[A]_p/[A]_s V_p \left(V_i + V_{add,s}\right)}{V_i \left(V_p + V_{LOS} + V_{add,p}\right)} + \frac{V_{LOS}\left(V_i + V_{add,s}\right)}{V_i \left(V_p + V_{LOS} + V_{add,p}\right)}$$

As $[A]_p/[A]_s = PC$, we can rearrange this equation to find an equation for P based on known volumes and the signal factor.

$$\frac{V_i \left(V_p + V_{LOS} + V_{add,p}\right)}{S} = PC \cdot V_p \left(V_i + V_{add,s}\right) + V_{LOS}\left(V_i + V_{add,s}\right) \tag{A.5}$$

$$PC = \frac{V_i \left(V_p + V_{LOS} + V_{add,p}\right)}{S V_p \left(V_i + V_{add,s}\right)} - \frac{V_{LOS}}{V_p} \tag{A.6}$$

A.7.2 Derivation of the partitioning coefficient formula used for 1-butanol partitioning measurements

This section outlines the derivation of the formula to determine the partitioning coefficient of molecules into 1-butanol.

An initial water based NMR sample of volume V_i contains a molecule A at a concentration of $[A]_i$. After adding butanol and equilibrating the bottom water phase has a phase V_w containing molecule A at a concentration $[A]_w$. The top butanol phase has a volume V_b and contains molecule A at a concentration $[A]_b$. Therefore, the partitioning coefficient of molecule A, PC, is given by $[A]_b/[A]_w$.

The total amount of molecules of A before and after the butanol addition remains constant, therefore:

$$V_i[A]_i = V_w[A]_w + V_b[A]_b \tag{A.7}$$

$[A]_b$ can be rewritten as $[A]_w P$ and inserted into equation A.7 giving:

$$V_i[A]_i = V_w[A]_w + V_b[A]_w P \tag{A.8}$$

Rearranging this equation obtains:

$$PC = \frac{V_i\left(\frac{[A]_i}{[A]_w}\right) - V_w}{V_b} \tag{A.9}$$

The concentration of A in each phase is related to the NMR signal obtained through $[A] \propto S \times \theta$, where θ is the proton 90^o pulse duration in that phase and S is the NMR signal intensity from molecule A. The correction by θ is necessary due to the possible change in composition of the water phase upon the addition of butanol.[486]

Therefore:

$$PC = \frac{V_i \left(\frac{S_i \times \theta_i}{S_w \times \theta_w} \right) - V_w}{V_b} \tag{A.10}$$

This equation can then be used to obtain the partitioning coefficient of molecule A in butanol.

A.8 Partitioning results

A.8.1 Partitioning coefficients obtained for the biomolecular condensates and organic solvents

The following results outline the partitioning coefficients (PC) obtained for the biomolecular condensates (DDX4^{N1}+/-polyA and EWSR1C+polyA) using the method established in A.4.1 as well as values measured for 1-butanol (using method A.4.2). For 1-butanol, n.d. refers to molecules that were excluded from the results, and for which an upper limit of exclusion has been estimated and shown in Figure 3 (but not used for fitting). For 1-octanol, literature values were used,[282] and an error of 10% (in logP space) was used for all values. n.d. refers to molecules for which experimental literature values were not found. n=3 for DDX4^{N1}, n=2 for DDX4^{N1}+polyA, EWSR1C+polyA and 1-butanol.

Partitioning Coefficient (PC)					
Molecule	DDX4^{N1}	DDX4^{N1} +polyA	EWSR1C +polyA	1-butanol	1-octanol ($\log_{10}$(PC))
ARG	1.654 ± 0.255	1.672 ± 0.023	0.726 ± 0.168	0.019 ± 0.015	-4.20 ± 0.42
HIS	0.936 ± 0.105	0.829 ± 0.057	0.426 ± 0.189	0.084 ± 0.079	-3.32 ± 0.332
LYS	0.985 ± 0.197	0.871 ± 0.031	0.394 ± 0.023	n.d.	-3.05 ± 0.305
ASP	0.609 ± 0.096	0.555 ± 0.0005	0.403 ± 0.127	n.d.	-3.89 ± 0.389
GLU	0.618 ± 0.056	0.536 ± 0.021	0.378 ± 0.204	0.011 ± 0.004	-3.69 ± 0.369
ASN	0.797 ± 0.127	0.700 ± 0.006	0.173 ± 0.173	0.029 ± 0.016	-3.82 ± 0.382
GLN	0.802 ± 0.212	0.688 ± 0.079	0.092 ± 0.066	0.038 ± 0.033	-3.64 ± 0.364
THR	0.761 ± 0.146	0.658 ± 0.091	0.350 ± 0.110	n.d.	-2.94 ± 0.294

Table A.11: Table 1 (of 2) of partitioning coefficients (PC) obtained for biomolecular condensates and organic solvents.

Partitioning Coefficient (PC)					
Molecule	$DDX4^{N1}$	$DDX4^{N1}$ +polyA	EWSR1C +polyA	1-butanol	1-octanol ($\log_{10}$(PC))
ALA	0.772 ± 0.142	0.691 ± 0.006	0.282 ± 0.156	0.026 ± 0.005	-2.85 ± 0.285
ILE	0.766 ± 0.115	0.692 ± 0.012	0.371 ± 0.139	0.151 ± 0.006	-1.7 ± 0.17
LEU	0.756 ± 0.130	0.690 ± 0.035	0.310 ± 0.187	0.199 ± 0.0005	-1.52 ± 0.152
VAL	0.740 ± 0.133	0.682 ± 0.004	0.341 ± 0.105	0.051 ± 0.009	-2.26 ± 0.226
MET	0.872 ± 0.127	0.761 ± 0.064	0.273 ± 0.273	0.120 ± 0.024	-1.87 ± 0.187
PHE	1.024 ± 0.149	0.917 ± 0.022	0.740 ± 0.272	0.299 ± 0.016	-1.38 ± 0.138
TRP	2.273 ± 0.297	2.220 ± 0.062	3.547 ± 0.381	0.624 ± 0.014	-1.06 ± 0.106
TYR	1.346 ± 0.146	1.202 ± 0.006	1.619 ± 0.443	0.153 ± 0.029	-2.26 ± 0.226
CYS	0.755 ± 0.367	0.370 ± 0.021	n.d.	n.d.	-2.49 ± 0.249
PRO	0.797 ± 0.158	0.692 ± 0.076	0.435 ± 0.206	0.027 ± 0.002	-2.54 ± 0.254
GLY	0.815 ± 0.194	0.874 ± 0.239	0.195 ± 0.158	n.d.	-3.21 ± 0.321
dG	5.194 ± 0.712	4.345 ± 1.203	4.133 ± 2.603	0.309 ± 0.024	-1.3 ± 0.13
cGMP	3.625 ± 0.497	3.498 ± 0.233	8.750 ± 1.198	0.010 ± 0.004	n.d.
GTP	4.161 ± 0.570	4.023 ± 0.029	33.765 ± 4.761	n.d.	n.d.
dA	3.726 ± 0.511	2.753 ± 0.698	4.876 ± 0.694	1.265 ± 0.162	-0.55 ± 0.055
cAMP	2.766 ± 0.379	2.445 ± 0.028	5.695 ± 0.140	0.053 ± 0.004	-2.96 ± 0.296
ATP	2.733 ± 0.375	1.916 ± 0.217	19.380 ± 1.701	n.d.	n.d.

Table A.12: Table 2 (of 2) of partitioning coefficients (PC) obtained for biomolecular condensates and organic solvents.

A.8.2 Free energy of partitioning measurements obtained for the biomolecular condensates and organic solvents

Values of ΔG(kJ/mol) were calculated using ΔG = -RTln(PC)/1000, where R is equal to 8.314 J/(mol·K), T=277 K and the PC values are in tables A.11 and A.12. Errors ($\sigma_{\Delta G}$) were calculated using $\sigma_{\Delta G} = RT\sigma_{PC}/PC$. For 1-butanol, n.d. refers to molecules that were excluded from the results, and for which an upper limit of exclusion has been estimated and shown in Figure 3 (but not used for fitting). For 1-octanol, an error of 10% (in logP space) was used for all values, and n.d. refers to molecules for which experimental literature values were not found. n=3 for $DDX4^{N1}$, n=2 for $DDX4^{N1}$+polyA, EWSR1C+polyA and 1-butanol.

$\Delta G_{partitioning}$ (kJ/mol)					
Molecule	$DDX4^{N1}$	$DDX4^{N1}$ +polyA	EWSR1C +polyA	1-butanol	1-octanol
ARG	-1.160 ± 0.356	-1.184 ± 0.032	0.737 ± 0.532	9.120 ± 1.859	23.960 ± 2.195
HIS	0.152 ± 0.259	0.432 ± 0.158	1.965 ± 1.022	5.703 ± 2.157	18.940 ± 1.735
LYS	0.036 ± 0.461	0.319 ± 0.083	2.144 ± 0.134	n.d.	17.400 ± 1.594
ASP	1.143 ± 0.362	1.358 ± 0.002	2.096 ± 0.724	n.d.	22.192 ± 2.033
GLU	1.111 ± 0.207	1.437 ± 0.090	2.239 ± 1.241	10.392 ± 0.876	21.051 ± 1.928
ASN	0.524 ± 0.368	0.820 ± 0.021	4.045 ± 2.304	8.153 ± 1.304	21.792 ± 1.996
GLN	0.507 ± 0.608	0.861 ± 0.267	5.507 ± 1.651	7.553 ± 2.025	20.766 ± 1.902
THR	0.628 ± 0.442	0.963 ± 0.320	2.416 ± 0.721	n.d.	16.772 ± 1.537

Table A.13: Table 1 (of 2) of partitioning free energies ($\Delta G_{partitioning}$) obtained for biomolecular condensates and organic solvents.

$\Delta G_{partitioning}$ (kJ/mol)					
Molecule	$DDX4^{N1}$	$DDX4^{N1}$ +polyA	EWSR1C +polyA	1-butanol	1-octanol
ALA	0.596 ± 0.423	0.851 ± 0.019	2.919 ± 1.278	8.418 ± 0.478	16.259 ± 1.489
ILE	0.615 ± 0.345	0.847 ± 0.036	2.284 ± 0.862	4.363 ± 0.094	9.698 ± 0.888
LEU	0.645 ± 0.398	0.856 ± 0.117	2.695 ± 1.391	3.724 ± 0.006	8.671 ± 0.794
VAL	0.694 ± 0.415	0.881 ± 0.012	2.481 ± 0.712	6.869 ± 0.400	12.893 ± 1.181
MET	0.316 ± 0.335	0.628 ± 0.193	2.995 ± 2.304	4.886 ± 0.456	10.668 ± 0.977
PHE	-0.054 ± 0.336	0.200 ± 0.056	0.692 ± 0.847	2.783 ± 0.125	7.873 ± 0.721
TRP	-1.892 ± 0.301	-1.837 ± 0.065	-2.918 ± 0.247	1.085 ± 0.051	6.047 ± 0.554
TYR	-0.685 ± 0.249	-0.425 ± 0.012	-1.110 ± 0.630	4.328 ± 0.431	12.893 ± 1.181
CYS	0.647 ± 1.116	2.292 ± 0.134	n.d.	n.d.	14.205 ± 1.301
PRO	0.523 ± 0.456	0.848 ± 0.253	1.919 ± 1.092	8.339 ± 0.165	14.490 ± 1.327
GLY	0.470 ± 0.549	0.310 ± 0.631	3.773 ± 1.873	n.d.	18.312 ± 1.678
dG	-3.796 ± 0.316	-3.385 ± 0.638	-3.270 ± 1.451	2.712 ± 0.181	7.416 ± 0.679
cGMP	-2.967 ± 0.316	-2.886 ± 0.154	-4.998 ± 0.316	10.533 ± 0.887	n.d.
GTP	-3.285 ± 0.316	-3.207 ± 0.017	-8.110 ± 0.325	n.d.	n.d.
dA	-3.031 ± 0.316	-2.333 ± 0.585	-3.651 ± 0.328	-0.541 ± 0.296	3.138 ± 0.287
cAMP	-2.344 ± 0.316	-2.060 ± 0.027	-4.009 ± 0.057	6.750 ± 0.153	16.886 ± 1.547
ATP	-2.317 ± 0.316	-1.498 ± 0.261	-6.830 ± 0.202	n.d.	n.d.

Table A.14: Table 2 (of 2) of partitioning free energies ($\Delta G_{partitioning}$) obtained for biomolecular condensates and organic solvents.

A.9 Seedless optimised pulses for HACA(CO) (figures A.3-A.10)

The following seedless optimisation scripts were used to create and optimise ^{13}C pulses of lengths 200μs, 2000μs and 5000μs within the HACA(CO) sequence on a sample by sample basis. The optimised seedless outputs are shown in figures A.8-A.5.

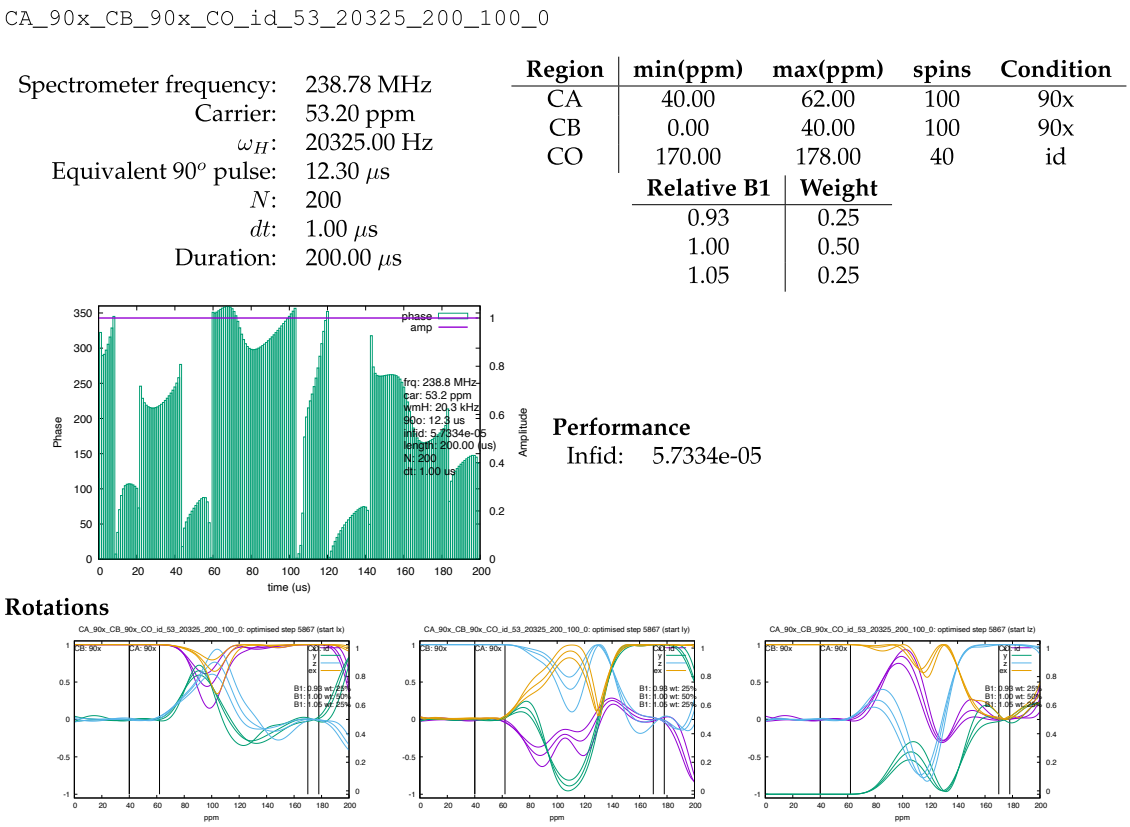


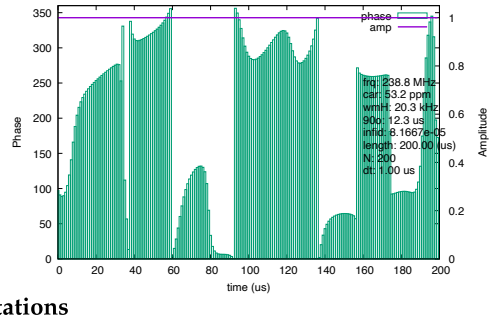
Figure A.3: An example seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 90^o pulse. The pulse provides unitary 90^o transformations to $^{13}C_\alpha$ and $^{13}C_\beta$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 20.3kHz pulse had a duration of 200μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_180x_CB_180x_CO_id_53_20325_200_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	53.20 ppm
ω_H:	20325.00 Hz
Equivalent 90^o pulse:	12.30 μs
N:	200
dt:	1.00 μs
Duration:	200.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	180x
CB	0.00	40.00	100	180x
CO	170.00	178.00	40	id

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance
Infid: 8.1667e-05

Rotations

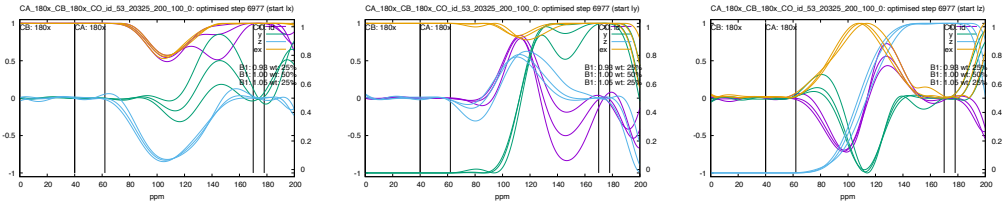


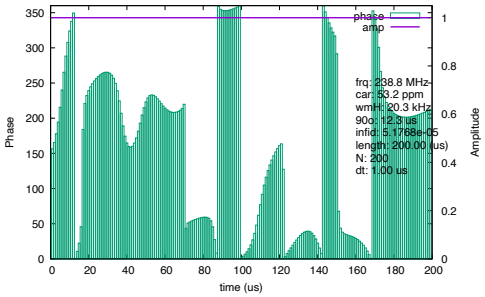
Figure A.4: An example seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 180^o pulse. The pulse provides unitary 180^o transformations to $^{13}C_\alpha$ and $^{13}C_\beta$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 20.3kHz pulse had a duration of 200μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_id_CB_id_CO_180x_53_20325_200_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	53.20 ppm
ω_H:	20325.00 Hz
Equivalent 90^o pulse:	12.30 μs
N:	200
dt:	1.00 μs
Duration:	200.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	id
CB	0.00	40.00	100	id
CO	170.00	178.00	40	180x

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance

Infid: 5.1768e-05

Rotations

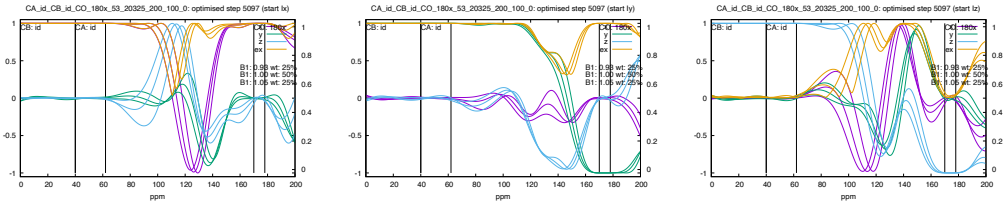


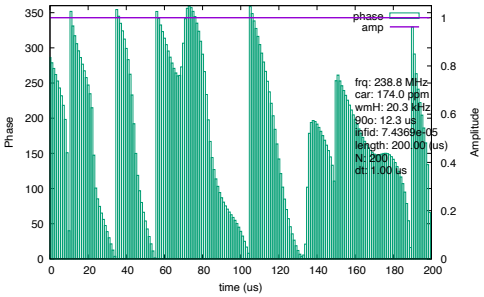
Figure A.5: An example seedless optimisation output for a $^{13}C_o$ selective pulse. The pulse provides a 180^o unitary transformation to $^{13}C_o$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 20.3kHz pulse had a duration of 200μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_180x_CB_180x_CO_id_174_20325_200_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	174.00 ppm
ω_H:	20325.00 Hz
Equivalent 90^o pulse:	12.30 μs
N:	200
dt:	1.00 μs
Duration:	200.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	180x
CB	0.00	40.00	100	180x
CO	170.00	178.00	40	id

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance
Infid: 7.4369e-05

Rotations

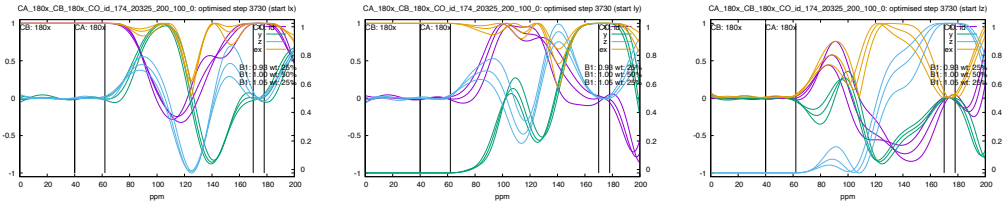


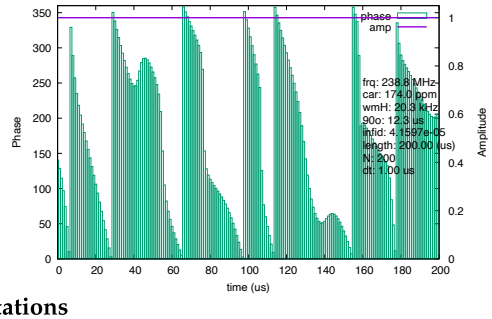
Figure A.6: An example seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_\beta$ selective 180^o pulse. The pulse provides unitary 180^o transformations to $^{13}C_\alpha$ and $^{13}C_\beta$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 20.3kHz pulse had a duration of 200μs with the carrier placed at 174.0ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_id_CB_id_CO_90x_174_20325_200_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	174.00 ppm
ω_H:	20325.00 Hz
Equivalent 90^o pulse:	12.30 μs
N:	200
dt:	1.00 μs
Duration:	200.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	id
CB	0.00	40.00	100	id
CO	170.00	178.00	40	90x

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance

Infid: 4.1597e-05

Rotations

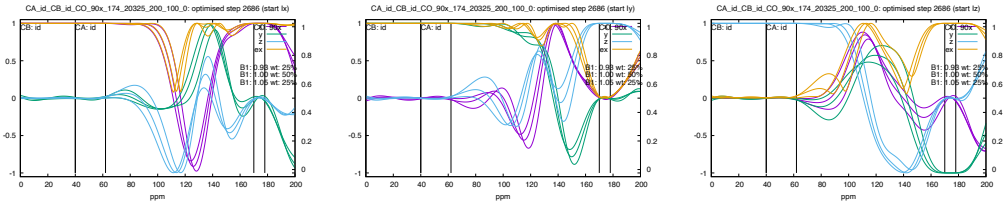


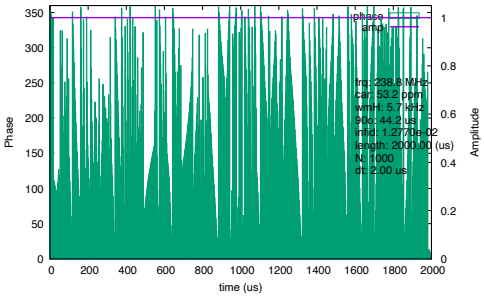
Figure A.7: An example seedless optimisation output for a $^{13}C_o$ selective pulse. The pulse provides a 90^o unitary transformation to $^{13}C_o$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 20.3kHz pulse had a duration of 200μs with the carrier placed at 174.0ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_180x_CB_IzIz_CO_IzIz_53_5661_2000_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	53.20 ppm
ω_H:	5661.00 Hz
Equivalent 90^o pulse:	44.16 μs
N:	1000
dt:	2.00 μs
Duration:	2000.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	180x
CB	0.00	40.00	100	Iz → Iz
CO	170.00	178.00	40	Iz → Iz

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance
Infid: 1.2770e-02

Rotations

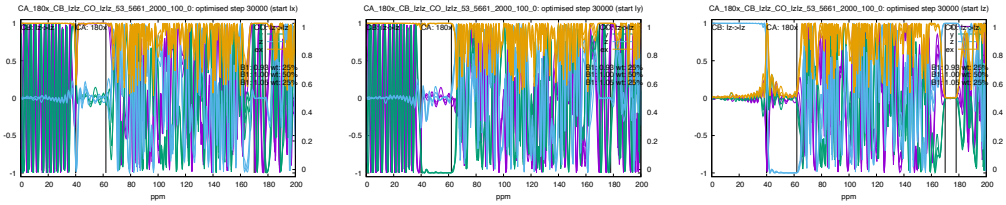


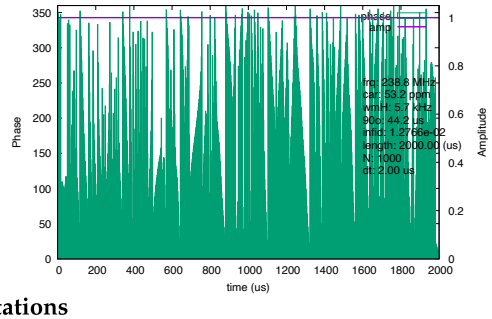
Figure A.8: An example seedless optimisation output for a $^{13}C_\alpha$ selective 180^o pulse. The pulse provides a unitary 180^o transformation to $^{13}C_\alpha$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 5.66kHz pulse had a duration of 2000μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

CA_180x_CB_IzIz_CO_Iz-Iz_53_5661_2000_100_0

Spectrometer frequency:	238.78 MHz
Carrier:	53.20 ppm
ω_H:	5661.00 Hz
Equivalent 90^o pulse:	44.16 μs
N:	1000
dt:	2.00 μs
Duration:	2000.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	40.00	62.00	100	180x
CB	0.00	40.00	100	Iz → Iz
CO	170.00	178.00	40	Iz → -Iz

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance

Infid: 1.2766e-02

Rotations

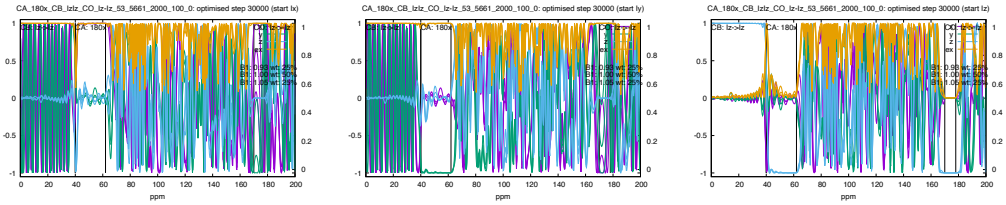


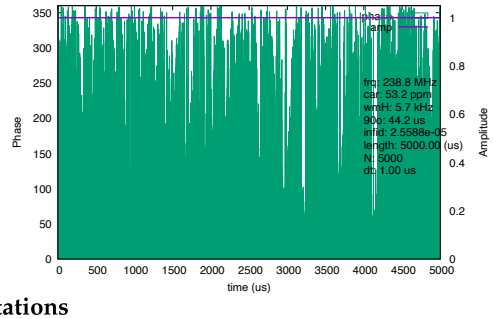
Figure A.9: An example seedless optimisation output for a $^{13}C_\alpha$+$^{13}C_o$ selective pulse. The pulse provides a unitary 180^o transformation to $^{13}C_\alpha$ spins, and z-inversion ($I_z \rightarrow -I_z$) to $^{13}C_o$ spins. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 5.66kHz pulse had a duration of 2000μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

`CA_id_CB_Iz-Iz_CO_IzIz_53_5661_5000_100_0`

Spectrometer frequency:	238.78 MHz
Carrier:	53.20 ppm
ω_H:	5661.00 Hz
Equivalent 90^o pulse:	44.16 μs
N:	5000
dt:	1.00 μs
Duration:	5000.00 μs

Region	min(ppm)	max(ppm)	spins	Condition
CA	50.00	62.00	100	id
CB	0.00	47.00	100	Iz → -Iz
CO	170.00	178.00	40	Iz → Iz

Relative B1	Weight
0.93	0.25
1.00	0.50
1.05	0.25



Performance
Infid: 2.5588e-05

Rotations

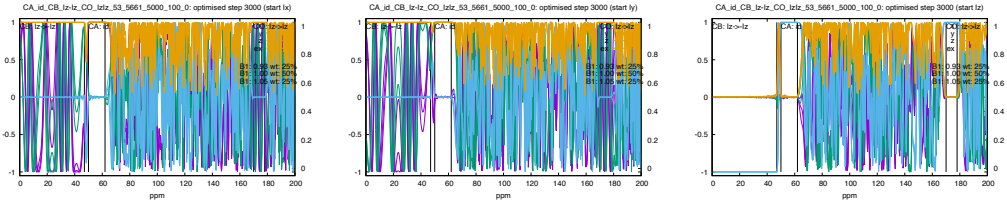


Figure A.10: An example seedless optimisation output for a $^{13}C_\beta$ selective pulse. The pulse provides z-inversion ($I_z \rightarrow -I_z$) to $^{13}C_\beta$ spins only. The pulse was optimised for a 950MHz spectrometer (238.78MHz for carbon). The 5.66kHz pulse had a duration of 5000μs with the carrier placed at 53.2ppm. The resulting pulse was applied in the HACA(CO) pulse sequence (Figure 3.14A)

A.10 TOCSY assignment of amino acids (figures A.11-A.17)

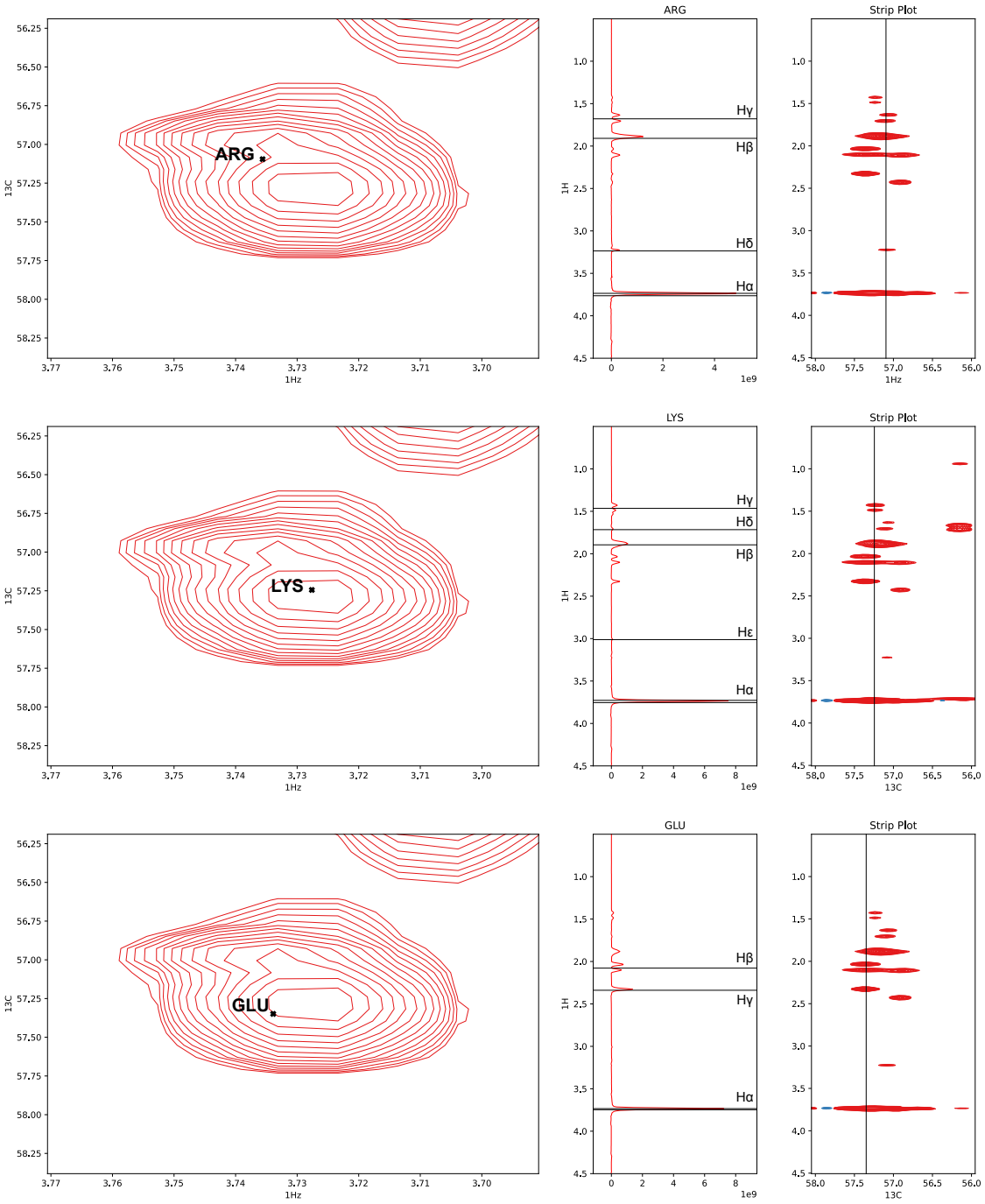


Figure A.11: The TOCSY based assignments of the H_α-C_α coherences for the amino acids ARG, LYS and GLU in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis.

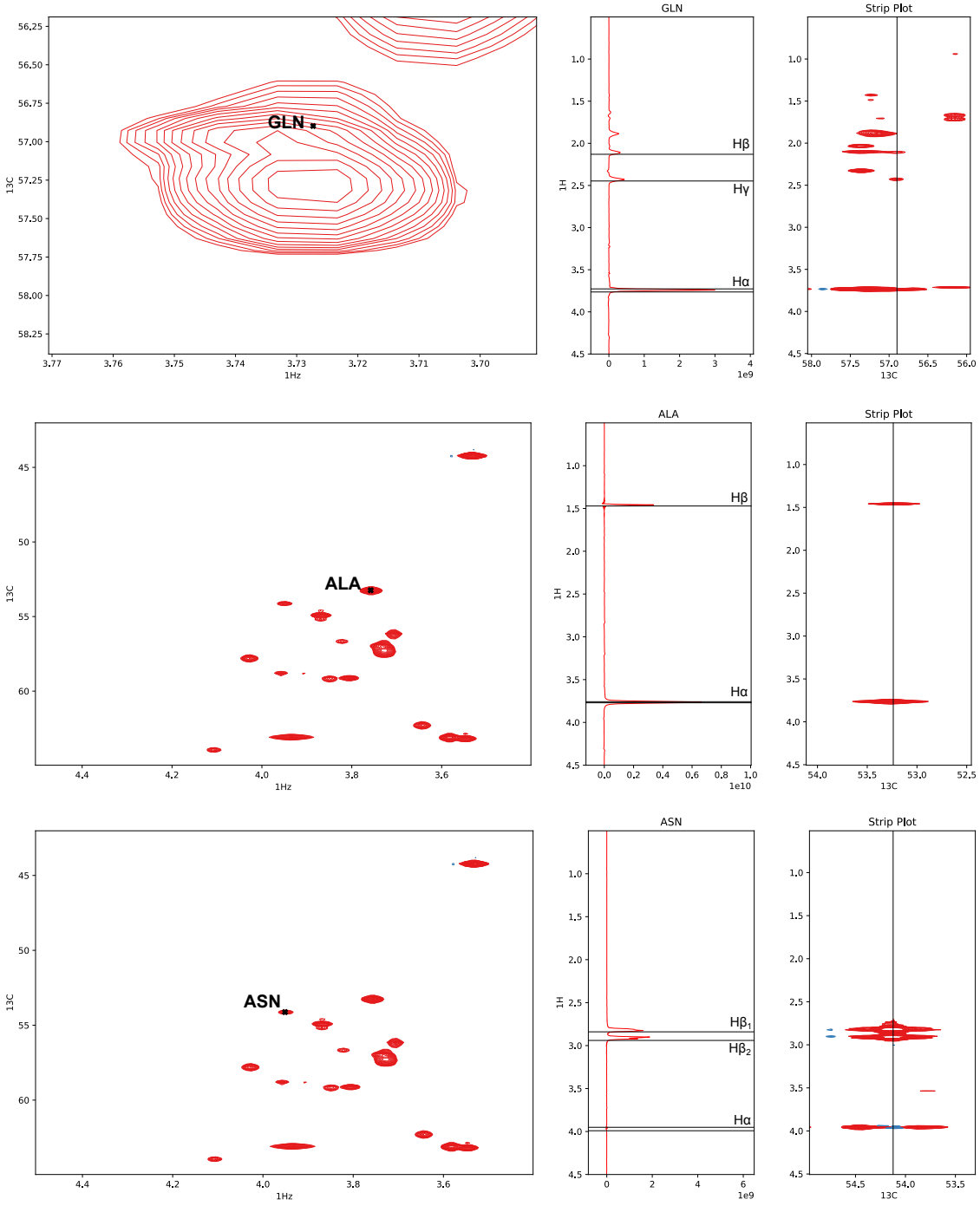


Figure A.12: The TOCSY based assignments of the H_α-C_α coherences for the amino acids GLN, ALA and ASN in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

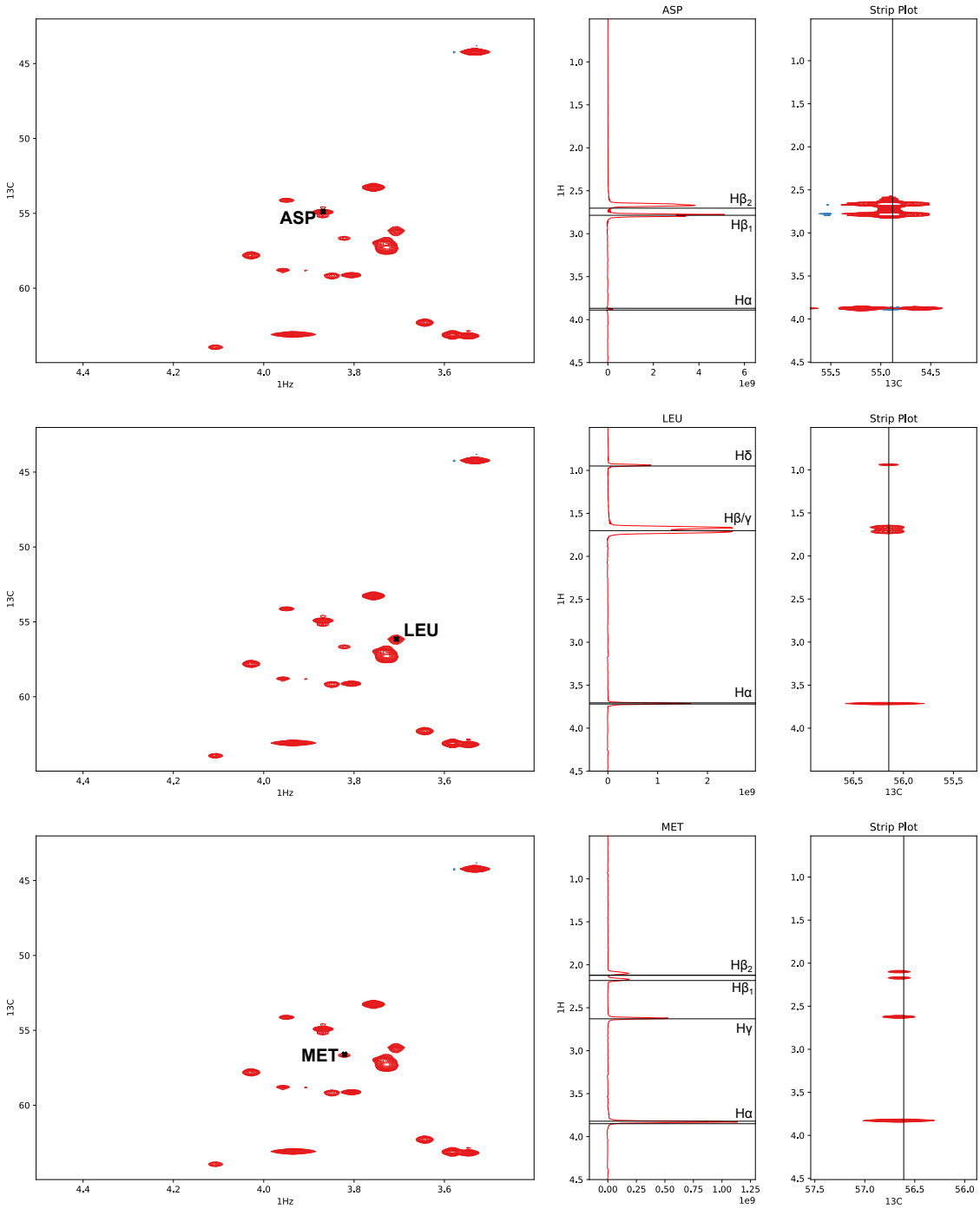


Figure A.13: The TOCSY based assignments of the H_α-C_α coherences for the amino acids ASP, LEU and MET in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

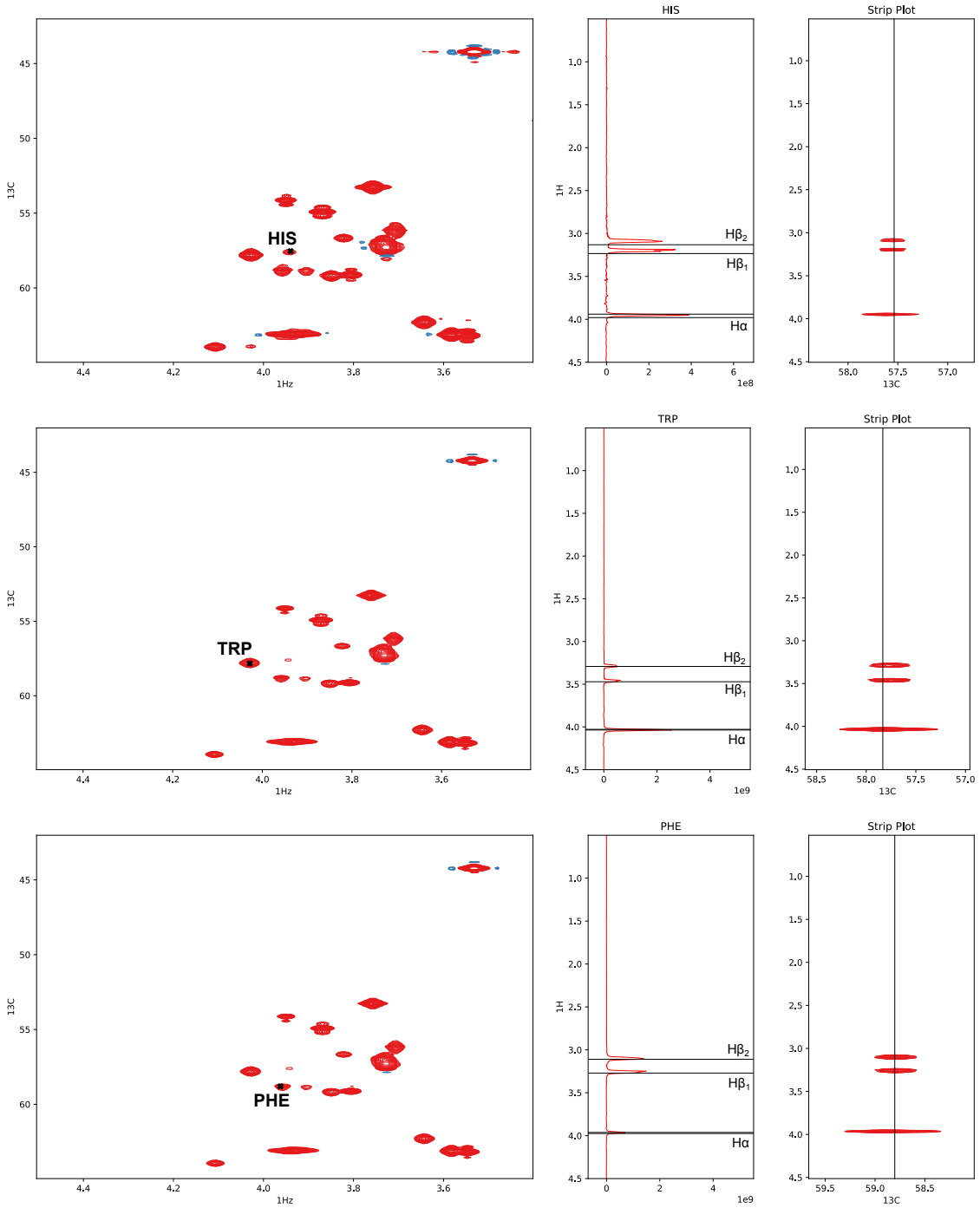


Figure A.14: The TOCSY based assignments of the H_α-C_α coherences for the amino acids HIS, TRP and PHE in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

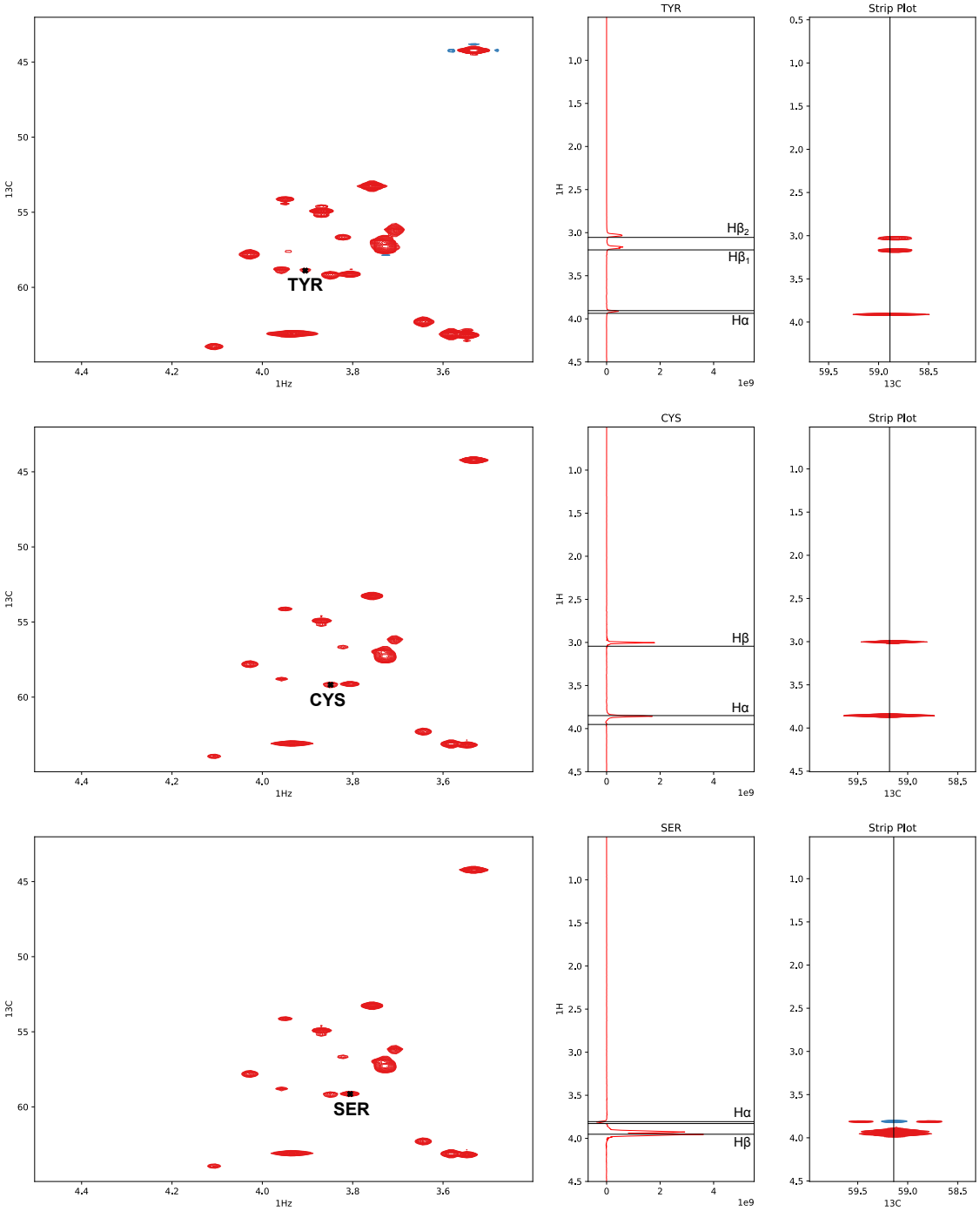


Figure A.15: The TOCSY based assignments of the H_α-C_α coherences for the amino acids TYR, CYS and SER in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

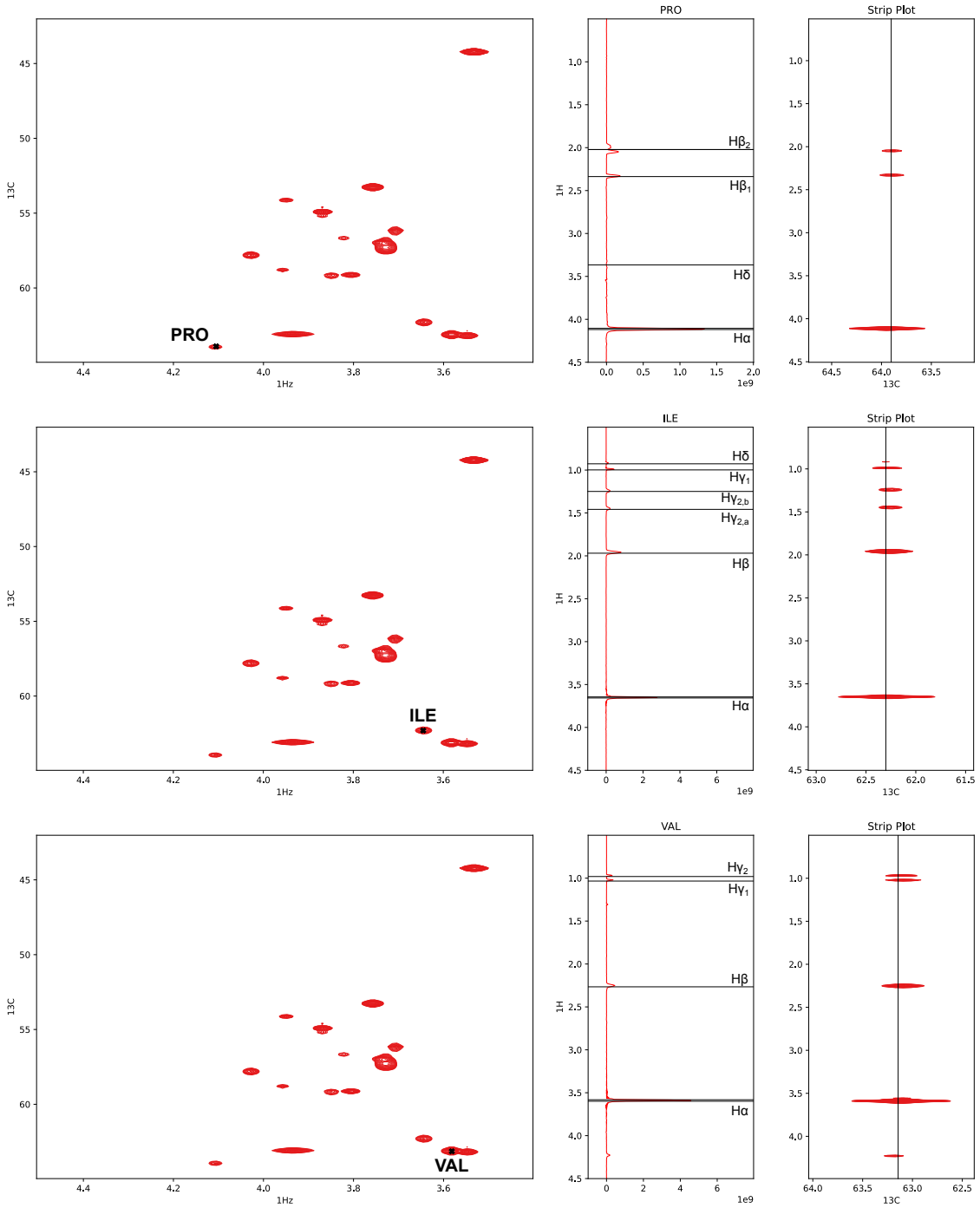


Figure A.16: The TOCSY based assignments of the H_α-C_α coherences for the amino acids PRO, ILE and VAL in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

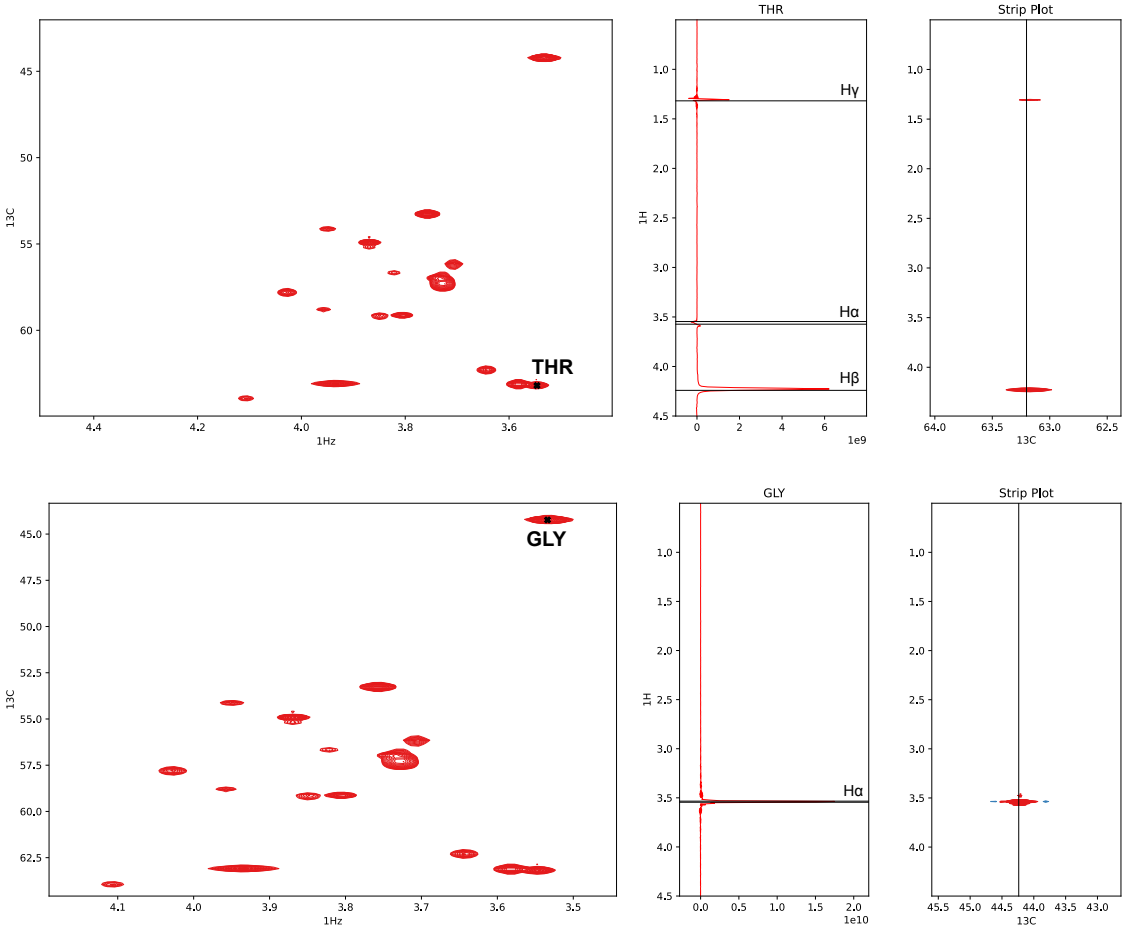


Figure A.17: The TOCSY based assignments of the H_α-C_α coherences for the amino acids GLY and THR in the ^{13}C,^{15}N amino acid mix used to measure amino acid partitioning in this thesis. Proton peak positions are shown and were compared to the expected proton free amino acid chemical shifts obtained from the BMRB (https://bmrb.io).

A.11 NMR Assignment of Nucleotides

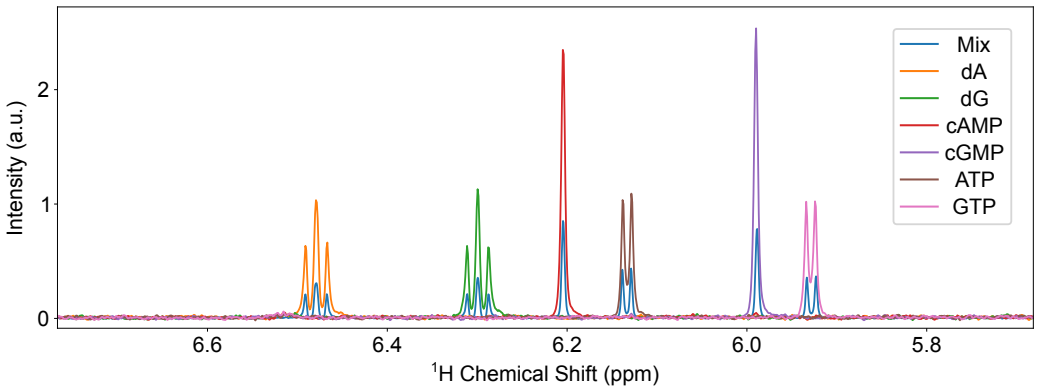


Figure A.18: The assignments of the nucleotide proton chemical shift resonances used in this paper to measure the partitioning of dA, dG, cAMP, cGMP, ATP, GTP. These given peaks were chosen as they were in a region of the proton spectrum far from any background protein or buffer component resonances.

A.12 Metabolite Chemical Properties

Molecule	Aromatic Rings	Phosphate Groups	Positive Charges	Negative Charges	Polar Surface Area (Å^2)	RB
ALA	0	0	1	1	63.3	1
ARG	0	0	2	1	125.22	5
ASN	0	0	1	1	106.41	3
ASP	0	0	1	2	100.62	3
CYS	0	0	1	2	63.32	2
GLN	0	0	1	1	100.62	4
GLU	0	0	1	2	106.41	4
GLY	0	0	1	1	63.32	1
HIS	1	0	1	1	92	3
ILE	0	0	1	1	63.32	3
LEU	0	0	1	1	63.32	3
LYS	0	0	2	1	89.34	5
MET	0	0	1	1	63.32	4
PHE	1	0	1	1	63.32	3
PRO	0	0	1	1	49.33	1
THR	0	0	1	1	83.55	2
TRP	2	0	1	1	79.11	3
TYR	1	0	1	1	83.55	3
VAL	0	0	1	1	63.32	2
cGMP	2	1	0	1	170.52	1
dG	2	0	0	0	134.99	2
GTP	2	3	0	4	294.81	8
dA	2	0	0	0	119.31	2
cAMP	2	1	0	1	154.84	1
ATP	2	3	0	4	279.13	8

Table A.15: Metabolite properties table one (of two) used for the metabolite partitioning fitting procedure. Data obtained from the Human Metabolome Database (HMDB).[282] RB represents rotatable bonds.

Molecule	α (Å^3)	R_m (m^3mol^{-1})	Molecular weight (Da)	Hydrogen bond acceptors	Hydrogen bond donors	Hydrogen bond donors and acceptors
ALA	8.49	20.5	89.09	3	2	5
ARG	17.8	53.92	174.2	6	5	11
ASN	11.68	28.35	132.12	4	3	7
ASP	11.28	26.53	133.1	5	3	8
CYS	11.41	28.22	121.158	3	3	6
GLN	13.87	33.11	146.14	5	3	8
GLU	13.32	31.29	147.12	4	3	7
GLY	6.65	16	75.07	3	2	5
HIS	14.67	38.06	155.15	4	3	7
ILE	14.11	34.09	131.17	3	2	5
LEU	14.16	34.17	131.17	3	2	5
LYS	15.84	37.81	146.19	4	3	7
MET	15.5	37.59	149.21	3	2	5
PHE	17.03	45.12	165.19	3	2	5
PRO	11.5	28.06	115.13	3	2	5
THR	11.08	26.46	119.12	4	3	7
TRP	21.05	56.2	204.23	3	3	6
TYR	18.01	47.1	181.19	4	3	7
VAL	12.19	29.49	117.15	3	2	5
cGMP	29.19	71.72	345.21	8	4	12
dG	25.24	63.11	267.24	7	4	11
GTP	39.81	97.24	523.18	14	8	22
dA	24.24	61.68	251.24	7	3	10
cAMP	28.42	70.29	329.21	8	3	11
ATP	38.92	95.81	507.181	14	7	21

Table A.16: Metabolite properties table two (of two) used for the metabolite partitioning fitting procedure. Data obtained from the Human Metabolome Database (HMDB).[282] α and R_m represent polarisability and refractivity.

A.13 Metabolite partitioning fitting results

A.13.1 Results of fitting metabolite properties to metabolite partitioning free energies (grouped properties fitting procedure)

No. molecular properties	No. tested models	No. fitting parameters (not joint)	No. fitting parameters (joint)	Total RMSD of winner (not joint)	Total RMSD of winner (joint)
1	12	10	8	2.4743	2.4765
2	38	15	12	1.7182	1.7215
3	52	20	16	1.5036	1.5077
4	33	25	20	1.4060	1.4105
5	**8**	**30**	**24**	**1.1621**	**1.1678**

No. molecular properties	Pass F-test vs lower complexity model	Joint wins F-test against not joint	VIF
1	N/A	True	1.000
2	True	True	1.754
3	True	True	3.481
4	True	True	4.307
5	**True**	**True**	**4.718**

Table A.17: A table of fitting results for global metabolite partitioning.Metabolite properties were separated into five groups using a clustering algorithm (Figure 3.21A). This prevented heavily correlated properties from being included in the same fit. The total number of possible molecular property combinations after grouping was 143, with a maximum of 5 properties able to be fit together. Joint/non-joint represent the results of the fitting procedure when the fitted coefficients for the partitioning into $DDX4^{N1}$ and $DDX4^{N1}$+polyA condensates are forced to be the same, or allowing them to differ, respectively. All statistical F-tests were performed using $p<0.05$. VIF (Variance Inflation Factor) is a measure of multicollinearity between the metabolite properties included in the fit; it is desirable to ensure the VIF is less than 5 to prevent overfitting.[287] The line in bold is highlighted as the winning model (with non-joint fitting not statistically justified).

A.13.2 Results of fitting metabolite properties to metabolite partitioning free energies (naïve fitting procedure)

No. molecular properties	No. tested models	No. fitting parameters (not joint)	No. fitting parameters (joint)	Total RMSD (not joint)	Total RMSD (joint)
1	12	10	8	2.4743	2.4765
2	66	15	12	1.6388	1.6422
3	220	20	16	1.0954	1.1008
4	495	25	20	0.9518	0.9582
5	792	30	24	0.8257	0.8332
6	924	35	28	0.7299	0.7395
7	**792**	**40**	**32**	**0.6257**	**0.6377**
8	495	45	36	0.6112	0.6236

No. molecular properties	Pass F-test vs lower complexity model (not joint)	Joint wins F-test against not joint	VIF (Variance Inflation Factor)
1	N/A	True	1.000
2	True	True	5.354
3	True	True	23.914
4	True	True	32.880
5	True	True	66.696
6	True	True	55.057
7	**True**	**True**	**74.819**
8	False	True	195.028

Table A.18: A table of fitting results for global metabolite partitioning without using metabolite property grouping (naïve fitting). The total number of possible molecular property combinations was 4096. Joint/non-joint represent the results of the fitting procedure when the fitted coefficients for the partitioning into $DDX4^{N1}$ and $DDX4^{N1}$+polyA condensates are forced to be the same, or allowing them to differ, respectively. All statistical F-tests were performed using $p<0.05$. VIF (Variance Inflation Factor) is a measure of multicollinearity between the metabolite properties included in the fit; it is desirable to ensure the VIF is less than 5 to prevent overfitting.[287] The line in bold is highlighted as the winning model (with non-joint fitting not statistically justified).

A.13.3 Fitted coefficients from the naïve fitting procedure

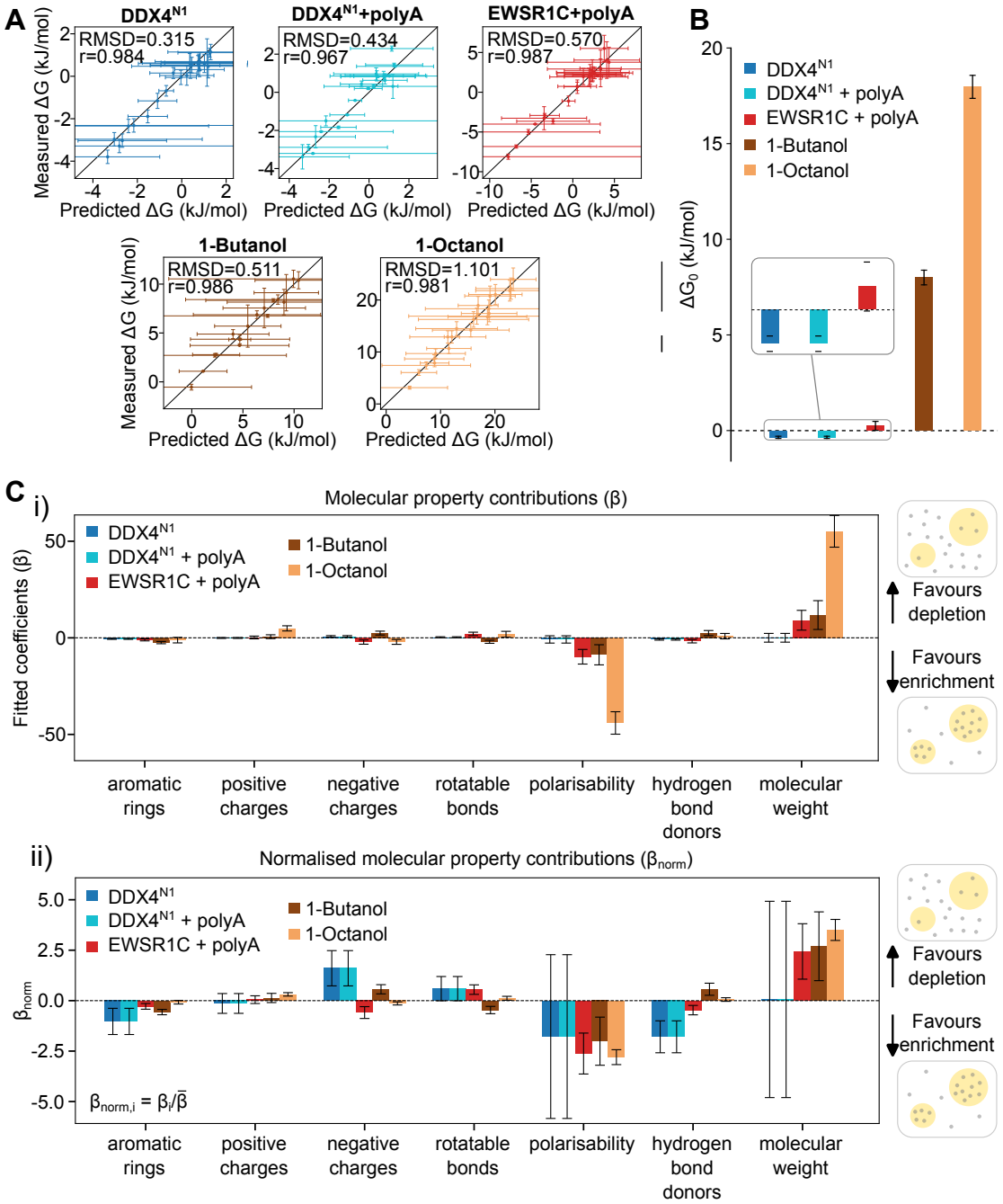


Figure A.19: **A** Plots showing the correlation between the fitted/predicted and measured partitioning free energies (ΔG) for each of the condensate and organic solvent systems for the winner of the naïve fitting procedure. **B** The fitted ΔG_0 values for each phase separating system, representing the free energy of partitioning for a hypothetical molecule whose properties are the mean of the properties for all the amino acids and nucleotides in the dataset. **C** The non-normalised (i) and normalised (ii) standardised coefficients obtained from the fitting procedure.

A.13.4 Fitted coefficients from the grouped fitting (using χ^2)

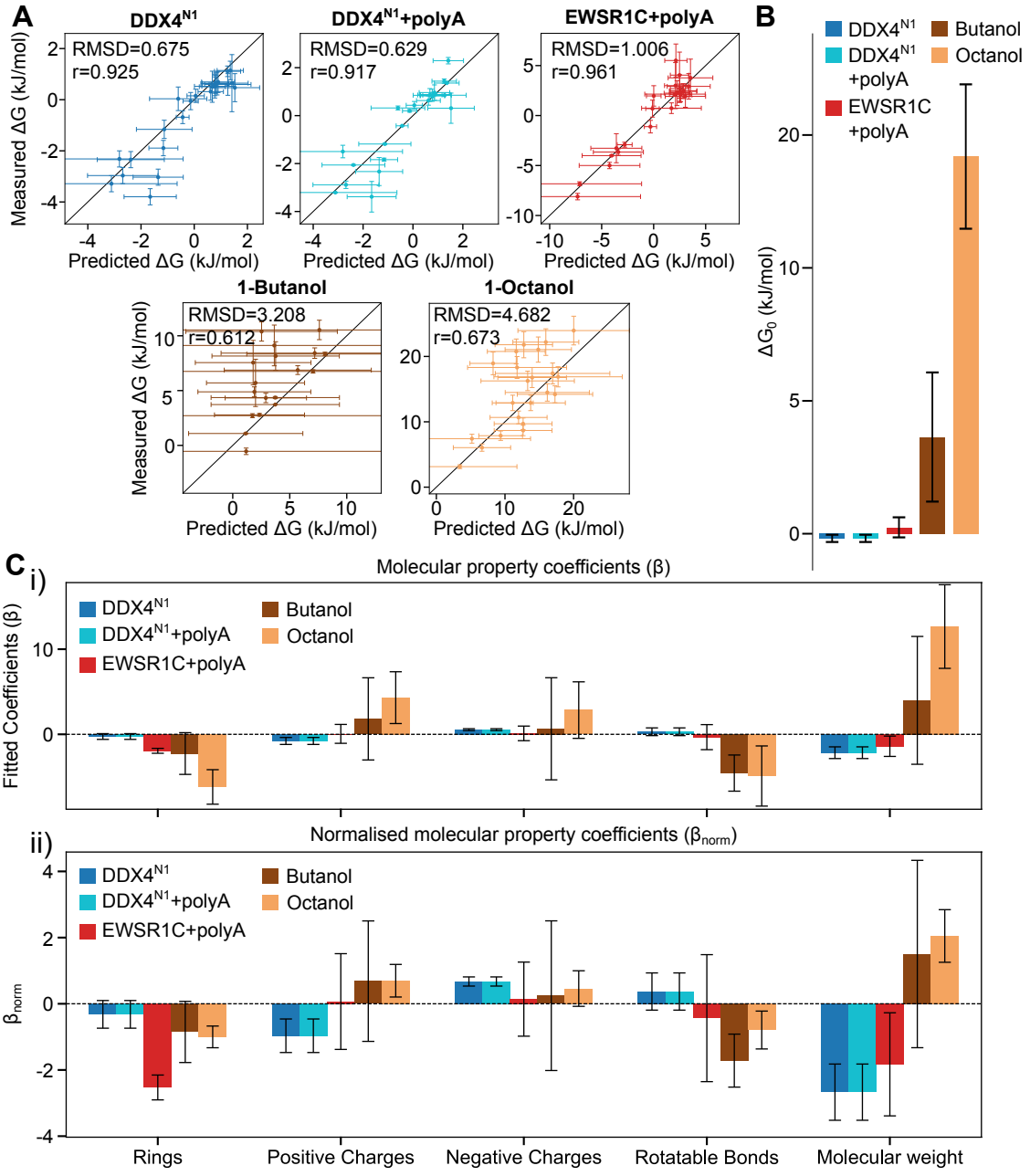


Figure A.20: **A** Plots showing the correlation between the fitted/predicted and measured partitioning free energies (ΔG) for each of the condensate and organic solvent systems for the winner of the grouped fitting procedure (minimising χ^2). **B** The fitted ΔG_0 values for each phase separating system, representing the free energy of partitioning for a hypothetical molecule whose properties are the mean of the properties for all the amino acids and nucleotides in the dataset. **C** The non-normalised (i) and normalised (ii) standardised coefficients obtained from the fitting procedure.

A.13.5 Bootstrap histograms showing fitted parameter errors for the naïve fitting procedure

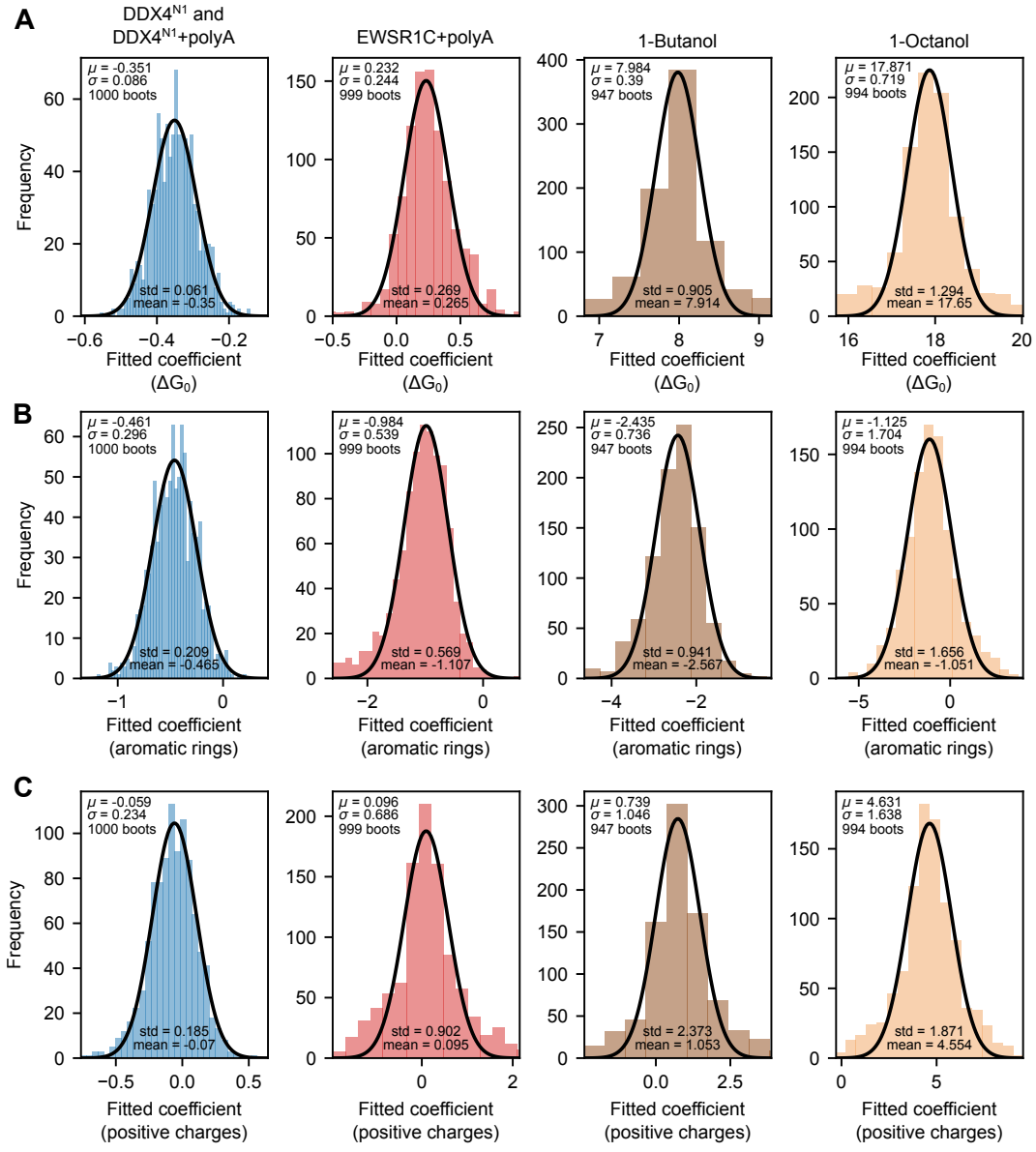


Figure A.21: Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure for ΔG_0 (**A**), aromatic rings (**B**) and positive charges (**C**). Fitted coefficient errors were taken as the numerical standard deviation (std). The histograms were also fitted to a gaussian curve with the fitted mean (μ) and standard deviation (σ) shown. A total of 1000 bootstraps were performed and fitted parameters further than +/-100 from the mean value were omitted from the histograms and error calculation. The number of bootstraps used to calculate the errors is shown in each case (boots).

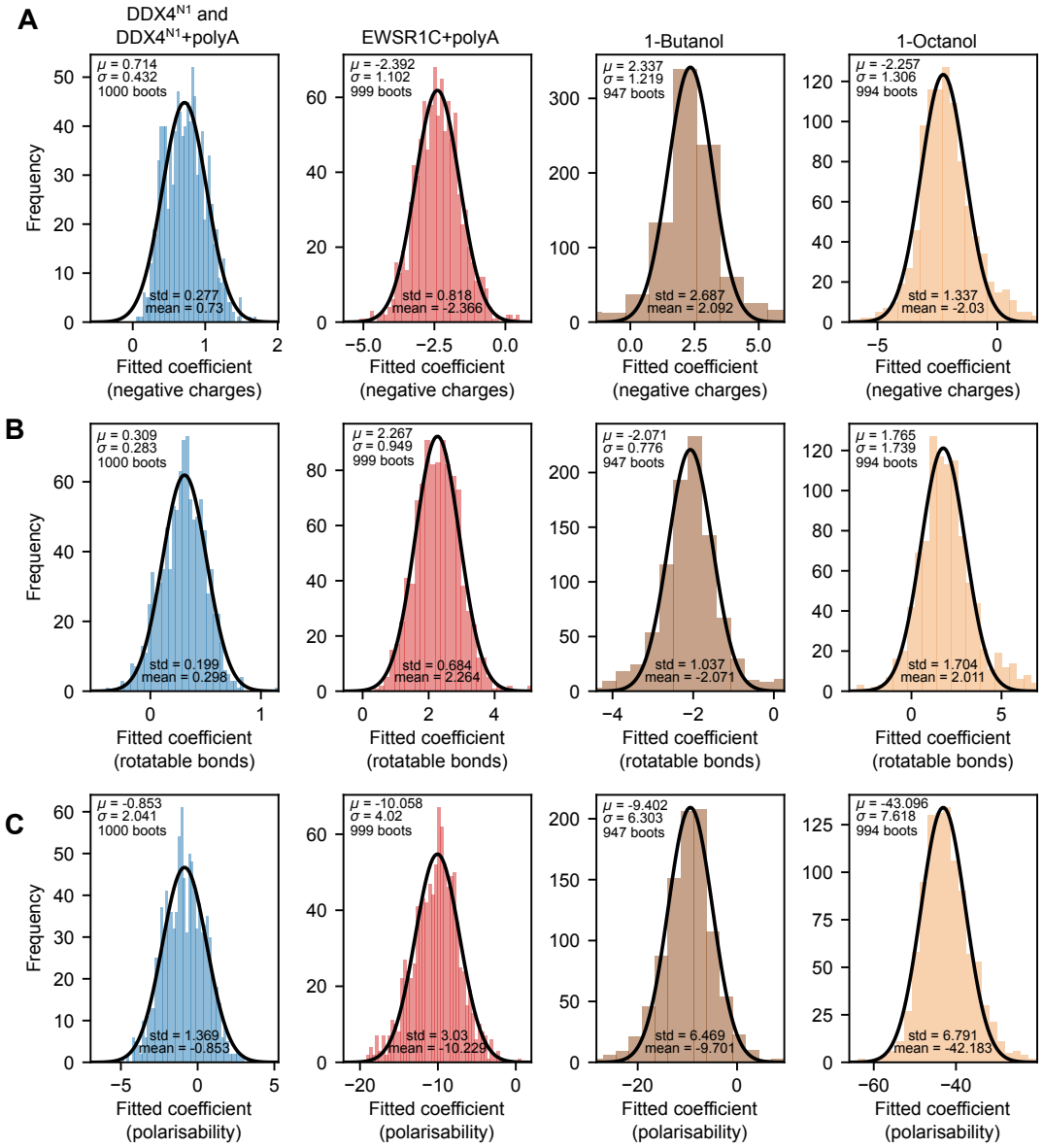


Figure A.22: Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure for negative charges (**A**), rotatable bonds (**B**) and polarisability (**C**). Fitted coefficient errors were taken as the numerical standard deviation (std). The histograms were also fitted to a gaussian curve with the fitted mean (μ) and standard deviation (σ) shown. A total of 1000 bootstraps were performed and fitted parameters further than +/-100 from the mean value were omitted from the histograms and error calculation. The number of bootstraps used to calculate the errors is shown in each case (boots).

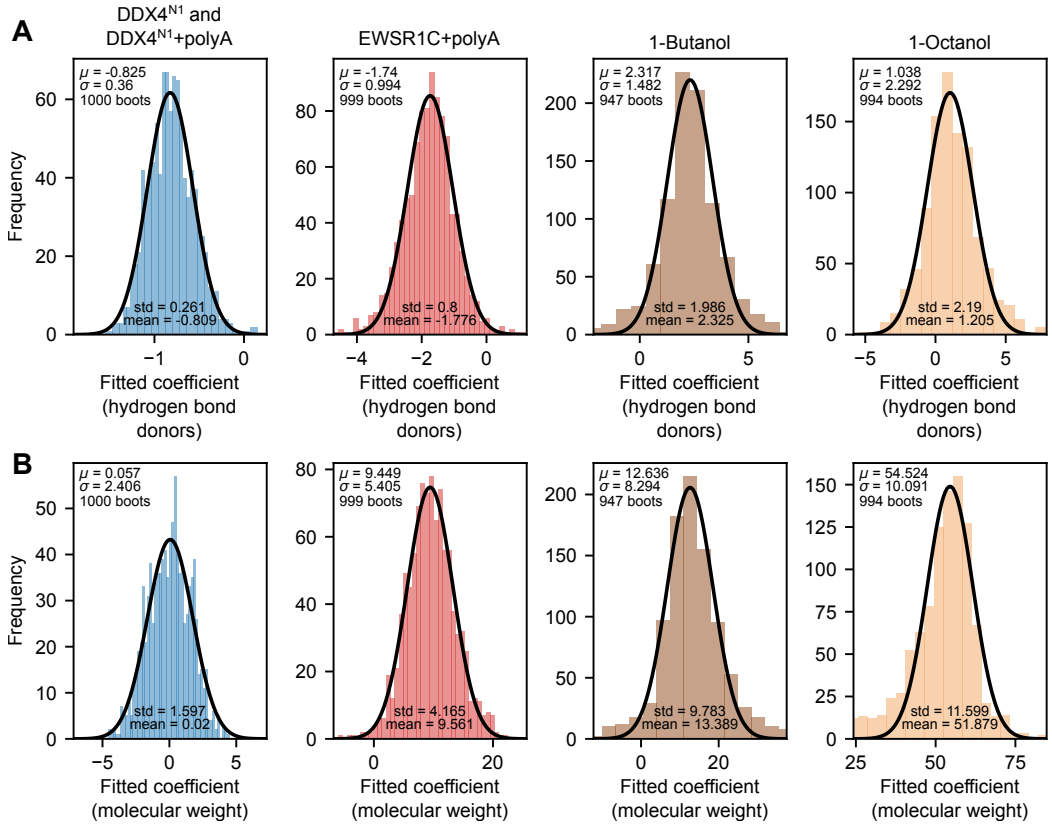


Figure A.23: Histogram of bootstrap fitted coefficients for the the winner of the naïve fitting procedure for hydrogen bond donors (**A**) and molecular weight (**B**). Fitted coefficient errors were taken as the numerical standard deviation (std). The histograms were also fitted to a gaussian curve with the fitted mean (μ) and standard deviation (σ) shown. A total of 1000 bootstraps were performed and fitted parameters further than +/-100 from the mean value were omitted from the histograms and error calculation. The number of bootstraps used to calculate the errors is shown in each case (boots).

A.13.6 Bootstrap histograms showing fitted parameter errors for the grouped fitting procedure

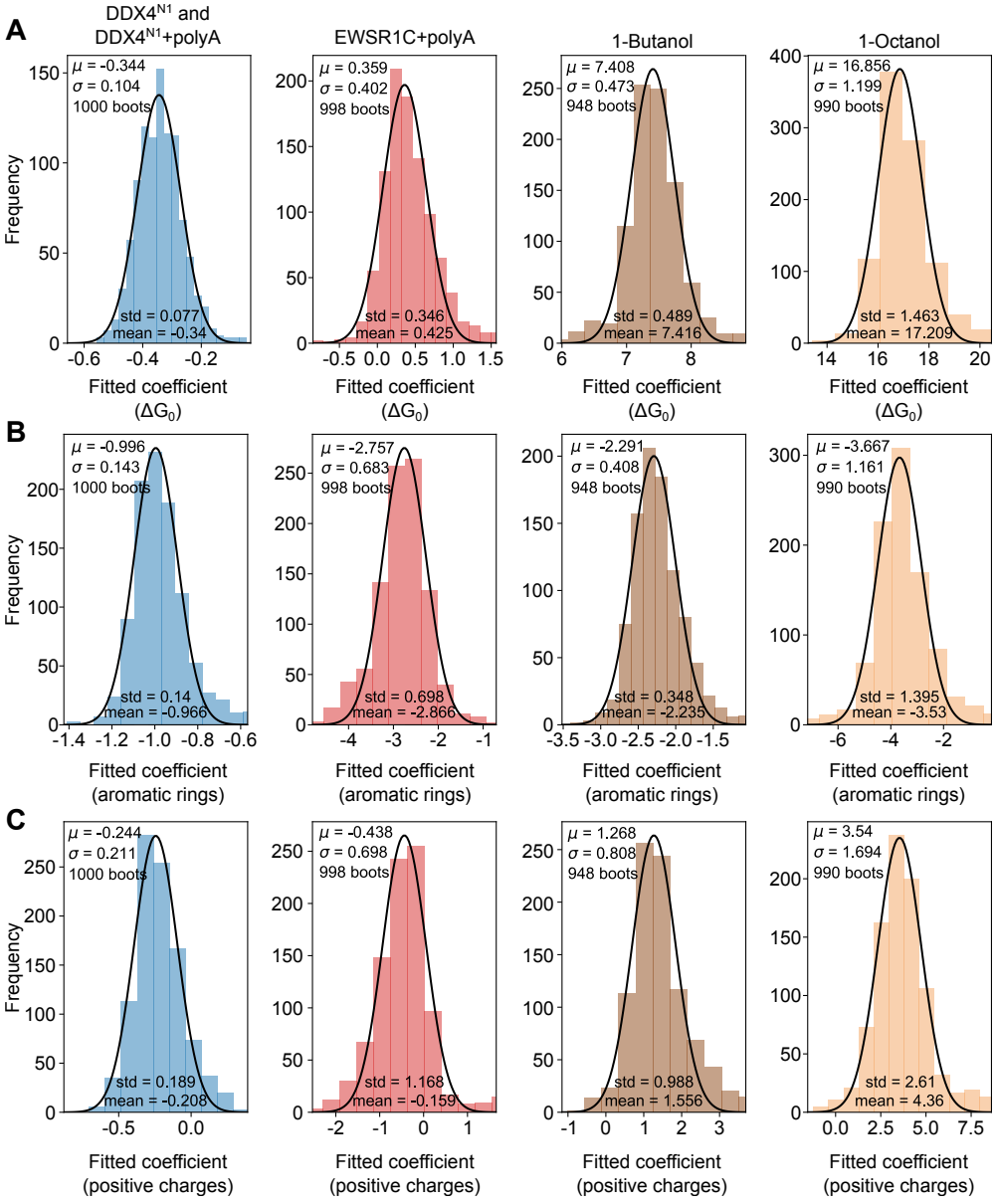


Figure A.24: Histogram of bootstrap fitted coefficients for the the winner of the grouped properties fitting procedure for ΔG_0 (**A**), aromatic rings (**B**) and positive charges (**C**). Fitted coefficient errors were taken as the numerical standard deviation (std). The histograms were also fitted to a gaussian curve with the fitted mean (μ) and standard deviation (σ) shown. A total of 1000 bootstraps were performed and fitted parameters further than +/-100 from the mean value were omitted from the histograms and error calculation. The number of bootstraps used to calculate the errors is shown in each case (boots).

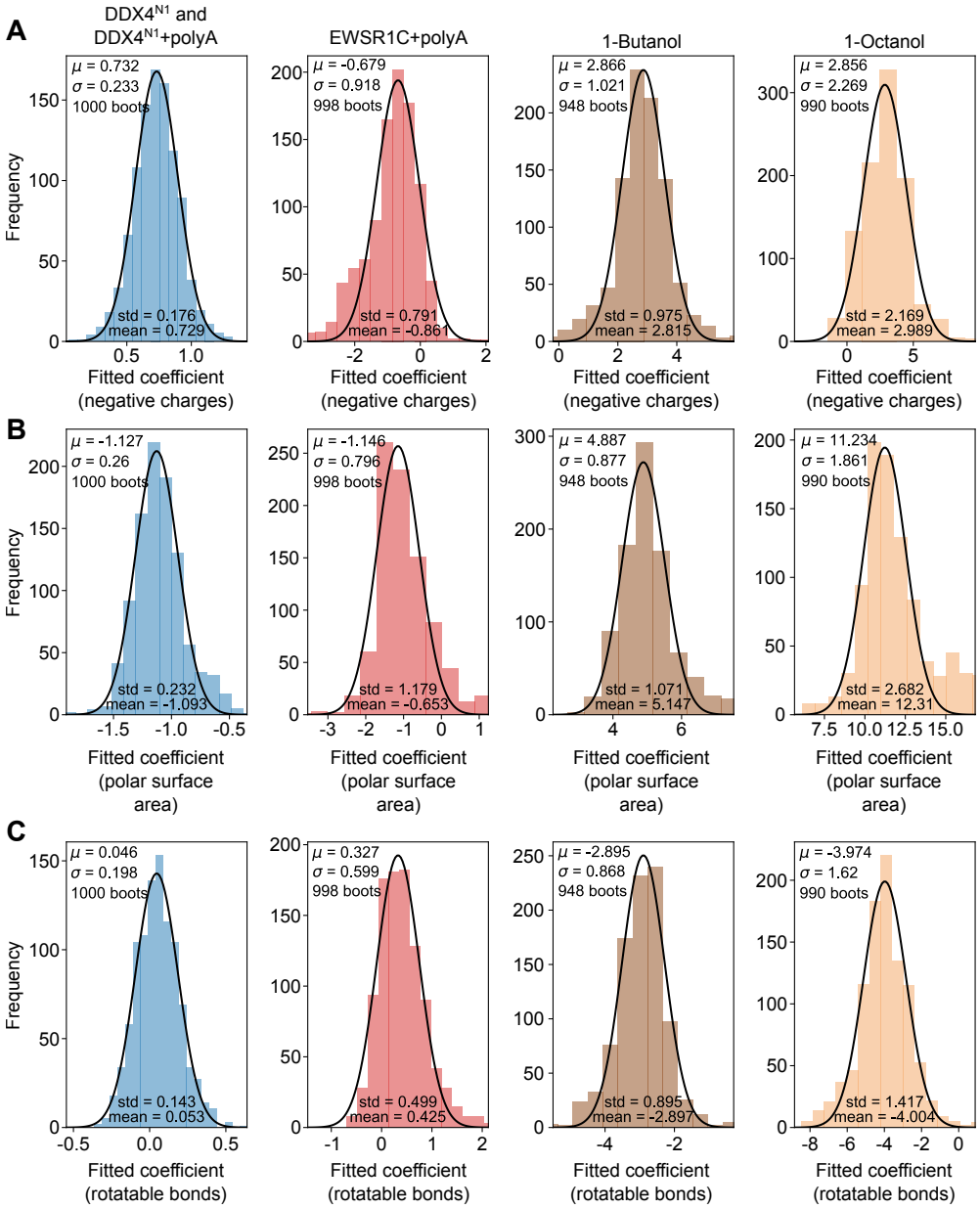


Figure A.25: Histogram of bootstrap fitted coefficients for the the winner of the grouped properties fitting procedure for negative charges (**A**), polar surface area (**B**) and rotatable (**C**). Fitted coefficient errors were taken as the numerical standard deviation (std). The histograms were also fitted to a gaussian curve with the fitted mean (μ) and standard deviation (σ) shown. A total of 1000 bootstraps were performed and fitted parameters further than +/-100 from the mean value were omitted from the histograms and error calculation. The number of bootstraps used to calculate the errors is shown in each case (boots).

A.14 Flory-Huggins Phase Separation Theory

This section outlines a derivation of the Gibbs free energy of mixing (ΔG_{mix}) for a polymer solution, a key expression used to model biological phase separation. This description of phase separation was originally provided by Flory and Huggins independently in 1942.[201,202] This derivation is based on the book 'Principles of polymer chemistry' by Paul J. Flory[487] and the book 'Polymers: chemistry and physics of modern materials' by Cowie *el al.*[488]

A.14.0.i Assumptions of the theory

Flory-Huggins theory assumes that a solution consists of N_{tot} lattice sites on a grid lattice. For a polymer solution, it is assumed that each polymer consists of r statistical segments, whose orientations are independent from one another, and that each statistical segment takes up the volume of a single lattice site. Therefore the total number of lattice sites (N_{tot}), is given by:

$$N_{tot} = N_1 + rN_2 \tag{A.11}$$

where N_1 is the number of solvent molecules and N_2 is the number of polymers in the system. Furthermore, it is assumed that each lattice site is surrounded by z other lattice sites (known as the coordination number). It is also assumed that the total volume of the system (and number of lattice sites) remains constant upon mixing.

To determine an expression for the free energy of mixing (ΔG_{mix}), expressions for the enthalpy (ΔH_{mix}) and entropy (ΔS_{mix}) of mixing need to be established and inserted into equation A.12.

$$\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix} \tag{A.12}$$

A.14.1 Calculation of entropy of mixing (ΔS_{mix})

An expression for the entropy of mixing of a polymer solution can be calculated by independently calculating expressions for the mixed and unmixed entropies (equation A.13).

$$\Delta S_{mix} = S_{mixed} - S_{unmixed} \tag{A.13}$$

Each entropy term can be calculated using the Boltzmann entropy formula $S=k_B\ln(\Omega)$, where k_B is the Boltzmann constant, and Ω represents the number of possible translational configurations of the system. Firstly, an expression of the mixed entropy will be determined (S_{mixed}). The mixed state will have an entropy according to the following expression:

$$S_{mixed} = k_B \left(\ln(\Omega_1) + \ln(\Omega_2)\right) \tag{A.14}$$

where Ω_1 and Ω_2 represent the the number of configurations of N_1 solvent molecules and N_2 polymer molecules on the lattice, respectively. The number of configurations of N_2 polymers on the lattice (Ω_2) will be considered first. If the lattice contains N_1 solvent molecules and i polymers, the number of lattice sites remaining is given by $N_{remaining} = N - ri$. Therefore the probability of a lattice site being available for the first statistical segment of an additional polymer (i+1) being added is given by p_i = (N-ri)/N. After the first statistical segment has been laid, the expected number of empty lattice sites available for the next statistical segment is given by:

$$zp_i = z\frac{N - ri}{N} \tag{A.15}$$

Each additional statistical segment being added (of which there is a total of r-2) has (z-1)p_i sites to be placed. Therefore, the total number of permutations of the

i+1 polymer on the lattice (ω_{i+1}) can be expressed as:

$$\begin{aligned}\omega_{i+1} &= \mathrm{N}_{\mathrm{remaining}} \times zp_i \times [(z-1)p_i]^{r-2} \\ &= (\mathrm{N} - ri) \times z\frac{\mathrm{N} - ri}{\mathrm{N}} \times \left[(z-1)\frac{(\mathrm{N} - ri)}{\mathrm{N}}\right]^{r-2}\end{aligned} \tag{A.16}$$

As the coordination number of a polymer solution is often large, it can be assumed that $z = z + 1$. Therefore the equation for ω_{i+1} (equation A.16) can be simplified to:

$$\omega_{i+1} = (\mathrm{N} - ri)^r \left[\frac{(z-1)}{N}\right]^{r-1} \tag{A.17}$$

Therefore:

$$\omega_i = (\mathrm{N} - r(i-1))^r \left[\frac{(z-1)}{N}\right]^{r-1} \tag{A.18}$$

The total number of configurations of N_2 polymers on the lattice can be calculated through:

$$\Omega_2 = \frac{\prod_{i=1}^{\mathrm{N}_2} \omega_i}{\mathrm{N}_2!} \tag{A.19}$$

where $\mathrm{N}_2!$ is required to prevent overcounting of configurations because all the polymers are indistinguishable from each other. Inserting equation A.18 into equation A.19 leads to:

$$\begin{aligned}\Omega_2 &= \frac{1}{\mathrm{N}_2!} \prod_{i=1}^{\mathrm{N}_2} (\mathrm{N} - r(i-1))^r \left[\frac{(z-1)}{\mathrm{N}}\right]^{r-1} \\ &= \frac{1}{\mathrm{N}_2!} \left(\frac{(z-1)}{\mathrm{N}}\right)^{\mathrm{N}_2(r-1)} \prod_{i=1}^{\mathrm{N}_2} (\mathrm{N} - r(i-1))^r\end{aligned} \tag{A.20}$$

The product on the right hand side of equation A.20 can be evaluated as follows:

$$\begin{aligned}\prod_{i=1}^{\mathrm{N}_2}(\mathrm{N}-r(i-1))^r &= \prod_{i=1}^{\mathrm{N}_2}\left[\left(\frac{\mathrm{N}-r(i-1)}{r}\right)r\right]^r \\ &= r^{r\mathrm{N}_2}\prod_{i=1}^{\mathrm{N}_2}\left(\frac{\mathrm{N}}{r}-i+1\right)^r \\ &= r^{r\mathrm{N}_2}\left[\left(\frac{\mathrm{N}}{r}\right)\left(\frac{\mathrm{N}}{r}-1\right)\left(\frac{\mathrm{N}}{r}-2\right)...\left(\frac{\mathrm{N}}{r}-\mathrm{N}_2-1\right)\right]^r \\ &= r^{r\mathrm{N}_2}\left[\frac{\left(\frac{\mathrm{N}}{r}\right)!}{\left(\frac{\mathrm{N}}{r}-\mathrm{N}_2\right)!}\right]^r \\ &= r^{r\mathrm{N}_2}\left[\frac{\left(\frac{\mathrm{N}}{r}\right)!}{\left(\frac{\mathrm{N}_1}{r}\right)!}\right]^r\end{aligned} \tag{A.21}$$

Inserting equation A.21 into equation A.20 leads to the final expression for the number of conformations of polymer (Ω_2) in the system:

$$\Omega_2 = \frac{1}{\mathrm{N}_2!}\left(\frac{(z-1)}{\mathrm{N}}\right)^{\mathrm{N}_2(r-1)} r^{r\mathrm{N}_2}\left[\frac{\left(\frac{\mathrm{N}}{r}\right)!}{\left(\frac{\mathrm{N}_1}{r}\right)!}\right]^r \tag{A.22}$$

As there is only one indistinguishable way to add the solvent to a given polymer lattice, the value of Ω_1 is assumed to be one, meaning $\ln(\Omega_1) = 0$. Therefore, the entropy of the mixed solution ($\mathrm{S}_{\mathrm{mixed}}$) is equal to $\mathrm{k_B}\ln(\Omega_2)$:

$$\begin{aligned}\mathrm{S}_{\mathrm{mixed}}/\mathrm{k_B} = \ln\left(\frac{1}{\mathrm{N}_2!}\right) + \mathrm{N}_2(r-1)\ln\left(\frac{z-1}{\mathrm{N}}\right) + r\mathrm{N}_2\ln(r) + r\ln\left(\left(\frac{\mathrm{N}}{r}\right)!\right) \\ - r\ln\left(\left(\frac{\mathrm{N}_1}{r}\right)!\right)\end{aligned} \tag{A.23}$$

As the values of N, N_1 and N_2 are large, Stirling's approximation can be applied to the logarithms ($\ln(x!) \approx x\ln(x) - x$). Therefore, equation A.23 can be manipulated in the following manner:

$$
\begin{aligned}
\mathrm{S_{mixed}/k_B} &= -\mathrm{N_2}\ln(\mathrm{N_2}) + \mathrm{N_2} + \mathrm{N}\ln\left(\frac{\mathrm{N}}{r}\right) - \mathrm{N} - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{r}\right) + \mathrm{N_1} \\
&\quad + r\mathrm{N_2}\ln(r) + \mathrm{N_2}(r-1)\ln(z-1) - \mathrm{N_2}(r-1)\ln(\mathrm{N}) \\
&= -\mathrm{N_2}\ln(\mathrm{N_2}) + \mathrm{N_2}\ln(\mathrm{N}) + \mathrm{N_2} + \mathrm{N_1} - \mathrm{N} + (\mathrm{N_1} + r\mathrm{N_2})\ln\left(\frac{\mathrm{N}}{r}\right) \\
&\quad - \mathrm{N_1}\ln\left(\frac{\mathrm{N}}{r}\right) + r\mathrm{N_2}\ln(r) - r\mathrm{N_2}\ln(\mathrm{N}) + \mathrm{N_2}(r-1)\ln(z-1) \\
&= -\mathrm{N_2}\ln\left(\frac{\mathrm{N_2}}{\mathrm{N}}\right) + \mathrm{N_2} - r\mathrm{N_2} + r\mathrm{N_2}\ln\left(\frac{\mathrm{N}}{r}\right) - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) \\
&\quad - r\mathrm{N_2}\ln\left(\frac{\mathrm{N}}{r}\right) + \mathrm{N_2}(r-1)\ln(z-1) \\
&= -\mathrm{N_2}\ln\left(\frac{\mathrm{N_2}}{\mathrm{N}}\right) - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) - \mathrm{N_2}(r-1) + \mathrm{N_2}(r-1)\ln(z-1) \\
&= -\mathrm{N_2}\ln\left(\frac{\mathrm{N_2}}{\mathrm{N}}\right) - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) + \mathrm{N_2}\left[(r-1)\ln(z-1) - (r-1)\right] \\
&= -\mathrm{N_2}\ln\left(\frac{\mathrm{N_2}}{\mathrm{N}}\right) - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) + \mathrm{N_2}\left[(r-1)\ln\left(\frac{z-1}{e}\right)\right]
\end{aligned}
\tag{A.24}
$$

Adding and subtracting $\mathrm{N_2}\ln(r)$ from the final expression in A.24 leads to the final expression for the mixed entropy:

$$
\mathrm{S_{mixed}/k_B} = -\mathrm{N_2}\ln\left(\frac{r\mathrm{N_2}}{\mathrm{N}}\right) - \mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) + \mathrm{N_2}\left[(r-1)\ln\left(\frac{z-1}{e}\right) + \ln(r)\right] \tag{A.25}
$$

Now that the mixed entropy term in equation A.13 has been determined, it is necessary to obtain the unmixed entropy term. The unmixed entropy of the solvent (1) and polymer (2) can be calculated by finding the limits of equation A.25 when $\mathrm{N_2}$ or $\mathrm{N_1}$ is equal to zero, respectively. Therefore:

$$\begin{aligned} \mathrm{S_{1,unmixed}/k_B} &= \mathrm{S_{mixed}(N_2 = 0)/k_B} = 0 \\ \mathrm{S_{2,unmixed}/k_B} &= \mathrm{S_{mixed}(N_1 = 0)/k_B} = \mathrm{N_2}\left[(r-1)\ln\left(\frac{z-1}{e}\right) + \ln(r)\right] \\ \mathrm{S_{unmixed}/k_B} &= \mathrm{N_2}\left[(r-1)\ln\left(\frac{z-1}{e}\right) + \ln(r)\right] \end{aligned} \tag{A.26}$$

The mixed (equation A.25) and unmixed (equation A.26) entropy terms can be combined as in equation A.13 to obtain an expression for the mixing entropy of the system:

$$\mathrm{\Delta S_{mix}/k_B} = -\left[\mathrm{N_1}\ln\left(\frac{\mathrm{N_1}}{\mathrm{N}}\right) + \mathrm{N_2}\ln\left(\frac{r\mathrm{N_2}}{\mathrm{N}}\right)\right] \tag{A.27}$$

The volume fractions of the solvent (ϕ_1) and polymer (ϕ_2) are represented by the following formulae:

$$\phi_1 = \frac{\mathrm{N_1}}{\mathrm{N}} \qquad \phi_2 = \frac{r\mathrm{N_2}}{\mathrm{N}} \tag{A.28}$$

Therefore, the entropy of mixing per lattice site can be expressed as follows:

$$\mathrm{\Delta S_{mix}/(Nk_B)} = -\left[\phi_1\ln(\phi_1) + \frac{\phi_2}{r}\ln(\phi_2)\right] \tag{A.29}$$

As the values of ϕ_1 and ϕ_2 add up to a sum of one, the expression can be simplified by setting $\phi_2 = \phi$ and $\phi_1 = 1 - \phi$. Therefore:

$$\mathrm{\Delta S_{mix}/(Nk_B)} = -\left[(1-\phi)\ln(1-\phi) + \frac{\phi}{r}\ln(\phi)\right] \tag{A.30}$$

A.14.2 Enthalpy of mixing

As a lattice model is used for Flory-Huggins theory, where the total volume is assumed to remain the same before and after mixing, with the lattice volume

remaining independent of pressure, it is assumed that enthalpy and internal energy are equivalent. Therefore, the enthalpy of mixing (ΔH_{mix}) is equivalent to the internal energy of mixing (ΔU_{mix}).

The internal energy of interactions between two solvent (1) molecules, two polymer (2) segments, and between a solvent and polymer segment are represented by u_{11}, u_{22} and u_{12}, respectively. When a solution containing a polymer volume fraction ϕ is unmixed, the total internal energy is given by:

$$U_{unmixed} = \frac{\phi}{2}u_{22} + \frac{(1-\phi)}{2}u_{11} \tag{A.31}$$

where the division by 2 is necessary to prevent double counting of interactions between partners. When mixed, an average single interaction experienced by a single solvent molecule (U_1) or a single polymer segment (U_2) can be represented by:

$$U_1 = (1-\phi)u_{11} + \phi u_{12} \qquad U_2 = \phi u_{22} + (1-\phi)u_{12} \tag{A.32}$$

The total mixed internal energy per lattice site (U_{mixed}) can therefore be determined using equation A.33, where z is the coordination number of each lattice site and a division by 2 is necessary to prevent double counting of interactions.

$$\begin{aligned} U_{mixed} &= z(\phi U_2 + (1-\phi)U_1)/2 \\ &= \frac{z}{2}\left[(\phi(\phi u_{22} + (1-\phi)u_{12}) + (1-\phi)((1-\phi)u_{11} + \phi u_{12})\right] \\ &= \frac{z}{2}\left[\phi^2 u_{22} + 2\phi(1-\phi)u_{12} + (1-\phi)^2 u_{12}\right] \end{aligned} \tag{A.33}$$

Subtracting equation A.31 from equation A.33 establishes an expression the internal energy of mixing per lattice site ($\Delta U_{mix}/N$):

$$\begin{aligned}\Delta \mathrm{U_{mix}/N} &= \frac{z}{2}\left[\phi^2 u_{22} + 2\phi(1-\phi)u_{12} + (1-\phi)^2 u_{11} - \phi u_{22} - (1-\phi)u_{11}\right] \\ &= \frac{z}{2}\left[(\phi^2-\phi)u_{22} + 2\phi(1-\phi)u_{12} - \phi(1-\phi)u_{11}\right] \\ &= z\phi(1-\phi)\left[u_{12} - \frac{(u_{11}+u_{22})}{2}\right] \\ &= \chi\phi(1-\phi)\end{aligned} \tag{A.34}$$

where χ (termed the Flory-Huggins interaction parameter) represents the difference in interactions of the solvent-polymer to the solvent-solvent and polymer-polymer interactions.

$$\chi = z\left[u_{12} - \frac{(u_{11}+u_{22})}{2}\right] \tag{A.35}$$

As the system is at constant volume and pressure, the enthalpy of mixing per lattice site can also be represented as:

$$\Delta \mathrm{H_{mix}/N} = \chi\phi(1-\phi) \tag{A.36}$$

A.14.3 Free energy of mixing ($\Delta \mathrm{G_{mix}}$)

Equations A.36 and A.30 can be inserted into equation A.12 to obtain the free energy of mixing per lattice site:

$$\Delta \mathrm{G_{mix}/N} = \chi\phi(1-\phi) + \mathrm{k_B T}\left[(1-\phi)\ln(1-\phi) + \frac{\phi}{r}\ln(\phi)\right] \tag{A.37}$$

or can be expressed per mole as:

$$\Delta \mathrm{G_{mix}} = \chi\phi(1-\phi) + \mathrm{RT}\left[(1-\phi)\ln(1-\phi) + \frac{\phi}{r}\ln(\phi)\right] \tag{A.38}$$

B Appendicies for chapter 4: probing the disordered adenovirus capsid protein VI using NMR

B.1 Protein sequences

B.1.1 pVI

MEDINFASLAPRHGSRPFMGNWQDIGTSNMSGGAFSWGSLWSGIKN
FGSTVKNYGSKAWNSSTGQMLRDKLKEQNFQQKVVDGLASGISGVVD
LANQAVQNKINSKLDPRPPVEEPPPAVETVLPDGRGEKRPRPDREETLVT
QIDEPPSYEEALKQGLPTTRPIAPMATGVLGQHTPVTLDLPPPADTQQKP
VLPGPSAVVVTRPSRASLRRAASGPRSLRPVASGNWQSTLNSIVGLGVQ
SLKRRRCF

B.1.2 Protein VI

AFSWGSLWSGIKNFGSTVKNYGSKAWNSSTGQMLRDKLKEQNFQQKV
VDGLASGISGVVDLANQAVQNKINSKLDPRPPVEEPPPAVETVLPDGRG
EKRPRPDREETLVTQIDEPPSYEEALKQGLPTTRPIAPMATGVLGQHTPV
TLDLPPPADTQQKPVLPGPSAVVVTRPSRASLRRAASGPRSLRPVASGN
WQSTLNSIVGL

B.1.3 Δ54-pVI

MGSKAWNSSTGQMLRDKLKEQNFQQKVVDGLASGISGVVDLANQAVQN

KINSKLDPRPPVEEPPPAVETVLPDGRGEKRPRPDREETLVTQIDEPPSYE

EALKQGLPTTRPIAPMATGVLGQHTPVTLDLPPPADTQQKPVLPGPSAVV

VTRPSRASLRRAASGPRSLRPVASGNWQSTLNSIVGLENLYFQ

B.2 Methods

B.2.1 NMR assignment experiments and processing recorded on Δ54-pVI

A 1.7 mM ^{15}N, ^{13}C labelled Δ54-pVI sample (20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O) was placed into a 3 mm NMR tube. NMR experiments were performed on a 950 MHz Oxford Instruments Company NMR magnet equipped with a Bruker® Avance III HD console and a 5 mm TCI (^{1}H, ^{13}C, ^{15}N) CryoProbe at a temperature of 5^{o}C.[†] A non-uniform sampling regime using 25% sparse sampling was used in the indirect dimension for all experiments. BEST-TROSY variants of the HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB and HN(CO)CACB experiments were recorded using the IBS NMRlib package.[365] A total of 256 complex points (if fully sampled) were obtained in all BEST-TROSY experiments for the ^{15}N dimension. For the HNCACB and HN(CO)CACB experiments, 160 complex points (if fully sampled) were obtained for the ^{13}C dimension. For all the other BEST-TROSY experiments, 64 complex points (if fully sampled) were obtained in the ^{13}C dimensions. Echo-AntiEcho and States-TPPI quadrature acqusition modes were used for the ^{15}N and ^{13}C spectral dimensions, respectively. The number of scans recorded for each experiment were as follows: 2 (HNCO), 8 (HN(CA)CO), 8 (HNCA), 8 (HN(CO)CA), 6 (HNCACB), 6 (HN(CO)CACB).

(H)N(CA)NH and (H)N(CO)(CA)NH experiments (hncannhgp3d, hncocannhgp3d) were performed using the IBS NMRlib package.[365] Non-uniform sampling using 25% sparse sampling was used in each indirect dimension and a total of 100 complex points (if fully sampled) were obtained for each ^{15}N dimension. Echo-Antiecho and States-TPPI were used to obtain quadrature for the two ^{15}N dimensions and ^{1}H decoupling was applied during acquisition. A total of 6 scans were performed for

[†]Experiments were performed in an equal collaboration between myself and Dr. Charles Buchanan

each experiment.

The (H)CC(CO)NH TOCSY experiment (hccconhgpwg3d3) was performed using a non-uniform sampling (25% sparse sampling) regime in each indirect dimension. A total of 108 and 128 complex points (if fully sampled) were obtained for the ^{15}N and ^{13}C dimensions, respectively. States-TPPI was used to obtain quadrature detection in all indirect dimensions. A total of 16 scans were obtained and ^{1}H decoupling was applied during acquisition. The acquisition time of all experiments was set to 89.7 ms, with relaxation delays of and 0.2 s and 1 s used for BEST-TROSY and non-BEST-TROSY experiments, respectively.

All assignment spectra were processed using the UniDecNMR[386] graphical user interface, based on nmrPipe[482] and nmrglue.[483] All data was referenced to the proton solvent signal (H_2O), equal to 4.96 ppm at 5^{o}C. Non-uniformly sampled data was reconstructed using the SMILE algorithm.[370]

B.2.2 Experiments to map the assignment of Δ54-pVI from 5^{o}C to 20^{o}C

BEST-TROSY HSQC experiments were performed on the 1.7 mM ^{15}N, ^{13}C labelled Δ54-pVI sample (20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O) increasing the temperature in approximately 2.5^{o}C increments between 5^{o}C and 20^{o}C (8^{o}C, 10^{o}C, 12.5^{o}C, 15^{o}C, 17.5^{o}C and 20^{o}C). The experiments were performed on a 950 MHz Oxford Instruments Company NMR magnet equipped with a Bruker® Avance III HD console and a 5 mm TCI (^{1}H, ^{13}C, ^{15}N) CryoProbe. The IBS NMRlib package[365] was used to setup the BEST-TROSY experiments. A total of 256 complex points were obtained in the ^{15}N dimension and quadrature was determined in an Echo-AntiEcho manner. The relaxation delay and acquisition times were set to 200 ms and 89.5 ms, respectively, and the number of scans was set to 10.

All spectra were processed using the UniDecNMR[386] graphical user interface, based on nmrPipe[482] and nmrglue.[483] The chemical shifts at each temperature were

referenced to the proton solvent signal (H_2O). The UniDecNMR[386] graphical user interface was used to map the ^{1}H-^{15}N correlation peaks sequentially from $5^{o}C$ to $20^{o}C$.

B.2.3 Chemical exchange experiments on Δ54-pVI

All chemical exchange experiments were performed on a 1 mM ^{15}N labelled Δ54-pVI sample (20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O). Approximately 230 μL of sample was loaded into a 5 mm Shigemi NMR tube assembly matched to D_2O. Experiments were performed on an Oxford Instruments 600MHz spectrometer equipped with a Varian 5 mm HCN triple resonance probe containing z-axis gradients at a temperature of $20^{o}C$

CPMG relaxation dispersion A ^{15}N CPMG experiment was performed using a constant relaxation time adaptation of the pulse sequence described by Loria et al[489] (Figure B.1). Anti-phase (N_xH_z) magnetisation was produced followed by a total relaxation time (T_{relax}) of 30 ms during which with the number of CPMG cycles (n) altered sequentially between 0 and 38. Each experimental CPMG increment consisted of $2\times[\tau_{CPMG}\text{-}180^{o}_{N}\text{-}\tau_{cp}]^{n}$ units separated by a P-element (INEPT), with the value of τ_{cp} equal to $T_{relax}/(4\times n)$. The CPMG frequency (ν_{CPMG}) could be determined using $\nu_{CPMG} = 1/(4\tau_{cp})$. Therefore, the 18 CPMG frequencies used were linearly spaced between 0 Hz and 1266 Hz. The P-element ensured that the measured ^{15}N relaxation rates represent an average of N_x and N_yH_z magnetisation relaxation rates, which will be interchanged during the P-element. This accounts for any CPMG frequency dependent J-modulation of the final signal intensity. For each CPMG frequency, 100 complex points were acquired in the ^{15}N chemical shift dimension using Rance-Kay hypercomplex data acqusition. The relaxation delay (d_1) was set to 3 s, and the acquisition time was set to 100 ms. During direct proton acquisition, ^{15}N decoupling was applied. A total of 16 scans were recorded for the

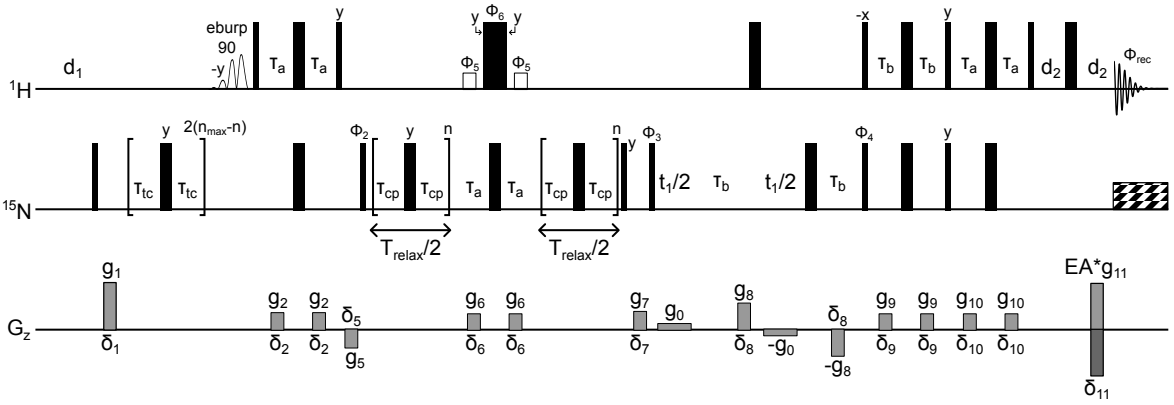


Figure B.1: The ^{15}N CPMG pulse sequence employed to measure relaxation dispersion for ^{15}N labelled Δ54-pVI. Values of τ_a and τ_b were set to 2.25 ms ($1/4J_{HN}$). The total transverse relaxation time (T_{relax}) was set to 30 ms and the number of CPMG cycles (n) was varied between 2 and 38 with the delays τ_{cp} altered to $T_{relax}/(4\times n)$. Temperature compensation was applied after the relaxation delay (d_1) using $2\times(n_{max}-n)$ CPMG cycles with delays set to $\tau_{tc}=T_{relax}/[4\times(n_{max}-n)]$. Narrow and wide black rectangles represent 90^o and 180^o pulses, respectively. The power level of ^{15} 180^o pulses in the CPMG trains were at reduced power compared to the other ^{15} pulses. White pulses (with black outline) signify low power (long duration) pulses centred on water. The value of d_2 was set to 500 μs. The phases of all pulses were set to x unless stated otherwise. The following phase cycle was employed: Φ_2=(x,-x), Φ_3=(y,y,y,y,-y,-y,-y,-y), Φ_4=(x), Φ_5=(x,x,-x,-x), Φ_6=(-x,-x,x,x), Φ_{rec}=(x,-x,x,-x,-x,x,-x,x). The gradient strengths (g) were set as follows in units of DAC: 500 (g_0), 5000 (g_1), 4000 (g_2), -6000 (g_5), 18000 (g_6), 15000 (g_7), 9000 (g_8), 3500 (g_9), 2500 (g_{10}), 18130 (g_{11}). The gradient durations (δ) were set as follows (in μs): 1000 (δ_1), 500 (δ_2), 1000 (δ_5), 500 (δ_6), 800 (δ_7), 1000 (δ_8), 400 (δ_9), 300 (δ_{10}), 100 (δ_{11}). Decoupling was applied to the ^{15}N channel during acquisition. Coherence selection was provided by ensuring $2g_8\delta_8\gamma_N = g_{11}\delta_{11}\gamma_H$. After every other t_1 increment, 180^o is added to Φ_3 and Φ_{rec} for States-TPPI acquisition.

experiment. Temperature compensation was applied after each relaxation delay to ensure that the total number of pulses/amount of heating applied to the sample remained constant for all CPMG frequencies used. The spectra were processed using SpinExplorer, a graphical user interface based on nmrPipe[482] and nmrglue.[483] The chemical shifts were referenced to the proton solvent signal (H_2O) and lorentzian-to-gaussian apodization was applied to both the ^{1}H and indirectly detected ^{15}N dimensions. Peak intensities (I) were fitted using FuDA[400] using GLORE (lorentzian and gaussian peak mixture) peak shapes for each CPMG frequency (n) and the $R_{2,eff}$ values were calculated using equation 4.5.

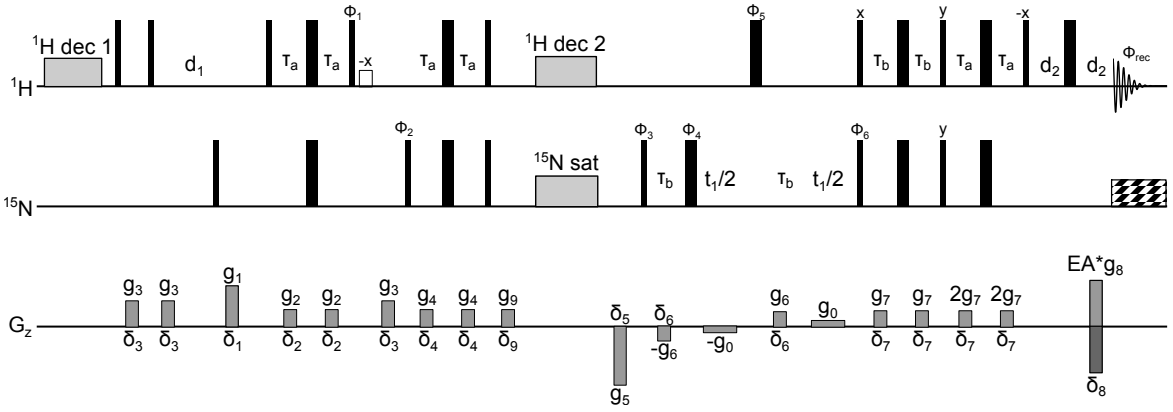


Figure B.2: The ^{15}N CEST pulse sequence employed to measure CEST profiles for ^{15}N labelled Δ54-pVI. Values of τ_a and τ_b were set to 2.25 ms ($1/4J_{HN}$). Nitrogen saturation (^{15}N sat) was applied at a range of chemical shifts for a duration of 500 ms at an effective field ($B_{1,eff}$) of 30 Hz. Proton decoupling (^{1}H dec 2) was applied during the saturation pulse using a 90_x-270_y-90_x composite pulse decoupling scheme. A reference plane was also recorded without the application of the ^{15}N saturation element. In this experiment, an additional proton decoupling element (^{1}H dec 1) was applied after the acquisition of each experiment to ensure constant heating for all experiments recorded. Narrow and wide black rectangles represent 90^o and 180^o pulses, respectively. White pulses (with black outline) signify low power (long duration) pulses centred on water intended for water suppression. The value of d_2 was set to 750 μs. The phases of all pulses were set to x unless stated otherwise. The following phase cycle was employed: Φ_1=(x), Φ_2=(x,-x), Φ_3=(y), Φ_4=(x,x,y,y,-x,-x,-y,-y), Φ_5=(x), Φ_6=(x), Φ_{rec}=(x,-x,-x,x). The gradient strengths (g) were set as follows in units of DAC: 500 (g_0), 5000 (g_1), 4000 (g_2), 10000 (g_3), 8000 (g_4), -25000 (g_5), 9000 (g_6), 4000 (g_7), 18320 (g_8). The gradient durations (δ) were set as follows (in μs): 1000 (δ_1), 500 (δ_2), 1000 (δ_3), 500 (δ_4), 1000 (δ_5), 2000 (δ_6), 500 (δ_7), 200 (δ_8). Decoupling was applied to the ^{15}N channel during acquisition. Coherence selection was provided by ensuring $2g_6\delta_6\gamma_N = g_8\delta_8\gamma_H$. After every other t_1 increment, 180^o is added to Φ_3 and Φ_{rec} for States-TPPI acquisition.

CEST A ^{15}N CEST experiment was performed using the pulse sequence described in Figure B.2 (based on the sequence in Vallurupalli *et al.*[31]). In-phase N_z magnetisation was produced followed by a the application of a saturation pulse on ^{15}N at 37 chemical shifts evenly spaced between 103 ppm and 133 ppm. The ^{15}N irradiation was applied for a total of 500 ms and had an effective field ($B_{1,eff}$) of 30 Hz. Proton decoupling (^{1}H dec 2) was applied during the ^{15}N irradiation using a 90_x270_y-90_x composite pulse decoupling scheme. A reference 2D ^{1}H-^{15}N spectrum without any continuous wave ^{15}N irradiation was also recorded. During the reference experiment, proton decoupling (^{1}H dec 1) was applied immediately

after acquisition before the relaxation delay (d_1) to ensure the heating of the sample during the reference experiment was the same as that during the experiments containing ^{15}N irradiation. The relaxation delay (d_1) and acquisition times were set to 1.5 s and 64 ms, respectively. A total of 4 scans were recorded for the whole experiment. In the indirectly detected (^{15}N, t_1) dimension, 128 complex points were obtained and processed using Rance-Kay processing. The spectra were processed using SpinExplorer, a graphical user interface based on nmrPipe[482] and nmrglue.[483] The chemical shifts were referenced to the proton solvent signal (H_2O) and lorentzian-to-gaussian apodization was applied to both the ^{1}H and the indirectly detected ^{15}N (t_1) dimensions. Peak intensities (I) were fitted using FuDA[400] using GLORE (lorentzian and gaussian peak mixture) peak shapes for each ^{15}N saturation pulse chemical shift and divided by the intensity of the reference spectrum (I_0) enabling the production of CEST profiles for each peak.

B.2.4 DDM experiments

A 5% (w/v) stock solution (98 mM) of n-dodecyl-β-D-maltoside (DDM) was prepared in 20 mM NaPi, 150 mM NaCl, pH 7.5.[†] The stock solution was added to 220μL samples of 200 μM ^{15}N-labelled Δ54-pVI (20 mM NaPi, 150 mM NaCl, pH 7.0, 10% D_2O) to give samples containing either 0.005% w/v (98 μM) or 0.01% w/v (196 μM) DDM. Therefore, samples below the critical micelle concentration (CMC) and above the critical micelle concentration of DDM were created (CMC = 130 μM[419]). A reference sample with no DDM added was also produced. The samples were loaded into Bruker® 3 mm NMR tubes and loaded into a 950 MHz Oxford Instruments Company NMR magnet equipped with a Bruker® Avance III HD console and a 5 mm TCI (^{1}H, ^{13}C, ^{15}N) CryoProbe. BEST-TROSY HSQC experiments were performed using the IBS NMRlib package[365] at a temperature of 20^{o}C. A total of 256 complex points were obtained in the

[†]This solution was provided by Susan Black, Oxford.

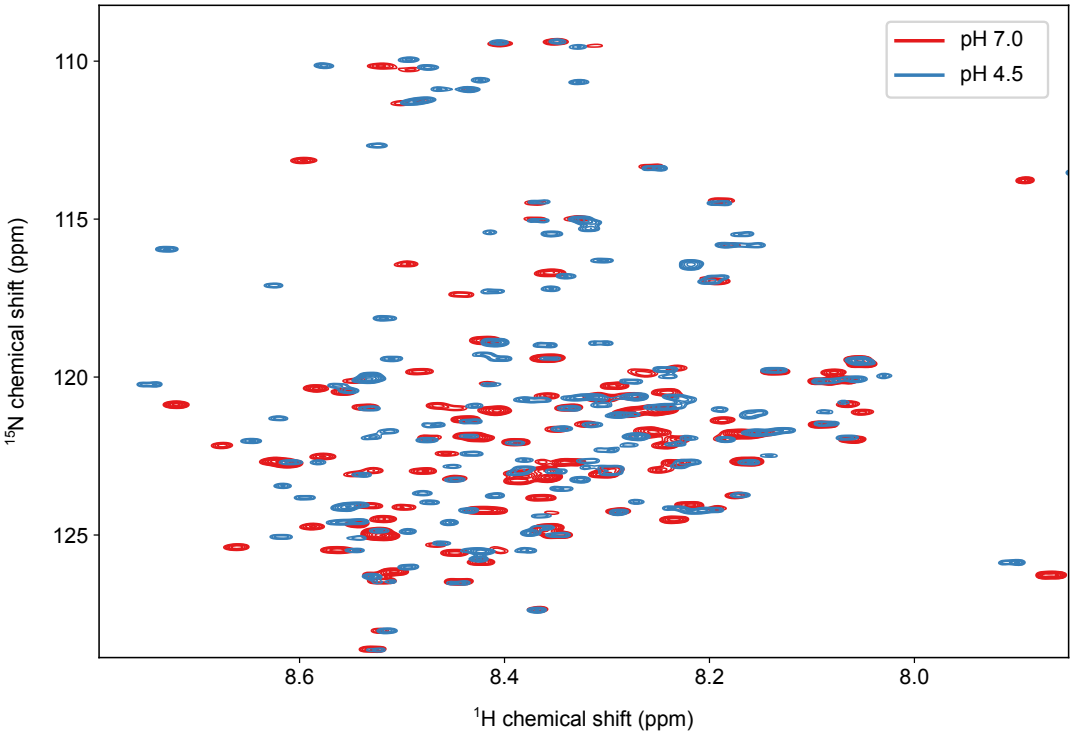


Figure B.3: The differences in the ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled Δ54-pVI at pH 7.0 and 4.5.

^{15}N dimension and quadrature was determined in an Echo-AntiEcho manner. The relaxation delay and acquisition times were set to 200 ms and 89.5 ms, respectively, and the number of scans was set to 10. The spectra were processed using SpinExplorer, a graphical user interface based on nmrPipe[482] and nmrglue.[483] The chemical shifts were referenced to the proton solvent signal (H_2O) and lorentzian-to-gaussian apodization was applied to both the ^{1}H and the indirectly detected ^{15}N (t_1) dimensions. Peak intensities (I) were fitted in FuDA[400] using a gaussian peak. Peaks whose signal intensity were hidden within the noise were discarded from downstream analysis.

B.2.5 pH dependence of Δ54-pVI NMR spectra

A 230 μL sample containing 1 mM of ^{15}N-labelled Δ54-pVI (20 mM NaPi, 150 mM NaCl, pH 7.0) was prepared in a 5 mm Shigemi tube assembly matched to D_2O. The sample was inserted into a 600 MHz Oxford Instruments NMR spectrometer

equiped with a 5 mm triple resonance (HCN) probe containing z-axis gradients and equilibrated at 10^{o}C. A ^{1}H-^{15}N sensitivity-enhanced HSQC spectrum (as described in Palmer *et al.*[490]) was performed using 240 complex points in the ^{15}N dimension, a relaxation delay of 1.5 s. A proton acquisition time of 100 ms was used, during which nitrogen decoupling was applied. The total number of scans was set to 8. After recording the spectra, the sample was removed, adjusted to pH 4.5 and all previous steps were repeated. Spectra were processed and overlaid using our in-house software SpinExplorer, based on nmrPipe[482] and nmrglue.[483] Each spectrum was referenced to the proton water chemical shift.

B.3 Chemical shift assignments of Δ54-pVI at 5^{o}C

residue number	residue type	HN (ppm)	N (ppm)	CO (ppm)	CA (ppm)	CB (ppm)
5	A		125.78	177.681	52.525	19.152
6	W	8.203	121.512	175.904	57.509	29.458
7	N	8.053	121.512	174.947	52.697	38.732
10	T	8.079	115.867	175.713	62.836	69.648
11	G	8.378	111.344	174.728	45.823	
12	Q	8.13	121.108	176.396	56.649	29.114
13	M	8.339	121.287	177.216	56.649	32.377
14	L	8.136	123.219	178.009	56.134	41.824
15	R	8.139	121.467	177.244	57.337	29.973
16	D	8.33	121.557	177.162	55.275	40.621
17	K	8.127	122.455	177.681	57.509	32.377
18	L	8.085	121.737	178.228	56.134	41.652
19	K	7.952	121.198	177.463	57.337	32.721
20	E	8.133	121.422	177.107	57.337	29.973
21	Q	8.263	120.973	176.123	56.478	29.114
22	N	8.336	119.67	175.357	53.384	38.56
23	F	8.161	121.422	175.959	58.54	39.076
24	Q	8.225	121.826	175.904	55.962	29.114
25	Q	8.25	122.051	175.931	55.962	29.114
26	K	8.285	123.444	176.834	56.649	32.893
27	V	8.206	123.579	176.615	62.836	
28	V	8.266	125.241	176.15	62.664	32.549
29	D	8.444	124.881	176.916	54.587	41.137
30	G	8.308	109.847	174.455	45.48	
31	L	8.073	122.141	177.681	55.275	42.167
32	A	8.263	125.241	177.927	52.697	18.981
33	S	8.244	115.402	175.248	58.54	63.808
34	G	8.339	111.314	174.181	45.308	
35	I	8.003	120.569	176.56	61.117	38.56
36	S	8.47	120.629	175.002	58.54	63.636
37	G	8.419	111.809	173.935	45.308	
38	V	7.977	120.569	176.505	62.664	32.549
39	V	8.269	125.6	175.685	62.492	
40	D	8.444	125.465	176.67	54.072	41.137

Table B.1: Chemical shift assignments of Δ54-pVI (1 of 5).

residue number	residue type	HN (ppm)	N (ppm)	CO (ppm)	CA (ppm)	CB (ppm)
41	L	8.349	124.657	178.173	55.962	41.824
42	A	8.257	123.579	178.42	53.384	18.637
43	N	8.098	117.394	175.767	53.556	38.56
44	Q	8.181	121.063	176.15	56.478	29.114
45	A	8.187	124.612	178.529	53.213	18.637
46	V	7.964	119.97	176.834	63.179	32.377
47	Q	8.355	123.623	176.123	56.134	29.114
48	N	8.457	120.329	175.466	53.556	38.56
49	K	8.165	122.336	176.998	56.821	32.721
50	I	8.149	122.68	176.423	61.633	38.389
51	N	8.536	123.114	175.494	53.384	39.076
52	S	8.273	117.649		58.712	63.808
53	K	8.254	123.399	176.451	56.478	32.721
59	P			176.837	62.836	31.862
60	V	8.238	121.242	176.095	62.148	32.893
61	E	8.489	126.004	175.959	55.962	30.316
62	E	8.517	125.211	173.962	54.072	29.629
65	P			176.509	62.492	32.034
66	A	8.435	125.465	177.873	52.182	18.981
67	V	8.225	120.838	176.177	62.148	32.893
68	E	8.578	125.915	176.478	56.306	30.316
69	T	8.371	117.963	174.291	61.977	69.819
70	V	8.295	124.387	175.904	61.977	32.893
71	S	8.603	122.725	173.115	56.306	63.293
72	P			177.259	63.695	32.034
73	E	8.492	120.808	177.298	56.993	29.801
74	G	8.406	110.715	174.427	45.48	
75	R	8.228	120.973	176.97	56.134	30.66
76	G	8.482	110.596	174.127	45.136	
77	E	8.317	121.377	176.533	56.306	30.316
78	K	8.463	123.504	176.369	56.134	32.721
79	R	8.419	124.567	174.127	53.9	29.973
82	P			176.603	63.179	32.034
83	D	8.47	120.928	176.15	54.416	40.793
84	R	8.161	120.928	176.068	55.962	30.832
85	E	8.486	122.935	176.533	56.478	30.145

Table B.2: Chemical shift assignments of Δ54-pVI (2 of 5).

residue number	residue type	HN (ppm)	N (ppm)	CO (ppm)	CA (ppm)	CB (ppm)
86	E	8.437	122.875	176.56	56.478	
87	T	8.279	117.304	174.291	61.977	69.648
88	L	8.349	126.379	177.162	55.103	42.339
89	V	8.263	123.264	176.314	62.32	32.721
90	T	8.282	120.059	174.127	61.805	69.819
91	Q	8.454	124.567	175.521	55.447	29.458
92	I	8.311	123.848	175.767	60.945	38.904
93	D	8.444	125.465		54.072	
96	P			176.744	62.664	31.862
97	S	8.422	116.945	174.373	57.852	64.152
98	Y	8.457	123.354	176.205	59.055	38.56
99	E	8.371	121.242	177.463	57.68	29.973
100	E	8.152	122.231	177.38	57.509	
101	A	8.333	124.702	178.857	53.556	18.465
102	L	7.987	120.299	178.584	55.962	41.995
103	K	7.955	121.557	177.162	57.165	32.721
104	Q	8.146	120.164	176.396	55.962	29.458
105	G	8.212	109.937	173.908	45.136	
106	L	8.076	123.039	175.357	53.041	41.48
107	P			177.166	63.008	32.034
108	T	8.295	115.028	174.755	61.633	69.819
109	T	8.127	117.379	174.099	61.633	69.819
110	R	8.393	125.81	174.017	53.9	30.145
111	P			176.65	62.836	32.034
112	I	8.317	122.59	175.822	60.774	38.56
113	A	8.412	130.841	175.248	50.291	18.122
114	P			176.931	62.836	32.034
115	M	8.457	121.467	176.123	55.275	33.064
116	A	8.454	126.768	177.791	52.353	19.152
117	T	8.181	113.83	175.166	61.805	69.991
118	G	8.39	111.674	174.127	45.136	
119	V	8.053	120.269	176.533	62.32	32.549
120	L	8.432	126.678	177.982	55.275	41.995
121	G	8.441	110.626	173.99	45.136	
122	Q	8.168	120.299	175.658	55.618	29.458
123	H	8.438	121.976	175.193	55.962	30.832
124	T	8.196	120.359	179.349	59.743	69.819
125	P			176.744	63.008	32.549

Table B.3: Chemical shift assignments of Δ54-pVI (3 of 5).

residue number	residue type	HN (ppm)	N (ppm)	CO (ppm)	CA (ppm)	CB (ppm)
126	V	8.368	121.916	176.396	62.32	33.064
127	T	8.295	120.059	174.099	61.633	69.819
128	L	8.374	126.049	176.697	54.759	42.511
129	D	8.339	122.366	175.548	54.244	40.965
130	L	8.152	124.477	174.728	52.697	41.652
133	P			176.791	62.664	32.034
134	A	8.466	125.106	177.709	52.525	19.152
135	D	8.403	120.299	176.752	54.244	40.965
136	T	8.107	114.938	174.783	61.977	69.648
137	Q	8.368	122.875	175.931	55.79	29.286
138	Q	8.393	122.41	175.767	55.618	
140	P			176.556	62.836	32.377
141	V	8.365	122.366	176.15	62.148	
142	L	8.466	129.194	175.111	52.697	41.48
143	P			177.119	63.008	32.206
144	G	8.276	109.817	178.857	44.449	
145	P			177.4	63.008	32.206
146	T	8.298	115.657	174.209	61.805	69.819
147	A	8.298	127.906	177.435	52.353	19.152
148	V	8.215	121.782	176.095	62.32	32.721
149	V	8.381	127.157	176.041	62.32	32.893
150	V	8.454	127.127	176.177	62.148	32.721
151	T	8.349	120.883	174.017	61.805	69.819
152	R	8.482	126.049	174.017	53.9	
156	A	8.4	124.836	178.529	53.384	18.809
157	S	8.231	115.747	175.111	59.055	63.465
158	L	8.098	124.552	177.736	55.618	42.167
159	R	8.158	122.051	176.615	56.478	30.488
160	R	8.273	123.039	176.341	56.306	30.66
161	A	8.308	125.87	177.654	52.525	19.152
162	A	8.289	123.938	177.736	52.525	19.152
163	S	8.215	115.417	174.564	58.196	64.152
164	G	8.225	111.045	178.83	44.621	
165	P			177.447	63.179	32.206
166	R	8.559	121.826	176.533	56.306	30.488
167	S	8.279	117.304	174.263	58.368	63.98
168	L	8.269	124.702	177.08	54.931	42.167

Table B.4: Chemical shift assignments of Δ54-pVI (4 of 5).

residue number	residue type	HN (ppm)	N (ppm)	CO (ppm)	CA (ppm)	CB (ppm)
168	L	8.269	124.702	177.08	54.931	42.167
169	R	8.209	123.399	174.017	53.9	30.145
170	P			176.884	63.008	31.862
171	V	8.266	121.497	176.123	62.32	32.721
172	A	8.447	128.535	177.599	52.353	19.152
173	S	8.289	115.956	175.002	58.368	63.808
174	G	8.346	111.045	173.798	45.308	
175	N	8.276	119.401	175.384	53.041	38.732
176	W	8.098	122.41	176.423	57.852	29.114
177	Q	8.044	122.231	175.931	56.134	
178	S	8.127	116.825	175.193	58.712	63.465
179	T	8.076	116.346		62.32	69.476
180	L	8.079	124.163	177.298	55.618	41.995
181	N	8.335	119.272	175.166	53.384	38.732
182	S	8.095	116.256	174.509	58.54	63.636
183	I	8.149	123.219	176.423	61.289	38.389
184	V	8.155	125.091	176.724	62.836	32.377
185	G	8.511	113.65	174.291	45.308	
186	L	8.076	122.231	177.709	55.275	42.167
187	E	8.622	121.317	176.314	57.165	29.801
188	N	8.333	119.311	174.947	53.213	38.732
189	L	7.977	122.41	176.779	55.275	42.339
190	Y	7.964	120.03	175.193	57.509	38.904
191	F	7.999	121.916	174.509	57.509	39.591
192	Q	7.774	126.648	173.279	57.337	30.488

Table B.5: Chemical shift assignments of Δ54-pVI (5 of 5).

C Appendices for chapter 5: Reducing NMR Limitations for Slowly Tumbling Biomolecules

C.1 General expression for a relaxation matrix element $R_{a,b}$

The following section outlines a general expression for a relaxation matrix element $R_{a,b}$ along with any modifications required for specific motional models. The equations in this section are from Eaton *et al.* (in preparation)[54] and based on those derived elsewhere.[84–86,94] Each element of the relaxation matrix in a given spin system can be evaluated as follows:[54]

$$\begin{aligned} R_{a,b} &= \langle \rho_a | \, \hat{\hat{R}} \, | \rho_b \rangle \\ &= \sum_{k,l} \sum_{i,j=1}^{9} \frac{6}{5} a_k a_l o_i o_j \delta_{\omega_i^k, \omega_j^l} \frac{\langle \rho_a [\hat{O}_j^l, [\hat{O}_i^{k\dagger}, \rho_b]] \rangle}{\langle \rho_a | \rho_b \rangle} J'_{(k,i),(l,j)} \end{aligned} \tag{C.1}$$

The sums over indices k and l represent sums over the set of relaxation interactions in the system, such as any dipolar or CSA mechanisms. The interaction parameters a_k are defined in table C.1 and depend on whether k is a dipolar or CSA interaction. The operators $\hat{O}^k$ are constituents of the Hamiltonian for each interaction k.$\hat{O}_i^k$ is the ith operator in table C.2, for interaction k, with o_i being its corresponding coefficient. In RelCalc, the double commutator $[\hat{O}_j^l, [\hat{O}_i^{k\dagger}, \rho_b]]$ is obtained from a pre-evaluated symbolic library, and $\langle \rho_a [\hat{O}_j^l, [\hat{O}_i^{k\dagger}, \rho_b]] \rangle$ is calculated by taking the trace of the resulting matrix within the angular brackets.[54] $\delta_{\omega_i^k, \omega_j^l}$ is a Kronecker delta which states that only cross terms between different operators (i, j) for relaxation mechanisms (k, l) with the same Larmor frequency can contribute to

k	a_k	A_k
Dipolar	$-\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi}$	$\frac{1}{r_{IS}^3}$
CSA	$\frac{1}{3}\gamma_I B_0 \Delta\delta_I$	1

Table C.1: Interaction constant definitions used to calculate relaxation rates in RelCalc. The factor of $1/r_{IS}^3$ (where r_{IS} is the internuclear IS distance) is explicitly factored out to allow treatment of dipolar interactions where this distance varies due to internal molecular motion such as the protons in a methyl group interacting with an external nucleus (EXT). γ_I is the gyromagnetic ratio of nucleus I, B_0 is the magnetic field strength, and $\Delta\delta_I$ is the axially symmetric CSA value for nucleus I ($\Delta\delta_I = \delta_{I,zz} - \delta_{I,xx/yy}$). Note: non-axially symmetric and asymmetric chemical shift tensors and the quadrupolar contribution to relaxation for quadrupolar nuclei can also be handled by RelCalc;[54] however, such systems are not included in this thesis. The constants μ_0 and $\hbar$ represent the magnetic permeability in a vacuum and Planck's constant.

i	o_i	$\hat{O}_i^{dipolar}$	$\hat{O}_i^{CSA}$	p_i	ω_i
1	$\frac{1}{2}$	$\hat{I}_-\hat{S}_-$		-2	$-\omega_I - \omega_S$
2	$\frac{1}{2}$	$\hat{I}_z\hat{S}_-$		-1	$-\omega_S$
3	$\frac{1}{2}$	$\hat{I}_-\hat{S}_z$	$\hat{I}_-$	-1	$-\omega_I$
4	$\frac{1}{2}\sqrt{\frac{8}{3}}$	$\hat{I}_z\hat{S}_z$	$\hat{I}_z$	0	0
5	$-\frac{1}{8}\sqrt{\frac{8}{3}}$	$\hat{I}_+\hat{S}_-$		0	$\omega_I - \omega_S$
6	$-\frac{1}{8}\sqrt{\frac{8}{3}}$	$\hat{I}_-\hat{S}_+$		0	$\omega_S - \omega_I$
7	$-\frac{1}{2}$	$\hat{I}_z\hat{S}_+$		1	ω_S
8	$-\frac{1}{2}$	$\hat{I}_+\hat{S}_z$	$\hat{I}_+$	1	ω_I
9	$\frac{1}{2}$	$\hat{I}_+\hat{S}_+$		2	$\omega_I + \omega_S$

Table C.2: A list of the operators required to describe the dipolar and CSA interactions, as well as their corresponding coefficients.[54] Note that the dipolar interaction is between two spins (I and S) and therefore requires spin operators that depend on both spins. On the other hand, the CSA interaction arises due to the presence of anisotropic electron density around a single spin. Therefore, it requires spin operators that depend only on one spin. Each operator $\hat{O}_i$ is associated with an irreducible spherical tensor $\hat{T}_{p_i}^{(2)}$, and it has an oscillation frequency of ω_i.

relaxation.

Each term $J'_{(k,i),(l,j)}$ depends on the exact motional model for the spin-system of interest. The expressions for $J'_{(k,i),(l,j)}$ are shown in table C.3 for the cases of a spin-system with isotropic global tumbling (ISO-ST), isotropic global tumbling with N-site hopping local motion about a rotational symmetry axis (ISO-ROT), and isotropic global tumbling with N-site hopping local motion about a rotational

Motional Model	$J'_{(k,i),(l,j)}$
ISO-ST	$A_k A_l J(\tau(0), \omega_{k,i}) P_0(\cos(\zeta_{k,l}))$
ISO-ROT	$A_k A_l \sum_{q=0}^{2} J(\tau(q), \omega_{k,i})(2-\delta_{q,0}) y_q^2 P_q^k P_q^l \cos(q(\phi_k - \phi_l))$
ISO-EXT	$\sum_{\lambda=0}^{\Lambda} J(\tau(\lambda), \omega_{k,i}) \sum_{\alpha,\beta}^{N} \frac{n_\lambda}{N^2} A_\beta^k A_\alpha^l \cos\big(\frac{2\pi}{N}\lambda(\beta-\alpha)\big) P_0(\cos(\zeta_{k\beta,l\alpha}))$

Table C.3: Spectral density calculations for different motional models in RelCalc. $\cos(\zeta_{k,l})$ represents the dot product between the normalised vectors along the interactions k and l in the molecular frame. For the ISO-EXT model, the number of time-scales Λ is floor(N/2 + 1) where N is the number of hopping sites sampled by the motion and the floor function rounds down to the nearest integer. For a methyl group hopping between sites (N = 3, Λ = 2). α and β can have values 0,1,2 representing the three hopping sites within the system. $\zeta_{k\beta,l\alpha}$ is the angle between the interaction vectors (k,l) with the external nucleus from hopping sites α and β. The value of n_λ is equal to 2-$\delta_{\lambda,0}$, where δ represents a Kronecker delta. ϕ_k and ϕ_l represent the azimuthal angles that orientate the interactions k and l in the molecular frame. The associated Legendre polynomials (P_q) are shown in table C.4 and constants $y_{|q|}$ are equal to $\sqrt{\frac{(2-q)!}{(2+q)!}}$.

symmetry axis with the presence of an external nucleus (ISO-EXT). The effects of global and local tumbling (N-site hopping) are included into a total correlation time $\tau(q)$:

$$\frac{1}{\tau(q)} = \frac{1}{\tau_c} + \frac{2}{\tau_m}(1 - \cos(2\pi q/N)) \tag{C.2}$$

with a global rotational correlation time τ_c and local motion governed by the timescale τ_m. The values of q can be 0,1 or 2. The spectral density function are defined as:

$$J(q,\omega) = \frac{\tau(q)}{1 + \omega^2 \tau(q)^2} \tag{C.3}$$

RelCalc is also able to calculate relaxation rates in the case of axially symmetric global tumbling (AX-ST, AX-ROT, AX-EXT); however, these motional models were not used in this thesis so their expressions are not shown here. In RelCalc, results for ST and ROT motions are expressed symbolically. In the case of EXT, the spectral density function is held in a function $C(\omega)$ and evaluated numerically when required.

q	P_q^k
0	$\frac{1}{2}(3\cos^2(\theta_k)-1)$
1	$3\cos(\theta_k)\sin(\theta_k)$
2	$3\sin^2(\theta_k)$

Table C.4: Associated Legendre polynomials for interaction k used to evaluate relaxation rates dependent on the polar angle θ_k which is one of the rotation angles rotating interaction k into the molecular frame.

C.2 Relaxation rate expressions

C.2.1 Relaxation rate expressions for a ^{15}N-^{1}H spin system

The equations in this section outline the shift basis transverse relaxation rate expressions within a ^{15}N-^{1}H spin system separated by a bond length of R_{H-N}. The equations were developed assuming a dipolar interaction between ^{15}N and ^{1}H (d) and chemical shift anisotropy on the ^{15}N (c_N) and ^{1}H (c_H) nuclei. Furthermore, the equations include the effects of dipolar interactions with an external nucleus X (X=H/D) aligned with the N-H bond a distance of R_{N-Xext}/R_{H-Xext} away is included. These additional dipolar interactions are governed by the dipolar constants f/g for the external proton and h/k for the external deuteron. Furthermore, isotropic global tumbling is assumed with a rotational correlation time τ_c and the Larmor frequencies of N/H/D are represented by $\omega_N/\omega_H/\omega_D$. The interaction constants and spectral density functions are defined as follows:

$$d = -\frac{\mu_0\hbar\gamma_N\gamma_H}{4\pi R_{H-N}^3} \quad c_N = \frac{1}{3}\gamma_N B_0 \Delta\delta_N \quad c_H = \frac{1}{3}\gamma_H B_0 \Delta\delta_H \quad f = -\frac{\mu_0\hbar\gamma_H^2}{4\pi R_{H-Hext}^3}$$
$$g = -\frac{\mu_0\hbar\gamma_N\gamma_H}{4\pi R_{N-Hext}^3} \quad h = -\frac{\mu_0\hbar\gamma_H\gamma_D}{4\pi R_{H-Dext}^3} \quad k = -\frac{\mu_0\hbar\gamma_N\gamma_D}{4\pi R_{N-Dext}^3} \quad J(\omega) = \frac{\tau_c}{1+\omega^2\tau_c^2} \tag{C.4}$$

where γ_A is the gyromagnetic ratio of nucleus A, $\Delta\delta_A$ is the chemical shift anisotropy of nucleus A, μ_0 is the magnetic permeability in a vacuum, and $\hbar$ is Planck's constant. The following equations are the shift basis transverse relaxation

rate expressions in the ^{15}N-^{1}H spin system described:

$$\begin{aligned}
R_{N_+H_\alpha} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_N+\omega_H) + \frac{3}{20}J(\omega_H)\right] \\
& + c_N^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] + c_H^2\left[\frac{3}{5}J(\omega_H)\right] + dc_N\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] \\
& + f^2\left[\frac{1}{20}\tau_c + \frac{3}{20}J(\omega_H) + \frac{3}{10}J(2\omega_H)\right] \\
& + g^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_H) + \frac{3}{10}J(\omega_N+\omega_H)\right] \\
& + h^2\left[\frac{2}{5}J(\omega_H) + \frac{2}{15}J(\omega_D-\omega_H) + \frac{4}{5}J(\omega_D+\omega_H)\right] \\
& + k^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_N) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_N) + \frac{4}{5}J(\omega_D+\omega_N)\right]
\end{aligned} \tag{C.5}$$

$$\begin{aligned}
R_{N_+H_\beta} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_N+\omega_H) + \frac{3}{20}J(\omega_H)\right] \\
& + c_N^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] + c_H^2\left[\frac{3}{5}J(\omega_H)\right] - dc_N\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] \\
& + f^2\left[\frac{1}{20}\tau_c + \frac{3}{20}J(\omega_H) + \frac{3}{10}J(2\omega_H)\right] \\
& + g^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_H) + \frac{3}{10}J(\omega_N+\omega_H)\right] \\
& + h^2\left[\frac{2}{5}J(\omega_H) + \frac{2}{15}J(\omega_D-\omega_H) + \frac{4}{5}J(\omega_D+\omega_H)\right] \\
& + k^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_N) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_N) + \frac{4}{5}J(\omega_D+\omega_N)\right]
\end{aligned} \tag{C.6}$$

$$\begin{aligned}
R_{N_+H_\alpha,N_+H_\beta} = {} & -c_H^2\left[\frac{3}{5}J(\omega_H)\right] + d^2\left[\frac{3}{20}J(\omega_H)\right] \\
& - f^2\left[\frac{1}{20}\tau_c + \frac{3}{20}J(\omega_H) + \frac{3}{10}J(2\omega_H)\right] \\
& - h^2\left[\frac{2}{5}J(\omega_H) + \frac{2}{15}J(\omega_D-\omega_H) + \frac{4}{5}J(\omega_D+\omega_H)\right]
\end{aligned} \tag{C.7}$$

$$
\begin{aligned}
R_{H_+N_\alpha} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_H) + \frac{3}{10}J(\omega_N+\omega_H) + \frac{3}{20}J(\omega_N)\right] \\
& + c_H^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] + c_N^2\left[\frac{3}{5}J(\omega_H)\right] + dc_H\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] \\
& + f^2\left[\frac{1}{4}\tau_c + \frac{9}{20}J(\omega_H) + \frac{3}{10}J(2\omega_H)\right] \\
& + g^2\left[\frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_N+\omega_H)\right] \\
& + h^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_H) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_H) + \frac{4}{5}J(\omega_D+\omega_H)\right] \\
& + k^2\left[\frac{2}{5}J(\omega_N) + \frac{2}{15}J(\omega_D-\omega_N) + \frac{4}{5}J(\omega_D+\omega_N)\right]
\end{aligned}
\tag{C.8}
$$

$$
\begin{aligned}
R_{H_+N_\beta} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_H) + \frac{3}{10}J(\omega_N+\omega_H) + \frac{3}{20}J(\omega_N)\right] \\
& + c_H^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] + c_N^2\left[\frac{3}{5}J(\omega_H)\right] - dc_H\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_N)\right] \\
& + f^2\left[\frac{1}{4}\tau_c + \frac{9}{20}J(\omega_H) + \frac{3}{10}J(2\omega_H)\right] \\
& + g^2\left[\frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_N+\omega_H)\right] \\
& + h^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_H) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_H) + \frac{4}{5}J(\omega_D+\omega_H)\right] \\
& + k^2\left[\frac{2}{5}J(\omega_N) + \frac{2}{15}J(\omega_D-\omega_N) + \frac{4}{5}J(\omega_D+\omega_N)\right]
\end{aligned}
\tag{C.9}
$$

$$
\begin{aligned}
R_{H_+N_\alpha,H_+N_\beta} = {} & -c_N^2\left[\frac{3}{5}J(\omega_N)\right] + d^2\left[\frac{3}{20}J(\omega_N)\right] \\
& - g^2\left[\frac{1}{20}J(\omega_N-\omega_H) + \frac{3}{20}J(\omega_N) + \frac{3}{10}J(\omega_N+\omega_H)\right] \\
& - k^2\left[\frac{2}{5}J(\omega_N) + \frac{2}{15}J(\omega_D-\omega_N) + \frac{4}{5}J(\omega_D+\omega_N)\right]
\end{aligned}
\tag{C.10}
$$

C.2.2 Relaxation rate expressions for a ^{13}C-^{19}F spin system

The equations in this section outline the shift basis transverse relaxation rate expressions within a ^{13}C-^{19}F spin system separated by a bond length of R_{C-F}. The equations were developed assuming a dipolar interaction between ^{13}C and ^{19}F nuclei (d) and chemical shift anisotropy on the ^{13}C (c_C) and ^{19}F (c_F) nuclei. Furthermore, the equations include the effects of dipolar interactions with an external nucleus X (X=H/D) aligned with the C-F bond a distance of R_{C-Xext}/R_{F-Xext} away is included. These additional dipolar interactions are governed by the dipolar constants p/q for the external proton and l/m for the external deuteron. Furthermore, isotropic global tumbling is assumed with a rotational correlation time τ_c and the Larmor frequencies of C/F/H/D are represented by $\omega_C/\omega_F/\omega_H/\omega_D$. The interaction constants and spectral density functions are defined as follows:

$$d = -\frac{\mu_0 \hbar \gamma_C \gamma_F}{4\pi R_{C-F}^3} \quad c_C = \frac{1}{3}\gamma_C B_0 \Delta\delta_C \quad c_F = \frac{1}{3}\gamma_F B_0 \Delta\delta_F \quad p = -\frac{\mu_0 \hbar \gamma_F \gamma_H}{4\pi R_{F-Hext}^3}$$
$$q = -\frac{\mu_0 \hbar \gamma_C \gamma_H}{4\pi R_{C-Hext}^3} \quad l = -\frac{\mu_0 \hbar \gamma_F \gamma_D}{4\pi R_{F-Dext}^3} \quad m = -\frac{\mu_0 \hbar \gamma_C \gamma_D}{4\pi R_{C-Dext}^3} \quad J(\omega) = \frac{\tau_c}{1+\omega^2\tau_c^2} \tag{C.11}$$

where γ_A is the gyromagnetic ratio of nucleus A, $\Delta\delta_A$ is the chemical shift anisotropy of nucleus A, μ_0 is the magnetic permeability in a vacuum, and $\hbar$ is Planck's constant. The following equations are the shift basis transverse relaxation

rate expressions in the ^{13}C-^{19}F spin system described:

$$\begin{aligned}
R_{C_+F_\alpha} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_F) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_C+\omega_F) + \frac{3}{20}J(\omega_F)\right] \\
& + c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] + c_F^2\left[\frac{3}{5}J(\omega_F)\right] + dc_C\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] \\
& + p^2\left[\frac{1}{20}J(\omega_H-\omega_F) + \frac{3}{20}J(\omega_F) + \frac{3}{10}J(\omega_H+\omega_F)\right] \\
& + q^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_H) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_H) + \frac{3}{10}J(\omega_C+\omega_H)\right] \\
& + l^2\left[\frac{2}{5}J(\omega_F) + \frac{2}{15}J(\omega_D-\omega_F) + \frac{4}{5}J(\omega_D+\omega_F)\right] \\
& + m^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_C) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_C) + \frac{4}{5}J(\omega_D+\omega_C)\right]
\end{aligned} \tag{C.12}$$

$$\begin{aligned}
R_{C_+F_\beta} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_F) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_C+\omega_F) + \frac{3}{20}J(\omega_F)\right] \\
& + c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] + c_F^2\left[\frac{3}{5}J(\omega_F)\right] - dc_C\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] \\
& + p^2\left[\frac{1}{20}J(\omega_H-\omega_F) + \frac{3}{20}J(\omega_F) + \frac{3}{10}J(\omega_H+\omega_F)\right] \\
& + q^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_H) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_H) + \frac{3}{10}J(\omega_C+\omega_H)\right] \\
& + l^2\left[\frac{2}{5}J(\omega_F) + \frac{2}{15}J(\omega_D-\omega_F) + \frac{4}{5}J(\omega_D+\omega_F)\right] \\
& + m^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_C) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_C) + \frac{4}{5}J(\omega_D+\omega_C)\right]
\end{aligned} \tag{C.13}$$

$$\begin{aligned}
R_{C_+F_\alpha,C_+F_\beta} = {} & -c_F^2\left[\frac{3}{5}J(\omega_F)\right] + d^2\left[\frac{3}{20}J(\omega_F)\right] \\
& - p^2\left[\frac{1}{20}J(\omega_H-\omega_F) + \frac{3}{20}J(\omega_F) + \frac{3}{10}J(\omega_H+\omega_F)\right] \\
& - l^2\left[\frac{2}{5}J(\omega_F) + \frac{2}{15}J(\omega_D-\omega_F) + \frac{4}{5}J(\omega_D+\omega_F)\right]
\end{aligned} \tag{C.14}$$

$$\begin{aligned}
R_{F_+C_\alpha} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_F) + \frac{3}{20}J(\omega_F) + \frac{3}{10}J(\omega_C+\omega_F) + \frac{3}{20}J(\omega_C)\right] \\
& + c_F^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] + c_C^2\left[\frac{3}{5}J(\omega_F)\right] + dc_F\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] \\
& + p^2\left[\frac{1}{4}J(\omega_H-\omega_F) + \frac{9}{20}J(\omega_F) + \frac{3}{10}J(\omega_H+\omega_F)\right] \\
& + q^2\left[\frac{1}{20}J(\omega_C-\omega_H) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_C+\omega_H)\right] \\
& + l^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_F) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_F) + \frac{4}{5}J(\omega_D+\omega_F)\right] \\
& + m^2\left[\frac{2}{5}J(\omega_C) + \frac{2}{15}J(\omega_D-\omega_C) + \frac{4}{5}J(\omega_D+\omega_C)\right]
\end{aligned} \tag{C.15}$$

$$\begin{aligned}
R_{F_+C_\beta} = {} & d^2\left[\frac{1}{5}\tau_c + \frac{1}{20}J(\omega_C-\omega_F) + \frac{3}{20}J(\omega_F) + \frac{3}{10}J(\omega_C+\omega_F) + \frac{3}{20}J(\omega_C)\right] \\
& + c_F^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] + c_C^2\left[\frac{3}{5}J(\omega_F)\right] - dc_F\left[\frac{4}{5}\tau_c + \frac{3}{5}J(\omega_C)\right] \\
& + p^2\left[\frac{1}{4}J(\omega_H-\omega_F) + \frac{9}{20}J(\omega_F) + \frac{3}{10}J(\omega_H+\omega_F)\right] \\
& + q^2\left[\frac{1}{20}J(\omega_C-\omega_H) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_C+\omega_H)\right] \\
& + l^2\left[\frac{8}{15}\tau_c + \frac{2}{5}J(\omega_F) + \frac{4}{5}J(\omega_D) + \frac{2}{15}J(\omega_D-\omega_F) + \frac{4}{5}J(\omega_D+\omega_F)\right] \\
& + m^2\left[\frac{2}{5}J(\omega_C) + \frac{2}{15}J(\omega_D-\omega_C) + \frac{4}{5}J(\omega_D+\omega_C)\right]
\end{aligned} \tag{C.16}$$

$$\begin{aligned}
R_{F_+C_\alpha,F_+C_\beta} = {} & -c_C^2\left[\frac{3}{5}J(\omega_C)\right] + d^2\left[\frac{3}{20}J(\omega_C)\right] \\
& - q^2\left[\frac{1}{20}J(\omega_C-\omega_H) + \frac{3}{20}J(\omega_C) + \frac{3}{10}J(\omega_C+\omega_H)\right] \\
& - m^2\left[\frac{2}{5}J(\omega_C) + \frac{2}{15}J(\omega_D-\omega_C) + \frac{4}{5}J(\omega_D+\omega_C)\right]
\end{aligned} \tag{C.17}$$

C.2.3 Auto-relaxation rate expressions for a ^{13}C-^{1}H$_3$ spin system

The equations in this section outline the shift basis transverse relaxation rate expressions within a $^{13}CH_3$ spin system with C_{3v} symmetry. The C-H bond length is denoted by R_{H-C} and the H-H internuclear distance (R_{H-H}) can be calculated using $\sqrt{3}R_{H-C}\sin(\beta)$ (where β is the internal angle of the $^{13}CH_3$ group, $\approx 109.5^o$. The equations were created using RelCalc assuming a dipolar interaction between ^{13}C and ^{1}H nuclei (d), a dipolar interaction between ^{1}H nuclei (e) and chemical shift anisotropy on the ^{13}C (c_C) and ^{1}H (c_H) nuclei. Furthermore, the equations include the effects of dipolar interactions with an external nucleus X (X=H/D) a distance of R_{ext} away at an angle of θ_{ext} to the 3-fold rotation axis. These additional dipolar interactions are governed by the dipolar constants f/g for the external proton/deuteron. The interaction constants are defined as follows:

$$\begin{gathered} d = -\frac{\mu_0 \hbar \gamma_C \gamma_H}{4\pi R_{H-C}^3} \quad e = -\frac{\mu_0 \hbar \gamma_H \gamma_H}{4\pi R_{H-H}^3} \quad c_C = \frac{1}{3}\gamma_C B_0 \Delta\delta_C \quad c_H = \frac{1}{3}\gamma_H B_0 \Delta\delta_H \\ f = -\frac{\mu_0 \hbar \gamma_H \gamma_H}{4\pi} \quad g = -\frac{\mu_0 \hbar \gamma_D \gamma_H}{4\pi} \end{gathered} \tag{C.18}$$

where γ_A is the gyromagnetic ratio of nucleus A, $\Delta\delta_A$ is the chemical shift anisotropy of nucleus A, μ_0 is the magnetic permeability in a vacuum, and $\hbar$ is Planck's constant.* Furthermore, the effects of global and local tumbling are included into a total correlation time $\tau(q)$:

$$\frac{1}{\tau(q)} = \frac{1}{\tau_c} + \frac{2}{\tau_m}(1 - \cos(2\pi q/3)) \tag{C.19}$$

with a global rotational correlation time τ_c and local motion about the 3-fold rotation axis governed by the timescale τ_m. The values of q can be 0,1 or 2. The

*The R_{ext} has been omitted from constants f and g as the distances are included within the modified spectral densities $C(\omega)$ in equation C.21.

spectral density functions are defined as:

$$J(q,\omega) = \frac{\tau(q)}{1+\omega^2\tau(q)^2} \tag{C.20}$$

Furthermore, due to the C-H bonds not being aligned with the symmetry axis of rotation, the relaxation rates in a $^{13}CH_3$ are governed by the second order associated Legendre polynomials (P_i, $i = 0, 1, 2$, table C.4), with θ set to the angle from the C-H bonds to the C axis of rotation (equal to approximately 109.5^o). Furthermore, due to the 3-fold rotation of the $^{13}CH_3$ group with the external nucleus (H/D) remaining fixed, dipolar interactions with the external nucleus are governed through modified spectral densities ($C(\omega)$):

$$C(\omega) = \sum_{\lambda=0}^{2} J(\lambda,\omega) \sum_{\alpha=0}^{2}\sum_{\beta=0}^{2} \frac{n_\lambda}{9} \frac{P_0(\cos(\zeta_{\alpha\beta}))}{r_\alpha^3 r_\beta^3} \cos\left(\frac{2\pi}{3}\lambda(\beta-\alpha)\right) \tag{C.21}$$

where α and β can have values 0,1,2 representing the three hopping sites within the $^{13}CH_3$ manifold. r_i represents the distance to the external proton from each hopping site, and $\zeta_{\alpha,\beta}$ is the angle between the interaction vectors with the external nucleus from hopping sites α and β. The value of n_λ is equal to 2-$\delta_{\lambda,0}$, where δ represents a Kronecker delta.

C.2.3.i Single quantum carbon auto-relaxation rates (shift basis)

The auto-relaxation rates of transverse carbon relaxation (C_+) have the following symbolic expressions:

$$
\begin{aligned}
&R_{C_+H_{\alpha\alpha\alpha}} = \\
&d^2P_0^2\left[\frac{9}{5}\tau_c + \frac{3}{20}J(0,\omega_C-\omega_H) + \frac{27}{20}J(0,\omega_C) + \frac{9}{20}J(0,\omega_H) + \frac{9}{10}J(0,\omega_C+\omega_H)\right] \\
&+d^2P_1^2\left[\frac{1}{20}J(1,\omega_C-\omega_H) + \frac{3}{20}J(1,\omega_H) + \frac{3}{10}J(1,\omega_C+\omega_H)\right] \\
&+d^2P_2^2\left[\frac{1}{80}J(2,\omega_C-\omega_H) + \frac{3}{80}J(2,\omega_H) + \frac{3}{40}J(2,\omega_C+\omega_H)\right] \\
&+e^2\left[\frac{9}{20}J(0,\omega_H) + \frac{27}{80}J(2,\omega_H) + \frac{9}{20}J(0,2\omega_H) + \frac{27}{20}J(2,2\omega_H)\right] \\
&+c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] + dc_CP_0\left[\frac{12}{5}\tau_c + \frac{9}{5}J(0,\omega_C)\right] \\
&+c_H^2P_0^2\left[\frac{9}{5}J(0,\omega_H)\right] + c_H^2P_1^2\left[\frac{3}{5}J(1,\omega_H)\right] + c_H^2P_2^2\left[\frac{3}{20}J(2,\omega_H)\right] - ec_HP_0\left[\frac{9}{5}J(0,\omega_H)\right] \\
&+ec_HP_2\left[\frac{9}{20}J(2,\omega_H)\right] + f^2\left[\frac{3}{20}C(0) + \frac{9}{20}C(\omega_H) + \frac{9}{10}C(2\omega_H)\right] \\
&+g^2\left[\frac{6}{5}C(\omega_H) + \frac{2}{5}C(\omega_D-\omega_H) + \frac{12}{5}C(\omega_D+\omega_H)\right]
\end{aligned}
\tag{C.22}
$$

$$
\begin{aligned}
&R_{C_+H_{\alpha\alpha\beta}} = \\
&d^2P_0^2\left[\frac{1}{5}\tau_c + \frac{3}{20}J(0,\omega_C-\omega_H) + \frac{3}{20}J(0,\omega_C) + \frac{9}{20}J(0,\omega_H) + \frac{9}{10}J(0,\omega_C+\omega_H)\right] \\
&+d^2P_1^2\left[\frac{4}{15}J(1,0) + \frac{1}{20}J(1,\omega_C-\omega_H) + \frac{3}{20}J(1,\omega_H) + +\frac{1}{5}J(1,\omega_C) + \frac{3}{10}J(1,\omega_C+\omega_H)\right] \\
&+d^2P_2^2\left[\frac{1}{5}J(2,0) + \frac{1}{80}J(2,\omega_C-\omega_H) + \frac{3}{80}J(2,\omega_H) + \frac{1}{20}J(2,\omega_C) + \frac{3}{40}J(2,\omega_C+\omega_H)\right] \\
&+e^2\left[\frac{9}{20}J(0,\omega_H) + \frac{27}{80}J(2,\omega_H) + \frac{9}{20}J(0,2\omega_H) + \frac{27}{20}J(2,2\omega_H)\right] \\
&+c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] + dc_CP_0\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] \\
&+c_H^2P_0^2\left[\frac{9}{5}J(0,\omega_H)\right] + c_H^2P_1^2\left[\frac{3}{5}J(1,\omega_H)\right] + c_H^2P_2^2\left[\frac{3}{20}J(2,\omega_H)\right] - ec_HP_0\left[\frac{3}{5}J(0,\omega_H)\right] \\
&+ec_HP_2\left[\frac{9}{5}J(2,\omega_H)\right] + f^2\left[\frac{3}{20}C(0) + \frac{9}{20}C(\omega_H) + \frac{9}{10}C(2\omega_H)\right] \\
&+g^2\left[\frac{6}{5}C(\omega_H) + \frac{2}{5}C(\omega_D-\omega_H) + \frac{12}{5}C(\omega_D+\omega_H)\right]
\end{aligned}
\tag{C.23}
$$

$$
\begin{aligned}
&R_{C_+H_{\alpha\beta\beta}} = \\
&d^2P_0^2\left[\frac{1}{5}\tau_c + \frac{3}{20}J(0,\omega_C-\omega_H) + \frac{3}{20}J(0,\omega_C) + \frac{9}{20}J(0,\omega_H) + \frac{9}{10}J(0,\omega_C+\omega_H)\right] \\
&+d^2P_1^2\left[\frac{4}{15}J(1,0) + \frac{1}{20}J(1,\omega_C-\omega_H) + \frac{3}{20}J(1,\omega_H) + +\frac{1}{5}J(1,\omega_C) + \frac{3}{10}J(1,\omega_C+\omega_H)\right] \\
&+d^2P_2^2\left[\frac{1}{5}J(2,0) + \frac{1}{80}J(2,\omega_C-\omega_H) + \frac{3}{80}J(2,\omega_H) + \frac{1}{20}J(2,\omega_C) + \frac{3}{40}J(2,\omega_C+\omega_H)\right] \\
&+e^2\left[\frac{9}{20}J(0,\omega_H) + \frac{27}{80}J(2,\omega_H) + \frac{9}{20}J(0,2\omega_H) + \frac{27}{20}J(2,2\omega_H)\right] \\
&+c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] - dc_CP_0\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] \\
&+c_H^2P_0^2\left[\frac{9}{5}J(0,\omega_H)\right] + c_H^2P_1^2\left[\frac{3}{5}J(1,\omega_H)\right] + c_H^2P_2^2\left[\frac{3}{20}J(2,\omega_H)\right] + ec_HP_0\left[\frac{3}{5}J(0,\omega_H)\right] \\
&-ec_HP_2\left[\frac{9}{5}J(2,\omega_H)\right] + f^2\left[\frac{3}{20}C(0) + \frac{9}{20}C(\omega_H) + \frac{9}{10}C(2\omega_H)\right] \\
&+g^2\left[\frac{6}{5}C(\omega_H) + \frac{2}{5}C(\omega_D-\omega_H) + \frac{12}{5}C(\omega_D+\omega_H)\right]
\end{aligned}
\tag{C.24}
$$

$$
\begin{aligned}
&R_{C_+H_{\beta\beta\beta}} = \\
&d^2P_0^2\left[\frac{9}{5}\tau_c + \frac{3}{20}J(0,\omega_C-\omega_H) + \frac{27}{20}J(0,\omega_C) + \frac{9}{20}J(0,\omega_H) + \frac{9}{10}J(0,\omega_C+\omega_H)\right] \\
&+d^2P_1^2\left[\frac{1}{20}J(1,\omega_C-\omega_H) + \frac{3}{20}J(1,\omega_H) + \frac{3}{10}J(1,\omega_C+\omega_H)\right] \\
&+d^2P_2^2\left[\frac{1}{80}J(2,\omega_C-\omega_H) + \frac{3}{80}J(2,\omega_H) + \frac{3}{40}J(2,\omega_C+\omega_H)\right] \\
&+e^2\left[\frac{9}{20}J(0,\omega_H) + \frac{27}{80}J(2,\omega_H) + \frac{9}{20}J(0,2\omega_H) + \frac{27}{20}J(2,2\omega_H)\right] \\
&+c_C^2\left[\frac{4}{5}\tau_c + \frac{3}{5}J(0,\omega_C)\right] - dc_CP_0\left[\frac{12}{5}\tau_c + \frac{9}{5}J(0,\omega_C)\right] \\
&+c_H^2P_0^2\left[\frac{9}{5}J(0,\omega_H)\right] + c_H^2P_1^2\left[\frac{3}{5}J(1,\omega_H)\right] + c_H^2P_2^2\left[\frac{3}{20}J(2,\omega_H)\right] + ec_HP_0\left[\frac{9}{5}J(0,\omega_H)\right] \\
&-ec_HP_2\left[\frac{9}{20}J(2,\omega_H)\right] + f^2\left[\frac{3}{20}C(0) + \frac{9}{20}C(\omega_H) + \frac{9}{10}C(2\omega_H)\right] \\
&+g^2\left[\frac{6}{5}C(\omega_H) + \frac{2}{5}C(\omega_D-\omega_H) + \frac{12}{5}C(\omega_D+\omega_H)\right]
\end{aligned}
\tag{C.25}
$$

C.2.3.ii Single quantum carbon cross-relaxation rates (shift basis)

$$\begin{aligned}
R_{C_+H_{\alpha\alpha\alpha},C_+H_{\alpha\alpha\beta}} = \; & +\left(\tfrac{27}{400}e-\tfrac{27}{25}c_XP_0\right)^2\left[-1J\left(0,\omega_H\right)\right] \\
& +c_X^2P_1^2\left[-\tfrac{\sqrt{3}}{5}J\left(1,\omega_H\right)\right] \\
& +c_X^2P_2^2\left[-\tfrac{\sqrt{3}}{20}J\left(2,\omega_H\right)\right] \\
& +d^2P_0^2\left[+\tfrac{\sqrt{27}}{20}J\left(0,\omega_H\right)\right] \\
& +d^2P_1^2\left[+\tfrac{\sqrt{3}}{20}J\left(1,\omega_H\right)\right] \\
& +d^2P_2^2\left[+\tfrac{\sqrt{3}}{80}J\left(2,\omega_H\right)\right] \\
& +e^2\left[-\tfrac{\sqrt{243}}{80}J\left(2,\omega_H\right)\right] \\
& +ec_XP_2\left[-\tfrac{\sqrt{27}}{20}J\left(2,\omega_H\right)\right] \\
& +f^2\left[-\tfrac{\sqrt{3}}{20}C\left(0\right)-\tfrac{\sqrt{27}}{20}C\left(\omega_H\right)-\tfrac{\sqrt{27}}{10}C\left(2\omega_H\right)\right]
\end{aligned} \tag{C.26}$$

$$\begin{aligned}
R_{C_+H_{\alpha\beta\beta},C_+H_{\alpha\alpha\beta}} = \; & +c_X^2P_0^2\left[-\tfrac{6}{5}J\left(0,\omega_H\right)\right] \\
& +c_X^2P_1^2\left[-\tfrac{2}{5}J\left(1,\omega_H\right)\right] \\
& +c_X^2P_2^2\left[-\tfrac{1}{10}J\left(2,\omega_H\right)\right] \\
& +d^2P_0^2\left[+\tfrac{3}{10}J\left(0,\omega_H\right)\right] \\
& +d^2P_1^2\left[+\tfrac{1}{10}J\left(1,\omega_H\right)\right] \\
& +d^2P_2^2\left[+\tfrac{1}{40}J\left(2,\omega_H\right)\right] \\
& +e^2\left[-\tfrac{27}{40}J\left(2,\omega_H\right)\right] \\
& +f^2\left[-\tfrac{1}{10}C\left(0\right)-\tfrac{3}{10}C\left(\omega_H\right)-\tfrac{3}{5}C\left(2\omega_H\right)\right]
\end{aligned} \tag{C.27}$$

$$
\begin{aligned}
R_{C_+H_{\beta\beta\beta},C_+H_{\alpha\beta\beta}} = \; & +\left(\tfrac{27}{400}e+\tfrac{27}{25}c_XP_0\right)^2\left[-1J\left(0,\omega_H\right)\right] \\
& +c_X^2P_1^2\left[-\tfrac{\sqrt{3}}{5}J\left(1,\omega_H\right)\right] \\
& +c_X^2P_2^2\left[-\tfrac{\sqrt{3}}{20}J\left(2,\omega_H\right)\right] \\
& +d^2P_0^2\left[+\tfrac{\sqrt{27}}{20}J\left(0,\omega_H\right)\right] \\
& +d^2P_1^2\left[+\tfrac{\sqrt{3}}{20}J\left(1,\omega_H\right)\right] \\
& +d^2P_2^2\left[+\tfrac{\sqrt{3}}{80}J\left(2,\omega_H\right)\right] \\
& +e^2\left[-\tfrac{\sqrt{243}}{80}J\left(2,\omega_H\right)\right] \\
& +ec_XP_2\left[+\tfrac{\sqrt{27}}{20}J\left(2,\omega_H\right)\right] \\
& +f^2\left[-\tfrac{\sqrt{3}}{20}C\left(0\right)-\tfrac{\sqrt{27}}{20}C\left(\omega_H\right)-\tfrac{\sqrt{27}}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.28}
$$

C.2.3.iii Single quantum carbon auto-relaxation rates (irreducible basis)

$$
\begin{aligned}
& R_{C_+|1\rangle\langle 1|} = \\
& \left(3dP_0+2c_A\right)^2\left[+\tfrac{3}{20}J\left(0,\omega_C\right)+\tfrac{1}{5}\tau_c\right]+\left(e-2c_XP_0\right)^2\left[+\tfrac{9}{20}J\left(0,\omega_H\right)\right] \\
& +c_X^2P_1^2\left[+\tfrac{3}{5}J\left(1,\omega_H\right)\right]+c_X^2P_2^2\left[+\tfrac{3}{20}J\left(2,\omega_H\right)\right] \\
& +d^2P_0^2\left[+\tfrac{3}{20}J\left(0,\omega_C-\omega_H\right)+\tfrac{9}{10}J\left(0,\omega_C+\omega_H\right)+\tfrac{9}{20}J\left(0,\omega_H\right)\right] \\
& +d^2P_1^2\left[+\tfrac{1}{20}J\left(1,\omega_C-\omega_H\right)+\tfrac{3}{10}J\left(1,\omega_C+\omega_H\right)+\tfrac{3}{20}J\left(1,\omega_H\right)\right] \\
& +d^2P_2^2\left[+\tfrac{1}{80}J\left(2,\omega_C-\omega_H\right)+\tfrac{3}{40}J\left(2,\omega_C+\omega_H\right)+\tfrac{3}{80}J\left(2,\omega_H\right)\right] \\
& +e^2\left[+\tfrac{27}{80}J\left(2,\omega_H\right)+\tfrac{9}{20}J\left(0,2\omega_H\right)+\tfrac{27}{20}J\left(2,2\omega_H\right)\right]+ec_XP_2\left[+\tfrac{9}{20}J\left(2,\omega_H\right)\right] \\
& +f^2\left[+\tfrac{3}{20}C\left(0\right)+\tfrac{9}{20}C\left(\omega_H\right)+\tfrac{9}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.29}
$$

$$
\begin{aligned}
&R_{C_+|2\rangle\langle 2|} = \\
&\left(\left(\tfrac{27}{140}\right)^{1/2} e - \left(\tfrac{21}{5}\right)^{1/2} c_X P_0\right)^2 \left[+1J\left(0,\omega_H\right)\right] + \left(dP_0 + 2c_A\right)^2 \left[+\tfrac{3}{20}J\left(0,\omega_C\right) + \tfrac{1}{5}\tau_c\right] \\
&+c_X^2 P_1^2 \left[+\tfrac{8}{15}J\left(1,0\right) + \tfrac{1}{5}J\left(1,\omega_H\right)\right] + c_X^2 P_2^2 \left[+\tfrac{2}{15}J\left(2,0\right) + \tfrac{1}{20}J\left(2,\omega_H\right)\right] \\
&+d^2 P_0^2 \left[+\tfrac{7}{20}J\left(0,\omega_C - \omega_H\right) + \tfrac{21}{10}J\left(0,\omega_C + \omega_H\right) + \tfrac{21}{20}J\left(0,\omega_H\right)\right] \\
&+d^2 P_1^2 \left[+\tfrac{2}{15}J\left(1,0\right) + \tfrac{1}{60}J\left(1,\omega_C - \omega_H\right) + \tfrac{1}{5}J\left(1,\omega_C\right) + \tfrac{1}{10}J\left(1,\omega_C + \omega_H\right) + \tfrac{1}{20}J\left(1,\omega_H\right)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{30}J\left(2,0\right) + \tfrac{1}{240}J\left(2,\omega_C - \omega_H\right) + \tfrac{1}{20}J\left(2,\omega_C\right) + \tfrac{1}{40}J\left(2,\omega_C + \omega_H\right) + \tfrac{1}{80}J\left(2,\omega_H\right)\right] \\
&+e^2 \left[+\tfrac{27}{40}J\left(2,0\right) + \tfrac{9}{35}J\left(0,\omega_H\right) + \tfrac{81}{80}J\left(2,\omega_H\right) + \tfrac{9}{20}J\left(0,2\omega_H\right)\right] \\
&+ec_X P_2 \left[+\tfrac{3}{5}J\left(2,0\right) - \tfrac{9}{20}J\left(2,\omega_H\right)\right] + f^2 \left[+\tfrac{7}{20}C\left(0\right) + \tfrac{21}{20}C\left(\omega_H\right) + \tfrac{21}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.30}
$$

$$
\begin{aligned}
&R_{C_+|3\rangle\langle 3|} = \\
&\left(\left(\tfrac{27}{140}\right)^{1/2} e + \left(\tfrac{21}{5}\right)^{1/2} c_X P_0\right)^2 \left[+1J\left(0,\omega_H\right)\right] + \left(dP_0 - 2c_A\right)^2 \left[+\tfrac{3}{20}J\left(0,\omega_C\right) + \tfrac{1}{5}\tau_c\right] \\
&+c_X^2 P_1^2 \left[+\tfrac{8}{15}J\left(1,0\right) + \tfrac{1}{5}J\left(1,\omega_H\right)\right] + c_X^2 P_2^2 \left[+\tfrac{2}{15}J\left(2,0\right) + \tfrac{1}{20}J\left(2,\omega_H\right)\right] \\
&+d^2 P_0^2 \left[+\tfrac{7}{20}J\left(0,\omega_C - \omega_H\right) + \tfrac{21}{10}J\left(0,\omega_C + \omega_H\right) + \tfrac{21}{20}J\left(0,\omega_H\right)\right] \\
&+d^2 P_1^2 \left[+\tfrac{2}{15}J\left(1,0\right) + \tfrac{1}{60}J\left(1,\omega_C - \omega_H\right) + \tfrac{1}{5}J\left(1,\omega_C\right) + \tfrac{1}{10}J\left(1,\omega_C + \omega_H\right) + \tfrac{1}{20}J\left(1,\omega_H\right)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{30}J\left(2,0\right) + \tfrac{1}{240}J\left(2,\omega_C - \omega_H\right) + \tfrac{1}{20}J\left(2,\omega_C\right) + \tfrac{1}{40}J\left(2,\omega_C + \omega_H\right) + \tfrac{1}{80}J\left(2,\omega_H\right)\right] \\
&+e^2 \left[+\tfrac{27}{40}J\left(2,0\right) + \tfrac{9}{35}J\left(0,\omega_H\right) + \tfrac{81}{80}J\left(2,\omega_H\right) + \tfrac{9}{20}J\left(0,2\omega_H\right)\right] \\
&+ec_X P_2 \left[-\tfrac{3}{5}J\left(2,0\right) + \tfrac{9}{20}J\left(2,\omega_H\right)\right] + f^2 \left[+\tfrac{7}{20}C\left(0\right) + \tfrac{21}{20}C\left(\omega_H\right) + \tfrac{21}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.31}
$$

$$
\begin{aligned}
&R_{C_+|4\rangle\langle 4|} = \\
&(3dP_0 - 2c_A)^2 \left[+\tfrac{3}{20}J(0,\omega_C) + \tfrac{1}{5}\tau_c\right] + (e + 2c_X P_0)^2 \left[+\tfrac{9}{20}J(0,\omega_H)\right] \\
&+c_X^2 P_1^2 \left[+\tfrac{3}{5}J(1,\omega_H)\right] + c_X^2 P_2^2 \left[+\tfrac{3}{20}J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{3}{20}J(0,\omega_C - \omega_H) + \tfrac{9}{10}J(0,\omega_C + \omega_H) + \tfrac{9}{20}J(0,\omega_H)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{20}J(1,\omega_C - \omega_H) + \tfrac{3}{10}J(1,\omega_C + \omega_H) + \tfrac{3}{20}J(1,\omega_H)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{80}J(2,\omega_C - \omega_H) + \tfrac{3}{40}J(2,\omega_C + \omega_H) + \tfrac{3}{80}J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{27}{80}J(2,\omega_H) + \tfrac{9}{20}J(0,2\omega_H) + \tfrac{27}{20}J(2,2\omega_H)\right] \\
&+ec_X P_2 \left[-\tfrac{9}{20}J(2,\omega_H)\right] + f^2 \left[+\tfrac{3}{20}C(0) + \tfrac{9}{20}C(\omega_H) + \tfrac{9}{10}C(2\omega_H)\right]
\end{aligned}
\tag{C.32}
$$

$$
\begin{aligned}
&R_{C_+|5\rangle\langle 5|}/R_{C_+|7\rangle\langle 7|} = \\
&(dP_0 + 2c_A)^2 \left[+\tfrac{3}{20}J(0,\omega_C) + \tfrac{1}{5}\tau_c\right] + c_X^2 P_0^2 \left[+\tfrac{3}{5}J(0,\omega_H)\right] \\
&+c_X^2 P_1^2 \left[+\tfrac{2}{5}J(1,0) + \tfrac{4}{5}J(1,\omega_H)\right] + c_X^2 P_2^2 \left[+\tfrac{1}{10}J(2,0) + \tfrac{1}{5}J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{20}J(0,\omega_C - \omega_H) + \tfrac{3}{10}J(0,\omega_C + \omega_H) + \tfrac{3}{20}J(0,\omega_H)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{6}J(1,0) + \tfrac{1}{15}J(1,\omega_C - \omega_H) + \tfrac{1}{5}J(1,\omega_C) + \tfrac{2}{5}J(1,\omega_C + \omega_H) + \tfrac{1}{5}J(1,\omega_H)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{24}J(2,0) + \tfrac{1}{60}J(2,\omega_C - \omega_H) + \tfrac{1}{20}J(2,\omega_C) + \tfrac{1}{10}J(2,\omega_C + \omega_H) + \tfrac{1}{20}J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{27}{80}J(2,0) + \tfrac{27}{40}J(2,\omega_H) + \tfrac{27}{40}J(2,2\omega_H)\right] \\
&+ec_X P_2 \left[+\tfrac{3}{10}J(2,0) + \tfrac{9}{20}J(2,\omega_H)\right] + f^2 \left[+\tfrac{1}{20}C(0) + \tfrac{3}{20}C(\omega_H) + \tfrac{3}{10}C(2\omega_H)\right]
\end{aligned}
\tag{C.33}
$$

$$
\begin{aligned}
&R_{C_+|6\rangle\langle 6|}/R_{C_+|8\rangle\langle 8|} = \\
&(dP_0 - 2c_A)^2 \left[+\tfrac{3}{20}J(0,\omega_C) + \tfrac{1}{5}\tau_c\right] + c_X^2 P_0^2 \left[+\tfrac{3}{5}J(0,\omega_H)\right] \\
&+c_X^2 P_1^2 \left[+\tfrac{2}{5}J(1,0) + \tfrac{4}{5}J(1,\omega_H)\right] + c_X^2 P_2^2 \left[+\tfrac{1}{10}J(2,0) + \tfrac{1}{5}J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{20}J(0,\omega_C - \omega_H) + \tfrac{3}{10}J(0,\omega_C + \omega_H) + \tfrac{3}{20}J(0,\omega_H)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{6}J(1,0) + \tfrac{1}{15}J(1,\omega_C - \omega_H) + \tfrac{1}{5}J(1,\omega_C) + \tfrac{2}{5}J(1,\omega_C + \omega_H) + \tfrac{1}{5}J(1,\omega_H)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{24}J(2,0) + \tfrac{1}{60}J(2,\omega_C - \omega_H) + \tfrac{1}{20}J(2,\omega_C) + \tfrac{1}{10}J(2,\omega_C + \omega_H) + \tfrac{1}{20}J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{27}{80}J(2,0) + \tfrac{27}{40}J(2,\omega_H) + \tfrac{27}{40}J(2,2\omega_H)\right] \\
&+ec_X P_2 \left[-\tfrac{3}{10}J(2,0) - \tfrac{9}{20}J(2,\omega_H)\right] + f^2 \left[+\tfrac{1}{20}C(0) + \tfrac{3}{20}C(\omega_H) + \tfrac{3}{10}C(2\omega_H)\right]
\end{aligned}
\tag{C.34}
$$

C.2.3.iv Single quantum proton auto-relaxation rates (shift basis)

These are expressions for the single quantum proton auto-relaxation rates (H_+):

$$
\begin{aligned}
&R_{H_+C_\alpha} = \\
&(dP_0 + 2c_X P_0)^2 \left[+\tfrac{3}{20}J(0,\omega_H) + \tfrac{1}{5}\tau_c\right] \\
&+ (dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{20}J(1,\omega_H) + \tfrac{1}{15}J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5}J(0,\omega_C)\right] + c_X^2 P_2^2 \left[+\tfrac{1}{15}J(2,0) + \tfrac{1}{20}J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{4}J(0,\omega_C - \omega_H) + \tfrac{3}{2}J(0,\omega_C + \omega_H) + \tfrac{9}{20}J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{30}J(1,\omega_C - \omega_H) + \tfrac{1}{5}J(1,\omega_C + \omega_H) + \tfrac{3}{20}J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{60}J(2,0) + \tfrac{1}{120}J(2,\omega_C - \omega_H) + \tfrac{1}{80}J(2,\omega_H) + \tfrac{1}{20}J(2,\omega_C + \omega_H) + \tfrac{3}{80}J(2,\omega_C)\right] \\
&+dc_X P_2^2 \left[+\tfrac{1}{15}J(2,0) + \tfrac{1}{20}J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{9}{40}\tau_c + \tfrac{27}{40}J(2,0) + \tfrac{3}{8}J(0,\omega_H) + \tfrac{9}{8}J(2,\omega_H) + \tfrac{3}{20}J(0,2\omega_H) + \tfrac{9}{20}J(2,2\omega_H)\right] \\
&+f^2 \left[+\tfrac{1}{4}C(0) + \tfrac{9}{20}C(\omega_H) + \tfrac{3}{10}C(2\omega_H)\right]
\end{aligned}
\tag{C.35}
$$

$$
\begin{aligned}
&R_{H_+C_\beta} = \\
&(dP_0 - 2c_X P_0)^2 \left[+\tfrac{3}{20} J(0,\omega_H) + \tfrac{1}{5}\tau_c\right] \\
&+ (dP_1 - 2c_X P_1)^2 \left[+\tfrac{1}{20} J(1,\omega_H) + \tfrac{1}{15} J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5} J(0,\omega_C)\right] + c_X^2 P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{20} J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{4} J(0,\omega_C - \omega_H) + \tfrac{3}{2} J(0,\omega_C + \omega_H) + \tfrac{9}{20} J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{30} J(1,\omega_C - \omega_H) + \tfrac{1}{5} J(1,\omega_C + \omega_H) + \tfrac{3}{20} J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{60} J(2,0) + \tfrac{1}{120} J(2,\omega_C - \omega_H) + \tfrac{1}{80} J(2,\omega_H) + \tfrac{1}{20} J(2,\omega_C + \omega_H) + \tfrac{3}{80} J(2,\omega_C)\right] \\
&+dc_X P_2^2 \left[-\tfrac{1}{15} J(2,0) - \tfrac{1}{20} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{9}{40}\tau_c + \tfrac{27}{40} J(2,0) + \tfrac{3}{8} J(0,\omega_H) + \tfrac{9}{8} J(2,\omega_H) + \tfrac{3}{20} J(0,2\omega_H) + \tfrac{9}{20} J(2,2\omega_H)\right] \\
&+f^2 \left[+\tfrac{1}{4} C(0) + \tfrac{9}{20} C(\omega_H) + \tfrac{3}{10} C(2\omega_H)\right]
\end{aligned}
\tag{C.36}
$$

C.2.3.v Single quantum proton auto-relaxation rates (irreducible basis)

$$
\begin{aligned}
&R_{C_\alpha|1\rangle\langle 2|} = \\
&(3e - 10c_X P_0)^2 \left[+\tfrac{3}{100} J(0,\omega_H)\right] + (dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{10} J(1,\omega_H) + \tfrac{1}{15} J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5} J(0,\omega_C)\right] + c_X^2 P_0^2 \left[+\tfrac{4}{5}\tau_c\right] + c_X^2 P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{10} J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{5}\tau_c + \tfrac{3}{20} J(0,\omega_C - \omega_H) + \tfrac{3}{4} J(0,\omega_H) + \tfrac{21}{10} J(0,\omega_C + \omega_H) + \tfrac{3}{4} J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{2}{5} J(1,\omega_C + \omega_H) + \tfrac{1}{10} J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{60} J(2,0) + \tfrac{1}{40} J(2,\omega_H) + \tfrac{1}{10} J(2,\omega_C + \omega_H) + \tfrac{1}{40} J(2,\omega_C)\right] \\
&+dc_A P_0 \left[+\tfrac{6}{5} J(0,\omega_C)\right] + dc_X P_0^2 \left[+\tfrac{4}{5}\tau_c + 3J(0,\omega_H)\right] \\
&+dc_X P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{10} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{9}{20}\tau_c + \tfrac{27}{80} J(2,0) + \tfrac{9}{50} J(0,\omega_H) + \tfrac{27}{40} J(2,\omega_H) + \tfrac{9}{20} J(0,2\omega_H) + \tfrac{27}{40} J(2,2\omega_H)\right] \\
&+ec_X P_0 \left[-\tfrac{6}{5}\tau_c\right] + ec_X P_2 \left[+\tfrac{3}{10} J(2,0)\right] \\
&+edP_0 \left[-\tfrac{3}{5}\tau_c - \tfrac{9}{10} J(0,\omega_H)\right] + edP_2 \left[+\tfrac{3}{20} J(2,0)\right] \\
&+f^2 \left[+\tfrac{9}{20} C(0) + \tfrac{21}{20} C(\omega_H) + \tfrac{3}{2} C(2\omega_H)\right]
\end{aligned}
\tag{C.37}
$$

$$
\begin{aligned}
&R_{C_\alpha|2\rangle\langle 3|} = \\
&(dP_0 + 2c_X P_0)^2 \left[+\tfrac{21}{20} J(0,\omega_H) + \tfrac{1}{5}\tau_c\right] + (dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{20} J(1,\omega_H) + \tfrac{2}{15} J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5} J(0,\omega_C)\right] + c_X^2 P_2^2 \left[+\tfrac{2}{15} J(2,0) + \tfrac{1}{20} J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{7}{20} J(0,\omega_C - \omega_H) + \tfrac{21}{10} J(0,\omega_C + \omega_H) + \tfrac{3}{20} J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{60} J(1,\omega_C - \omega_H) + \tfrac{1}{10} J(1,\omega_C + \omega_H) + \tfrac{1}{5} J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{30} J(2,0) + \tfrac{1}{240} J(2,\omega_C - \omega_H) + \tfrac{1}{80} J(2,\omega_H) + \tfrac{1}{40} J(2,\omega_C + \omega_H) + \tfrac{1}{20} J(2,\omega_C)\right] \\
&+dc_X P_2^2 \left[+\tfrac{2}{15} J(2,0) + \tfrac{1}{20} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{27}{40} J(2,0) + \tfrac{9}{20} J(0,\omega_H) + \tfrac{81}{80} J(2,\omega_H) + \tfrac{9}{20} J(0,2\omega_H)\right] \\
&+f^2 \left[+\tfrac{11}{20} C(0) + \tfrac{27}{20} C(\omega_H) + \tfrac{21}{10} C(2\omega_H)\right]
\end{aligned}
\tag{C.38}
$$

$$
\begin{aligned}
&R_{C_\alpha|3\rangle\langle 4|} = \\
&(3e + 10c_X P_0)^2 \left[+\tfrac{3}{100} J(0,\omega_H)\right] + (dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{10} J(1,\omega_H) + \tfrac{1}{15} J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5} J(0,\omega_C)\right] + c_X^2 P_0^2 \left[+\tfrac{4}{5}\tau_c\right] + c_X^2 P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{10} J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{5}\tau_c + \tfrac{7}{20} J(0,\omega_C - \omega_H) + \tfrac{3}{4} J(0,\omega_H) + \tfrac{9}{10} J(0,\omega_C + \omega_H) + \tfrac{3}{4} J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{15} J(1,\omega_C - \omega_H) + \tfrac{1}{10} J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{60} J(2,0) + \tfrac{1}{60} J(2,\omega_C - \omega_H) + \tfrac{1}{40} J(2,\omega_H) + \tfrac{1}{40} J(2,\omega_C)\right] \\
&+dc_A P_0 \left[-\tfrac{6}{5} J(0,\omega_C)\right] + dc_X P_0^2 \left[+\tfrac{4}{5}\tau_c + 3J(0,\omega_H)\right] \\
&+dc_X P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{10} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{9}{20}\tau_c + \tfrac{27}{80} J(2,0) + \tfrac{9}{50} J(0,\omega_H) + \tfrac{27}{40} J(2,\omega_H) + \tfrac{9}{20} J(0,2\omega_H) + \tfrac{27}{40} J(2,2\omega_H)\right] \\
&+ec_X P_0 \left[+\tfrac{6}{5}\tau_c\right] + ec_X P_2 \left[-\tfrac{3}{10} J(2,0)\right] + edP_0 \left[+\tfrac{3}{5}\tau_c + \tfrac{9}{10} J(0,\omega_H)\right] \\
&+edP_2 \left[-\tfrac{3}{20} J(2,0)\right] + f^2 \left[+\tfrac{9}{20} C(0) + \tfrac{21}{20} C(\omega_H) + \tfrac{3}{2} C(2\omega_H)\right]
\end{aligned}
\tag{C.39}
$$

$$
\begin{aligned}
&R_{C_\alpha|5\rangle\langle 6|}/R_{C_\alpha|7\rangle\langle 8|} = \\
&(dP_0 + 2c_X P_0)^2 \left[+\tfrac{3}{20} J(0,\omega_H) + \tfrac{1}{5}\tau_c\right] + (dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{5} J(1,\omega_H) + \tfrac{1}{6} J(1,0)\right] \\
&+c_A^2 \left[+\tfrac{3}{5} J(0,\omega_C)\right] + c_X^2 P_2^2 \left[+\tfrac{1}{6} J(2,0) + \tfrac{1}{5} J(2,\omega_H)\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{20} J(0,\omega_C - \omega_H) + \tfrac{3}{10} J(0,\omega_C + \omega_H) + \tfrac{3}{20} J(0,\omega_C)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{15} J(1,\omega_C - \omega_H) + \tfrac{2}{5} J(1,\omega_C + \omega_H) + \tfrac{1}{5} J(1,\omega_C)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{24} J(2,0) + \tfrac{1}{60} J(2,\omega_C - \omega_H) + \tfrac{1}{20} J(2,\omega_H) + \tfrac{1}{10} J(2,\omega_C + \omega_H) + \tfrac{1}{20} J(2,\omega_C)\right] \\
&+dc_X P_2^2 \left[+\tfrac{1}{6} J(2,0) + \tfrac{1}{5} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{27}{80} J(2,0) + \tfrac{27}{40} J(2,\omega_H) + \tfrac{27}{40} J(2,2\omega_H)\right] \\
&+f^2 \left[+\tfrac{1}{4} C(0) + \tfrac{9}{20} C(\omega_H) + \tfrac{3}{10} C(2\omega_H)\right]
\end{aligned}
\tag{C.40}
$$

C.2.3.vi Multiple quantum auto-relaxation rates (irreducible basis)

$$
\begin{aligned}
&R_{C_+|1\rangle\langle 2|} = \\
&+(3e - 10c_X P_0)^2 \left[+\tfrac{3}{100} J(0,\omega_H)\right] + (dP_0 + c_A)^2 \left[+\tfrac{3}{5} J(0,\omega_C) + \tfrac{4}{5}\tau_c\right] \\
&+(dP_1 + 2c_X P_1)^2 \left[+\tfrac{1}{15} J(1,0)\right] + (dP_1 + 4c_X P_1)^2 \left[+\tfrac{1}{40} J(1,\omega_H)\right] \\
&+c_X^2 P_0^2 \left[+\tfrac{4}{5}\tau_c\right] + c_X^2 P_2^2 \left[+\tfrac{1}{15} J(2,0) + \tfrac{1}{10} J(2,\omega_H)\right] + c_X c_A P_0 \left[+\tfrac{8}{5}\tau_c\right] \\
&+d^2 P_0^2 \left[+\tfrac{1}{5} J(0,\omega_C - \omega_H) + \tfrac{3}{20} J(0,\omega_C) + \tfrac{9}{5} J(0,\omega_C + \omega_H) + \tfrac{3}{4} J(0,\omega_H)\right] \\
&+d^2 P_1^2 \left[+\tfrac{1}{60} J(1,\omega_C - \omega_H) + \tfrac{1}{10} J(1,\omega_C) + \tfrac{3}{10} J(1,\omega_C + \omega_H) + \tfrac{3}{40} J(1,\omega_H)\right] \\
&+d^2 P_2^2 \left[+\tfrac{1}{60} J(2,0) + \tfrac{1}{240} J(2,\omega_C - \omega_H) + \tfrac{1}{40} J(2,\omega_C) + \tfrac{3}{40} J(2,\omega_C + \omega_H) + \tfrac{1}{40} J(2,\omega_H)\right] \\
&+dc_X P_0^2 \left[+\tfrac{8}{5}\tau_c - \tfrac{6}{5} J(0,\omega_H)\right] + dc_X P_2^2 \left[-\tfrac{1}{15} J(2,0) + \tfrac{1}{20} J(2,\omega_H)\right] \\
&+e^2 \left[+\tfrac{9}{20}\tau_c + \tfrac{27}{80} J(2,0) + \tfrac{9}{50} J(0,\omega_H) + \tfrac{27}{40} J(2,\omega_H) + \tfrac{9}{20} J(0,2\omega_H) + \tfrac{27}{40} J(2,2\omega_H)\right] \\
&+ec_A \left[-\tfrac{6}{5}\tau_c\right] + ec_X P_0 \left[-\tfrac{6}{5}\tau_c\right] + ec_X P_2 \left[+\tfrac{3}{10} J(2,0)\right] + edP_0 \left[-\tfrac{6}{5}\tau_c\right] \\
&+edP_2 \left[-\tfrac{3}{20} J(2,0) + \tfrac{9}{40} J(2,\omega_H)\right] \\
&+f^2 \left[+\tfrac{9}{20} C(0) + \tfrac{21}{20} C(\omega_H) + \tfrac{3}{2} C(2\omega_H)\right]
\end{aligned}
\tag{C.41}
$$

$$
\begin{aligned}
&R_{C_+|2\rangle\langle 3|} = \\
&+c_A^2\left[+\tfrac{4}{5}\tau_c+\tfrac{3}{5}J\left(0,\omega_C\right)\right]+c_X^2P_0^2\left[+\tfrac{4}{5}\tau_c+\tfrac{21}{5}J\left(0,\omega_H\right)\right] \\
&+c_X^2P_1^2\left[+\tfrac{8}{15}J\left(1,0\right)+\tfrac{1}{5}J\left(1,\omega_H\right)\right]+c_X^2P_2^2\left[+\tfrac{2}{15}J\left(2,0\right)+\tfrac{1}{20}J\left(2,\omega_H\right)\right] \\
&+c_Xc_AP_0\left[+\tfrac{8}{5}\tau_c\right] \\
&+d^2P_0^2\left[+\tfrac{3}{10}J\left(0,\omega_C-\omega_H\right)+\tfrac{3}{20}J\left(0,\omega_C\right)+\tfrac{12}{5}J\left(0,\omega_C+\omega_H\right)+\tfrac{21}{20}J\left(0,\omega_H\right)\right] \\
&+d^2P_1^2\left[+\tfrac{2}{15}J\left(1,0\right)+\tfrac{1}{5}J\left(1,\omega_C\right)+\tfrac{1}{5}J\left(1,\omega_C+\omega_H\right)+\tfrac{1}{20}J\left(1,\omega_H\right)\right] \\
&+d^2P_2^2\left[+\tfrac{1}{30}J\left(2,0\right)+\tfrac{1}{20}J\left(2,\omega_C\right)+\tfrac{1}{20}J\left(2,\omega_C+\omega_H\right)+\tfrac{1}{80}J\left(2,\omega_H\right)\right] \\
&+e^2\left[+\tfrac{27}{40}J\left(2,0\right)+\tfrac{9}{20}J\left(0,\omega_H\right)+\tfrac{81}{80}J\left(2,\omega_H\right)+\tfrac{9}{20}J\left(0,2\omega_H\right)\right] \\
&+edP_0\left[-\tfrac{9}{10}J\left(0,\omega_H\right)\right]+edP_2\left[+\tfrac{3}{10}J\left(2,0\right)-\tfrac{9}{40}J\left(2,\omega_H\right)\right] \\
&+f^2\left[+\tfrac{11}{20}C\left(0\right)+\tfrac{27}{20}C\left(\omega_H\right)+\tfrac{21}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.42}
$$

$$
\begin{aligned}
&R_{C_+|3\rangle\langle 4|} = \\
&+\left(3e+10c_XP_0\right)^2\left[+\tfrac{3}{100}J\left(0,\omega_H\right)\right]+\left(dP_0-c_A\right)^2\left[+\tfrac{3}{5}J\left(0,\omega_C\right)+\tfrac{4}{5}\tau_c\right] \\
&+\left(dP_1-2c_XP_1\right)^2\left[+\tfrac{1}{15}J\left(1,0\right)\right]+\left(dP_1-4c_XP_1\right)^2\left[+\tfrac{1}{40}J\left(1,\omega_H\right)\right] \\
&+c_X^2P_0^2\left[+\tfrac{4}{5}\tau_c\right]+c_X^2P_2^2\left[+\tfrac{1}{15}J\left(2,0\right)+\tfrac{1}{10}J\left(2,\omega_H\right)\right]+c_Xc_AP_0\left[+\tfrac{8}{5}\tau_c\right] \\
&+d^2P_0^2\left[+\tfrac{1}{5}J\left(0,\omega_C-\omega_H\right)+\tfrac{3}{20}J\left(0,\omega_C\right)+\tfrac{9}{5}J\left(0,\omega_C+\omega_H\right)+\tfrac{3}{4}J\left(0,\omega_H\right)\right] \\
&+d^2P_1^2\left[+\tfrac{1}{60}J\left(1,\omega_C-\omega_H\right)+\tfrac{1}{10}J\left(1,\omega_C\right)+\tfrac{3}{10}J\left(1,\omega_C+\omega_H\right)+\tfrac{3}{40}J\left(1,\omega_H\right)\right] \\
&+d^2P_2^2\left[+\tfrac{1}{60}J\left(2,0\right)+\tfrac{1}{240}J\left(2,\omega_C-\omega_H\right)+\tfrac{1}{40}J\left(2,\omega_C\right)+\tfrac{3}{40}J\left(2,\omega_C+\omega_H\right)+\tfrac{1}{40}J\left(2,\omega_H\right)\right] \\
&+dc_XP_0^2\left[-\tfrac{8}{5}\tau_c+\tfrac{6}{5}J\left(0,\omega_H\right)\right]+dc_XP_2^2\left[+\tfrac{1}{15}J\left(2,0\right)-\tfrac{1}{20}J\left(2,\omega_H\right)\right] \\
&+e^2\left[+\tfrac{9}{20}\tau_c+\tfrac{27}{80}J\left(2,0\right)+\tfrac{9}{50}J\left(0,\omega_H\right)+\tfrac{27}{40}J\left(2,\omega_H\right)+\tfrac{9}{20}J\left(0,2\omega_H\right)+\tfrac{27}{40}J\left(2,2\omega_H\right)\right] \\
&+ec_A\left[+\tfrac{6}{5}\tau_c\right]+ec_XP_0\left[+\tfrac{6}{5}\tau_c\right]+ec_XP_2\left[-\tfrac{3}{10}J\left(2,0\right)\right]+edP_0\left[-\tfrac{6}{5}\tau_c\right] \\
&+edP_2\left[-\tfrac{3}{20}J\left(2,0\right)+\tfrac{9}{40}J\left(2,\omega_H\right)\right]+f^2\left[+\tfrac{9}{20}C\left(0\right)+\tfrac{21}{20}C\left(\omega_H\right)+\tfrac{3}{2}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.43}
$$

$$
\begin{aligned}
&R_{C_+|5\rangle\langle 6|}/R_{C_+|7\rangle\langle 8|} = \\
&c_A^2\left[+\tfrac{4}{5}\tau_c + \tfrac{3}{5}J\left(0,\omega_C\right)\right] + c_X^2 P_0^2\left[+\tfrac{4}{5}\tau_c + \tfrac{3}{5}J\left(0,\omega_H\right)\right] \\
&+c_X^2 P_1^2\left[+\tfrac{2}{3}J\left(1,0\right) + \tfrac{4}{5}J\left(1,\omega_H\right)\right] + c_X^2 P_2^2\left[+\tfrac{1}{6}J\left(2,0\right) + \tfrac{1}{5}J\left(2,\omega_H\right)\right] + c_X c_A P_0\left[+\tfrac{8}{5}\tau_c\right] \\
&+d^2 P_0^2\left[+\tfrac{3}{20}J\left(0,\omega_C\right) + \tfrac{3}{5}J\left(0,\omega_C+\omega_H\right) + \tfrac{3}{20}J\left(0,\omega_H\right)\right] \\
&+d^2 P_1^2\left[+\tfrac{1}{10}J\left(1,0\right) + \tfrac{1}{20}J\left(1,\omega_C-\omega_H\right) + \tfrac{1}{5}J\left(1,\omega_C\right) + \tfrac{1}{2}J\left(1,\omega_C+\omega_H\right) + \tfrac{1}{5}J\left(1,\omega_H\right)\right] \\
&+d^2 P_2^2\left[+\tfrac{1}{40}J\left(2,0\right) + \tfrac{1}{80}J\left(2,\omega_C-\omega_H\right) + \tfrac{1}{20}J\left(2,\omega_C\right) + \tfrac{1}{8}J\left(2,\omega_C+\omega_H\right) + \tfrac{1}{20}J\left(2,\omega_H\right)\right] \\
&+e^2\left[+\tfrac{27}{80}J\left(2,0\right) + \tfrac{27}{40}J\left(2,\omega_H\right) + \tfrac{27}{40}J\left(2,2\omega_H\right)\right] \\
&+edP_2\left[+\tfrac{3}{20}J\left(2,0\right) + \tfrac{9}{40}J\left(2,\omega_H\right)\right] + f^2\left[+\tfrac{1}{4}C\left(0\right) + \tfrac{9}{20}C\left(\omega_H\right) + \tfrac{3}{10}C\left(2\omega_H\right)\right]
\end{aligned}
\tag{C.44}
$$

C.2.4 Relaxation rate expressions for a ^{13}C-^{19}F$_3$ spin system

The expressions for single quantum auto-relaxation rates in a ^{13}C-^{19}F$_3$ spin system are very similar to those in appendix C.2.3. The interaction constants for the ^{13}C-^{19}F$_3$ system are defined as:

$$
\begin{gathered}
d = -\frac{\mu_0\hbar\gamma_C\gamma_F}{4\pi R_{F-C}^3} \quad e = -\frac{\mu_0\hbar\gamma_F\gamma_F}{4\pi R_{F-F}^3} \quad c_C = \frac{1}{3}\gamma_C B_0 \Delta\delta_C \quad c_F = \frac{1}{3}\gamma_F B_0 \Delta\delta_F \\
f = -\frac{\mu_0\hbar\gamma_F\gamma_H}{4\pi} \quad g = -\frac{\mu_0\hbar\gamma_F\gamma_D}{4\pi}
\end{gathered}
\tag{C.45}
$$

Furthermore, all ω_H values in expressions need to be changed to ω_F. In addition the terms in appendix C.2.3 governed by the constant f^2 need to be changed, where $C(0)$ and $C(2\omega_H)$ terms are to be replaced by $C(\omega_H - \omega_F)$ and $C(\omega_H + \omega_F)$, respectively.

C.3 Methods

C.3.1 Pulse sequence simulations

C.3.1.i Pulse sequence simulations for ^{15}N-^{1}H spin systems

SimSeq was used to perform simulations of sensitivity enhanced HSQC (Figure C.1A) and TROSY (Figure C.1B) pulse sequences on a ^{15}N-^{1}H spin system. The full Liouvillian relaxation matrix for the spin system was calculated using RelCalc assuming an N-H bond length of 1 Å, chemical shift anisotropy values of $\Delta\delta_N$=-160 ppm and $\Delta\delta_N$=-16 ppm. The effects of an external proton/deuteron a distance of 2.0 Å away were also included. The chemical shifts of the ^{15}N and ^{1}H nuclei were set to 120 ppm and 8 ppm, respectively, and the ^{15}N-^{1}H coupling constant was set to -93.3 Hz.

The simulations were performed according to the sequences in Figure C.1 with 4 dummy scans and 8 transients. The acquisition times for the direct and indirect dimensions were set to 1 s and 0.5 s, respectively. The simulations assumed complete return to equilibrium before the next transient was recorded. All pulses were applied on resonance to the ^{15}N and ^{1}H nuclei and the sweep widths (SW) were set to 0.7 ppm and 3.0 ppm for the direct and indirect dimensions, respectively. The first t_1 increment was set to $1/\mathrm{SW}_N$(Hz). Exponential line-broadening values of 20 Hz and 2 Hz were applied to the proton and ^{15}N dimensions, respectively. Zero-filling was used to double the FID size in each dimension. Zero order ($P_0 = 90^o$) and first order ($P_1 = -180^o$) phase corrections were applied to the indirect dimensions. Zero order phase corrections of 180^o and -90^o were applied to the direct dimension for the sensitivity-enhanced HSQC and TROSY spectra, respectively.

Simulations were performed for a 600 MHz spectrometer at 50 rotational correlation times (τ_c) between 0.1 ns and 1000 ns (evenly separated in logarithmic base-10 space) including the effects of both external protons and deuterons. The

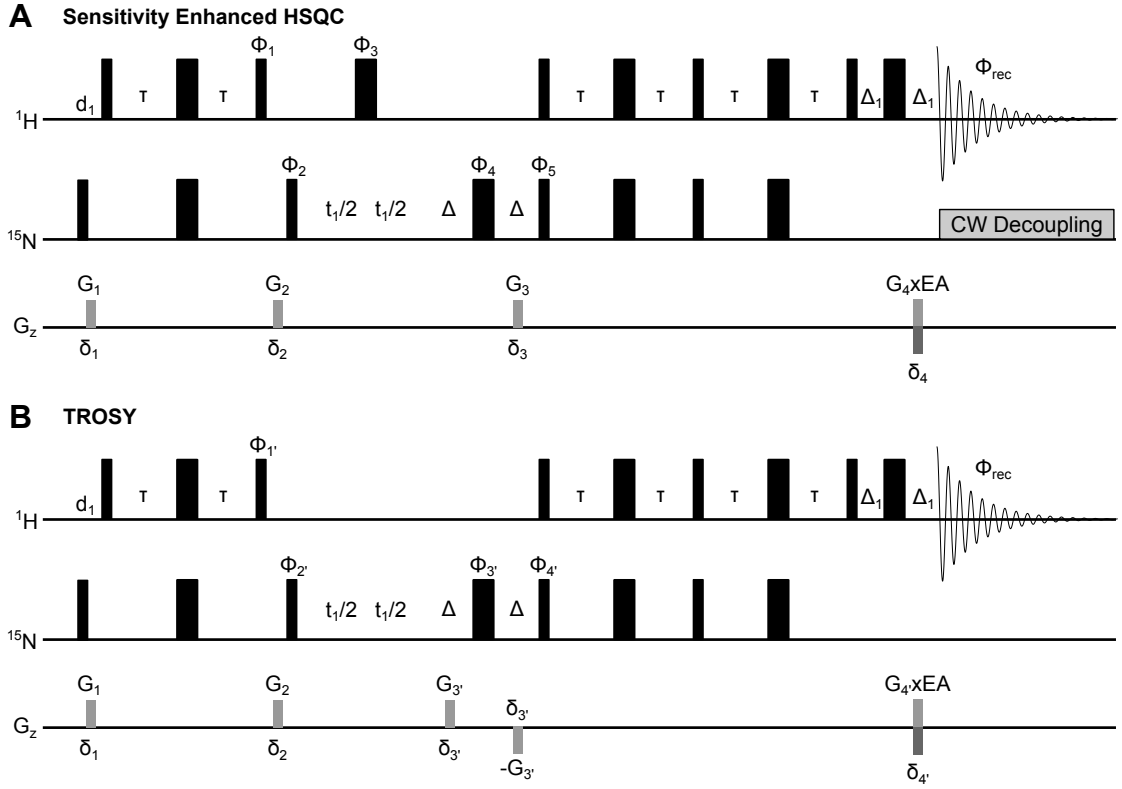


Figure C.1: Pulse sequences used to simulate the **A** sensitivity-enhanced HSQC[491] and **B** TROSY[53] spectra of a ^{15}N-^{1}H group using SimSeq. Narrow and wide black rectangles represent 90^o and 180^o pulses, respectively. The values of τ were set to 2.71 ms ($\approx 1/(4|J|)$) and Δ/Δ_1 were set to 1.38 ms and 500 μs, respectively, for both sequences. The phase cycle for the sensitivity-enhanced HSQC was set as follows: $\Phi_1 = y$, $\Phi_2 = y$, $\Phi_3 = y$,$\Phi_4 = (x, y, -x, -y)$, $\Phi_5 = x$, $\Phi_{rec} = (y, -y)$. All other pulses had phase x. Quadrature detection in the indirect dimension of the sensitivity-enhanced HSQC sequence was performed through changing the phase of Φ_5 to y and inverting the sign of EA from -1 to 1. Continuous wave (CW) ^{15}N decoupling was provided during acquisition. The gradients strengths (G) and durations (δ) in the sensitivity-enhanced HSQC sequence were set to ensure coherence selection ($\gamma_N G_3 \delta_3 = \gamma_H G_4 \delta_4$). The phase cycle for the TROSY sequence was set as follows: $\Phi_{1'} = (y, -x, -y, x)$, $\Phi_{2'} = (y, x, -y, -x)$, $\Phi_{3'} = y$,$\Phi_{4'} = y$, $\Phi_{rec'} = (-y, -x, y, x)$. All other pulses had phase x. Quadrature detection in the indirect dimension of the TROSY was performed through 180^o incrementation of $\Phi_{3'}$ and $\Phi_{4'}$, changing from $\Phi_{1'}$ to $\Phi_{2'}$ and inverting the sign of EA from 1 to -1. The gradients strengths (G) and durations (δ) in the TROSY sequence were set to ensure coherence selection ($2\gamma_N G_{3'} \delta_{3'} = \gamma_H G_{4'} \delta_{4'}$). G_1 and G_2 purge gradients were used to remove unwanted coherences in both spectra. In both sequences the data was processed in a Rance-Kay/Echo-AntiEcho manner to produce sensitivity enhancement of a factor of 2. Note that SimSeq performs simulations in the absence of noise. If noise were included, the signal to noise enhancement by sensitivity enhancement would by $\sqrt{2}$.

TROSY simulations were then repeated (for τ_c values of 1 ns and 40 ns) at 11 spectrometer frequencies between 500 MHz and 1500 MHz.

C.3.1.ii Pulse sequence simulations for ^{13}C-^{1}H$_3$ spin systems

SimSeq was used to perform simulations of HSQC (Figure C.2A), HMQC (Figure C.2B) and ddHMQC (Figure C.2C) pulse sequences on a $^{13}CH_3$ spin system. The full Liouvillian relaxation matrix for the spin system was calculated using RelCalc assuming a C-H bond length of 1.09 Å, chemical shift anisotropy values of $\Delta\delta_C$=20 ppm and $\Delta\delta_H$=1 ppm. Rotation about the 3-fold rotational axis was assumed to have a timescale τ_m of 10 ps. The effects of an external proton/deuteron a distance of 2.0 Å away an angle of 60^o to the 3-fold rotation axis were also included. The chemical shifts of the ^{13}C and ^{1}H nuclei were set to 15 ppm and 2.0 ppm, respectively, and the ^{13}C-^{1}H coupling constant was set to 139 Hz.

The 1D pulse acquire carbon pulse sequence simulations were performed with 4 dummy scans and 148 transients. 2D heteronuclear pulse sequence simulations were performed according to the pulse sequences in Figure C.2, with 4 dummy scans and 2 transients. The acquisition times for the direct and indirect dimensions were set to 1 s and 50 ms, respectively. The simulations assumed complete return to equilibrium before the next transient was recorded. All pulses were applied on resonance to the ^{13}C and ^{1}H nuclei and the sweep widths (SW) were set to 5.0 ppm in both dimensions. The first t_1 increment was set to $1/\mathrm{SW}_C$(Hz). For the ddHMQC, each FID was modified by division of $\sin(\pi Jt)$ for $0 < t < 1/(2J)$. Zero-filling was used to double the FID size in each dimension for all spectra. Exponential line-broadening values of 20 Hz and 2 Hz were applied to the proton and ^{13}C dimensions, respectively. Zero order ($P_0 = 90^o$) and first order ($P_1 = -180^o$) phase corrections were applied to the indirect dimensions.

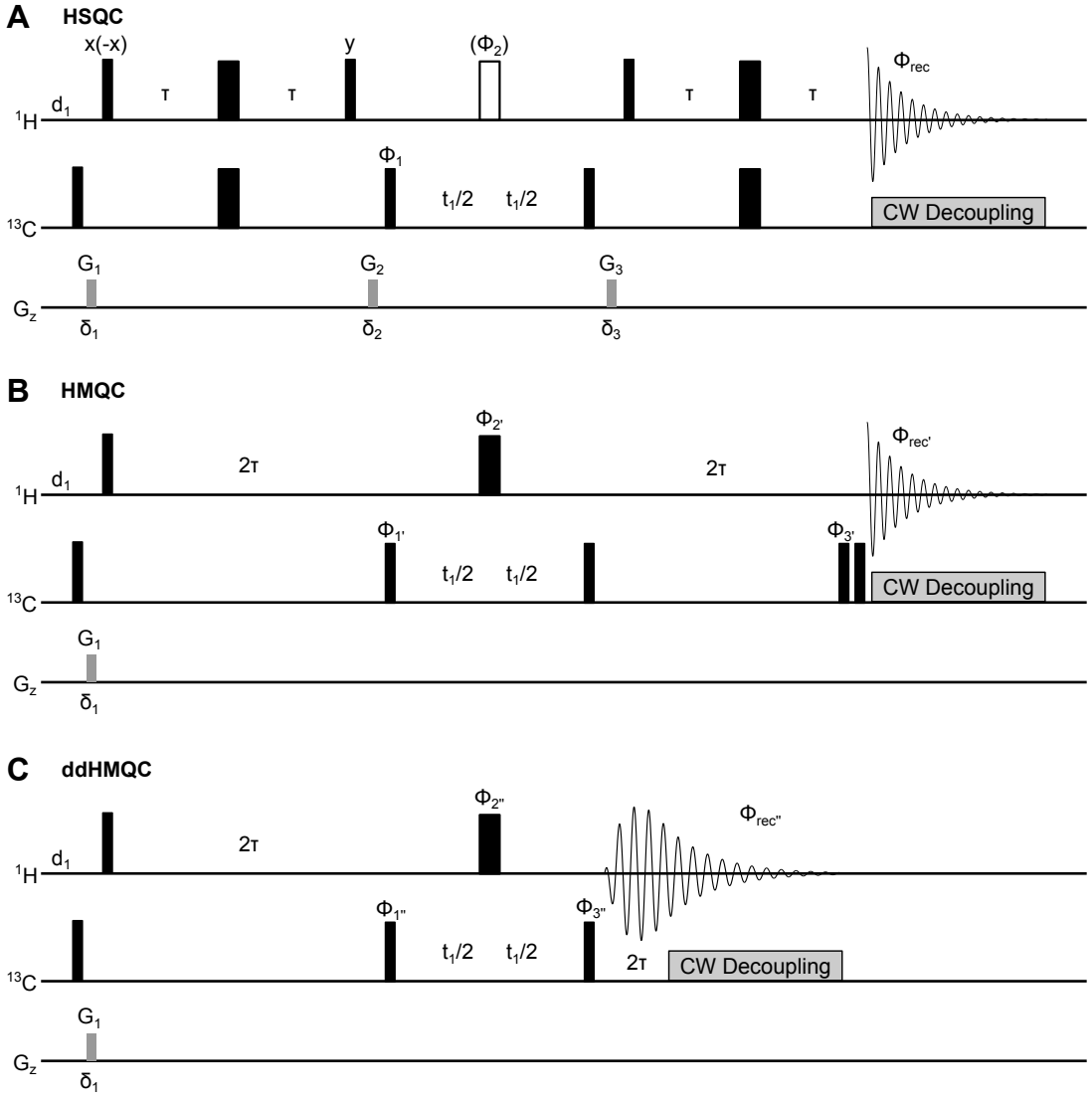


Figure C.2: Pulse sequences used to simulate the **A** HSQC,[108] **B** HMQC[440–442] and **C** ddHMQC[59] spectra of a $^{13}CH_3$ group using SimSeq. Narrow and wide black rectangles represent 90^o and 180^o pulses, respectively. The values of τ were set to 1.90 ms ($\approx 1/(4|J|)$) in all sequences. Optional proton decoupling was applied during the HSQC by adding the 180^o pulse with phase Φ_2 and changing the phase of the first proton pulse to $-x$. The phase cycle for the HSQC was set as follows: $\Phi_1 = (x, -x)$, $\Phi_2 = (x, x, -x, -x)$, $\Phi_{rec} = (x, -x)$. All other pulses had phase x. Quadrature detection in the indirect dimension of the HSQC sequence was performed by incrementing Φ_1 by 90^o. The phase cycle for the HMQC sequence was set as follows: $\Phi_{1'} = (x, -x)$, $\Phi_{2'} = (x, x, y, y, -x, -x, -y, -y)$, $\Phi_{3'} = (x, x, -x, -x)$,$\Phi_{rec'} = (y, -y, -y, y)$. All other pulses had phase x. For the ddHMQC, the phase cycle was as follows: $\Phi_{1''} = (x, -x)$, $\Phi_{2''} = x$, $\Phi_{3''} = x$,$\Phi_{rec''} = (-y, y)$. Quadrature detection in the indirect dimension of the HMQC/ddHMQC was performed through 90^o incrementation of $\Phi_{1'}/\Phi_{1''}$. Continuous wave (CW) ^{13}C decoupling was provided during acquisition for all experiments, being delayed by a duration 2τ in the ddHMQC.

C.3.1.iii Pulse sequence simulations for ^{13}C-^{19}F$_3$ spin systems

SimSeq was used to perform simulations of HSQC (Figure C.2A) and HMQC (Figure C.2B) pulse sequences on a $^{13}CF_3$ spin system. The complete Liouvillian relaxation matrix for the spin system was calculated using RelCalc assuming a C-F bond length of 1.38 Å, fitted chemical shift anisotropy values of $\Delta\delta_C$=47.1 ppm and $\Delta\delta_F$=161.2 ppm. Rotation about the 3-fold rotational axis was set to the fitted timescale τ_m of 224 ps. The chemical shifts of the ^{13}C and ^{19}F nuclei were set to 130 ppm and -40 ppm, respectively, and the ^{13}C-^{19}F coupling constant was set to -305 Hz.

The 1D pulse acquire carbon pulse sequence simulations were performed with 4 dummy scans and 148 transients. 2D heteronuclear pulse sequence simulations were performed according to the sequences in Figure C.2 with 4 dummy scans and 2 transients. The only change from the $^{13}CH_3$ simulations was the τ delays which were set to 0.82 ms. The acquisition times for the direct and indirect dimensions were set to 1 s and 50 ms, respectively. The simulations assumed complete return to equilibrium before the next transient was recorded. All pulses were applied on resonance to the ^{13}C and ^{19}F nuclei and the sweep widths (SW) were set to 5.0 ppm in both dimensions. The first t_1 increment was set to $1/\mathrm{SW}_C$(Hz). Zero-filling was used to double the FID size in each dimension for all spectra. Exponential line-broadening values of 20 Hz and 2 Hz were applied to the fluorine and ^{13}C dimensions, respectively. Zero order ($P_0 = 90^o$) and first order ($P_1 = -180^o$) phase corrections were applied to the indirect dimensions.

C.3.2 Experimental NMR methods and analysis

C.3.2.i NMR sample preparation and spectrometer type

The pure $^{13}CF_3$ labelled HSP16.5 and background deuterated $^{13}CH_3$ labelled HSP16.5 samples were placed in a sodium phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH=7.0) containing 1 mM EDTA, 2 mM NaN_3, with 10% and 100% D_2O percentages. Additionally, 100 μM 1,1,1-trifluoroacetic acid (TFA) was added to the $^{13}CF_3$ labelled sample. The final concentrations of protein in the $^{13}CF_3$ and $^{13}HF_3$ labelled samples were approximately 4 mM (170 μM of HSP16.5 oligomers) and 1 mM (41 μM oligomers), respectively[†]. The $^{13}CF_3$-labelled HSP16.5 sample was placed in a Bruker Shigemi tube matched to D_2O. The $^{13}CH_3$-labelled HSP16.5 sample was placed in a Bruker 3 mm tube. NMR spectra for this project were performed on a Bruker AVANCE NEO 600 MHz spectrometer with CPRHe-QR-1H/19F/13C/15N-5mm-Z helium-cooled cryoprobe. All spectra were processed and analysed using SpinExplorer, a python-based graphical interface used to process data based on NMRPipe[482] and nmrGlue[483] in python. All ^{19}F chemical shifts were referenced to TFA with a chemical shift of -76.55 ppm relative to $CFCl_3$.

C.3.2.ii Fluorine and proton 1D NMR experiments

Fluorine 1D experiments were performed using the Bruker standard pulse-acquire (zg) experiment. Proton 1D experiments were performed using the Bruker standard pulse sequence with excitation sculpting water suppression (zgesgp). The experiments were performed at 5^{o}C and the fluorine and proton carrier frequencies were set to -41.000 ppm and 4.961 ppm. The acquisition times were set to 0.1 s and 1.7 s for fluorine and proton experiments. The number of transients for the fluorine and proton experiments were set to 256 and 128.

[†]NMR sample preparation was performed by Dr. Satoshi Kishigami.

C.3.2.iii Diffusion NMR experiments

Fluorine and proton stimulated echo longitudinal encode-decode diffusion experiments were performed on the labelled HSP16.5 samples (Bruker sequences stebpgp1s/stebpgp1s19) utilising bipolar gradients at 20^{o}C. The maximum spectrometer gradient (100%) was calibrated as 48.15 Gcm^{-1} by measuring the diffusion coefficient of water at 20^{o}C. The gradient files used were SMSQ10.100 with an integral factor of 0.9. For $^{13}CF_3$ HSP16.5 a total of 4 dummy scans and 248 transients were recorded for each gradient value and the relaxation delay between repeat scans was set to 1.5 s. For $^{13}CH_3$ HSP16.5, 32 transients were used instead. The resulting pseudo-2D spectra were processed using exponential line broadening of 5 Hz and zero filling. The diffusion data was analysed using SpinExplorer to calculate the diffusion coefficient.

C.3.2.iv Single quantum carbon NMR relaxation experiments

The pulse sequence used to measure single quantum carbon relaxation rates on the labelled HSP16.5 samples was shown in Figure 5.17A. For the $^{13}CF_3$ labelled system, the carbon carrier frequency was set to 130.80 ppm. The fluorine carrier frequencies were set to -41.930 ppm and -41.000 ppm for the experiments at 5^{o}C and 20^{o}C, respectively. A total of 128 and 64 complex points were obtained in the indirect dimension for experiments at 5^{o}C and 20^{o}C, respectively. A total of 128 dummy scans and 64 transients were recorded for each spectrum.

For the $^{13}CH_3$ labelled system at 20^{o}C, the carbon carrier frequency was set to 16.479 ppm. The proton carrier frequency was centred on water at 4.964 ppm. A total of 256 complex points were obtained in the indirect dimension. The spectrum was recorded in 2 parts containing 16 transients each. These were added together to give the total spectrum. The number of dummy scans was set to 32.

In all experiments, the relaxation delay time between repeat scans was set to 1.5 s. DIPSI-2 carbon decoupling was applied during acquisition for a total of

100 ms. During processing, linear prediction was performed in the indirect dimension for the data recorded on the $^{13}CF_3$ labelled system at 20^{o}C. In both dimensions, Lorentz-to-Gauss apodization was applied to produce gaussian shaped peaks. The peaks corresponding to each single quantum carbon coherence were then fitted using FuDA[485] at each relaxation delay (T_{relax}) to obtain peak intensities. These intensities were fitted to single exponential decays to obtain the relaxation rates for each peak.

C.3.2.v Single quantum proton NMR relaxation experiments

The pulse sequence used to measure single quantum proton relaxation rates on the $^{13}CH_3$ labelled HSP16.5 sample was shown in Figure 5.17B. The carbon carrier frequency was set to approximately 16 ppm. The proton carrier frequency was centred on water at 4.964 ppm. During the experiments, the number of CPMG cycles (n) was varied in units of 4 between 4 and 28. The CPMG delay τ_{cp} was set to 200 μs. The spectrum was recorded in 4 parts containing 512 transients each. These were added together to give the total spectrum equalling 2048 transients. The relaxation delay time between repeat scans was set to 1.5 s and the acquisition time was set to 0.5 s. During processing, digital solvent suppression was applied to the FID. Exponential line broadening of 1.0 Hz and zero filling was applied. The processed relaxation data was analysed using SpinExplorer to determine the relaxation rate of each coherence.

C.3.2.vi Single quantum fluorine NMR relaxation experiments

A fluorine CPMG experiment used to measure single quantum fluorine relaxation rates on the $^{13}CF_3$ labelled HSP16.5 sample (Figure 5.17C). The fluorine carrier frequencies were set to -42.487 ppm and -41.800 ppm for the experiments at 5^{o}C and 20^{o}C, respectively. During the experiments, the number of CPMG cycles (n) was varied in units of 4 between 4 and 20. The CPMG delay τ_{cp} was set to 50 μs and 1.25 ms for the experiments at 5^{o}C and 20^{o}C. A total of 1024 and 2048 transients

were obtained for the spectra at 5^{o}C and 20^{o}C. The relaxation delay time between repeat scans was set to 2.5 s and the acquisition time was set to 0.1 s. Exponential line broadening of 25.0 Hz and 20.0 Hz was applied to the 5^{o}C and 20^{o}C data to increase the signal-to-noise ratio and zero filling was applied. The processed relaxation data was analysed using SpinExplorer to determine the relaxation rate of each coherence.

C.3.2.vii Two-dimensional NMR experiments

HEHAHA The 2D HEHAHA experiments were performed using the sequence described in Figure 5.22A at 5^{o}C. For $^{13}CF_3$ labelled HSP16.5 sample, the fluorine and carbon carrier frequencies were set to -41.000 ppm and 130.8 ppm. The proton and carbon carrier frequencies for the $^{13}CH_3$ labelled system were 1.976 ppm and 14.48 ppm. A pair of 4.16 kHz $^{19}F/^{1}H$ and ^{13}C DIPSI-2x elements were applied for durations of 1.96 ms or 4.3 ms (for ^{1}H or ^{19}F labelled systems) to achieve cross polarisation. A total of 64 and 128 complex points were obtained in the indirect dimensions for the $^{13}CF_3$ and $^{13}CH_3$ labelled samples. The number of transients for the experiments on the $^{13}CH_3$ and $^{13}CF_3$ labelled systems were 32 and 128. The number of dummy scans was set to 32 and 4 for each system. The relaxation delay was set to 1.2 s for both cases. Exponential line broadening of 5 Hz and 50 Hz was applied to the direct and indirect dimensions, respectively, and a single round of zero filling was applied in each dimension.

HSQC/HMQC The 2D HSQC and HMQC experiments were performed using the sequences described in Figure C.3 at 5^{o}C. For $^{13}CF_3$ labelled HSP16.5 sample, the fluorine and carbon carrier frequencies were set to -41.000 ppm and 130.8 ppm. The proton and carbon carrier frequencies for the $^{13}CH_3$ labelled system were 4.961 ppm and 14.48 ppm. DIPSI-2x ^{13}C decoupling was provided during acquisition. A total of 64 and 128 complex points were obtained in the indirect dimensions for the $^{13}CF_3$ and $^{13}CH_3$ labelled samples. The number of transients for the experiments

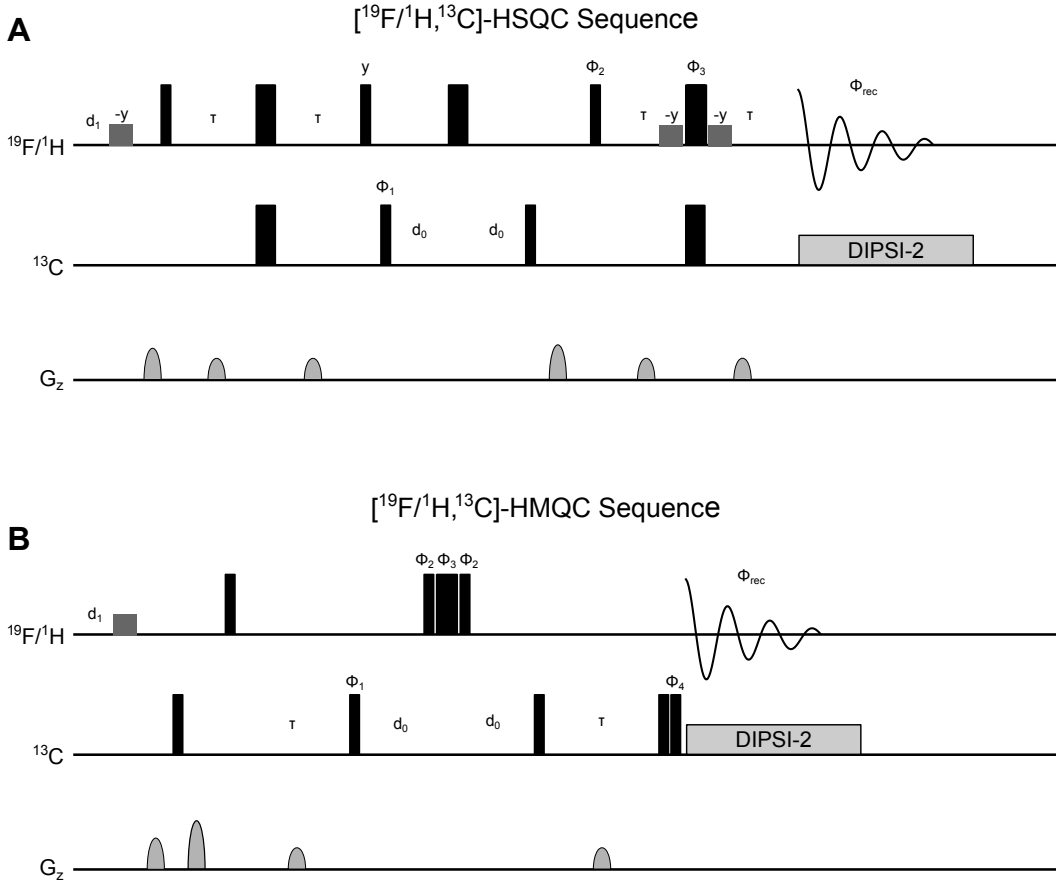


Figure C.3: The (**A**) HSQC and (**B**) HMQC pulse sequences applied to $^{13}CF_3$ and $^{13}CH_3$ labelled HSP16.5 samples. The phase cycles used were as follows, Φ_1=(x,x,x,x,-x,-x,-x,-x), Φ_2=(x,-x), Φ_3=(x,x,-x,-x), Φ_{rec}=(x,-x,x,-x,-x,x,-x,x). For the HMQC: Φ_1=(x,-x), Φ_2=(x,x,y,y,-x,-x,-y,-y), Φ_3=(y,y,-x,-x,-y,-y,x,x),Φ_4=(x,x,-x,-x), Φ_{rec}=(x,-x,-x,x). The delays τ were set to $1/|4J|$, where J is the ^{13}C-$^{19}F/^{1}H$ coupling constant and had a values of -305 Hz/139 Hz, respectively. Quadrature detection was performed through 90^o incrementation of Φ_1 for each sequence. Water selective proton pulses (gray) were applied for the $^{13}CH_3$ measurements.

on the $^{13}CH_3$ and $^{13}CF_3$ labelled systems were 32 and 128. The number of dummy scans was set to 32 and 4 for each system. The relaxation delay was set to 1.2 s for both cases. Digital solvent suppression was applied to the FID for the $^{13}CH_3$ experiments. Exponential line broadening of 5 Hz and 50 Hz was applied to the direct and indirect dimensions, respectively, and a single round of zero filling was applied in each dimension.

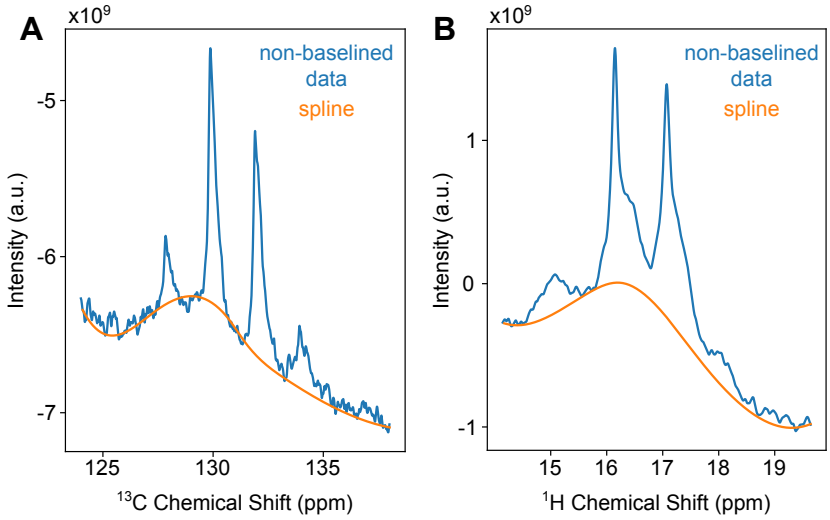


Figure C.4: Spline baselining used on the processed 1D carbon experiments for (**A**) $^{13}CF_3$ and (**B**) $^{13}CF_3$ labelled HSP16.5

C.3.2.viii Carbon pulse-acquire NMR experiments

Carbon 1D pulse-acquire spectra were obtained using the Bruker standard (zg) sequence on a non-broadband (proton optimised) NMR probe at 25^{o}C. The acquisition times were set to 1.83 s and the relaxation delay was set to 2 s. A total of 72000 and 6000 transients were recorded for the $^{13}CF_3$ and $^{13}CH_3$ samples. The spectra were processed using 10 Hz of exponential line broadening. Spline baselining was used to achieve a flat baseline after processing (Figure C.4). In subsequent peak fitting, the differences in line-broadening between the spectra on $^{13}CF_3$ and $^{13}CH_3$ samples was taken into account.

C.4 Justification of the inclusion of external protons for the fitting of $^{13}CH_3$ relaxation rates

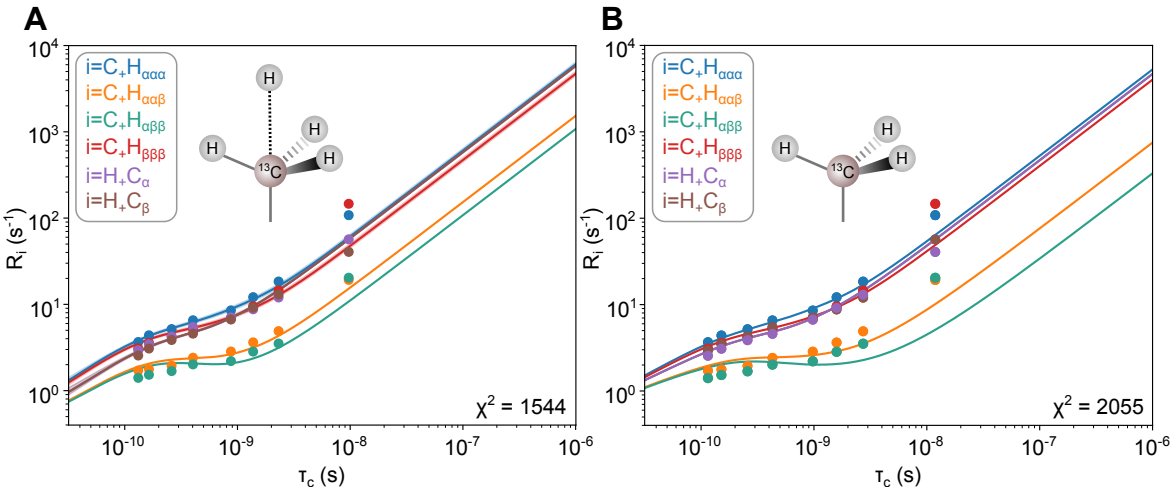


Figure C.5: Fits of the experimentally determined relaxation rates for $^{13}CH_3$ single quantum coherences to symbolic rate expressions determined using RelCalc (equations in appendix C.2.3) with (**A**) and without (**B**) the inclusion of external protons. The C-H bond length was taken to be 1.09 Å. From the fit, estimates of τ_c for each set of relaxation rates were determined, along with the fitted values of $\Delta\sigma_C$ for ^{13}C, $\Delta\sigma_H$ for ^1H/, as well the methyl tumbling rate τ_m. Using the obtained χ^2 values from each fit, the fit with external protons (12 fitting parameters) was statistically justified over that without external protons (11 fitting parameters) using an F-test with a p-value of 0.05.

C.5 Bootstrap histograms from the global $^{13}CF_3$ and $^{13}CH_3$ spin systems fitting procedures

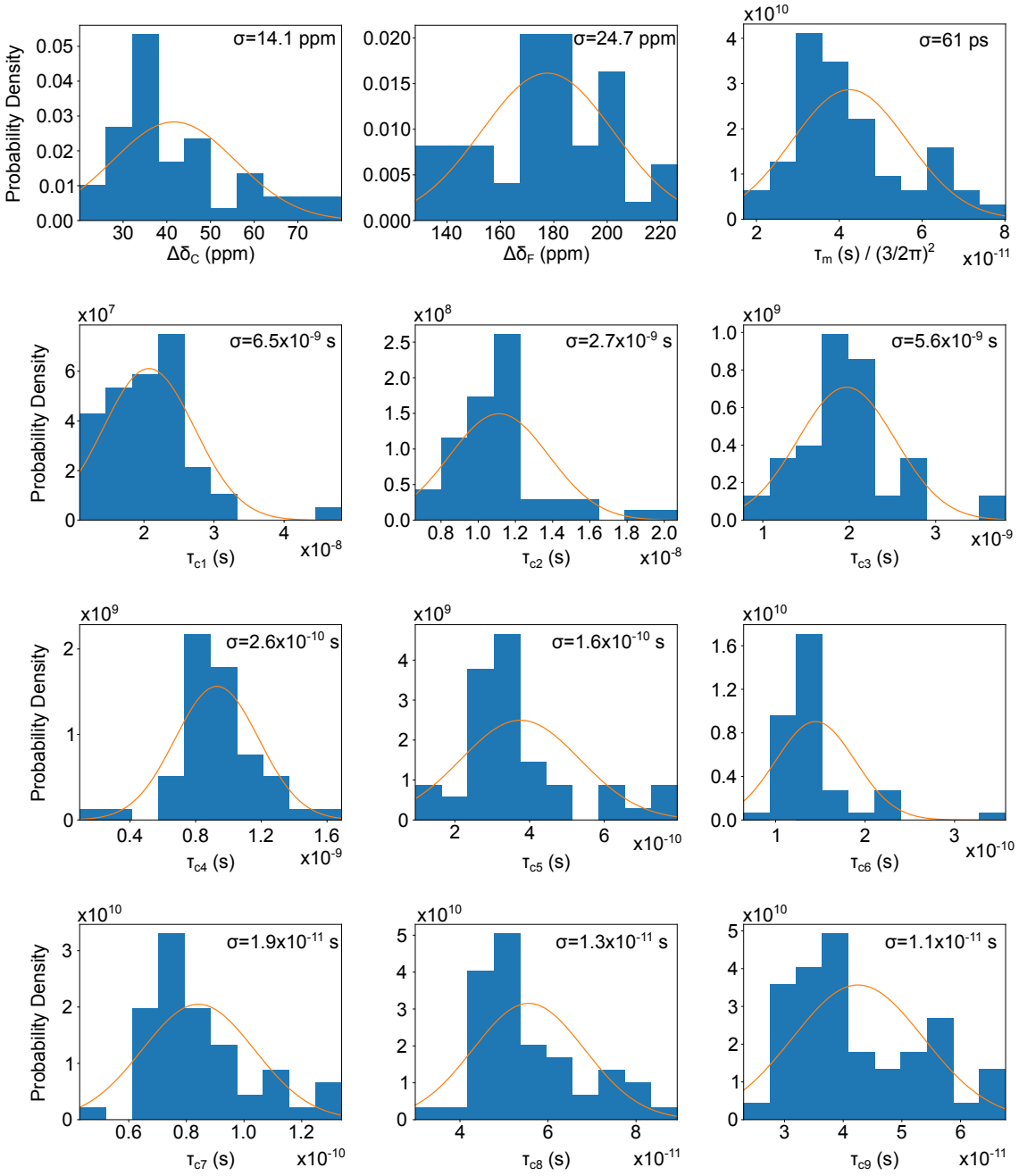


Figure C.6: Histograms of fitting parameters ($\Delta\delta_C$, $\Delta\delta_F$, τ_m and τ_{c1}-τ_{c9}) from the global fitting of all $^{13}CF_3$ relaxation rates to the symbolic rate expressions in appendix C.2.3. The histograms represent the fitted parameters obtained when re-sampling the obtained data with replacement and re-fitting 50 times. The standard deviations of the histograms are used as the parameter errors.

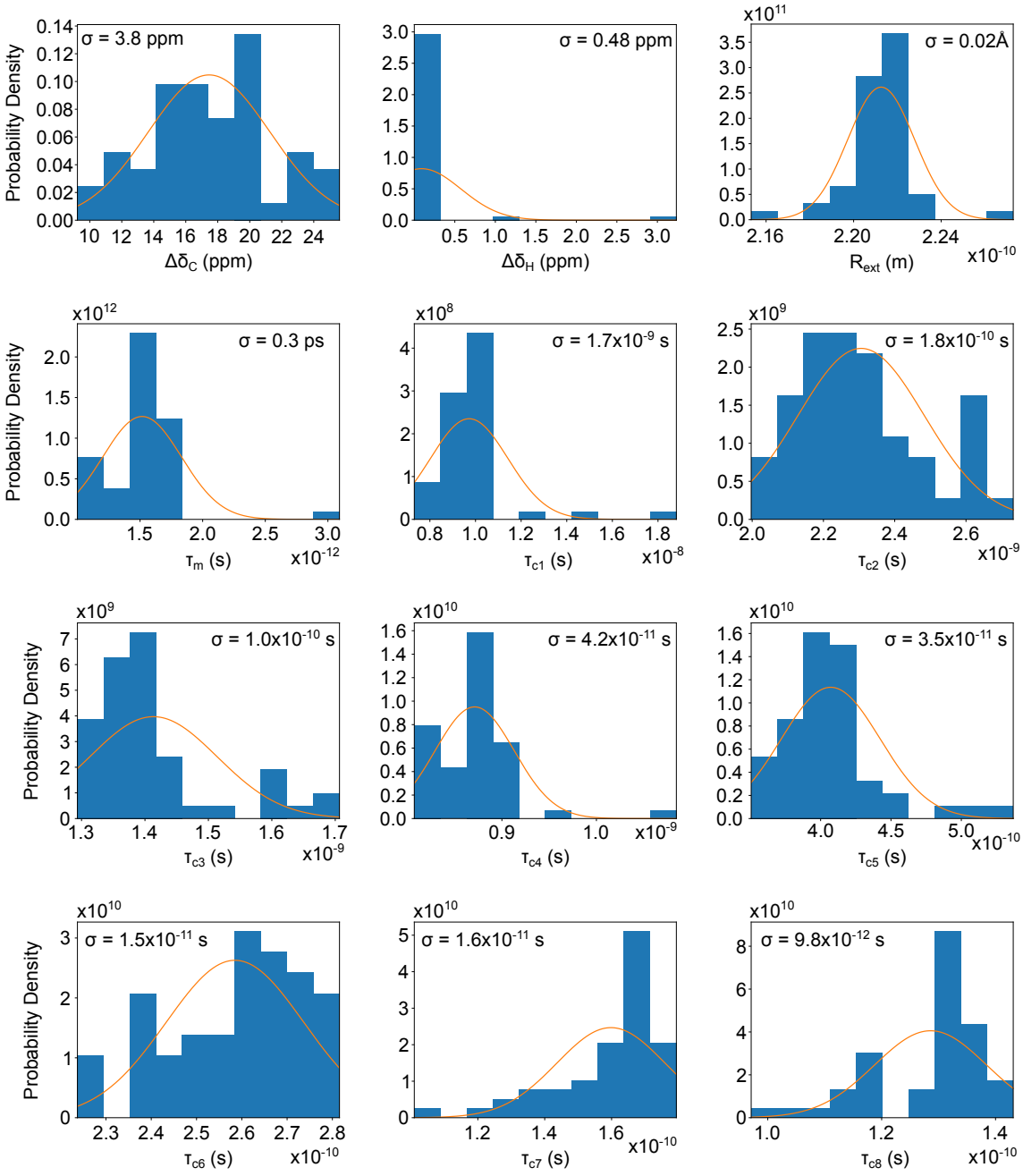


Figure C.7: Histograms of fitting parameters ($\Delta\delta_C$, $\Delta\delta_H$, τ_m and τ_{c1}-τ_{c9}) from the global fitting of all $^{13}CH_3$ relaxation rates to the symbolic rate expressions in appendix C.2.3. The histograms represent the fitted parameters obtained when re-sampling the obtained data with replacement and re-fitting 50 times. The standard deviations of the histograms are used as the parameter errors.

C.6 Origin of the 3:1:1:3 quartet peak ratio for ^{13}C in a HSQC on $^{13}CH_3$ groups

The following section derives an expression for an $^{13}CH_3$ spin system, showing that the quartet peak has a 3:1:1:3 intensity ratio in the indirectly detected dimension (^{13}C) of a HSQC experiment, in the absence of relaxation and decoupling.[492] This is the opposite to the 1:3:3:1 multiplet intensity ratio expected for the one-dimensional pulse acquire spectrum on ^{13}C.

Before the indirect detection delay (t_1) in a HSQC, the prepared magnetisation $\rho(0)$ will be equal to $2\hat{C}_x\hat{H}^1_z$ (or equally probable $2\hat{C}_x\hat{H}^2_z$ and $2\hat{C}_x\hat{H}^3_z$) where 1-3 are the indexes of the H nuclei. During the delay t_1, the acting Hamiltonian $\hat{H}$ on the system is equal to:

$$\hat{H} = \omega_H\hat{H}_z + \omega_C\hat{C}_z + 2\pi J\left[\hat{C}_z\hat{H}^1_z + \hat{C}_z\hat{H}^2_z + \hat{C}_z\hat{H}^3_z\right] \tag{C.46}$$

In the following discussion it will be assumed that the rotating frame frequencies ω_H and ω_C are equal to zero (i.e. on resonance). The magnetisation at time t_1 can be expressed as:

$$\begin{aligned}\rho(t_1) &= e^{-i\hat{H}t_1}\hat{C}_x\hat{H}^1_z e^{i\hat{H}t_1} \\ &= e^{-i2\pi Jt_1\hat{C}_z\hat{H}^3_z}e^{-i2\pi Jt_1\hat{C}_z\hat{H}^2_z}e^{-i2\pi Jt_1\hat{C}_z\hat{H}^1_z}2\hat{C}_x\hat{H}^1_z e^{i2\pi Jt_1\hat{C}_z\hat{H}^1_z}e^{i2\pi Jt_1\hat{C}_z\hat{H}^2_z}e^{i2\pi Jt_1\hat{C}_z\hat{H}^3_z}\end{aligned} \tag{C.47}$$

Expanding the inner section of equation C.47 leads to:

$$\begin{aligned}&e^{-i2\pi Jt_1\hat{C}_z\hat{H}^1_z}2\hat{C}_x\hat{H}^1_z e^{i2\pi Jt_1\hat{C}_z\hat{H}^1_z} \\ &= 2\hat{C}_x\hat{H}^1_z\cos(\pi Jt_1) - i\left[2\hat{C}_x\hat{H}^1_z, 2\hat{C}_z\hat{H}^1_z\right]\sin(\pi Jt_1) \\ &= 2\hat{C}_x\hat{H}^1_z\cos(\pi Jt_1) - \hat{C}_y\sin(\pi Jt_1)\end{aligned} \tag{C.48}$$

Applying the coupling term to the second spin leads to:

$$e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\left[2\hat{C}_x\hat{H}_z^1\cos(\pi Jt_1)-\hat{C}_y\sin(\pi Jt_1)\right]e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2}$$
$$=\cos(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}2\hat{C}_x\hat{H}_z^1e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2}-\sin(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\hat{C}_ye^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2} \tag{C.49}$$

The first term in equation C.49 can be expanded as:

$$\begin{aligned}&\cos(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}2\hat{C}_x\hat{H}_z^1e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\\&=\cos(\pi Jt_1)\left(2\hat{C}_x\hat{H}_z^1\cos(\pi Jt_1)-i\left[2\hat{C}_x\hat{H}_z^1,2\hat{C}_x\hat{H}_z^2\right]\sin(\pi Jt_1)\right)\\&=\cos(\pi Jt_1)\left(2\hat{C}_x\hat{H}_z^1\cos(\pi Jt_1)-\hat{C}_y\hat{H}_z^1\hat{H}_z^2\sin(\pi Jt_1)\right)\end{aligned} \tag{C.50}$$

The final term in equation C.50 can be ignored as it will be removed using phase cycling and not be observed leading to:

$$\begin{aligned}&\cos(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}2\hat{C}_x\hat{H}_z^1e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\\&=\cos^2(\pi Jt_1)2\hat{C}_x\hat{H}_z^1\end{aligned} \tag{C.51}$$

The second term in equation C.49 can be expanded as:

$$\begin{aligned}&\sin(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\hat{C}_ye^{i2\pi Jt_1\hat{C}_z\hat{H}_z^2}\\&=\sin(\pi Jt_1)\left(\hat{C}_y\cos(\pi Jt_1)-i\left[\hat{C}_y,2\hat{C}_z\hat{H}_z^2\right]\sin(\pi Jt_1)\right)\\&=\sin(\pi Jt_1)\left(\hat{C}_y\cos(\pi Jt_1)-2\hat{C}_x\hat{H}_z^2\sin(\pi Jt_1)\right)\end{aligned} \tag{C.52}$$

Therefore, after inserting equations C.51 and C.52 into equation C.49, the total observed magnetisation after the application of the coupling to the second spin is equal to:

$$\cos^2(\pi Jt_1)2\hat{C}_x\hat{H}_z^1-\sin(\pi Jt_1)\left(\hat{C}_y\cos(\pi Jt_1)-2\hat{C}_x\hat{H}_z^2\sin(\pi Jt_1)\right) \tag{C.53}$$

Applying the third scalar coupling term to the first part of equation C.53 leads to the following observed magnetisation terms:

$$\begin{aligned}
&\cos^2(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^3}2\hat{C}_x\hat{H}_z^1e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\\
&=\cos^2(\pi Jt_1)\left(2\hat{C}_x\hat{H}_z^1\cos(\pi Jt_1)-i\left[2\hat{C}_x\hat{H}_z^1,2\hat{C}_z\hat{H}_z^3\right]\sin(\pi Jt_1)\right)\\
&=\cos^2(\pi Jt_1)\left(2\hat{C}_x\hat{H}_z^1\cos(\pi Jt_1)-\hat{C}_y\hat{H}_z^1\hat{H}_z^3\sin(\pi Jt_1)\right)\\
&=\cos^3(\pi Jt_1)2\hat{C}_x\hat{H}_z^1
\end{aligned} \tag{C.54}$$

where the term containing $\hat{C}_y\hat{H}_z^1\hat{H}_z^3$ has been neglected as it is not observed in the experiment.

Applying the third scalar coupling term to the second part of equation C.53 leads to the following observed magnetisation terms:

$$\begin{aligned}
&\sin(\pi Jt_1)e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\left(\hat{C}_y\cos(\pi Jt_1)-2\hat{C}_x\hat{H}_z^2\sin(\pi Jt_1)\right)e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\\
&e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\hat{C}_ye^{i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\\
&=\hat{C}_y\cos(\pi Jt_1)-i\left[\hat{C}_y,2\hat{C}_z\hat{H}_z^3\right]\sin(\pi Jt_1)\\
&=\hat{C}_y\cos(\pi Jt_1)-2\hat{C}_x\hat{H}_z^3\sin(\pi Jt_1)\\
&e^{-i2\pi Jt_1\hat{C}_z\hat{H}_z^3}2\hat{C}_x\hat{H}_z^2e^{i2\pi Jt_1\hat{C}_z\hat{H}_z^3}\\
&=2\hat{C}_x\hat{H}_z^2\cos(\pi Jt_1)-i\left[2\hat{C}_x\hat{H}_z^2,2\hat{C}_z\hat{H}_z^3\right]\\
&=2\hat{C}_x\hat{H}_z^2\cos(\pi Jt_1)-\hat{C}_y\hat{H}_z^2\hat{H}_z^3\sin(\pi Jt_1)\\
&=2\hat{C}_x\hat{H}_z^2\cos(\pi Jt_1)\qquad\text{Ignoring non-observed extra term}
\end{aligned} \tag{C.55}$$

Therefore, after applying the third scalar coupling term to the second part of equation C.53 leads to the following observed magnetisation:

$$\begin{aligned}&\sin(\pi Jt_1)\cos^2(\pi Jt_1)\hat{C}_y - \sin^2(\pi Jt_1)\cos(\pi Jt_1)2\hat{C}_x\hat{H}_z^3 \\ &- \sin^2(\pi Jt_1)\cos(\pi Jt_1)2\hat{C}_x\hat{H}_z^2\end{aligned} \tag{C.56}$$

The total observed magnetisation at the end of the indirect detection period t_1 can be constructed through adding the final terms in equation C.54 to those in equation C.56:

$$\begin{aligned}\rho(t_1) = &\cos^3(\pi Jt_1)2\hat{C}_x\hat{H}_z^1 + \sin(\pi Jt_1)\cos^2(\pi Jt_1)\hat{C}_y \\ &- \sin^2(\pi Jt_1)\cos(\pi Jt_1)2\hat{C}_x\hat{H}_z^3 - \sin^2(\pi Jt_1)\cos(\pi Jt_1)2\hat{C}_x\hat{H}_z^2\end{aligned} \tag{C.57}$$

Therefore, the time-modulation of the anti-phase magnetisation ($2\hat{C}_x\hat{H}_z^i$) created at the end of t_1 is given by:

$$\begin{aligned}&\cos^3(\pi Jt_1) - 2\sin^2(\pi Jt_1)\cos(\pi Jt_1) = \\ &\frac{3}{8}\cos(3\pi Jt_1) + \frac{1}{8}\cos(\pi Jt_1) + \frac{1}{8}\cos(-\pi Jt_1) + \frac{3}{8}\cos(-3\pi Jt_1)\end{aligned} \tag{C.58}$$

giving the observed 3:1:1:3 ratio.