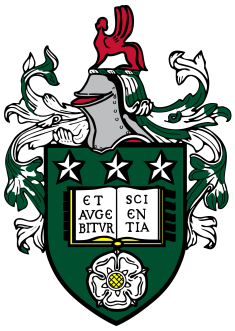
Dynamic structures of rare conformations involved in amyloid formation



Hugh Ian Smith

Astbury Centre for Structural Molecular Biology

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

September 2018

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. The right of Hugh Ian Smith to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Hugh Ian Smith

September 2018

Acknowledgements

Many people have contributed to the work in this thesis in various ways, whom I would like to thank.

Sheena Radford and Alex Breeze, my supervisors, have given continuous support and commitment to this project and my personal development. Arnout Kavlerda and Gary Thompson have provided valuable help with NMR; I would like to thank them for their time and effort in helping me to understand NMR in theory and in practice.

Theo Karamanos introduced me to NMR and amyloid and mentored me during my PhD studies; I would like to thank Theo for dedicating his time and effort into training me in NMR techniques and in the lab.

Max Bonomi helped design the metadyanmic-metainference molecular dynamics simulations in this thesis. I would like to thank him for his tuition in practical aspects of molecular dynamics simulations, and for always being on hand to help with the various errors and bugs I encountered along the way. I would also like to thank Michele Vendruscolo for his advice and input into this work.

Janet Saunders helped carry out the split-beta-lactamase assays in this thesis and provided instruction in how to perform these experiments. I would like to thank Janet for her enthusiasm in driving these experiments forward.

Owen Cornwell performed all of the mass spectrometry experiments presented in this thesis. Emma Cawood provided scripts for analysing aggregation kinetics data. I would like to thank Owen, Emma and Nicolas Guthertz for their help, advice and support throughout the project.

I would like to thank Nasir Khan for his technical support with CD spectroscopy and for his hard work to keep the lab running smoothly. I would also like to thank everyone in the Radford and Brockwell labs, past and present, for their friendship and support.

Finally, I would like to thank my family and Harriet, my best possible friend, for their love and support.

Abstract

In work described in this thesis the conformational landscape of a recently identified β_2m variant, D76N, was examined with NMR and state of the art molecular dynamics simulations. The D76N β_2m variant was recently identified as the first naturally occurring amyloidogenic variant of β_2m, a 99 residue, seven stranded Ig domain protein. The D76N variant of β_2m causes a hereditary, late onset and fatal systemic amyloid disease. Affected individuals are heterozygous for the D76N mutation and thus also express wild-type (WT) β_2m in the serum. Co-aggregation of D76N and WT β_2m is not observed *in vivo*, and D76N amyloidosis follows a distinctive disease progression from the WT β_2m associated disease, Dialysis-Related Amyloidosis, which occurs in patients undergoing long-term haemodialysis.

This work aims to understand the mechanism by which the single amino acid substitution, D76N, results in a dramatically more aggregation-prone protein while maintaining the same overall native structure as WT β_2m. In this work, the structure, stability and dynamics of the D76N variant are studied with NMR spectroscopy and a range of complementary biophysical techniques. The folding intermediates of D76N and WT β_2m are also probed with real-time NMR and circular dichroism spectroscopy, revealing structural, dynamic and kinetic insights into the folding intermediates present within the D76N and WT β_2m folding pathways and their roles in aggregation. The study of the folding intermediates demonstrated that the I_T-state, which is a known precursor for β_2m aggregation, is not responsible for the rapid aggregation of the D76N variant. Instead, a low-populated state, unique to D76N, in equilibrium with the native state must be responsible for D76N aggregation, consistent with the different disease progression associated with D76N. The NMR and data-guided molecular dynamics simulations presented in this thesis reveal the effects of the D76N substitution on the conformational landscape of the D76N variant and detail the structures of the low-populated states in equilibrium with the native state. Additionally, saturation mutagenesis of position 76 in D76N, carried out with a split-β-lactamase assay have demonstrated the exquisite specificity for an aspartate residue at position 76 in the β_2m structural ensemble.

Table of contents

List of figures xiii

List of tables xix

Nomenclature xxi

1 Introduction 1

1.1 The energy landscape description of protein structure and dynamics . . . 1

1.2 Probing 'excited' states using NMR 4

1.2.1 Chemical exchange saturation transfer 4

1.2.2 Carr-Purcell-Meiboom-Gill sequence relaxation-dispersion 5

1.2.3 Hydrogen exchange experiments 7

1.2.4 Paramagnetic relaxation enhancement 8

1.2.5 Residual dipolar couplings 9

1.2.6 Heteronuclear relaxation experiments and protein dynamics . . . 11

1.3 The role of excited states in protein function and disease 12

1.3.1 SOD1 excited states 12

1.3.2 Functionally important 'invisible' states in enzyme energy landscapes 17

1.3.3 Excited states of human lysozyme variants cause amyloidosis . . . 19

1.3.4 TTR aggregation proceeds via and excited state 22

1.4 Amyloid diseases share a common underlying architecture 23

1.5 β_2-microglobin 26

1.5.1 β_2-microglobin and dialysis related amyloidosis 26

1.5.2 D76N β_2-microglobin associated systemic amyloid disease 33
1.6 Aims and scope 40

2 Materials and methods 43
2.1 Chemicals 43
2.1.1 *E.coli* bacterial strains 43
2.2 Molecular biology 44
2.2.1 Site-directed mutagenesis 44
2.3 Protein preparation and general protein analysis 45
2.3.1 Expression plasmid 45
2.3.2 Bacterial culture media 47
2.3.3 Expression and purification 47
2.3.4 SDS-PAGE 49
2.3.5 Mass spectrometry 50
2.3.6 Estimation of protein concentration using UV absorption at 280 nm 51
2.4 Circular dichroism spectroscopy 52
2.4.1 Background 52
2.4.2 Thermal denaturation, monitored by CD 53
2.4.3 Real-time refolding, monitored by CD 54
2.5 Equilibrium unfolding experiments 56
2.6 Aggregation assays 57
2.6.1 *In vitro* fibrillation assays 57
2.6.2 Electron microscopy 58
2.6.3 Fibril yield assay 58
2.7 Nuclear Magnetic Resonance spectroscopy 59
2.7.1 Data processing 59
2.7.2 Backbone Assignment 59
2.7.3 Chemical shift perturbation (CSP) analysis 59
2.7.4 Real-time refolding experiments monitored using NMR 60
2.7.5 HDX-NMR 60
2.7.6 ^{15}N transverse relaxation rate determination 61

2.7.7 ^{15}N relaxation dispersion experiments . . . 62
2.7.8 Measurement of residual dipolar couplings . . . 63
2.8 Split-β-lactamase assay . . . 64
2.9 Metadynamic-metainference molecular dynamics simulations . . . 67
2.9.1 Experimental setup . . . 67
2.9.2 Generation of free energy surfaces . . . 68
2.9.3 Clustering analysis . . . 70

3 Global properties and native structure of D76N β_2m 71
3.1 Introduction . . . 71
3.2 Results . . . 74
3.2.1 Stability and aggregation of D76N β_2m . . . 74
3.2.2 Assessing the residue level effect of the D76N substitution using NMR . . . 88
3.2.3 The role of residue type at position 76 in structure stability and aggregation of β_2m . . . 99
3.3 Discussion . . . 115

4 Characterisation of the D76N I_T-state using real-time NMR 119
4.1 Introduction . . . 119
4.2 Results . . . 121
4.2.1 Real-time refolding of β_2m variants, monitored using far-UV CD . 121
4.2.2 Real-time refolding of WT and D76N β_2m, followed by NMR . . 133
4.2.3 Generation of a trapped D76N I_T-state mimic (ΔN6-D76N) . . . 149
4.3 Discussion . . . 155

5 Conformational dynamics of D76N 157
5.1 Introduction . . . 157
5.2 Results . . . 159
5.2.1 Exploring the role of protein dynamics in D76N aggregation with NMR . . . 159

5.2.2 Metadynamic-metainference molecular dynamics simulations as a tool to explore the conformational energy landscape of D76N . . 182
5.3 Discussion 204

6 Summary and conclusions 207
6.1 Success and limitations 207
6.1.1 What is the role of the I_T-state in D76N aggregation? 207
6.1.2 How important is the identity of residue 76 for aggregation? . . . 209
6.1.3 Is there an I_T-state independent excited state involved in D76N aggregation? What is the nature of this state? 210
6.2 Conclusions 212

References 217

Appendix A 243

Appendix B 245
B.1 CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz . 245
B.2 CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz . . 255
B.3 CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz . 265
B.4 CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz . 275

Appendix C 285

List of figures

1.1 The energy landscape and protein dynamics 2
1.2 Timescales of protein dynamics 3
1.3 Accessing excited states with CEST NMR. 5
1.4 Probing low-populated states with CPMG-RD NMR. 6
1.5 PREs to measure weak, transient states and interactions. 9
1.6 RDCs depend on the the angle between the intermolecular bond and the external magnetic field. 10
1.7 T_1 and T_2 relaxation to measure protein dynamics. 11
1.8 Crystal structure of the mature SOD1 homodimer 13
1.9 The four transiently populated states in the immature SOD1 conformational landscape. 14
1.10 CEST profiles differentiate between to independent SOD1 processes. . . 15
1.11 SOD1 oligomerisation may proceed via a non-native dimer. 16
1.12 The dynamic landscape of DHFR catalysis, revealed by CPMG-RD NMR. . 18
1.13 Familial lysozyme mutation sites. 20
1.14 The proposed mechanism of lysozyme amyloid formation. 21
1.15 The crystal structure of human TTR. 23
1.16 Common hallmarks of amyloid. 24
1.17 The chemical structure of Thioflavin-T. 25
1.18 Nucleation dependent mechanism of amyloid formation *in vitro*. 25
1.19 The structure of the MHC-I complex and monomeric β_2m. 27
1.20 Localisation of β_2m amyloid deposits in dialysis related amyloidosis. 28

1.21 ΔN6 as a structural mimic of the I_T-state. 30
1.22 ΔN6 enhances WT β_2m aggregation propensity *in vitro*. 31
1.23 The inhibitory ΔN6-mβ_2m complex. 32
1.24 Theoretical energy landscapes for β_2m folding and aggregation. 33
1.25 WT vs D76N disease progression schematic. 34
1.26 Crystallographic B-factors for WT and D76N crystal structures. 36
1.27 Hydrogen bonding networks involving residue 76 in β_2m. 37
1.28 Simulated structures of the D76N I_T-state. 38

2.1 The pINK plasmid, containing the WT β_2m gene. 46
2.2 Molecular weight calibrants gel filtration. 49
2.3 Examples of different CD spectra produced by various secondary structure elements. 53
2.4 Example schematic of the 48 well plate used for MIC assay. 65
2.5 Plasmid map of the tripartite fusion construct used in the split-β-lactamase assay. 66

3.1 WT and D76N crystal structures. 72
3.2 Purification of recombinant D76N and WT protein. 75
3.3 WT and D76N Mass spectra. 76
3.4 D76N and WT structure and thermal stability, monitored by far-UV CD spectroscopy at pH 7.4. 78
3.5 Equilibrium urea denaturation of D76N and WT β_2m, followed by tryptophan fluorescence. 80
3.6 D76N aggregation kinetics at pH 7.4 and pH 6.2. 82
3.7 Comparison of the aggregation rates of D76N and ΔN6. 84
3.8 Inhibition of D76N aggregation by murine-β_2m. 85
3.9 Thermal stability of D76N and WT β_2m, monitored by far-UV CD spectroscopy at pH 6.2. 86
3.10 Backbone resonance assignment of D76N. 89
3.11 ^{1}H-^{15}N HSQC spectrum of D76N, showing resonance assignments. . . . 90

3.12 Comparison of the ^{1}H-^{15}N HSQC spectra of WT and D76N at pH 7.4. . 96
3.13 Comparison of WT and D76N ^{1}H-^{15}N HSQC spectra at pH 6.2. 98
3.14 Split β-lactamase assay schematic. 100
3.15 Split-β-lactamase assay for D76N and WT β_2m. 101
3.16 Split-β-lactamase assay for D76X constructs. 102
3.17 Split-β-lactamase assay for D76X constructs. 103
3.18 Comparison of all the MCD_{GROWTH} value for D76X variants measured using the split-β-lactamase scores. 105
3.19 Correlation between split-β-lactamase assay score with hydrophobicity and residue size. 106
3.20 Purification of D76E and D76A variants. 108
3.21 D76E and D76A Mass spectra. 109
3.22 D76E and D76A temperature ramp, monitored by far-UV CD spectroscopy at pH 6.2. 110
3.23 Comparison of the aggregation rates for D76E and D76A with D76N. . . 112
3.24 Comparison of D76E and D76A variants ^{1}H-^{15}N HSQC, with WT and D76N β_2m. 113
3.25 CSP analysis of D76E and D76A, in comparison to WT β_2m at pH 6.2. . . 114
3.26 Positions of the seven aspartate resides in β_2m. 117

4.1 Structural rearrangements involved in the I_T-state to native transition. . . 121
4.2 Schematic describing the protocol used for real time refolding reaction studies, monitored by far-UV CD. 122
4.3 Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy. 123
4.4 Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy as a function of time at 216 nm. 125
4.5 Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy as a function of time at 235 nm. 126
4.6 Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy. 128
4.7 Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy as a function of time at 216 nm. 129

4.8 Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy as a function of time at 235 nm. 131
4.9 Schematic of the real-time refolding reaction, followed by NMR. 133
4.10 Real time refolding of WT β_2m, monitored by NMR. 135
4.11 Real time refolding of D76N β_2m, monitored by NMR. 136
4.12 ^{1}H-^{15}N assignments for the WT β_2m 90 s refolding spectrum, used for CSP and refolding kinetic analysis. 138
4.13 ^{1}H-^{15}N assignments for the D76N β_2m 90 s refolding spectrum, used for CSP and refolding kinetic analysis. 139
4.14 Chemical shift differences between spectra recorded at 90 s and 180 min, during the real-time refolding reaction at pH 7.4. 141
4.15 Comparison of WT and D76N ^{1}H-^{15}N HMQC spectra recorded at 90 s, during the real-time refolding reaction. 143
4.16 Single residue refolding rates for N-state peaks. 145
4.17 Single residue refolding rates for I-state peaks. 146
4.18 Comparison of pure-I_T-state spectra and refolding spectra. 148
4.19 Purification of ΔN6-D76N. 150
4.20 ΔN6-D76N and ΔN6-WT temperature ramp, monitored by far-UV CD spectroscopy. 151
4.21 Comparison of the ^{1}H-^{15}N HMQC spectrum of ΔN6-D76N and the D76N I_T-state at 293K. 152
4.22 ΔN6-D76N aggregation assay. 154

5.1 WT and D76N spectra after initiation of hydrogen-deuterium exchange. . 160
5.2 Protected core residues for D76N and WT β_2m mapped onto the protein structure. 161
5.3 Representative HD-exchange plots. 162
5.4 Comparison of WT and D76N protection factors at pH 7.4. 164
5.5 Transverse relaxation rates for D76N and WT β_2m at pH 7.4. 167
5.6 Representative transverse relaxation decay curves for D76N and WT β_2m. 168
5.7 Areas of perturbed dynamic behaviour in D76N compared to WT at pH 7.4. 169

5.8 Concentration dependence of D76N transverse relaxation rates. 170
5.9 Measuring D76N dynamics, compared to WT β_2m at pH 6.2. 171
5.10 Areas of perturbed dynamic behaviour in D76N compared to WT at pH 6.2. 172
5.11 Transverse relaxation rates of the D76E and D76A variants, compared to WT β_2m at pH 6.2. 173
5.12 Representative relaxation dispersion profiles for exchanging residues in D76N and WT β_2m at pH 7.4. 175
5.13 Estimated R_{ex} values for WT and D76N, at pH 7.4 and 600 MHz 176
5.14 Exchanging residues at pH 7.4, mapped on to WT and D76N crystal structures. 177
5.15 Representative relaxation dispersion profiles for exchanging residues in D76N at pH 7.4 and pH 6.2. 178
5.16 Estimated R_{ex} values for D76N at pH 7.4 and pH 6.2 at 750 MHz 179
5.17 The effect of pH on the D76N conformational exchange. 180
5.18 Measurement of D76N RDCs at pH 7.4. 183
5.19 Collective variables used in MM-MD simulations. 186
5.20 Conformational differences in D76N crystal structures. 187
5.21 Monodimensional free energy landscape as a function of the AB-loop CV. 188
5.22 Monodimensional free energy landscape as a function of the D-strand-APBR CV. 189
5.23 Monodimensional free energy landscape as a function of the Pro32 CV. . 190
5.24 Monodimensional free energy landscape as a function of the N-term CV. . 191
5.25 Monodimensional free energy landscapes as a function of the Pro32:N-term and Pro32:DE-loop CVs. 192
5.26 Monodimensional free energy landscapes as a function of the Distance76-42 and Distance76-73 CVs. 194
5.27 Free energy surface as a function of the 'AB-loop' and 'D-strand-APBR' CVs. 195
5.28 Free energy surface as a function of the 'Distance76-42' and 'Distance76-73' CVs. 196

5.29 Free energy surface as a function of the 'Pro32' and 'Pro32:N-term' CVs. 197
5.30 Free energy surface as a function of the 'Pro32' and 'D-strand-APBR' CVs. 198
5.31 Populations of the predominant D76N species identified in the clustering analysis. 200
5.32 Visualisation of the predominant D76N species identified in the clustering analysis. 202
5.33 D76N species identified in the clustering analysis replicate structural heterogeneity in D76N crystal structures. 203
5.34 WT β_2m structural heterogeneity. 205

6.1 Hypothetical conformational free energy landscape for WT and D76N β_2m 214

A.1 Single residue refolding rates for N-state peaks, without filtering fits based on errors. 243
A.2 Single residue refolding rates for I_T-state peaks, without filtering fits based on errors. 244

C.1 Protein sequence of WT β_2m. 285

List of tables

2.1 Primers used for site-directed mutagenesis. 44
2.2 Components made up to 1 L to make HCDMI media prior to autoclaving. 47
2.3 Components added to 1 L HCDMI media after autoclaving. (*: isotopically labelled as appropriate) 47
2.4 30% Tris-tricine buffered SDS-PAGE gel components 50
2.5 Instrument conditions used for mass spectrometry analysis of purified β_2m 51
2.6 Metadynamics settings used for each CV 68

3.1 Apparent Tm values, for WT and D76N β_2m in 25 mM sodium phosphate buffer at pH 7.4 and pH 6.2 87
3.2 Backbone chemical shift assignments for D76N β_2m. 91
3.3 Apparent Tm values, for WT, D76N, D76E and D76A β_2m in 25 mM sodium phosphate buffer at pH 7.4 and pH 6.2 111
3.4 Aggregation T_{half} values, for D76N, D76E and D76A β_2m in 25 mM sodium phosphate buffer pH 6.2 112
3.5 Summary of $\Delta\Delta G^{\circ}_{UN}$ and aggregation propensity of β_2m variants 115

4.1 Refolding rate constants, measured by far-UV CD at 216 nm and at 235 nm for all β_2m variants at pH 7.4. 127
4.2 Refolding rate constants, measured by far-UV CD at 216 nm and at 235 nm for all β_2m variants at pH 6.2. 132
4.3 $T_{m,app}$ values for WT and D76N (full length and ΔN6). 152

5.1 Per-residue RDC values for D76N. 184

5.2 Collective variables used in D76N MM-MD simulations 185
5.3 The number of frames per cluster generated with a 2Å cut-off 199
5.4 Equilibrium population of clusters 1-10, weighted based on metadynamics bias potential 200

B.1 Exchanging residues in D76N at pH 7.4, measured at 600 MHz. 245
B.2 Exchanging residues in WT at pH 7.4, measured at 600 MHz. 255
B.3 Exchanging residues in D76N at pH 7.4, measured at 750 MHz. 265
B.4 Exchanging residues in D76N at pH 6.2, measured at 750 MHz. 275

Nomenclature

Acronyms / Abbreviations

ALS Amyotrophic Lateral Sclerosis

AUC Area Under Curve

β_2m β_2 microglobulin (human)

CD Circular Dichroism

CEST Chemical Exchange Saturation Transfer

CPMG-RD Carr-Purcell-Meiboom-Gill sequence relaxation-dispersion

CS Chemical Shift

CSP Chemical Shift Perturbation

CV Collective Variable

DHFR Dihydrofolate Reductase

DNA Deoxyribonucleic Acid

DRA Dialysis Related Amyloidosis

DSS 4,4-dimethyl-4-silapentane-1-sulfonic acid

DTT Dithiothreitol

EDTA ethylenediaminetetraacetic acid

EM Electron Microscopy

ESI-MS Electrospray Ionisiation Mass Spectrometry

FTIR Fourier-Transform Infrared

GdnHCl Guanidine Hydrochloride

H-bond Hydrogen Bond

HDX Hydrogen-Deuterium Exchange

HMQC Heteronuclear Multiple Quantum Correlation

HMX Hydrogen Magnetisation Exchange

HSQC Heteronuclear Single Quantum Coherence

HX Hydrogen Exchange

IPAP In-Phase and Anti-Phase

IPTG isopropyl β-D-1-thiogalactopyranoside

mβ_2m murine β_2 microglobulin

MCD Maximal Cell Dilution

MD Molecular Dynamics

MHC-I major histocompatibility complex class I

MM-MD Metadynamic-Metainference Molecular Dynamics

MRE Mean Residue Ellipticity

MTSL (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate

MWCO Molecular Weight Cut-Off

NMR Nuclear Magnetic Resonance

NOE Nuclear Overhauser Effect

PDB Protein Data Bank

PMSF phenylmethanesulfonyl fluoride

ppm Parts Per Million

PRE Paramagnetic Relaxation Enhancement

RDC Residual Dipolar Coupling

RMSD Root-Mean-Square Deviation

SDS-PAGE Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SOD1 Superoxide Dismutase 1

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

SVD Single Value Decomposition

TFE Tetrafluoroethylene

ThT Thioflavin-T

TOF Time Of Flight

TTR Transthyretin

WT Wild-Type

Chapter 1

Introduction

1.1 The energy landscape description of protein structure and dynamics

Proteins are dynamic, metastable entities which sample a range of conformations in solution where the ensemble of structures that a protein occupies is often exquisitely evolved to strike a delicate balance between function and stability [1, 2, 3, 4]. In fact, most globular proteins are marginally stable [5] and the inherent flexibility of globular proteins as a result of marginal stability confers a strong evolutionary advantage [5, 6].

The energy landscape model of protein structure and dynamics describes the ensemble of structures that a protein visits at equilibrium as a result of thermal fluctuations. The energy landscape itself is highly multidimensional and contains a complete description of the thermodynamics (the relative populations of conformational states) and kinetics (the rates of interconversion between states, based upon the height of the energy barriers between them) of the system [7, 8].

The schematic in Figure 1.1 shows a mono-dimensional energy landscape as a function of a single, 'conformational coordinate'. There are two states 'A' and 'B' in the energy landscape which are separated by a high energy barrier. Within each of the states 'A' and 'B', there are further sub-states separated by much lower energy barriers, which are also much closer together in terms of the 'conformational coordinate' [8]. All of the states in the energy landscape are accessible through thermal fluctuations at equilibrium.

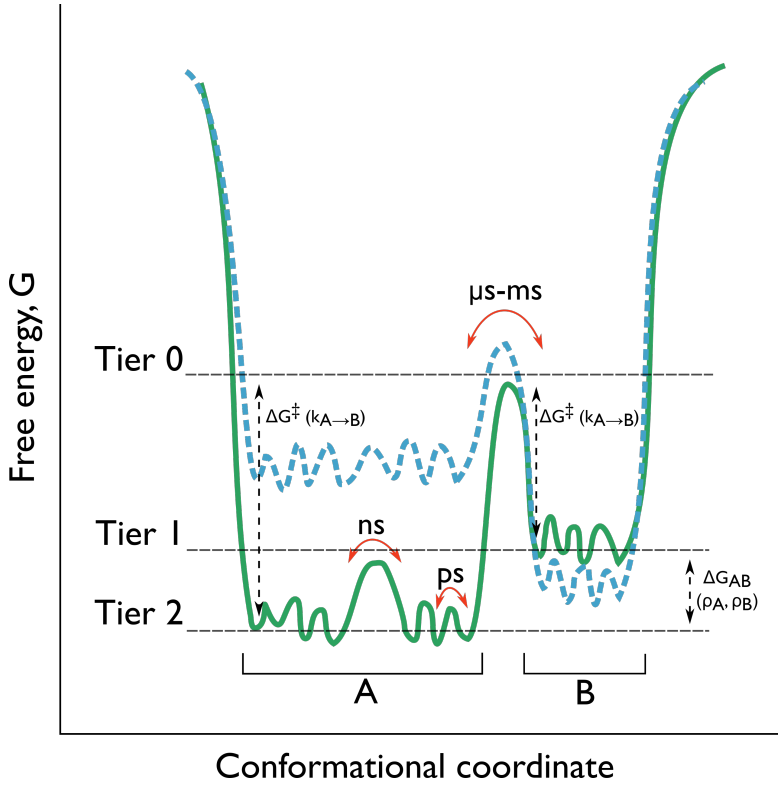


Fig. 1.1: The energy landscape and protein dynamics.
The mono-dimensional view of the multidimensional energy landscape (green) of a protein in terms of a 'conformational coordinate'. There are two Tier 0 states (A and B), which have a free energy difference of ΔG_{AB}, which defines their relative populations (thermodynamics). The rate of interconversion between the two states (kinetics) is given by the energy barrier between them $\Delta G^{\ddagger}$. Tier 1 and 2 states represent the fast-timescale dynamics within each Tier 0 state. The dashed blue landscape is an example of how a landscape can change based upon factors such as ligand binding or mutations, which can alter the relative populations and rate of interconversion between states. Adapted from [8]

Motions between states 'A' and 'B' are often referred to as Tier 0 or 'slow' motions and occur on a μs-ms timescale due to the high energy barriers between the states, which limits the rate of interconversion [7, 8]. Within each Tier 0 state ('A' and 'B' in Figure 1.1), there are conformational changes which proceed on a 'fast' timescale. These states are referred to as Tier 1 and 2 states and are separated by much lower energy barriers, resulting in faster exchange (ns-ps timescale) [8, 9].

Typically Tier 0 dynamics (i.e conversion between 'A' and 'B' in Figure 1.1) correspond to collective domain motions, while Tier 1 and 2 transitions involve local flexibility cat-

egorised by bond vibration, methyl rotation, side chain rotamerisation and loop motion (Figure 1.2) [7, 8].

While Tier 1 and 2 sub-states tend to have similar free energies to other substates within the same Tier 0 state. The energy differences between two Tier 0 states tend to be greater. This makes the study of some higher energy Tier 0 states difficult because the relative population difference between two states is proportional to the exponential of the free energy difference between the two states [10]. This can result in a secondary Tier 0 state being populated to only a few %. These low-populated states are often referred to as 'invisible' or 'excited' states [10, 11, 9, 12] and are often very important in protein function [9] and in protein dysfunction, leading to disease [10, 8, 13]

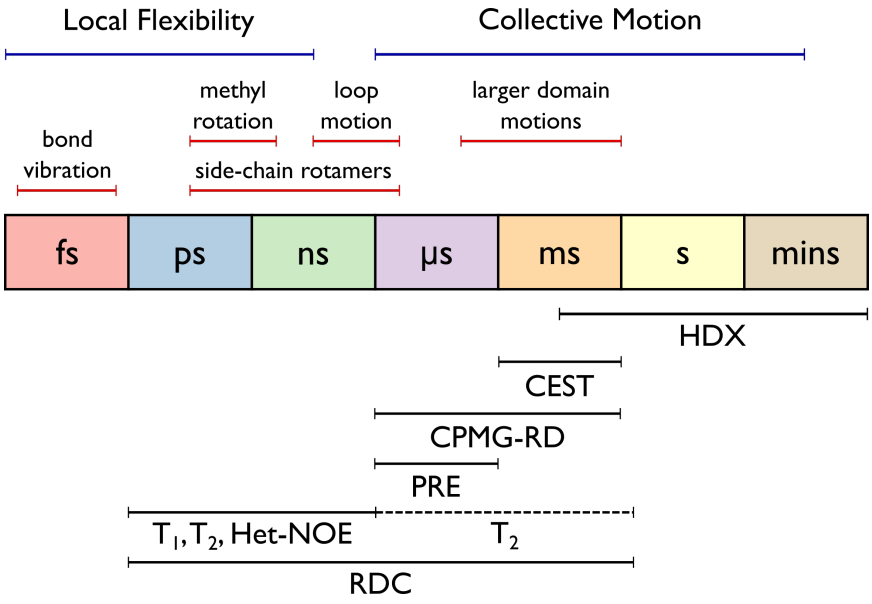


Fig. 1.2: Timescales of protein dynamics.
Protein motions occur on a range of timescales and can be accessed with many different NMR techniques. Adapted from [8]

The energy landscape of a protein is also highly sensitive to factors such as temperature and other physical changes in conditions such as pH [8, 14, 15]. Ligand binding and mutations can also influence the energy landscape which may change the equilibrium between states and the rate of interconversion between them [8, 16, 17]. These changes, which may change the population of a certain sub-state can be useful for function, but can also be detrimental and cause disease. Therefore, structural information about the con-

formational landscape of a protein, and the structures, populations and kinetics of certain sub-states within the energy landscape are essential in understanding protein function and disease.

1.2 Probing 'excited' states using NMR

There are many experimental techniques able to probe conformational protein dynamics on different timescales. Some techniques can provide atomic resolution of structures from different states (X-ray crystallography, Cryo-EM) however kinetic data is often lacking with these techniques [8, 18]. Kinetic data can be obtained at low resolution via a range of methods including fluorescence techniques, spectroscopic techniques (circular dichroism, raman, infrared). However experimentally, nuclear magnetic resonance (NMR) spectroscopy can provide both structural and kinetic information at atomic resolution about 'fast' (ps-ns) and 'slow' (μs-ms) timescales and very slow timescales (minutes) at atomic resolution [10, 18, 11, 9, 19, 20, 21, 12, 22, 23]. Chemical exchange saturation transfer (CEST) [24, 25, 26, 22] experiments and Carr-Purcell-Meiboom-Gill sequence relaxation-dispersion (CPMG-RD) NMR experiments [27, 28, 9, 29, 30] are particularly powerful tools for studying 'invisible' states in proteins (populations above approximately 0.5% [12]), and are effective on ms (CEST) and μs-ms (CPMG-RD) timescales. Molecular Dynamics (MD) simulations can provide information on fast timescales, and recently on slower timescales with through the use of metadynamics [31] and is particularly powerful in combination with NMR data [32, 33].

1.2.1 Chemical exchange saturation transfer

CEST experiments are saturation transfer experiments which can probe resonances belonging to a minor 'invisible' state, not visible in the NMR spectrum (Figure 1.3a). CEST experiments are carried out by applying saturating continuous wave radiofrequency fields at a range of offset frequencies, spanning the entire spectrum in uniform increments [12]. In cases where the saturating field interacts with the resonance of the 'invisible' state, the intensity of the major state will be affected by exchange with the saturated minor state, resulting in a loss of intensity in the major state resonance. This saturation transfer only

works when the rate of exchange is between approximately 50 s^{-1} and 400 s^{-1} [10]. CEST data are visualised by plotting the ratio between the peak intensity of the unsaturated resonance vs the saturated resonance, for a single resonance against the offset frequency of the saturating field (Figure 1.3b) [10, 12]. From this CEST data, information on the population (thermodynamics), exchange rates (kinetics) and chemical shift (structure) of the minor state can be obtained [10, 12].

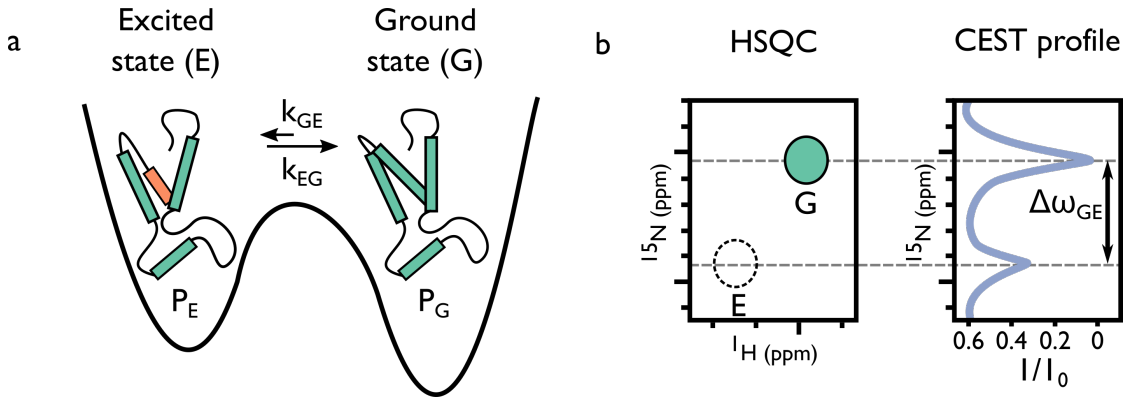


Fig. 1.3: Accessing excited states with CEST NMR.
The energy landscape in (a) contains a major 'ground state' (G) and an invisible 'excited state' (E). The resonances for the excited state can not be observed in the HSQC spectrum (b) (black, dashed peak) yet the position of the peak as well as the relative population and rate of interconversion of the excited state can be determined with CEST. In the CEST profile (b) the ratio of the saturated peak intensity to unsaturated peak intensity (I/I_0) of the major peak (green) is plotted as a function of the ^{15}N position of the saturation field applied. There is a major decrease in I/I_0 at the ^{15}N position of the major state and a minor decrease in I/I_0 at the ^{15}N position of the minor peak. Adapted from [10].

1.2.2 Carr-Purcell-Meiboom-Gill sequence relaxation-dispersion

CPMG experiments, like CEST experiments, can provide detailed information about 'invisible' species present in the conformational energy landscape of a protein. CPMG-RD experiments are effective on faster timescales [10] than CEST experiments with exchange rates approximately between 200 s^{-1} to 6000 s^{-1} [23] (providing there is a chemical shift difference between the states) [9, 34, 12]. In the schematic presented in Figure 1.4a, the dominant state 'A' is exchanging with a minor state 'B', which is not visible in the NMR spectrum in Figure 1.4a. State 'B' is invisible because it is only populated to 1% and the

relative peak intensity of peaks 'A' and 'B' is proportional to the relative populations of the two states [34].

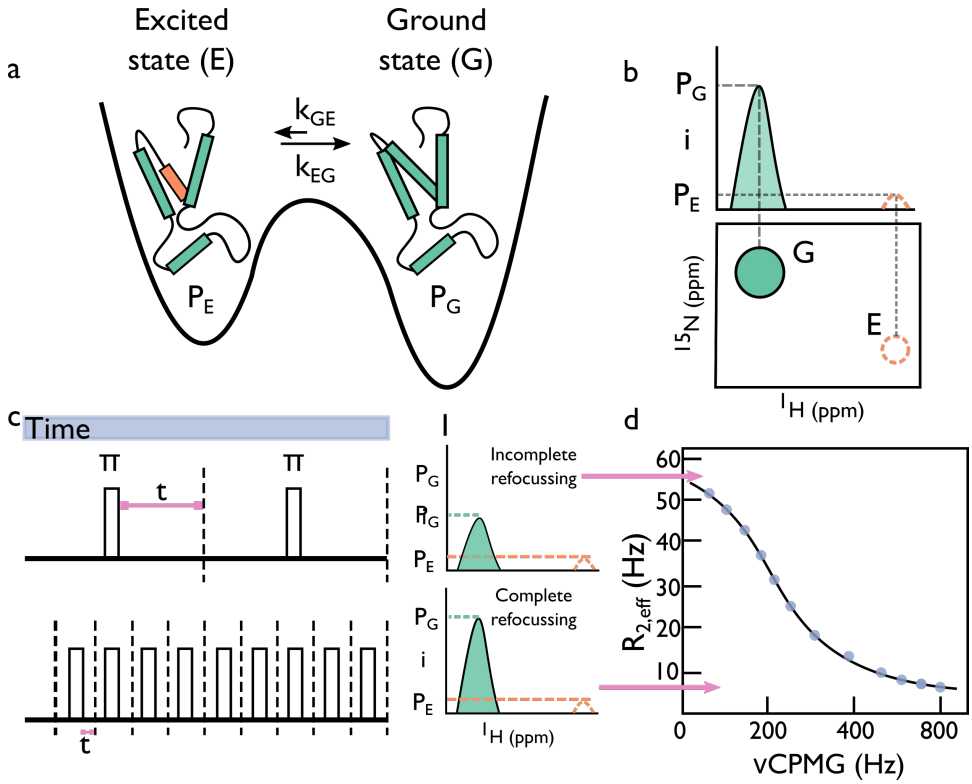


Fig. 1.4: Probing low-populated states with CPMG-RD NMR.
The protein in (a) is exchanging with a low-populated state which is populated to 1%. This state is not visible in the HSQC spectrum (b) as the relative peak intensities are proportional to the relative populations of the two states (as long as the rate of exchange is less than the chemical shift difference between the two states and that there is a chemical shift difference between them). When the refocussing pulses in (c) are applied, the spin will be refocussed after the time (t). However, if exchange is active during this period (t) then refocussing will be incomplete, resulting in lower peak intensities due to line broadening. The dispersion curve (d) is generated by plotting the peak intensity against the refocussing frequency. Adapted from [34, 10].

In the CPMG-RD experiment in Figure 1.4b the effective linewidth ($R_{2,eff}$) for each resonance is measured as a function of the frequency at which 'refocussing' pulses are applied (νCPMG) [34, 9, 10]. If exchange is active during the refocussing period, incomplete refocusing will occur, resulting in lower peak intensities (greater $R_{2,eff}$ values). As the frequency of the refocusing pulses in increased (greater νCPMG), the time between refocusing pulses is shorter; therefore, less incomplete refocussing will occur.

The $R_{2,eff}$ is plotted as a function of the νCPMG (Figure 1.4b) to produce a dispersion curve which can be fitted to appropriate models [35, 36] to extract values for relative populations, rates of exchange and chemical shift differences between the two states. The direction of the chemical shift difference between the two states can be obtained by either the comparison of spectra from HSQC vs HMQC experiments [37] or from measurement at different fields [38].

1.2.3 Hydrogen exchange experiments

Hydrogen exchange (HX) NMR experiments probe the rate of exchange of hydrogen atoms in the main chain of protein structure with the surrounding solvent. HX-NMR experiments provide information on ms and slower protein dynamics [34, 39, 40, 41]. Experimentally, HX experiments are often carried out by introducing a protonated protein into a deuterated solution and monitoring the corresponding loss of peaks in an NMR spectrum, such as an ^{1}H-^{15}N-SOFAST-HMQC, over time (Hydrogen-Deuterium exchange, HDX). The measured rates of exchange can then be converted into protection factors (Equation 2.8) which measure the difference in exchange rate of residues in the context of the protein fold, compared to an unfolded polypeptide chain [34, 39, 40, 41]. Hydrogen exchange can also be measured by monitoring the transfer of ^{1}H magnetisation from the solvent to the protein (Hydrogen Magnetisation Exchange, HMX), which is sensitive to faster timescales than HDX experiments [42, 43, 44, 45, 46]

Hydrogen exchange typically proceeds through two limiting kinetic mechanisms. One process, known as EX2 is described in Equation 1.1 and is caused by small, local and uncooperative fluctuations in the protein structure [34].

$$k_{obs} = \frac{k_{op}k_{ch}}{k_{cl} + k_{ch}} \tag{1.1}$$

Where k_{obs} is the observed rate of exchange. k_{op} is rate of protein opening, or unfolding. k_{ch} is the intrinsic rate of exchange and k_{ch} is the rate of protein refolding [34].

The other limiting kinetic mechanisms, known as EX1 (Equation 1.2), is caused by global unfolding and other large scale, cooperative motions in the protein [34]

$$k_{obs} = k_{op} \tag{1.2}$$

Where k_{obs} is the observed rate of exchange. k_{op} is rate of protein opening, or unfolding [34].

Hydrogen exchange experiments are particularly sensitive to rare states as brief excursions to a low-populated state, can have profound effects on the rates of exchange for a residue which is normally protected from exchange in the native state which becomes solvent exposed in the excited state. [12]. While HX-NMR experiments have been used extensively to study protein folding pathways [40, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57], HX-NMR has also been utilised to provide residue level information on excited states for HIV-1 integrase [12, 43], which show that the enzyme may sample the DNA-bound state, even in the absence of DNA [43]. Hydrogen exchange experiments, measured by both NMR and mass spectrometry, were used to reveal the excited state involved in lysozyme associated amyloidosis and the effect of disease associated point mutations on the energy landscape of the protein [58, 59].

1.2.4 Paramagnetic relaxation enhancement

The presence of a paramagnet, such as MTSL ((1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate) enhances the relaxation rate of nearby nuclei (within 35 Å) with an r^{-6} distance dependence [60, 10, 61]. The PRE effect is due to the magnetic dipolar interactions between the unpaired electron on the paramagnet and the atomic nuclei in question [62, 63]. The addition of a paramagnetic label like MTSL to a protein, usually via a disulphide bond with a cysteine residue can be used to measure intermolecular and intramolecular PREs (Figure 1.5) [60]. The PRE effect is measured with NMR through relaxation experiments, if the signal is good enough, or by HSQC experiments. In both cases, the measurements are carried out in

the presence and then the absence of the paramagnetic label. The PRE effect will result in faster relaxation rates or lower peak intensity [60, 10]. The PRE data are useful in structural calculations, identification of weak, transient interactions and to measure transient structure in intrinsically disordered proteins [60, 10, 64, 61]. PRE experiments can provide information on ms timescales [23, 10] so can be effective on timescales which are not accessible by CPMG-RD and CEST experiments [23, 10, 9]. PRE experiments, are complimentary to CPMG-RD and CEST as they are also sensitive to very low populations (1%) [10].

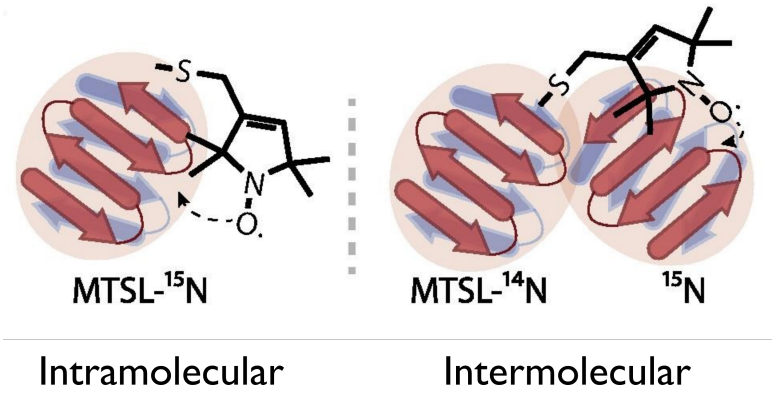


Fig. 1.5: PREs to measure weak, transient states and interactions.
The MTSL spin label will cause faster relaxation rates or lower peak intensity in nearby residues. Intramolecular PREs are sensitive to the presence of rare conformations in the protein. The intermolecular PRE scheme is sensitive to weak, transient interactions between the spin labelled ^{14}N protein with the ^{15}N protein which is observed in the NMR experiment. Adapted from [34].

1.2.5 Residual dipolar couplings

Residual dipolar couplings (RDCs) between two atoms are dependent on the angle between the intermolecular bond and the external magnetic field (Figure 1.6) [65, 66, 67]. Usually, in isotropic conditions, molecular tumbling averages RDCs observed in proteins to zero. However, the introduction of weak alignment, usually through the addition of alignment media such as filamentous phage, liquid crystalline media or mechanically strained gels [68] allows direct measurement of the RDC [69, 66].

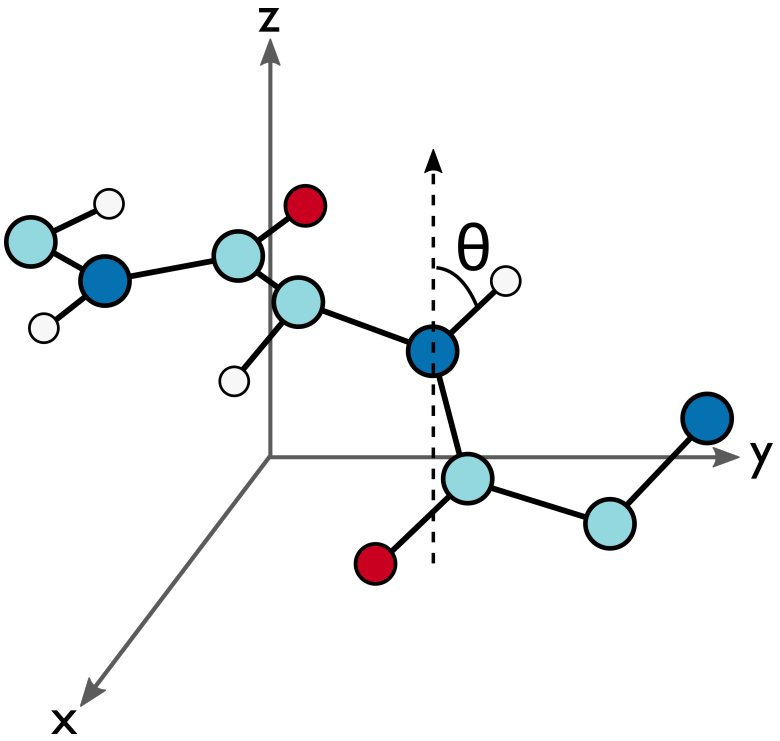


Fig. 1.6: RDCs depend on the the angle between the intermolecular bond and the external magnetic field.
In this schematic the θ angle is the angle between the HN intermolecular bond and the external magnetic field for the peptide illustrated as balls and sticks. The atoms are colour coded where H is white, N is blue, O is red and C is cyan, bonds are shown as black lines. Adapted from [67].

RDCs contain rich structural and dynamic information [67, 34] over a wide range of timescales (ps-ms) [23] and are therefore a powerful tool in structural biology. RDCs have been used for structural calculations [66, 70, 71, 72, 73, 74] and to interpret dynamics [75, 76, 77, 78, 79] in proteins. The calculation of structure and dynamics can be achieved through a variety of methods, typically requiring the definition of an alignment tensor. Alignment tensors can be predicted from structural properties of the alignment medium and the protein [80, 81, 82], or via fitting methods based upon the single value decomposition (SVD) method [83]. These alignment tensor dependent methods become increasingly complex in cases where many conformational states are present [67]. However, new tensor-free methods, in combination with replica-averaged molecular dynamics approaches have been used to generate accurate structural ensembles, by using the direct relationship between the RDCs and the angle between the intermolecular bond and

the external magnetic field [67, 33]. The combination of multiple replica simulations and RDC data has been used successfully to generate structural ensembles which are in good agreement with the experimental data [78, 79, 67, 33].

1.2.6 Heteronuclear relaxation experiments and protein dynamics

T_1 and T_2 heteronuclear relaxation experiments are both able to probe conformational dynamics on ps-ns timescales, while T_2 heteronuclear relaxation experiments are also sensitive to conformational exchange on µs-ms timescales [84, 85, 86, 23]. A schematic describing how the experiments are carried out is presented in Figure 1.7. T_1 relaxation experiments measure the loss of magnetisation from the xy plane back to the z axis as a function of a certain delay period, while T_2 relaxation experiments measure the loss of coherence of the magnetisation in the xy plane. In both cases, the data are plotted as the loss of signal intensity vs the relaxation delay, which is used to measure the relaxation rate.

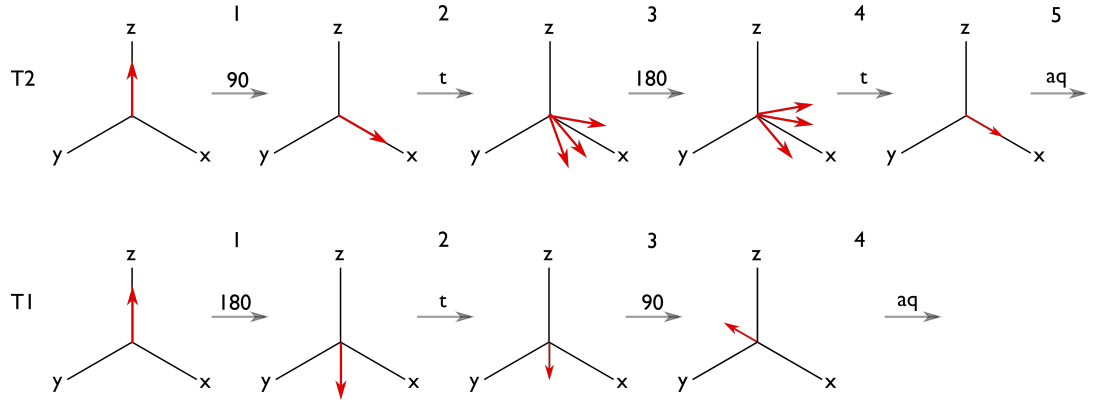


Fig. 1.7: T_1 and T_2 relaxation to measure protein dynamics.
In the T_2 experiment, a 90° pulse is applied (step 1), which brings the net magnetisation into the xy plane. The magnetisation is allowed to precess for time t (step 2) before a 180° refocussing pulse is applied (step 3) and the magnetisation is allowed to precess for a further time t (step 4). After step 4 the magnetisation is all in the x axis for detection. In the T_2 experiment the steps 2-4 are repeated for a set relaxation period. The peak intensity as a function of different relaxation period durations is plotted and rates are extracted. In the T_1 relaxation experiments, a 180° pulse puts the magnetisation into -z axis (step 1) after time, t (step 2) a 90° pulse is applied (step 3) to bring the magnetisation into the -x axis for detection. The peak intensity is plotted as a function of different relaxation times (t) and decay rates are extracted.

Per-residue T_1 and T_2 relaxation rates are used together with heteronuclear-NOE measurements to calculate order parameters for a protein. The order parameter describes the degree of flexibility in the bond vector on ps-ns timescales, where a value of zero would equate to unrestricted flexibility and a value on one would mean the bond vector is entirely restricted [85, 86, 23, 87]. These relaxation experiments are sensitive to many effects such as protein-protein interactions [88] and areas of local flexibility [84, 85, 86, 23].

1.3 The role of excited states in protein function and disease

Through the use of the experimental approaches presented in the previous section (Section 1.2), it is possible to study conformational changes on a wide range of timescales and to probe the conformational energy landscape of a protein. These experiments can thus reveal low-populated 'invisible' or 'excited' states, which exchange with the major, native state on μs-ms timescales. 'Excited' states are often important in protein function and also in disease. Various examples of the importance of such states are presented in this section.

1.3.1 SOD1 excited states

SOD1 is an antioxidant enzyme which is involved in the amyotrophic neurodegenerative disease known as amyotrophic lateral sclerosis (ALS) [89]. ALS is a complex disease. However, SOD1 oligomerisation and aggregation is a factor in familial forms of ALS where 20% of cases involve SOD1 mutations [89, 90]. Wild-type SOD1 is also implicated in sporadic forms of ALS as SOD1 has been shown to aggregate and cause toxicity *in vitro* and *in vivo* [91, 92, 93, 94, 95]. SOD1 inclusions are observed in both familial and sporadic cases of ALS, which suggests that there may be a common pathway or species involved in both forms of the disease [90, 95, 96].

SOD1 has an elaborate scheme of post-translational modifications required to reach the mature functional enzyme, which exists as a homodimer. The 153 residue, imma-

ture, monomeric protein must bind a Cu and Zn ion and form an disulphide bond with the other monomer in the homodimeric complex to reach the mature form [90, 97, 98] (Figure 1.8). The mature SOD1 enzyme is highly stable with a melting temperature between 85-96°C [97, 99] and remains enzymatically active in either 8 M urea or 4M GdnHCl [97, 100].

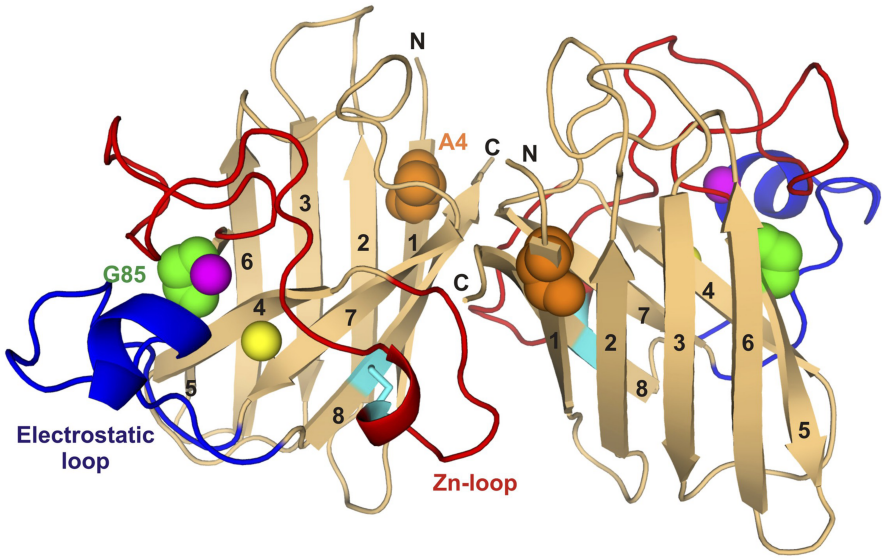


Fig. 1.8: Crystal structure of the mature SOD1 homodimer.
The crystal structure of the mature SOD1 homodimer (PDB:1HL5 [101]). Each strand in the 8-stranded β-barrel is numbered in both monomers. The Zn-loop is red and the electrostatic loop is blue. The bound Cu (yellow), Zn (magenta) and intermolecular disulphide bond (cyan) are required for maturity. The non-conserved Cys residues at positions 85 (green) and 5 (orange) where mutated out in the NMR study described here [90]. Figure was adapted from [90].

The immature SOD1, lacking the bound Cu and Zn ions, and without the intermolecular disulphide bond, remains predominantly monomeric, is aggregation prone [102, 103, 104] and may be important in causing disease, particularly when there are failures in protein quality control, which contribute to ALS [89]. Extensive NMR studies of the immature SOD1 protein, using CEST and CPMG-RD experiments to study various 'hidden' states [90] have shown that immature SOD1 has a complex free energy landscape. The immature SOD1 protein explores four low-populated states (Figure 1.9). One of these states involves local folding in the monomeric form of the protein, and the three other states are all homodimeric species. Two of the dimeric states are dysfunctional arrange-

ments, possibly linked to oligomerisation and aggregation, whereas one of the dimeric arrangements resembles the mature homodimeric SOD1 discussed earlier.

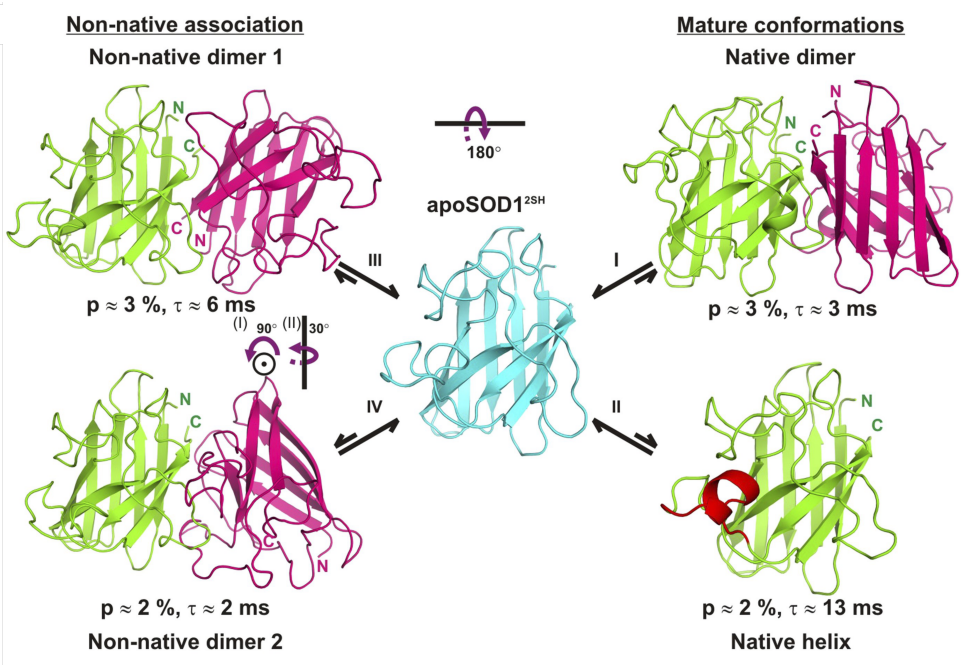


Fig. 1.9: The four transiently populated states in the immature SOD1 conformational landscape.
Immature SOD1 transiently populates two states resembling 'Mature' conformations (Native dimer and Native helix) and two 'Non-native' states which form dimeric structures inconsistent with the mature homodimer. The populations of each state at equilibrium (p) and the lifetimes (τ) are shown. Figure adapted from [90].

CEST experiments were utilised in this work, to identify processes I and II (Figure 1.10) in the immature SOD1 landscape. The CEST data for the immature SOD1 enzyme in Figure 1.10 [90] show that some residues (such as Thr135) display a concentration dependence, which is indicative of oligomerisation. The data for residue Gly61, however, does not display a concentration dependence so is likely to be due to exchange with an invisible monomeric species. Analysis of the chemical shift data for the CEST data showed that chemical shifts for residues in both processes are consistent with the mature dimer structure [90, 10]. This chemical shift information showed that the monomeric protein was able to form a dimer-like conformation and that the dimer itself could pre-form in the immature enzyme, but presumably requires the bound metal ions and intermolecular disulphide bond to remain in the mature configuration.

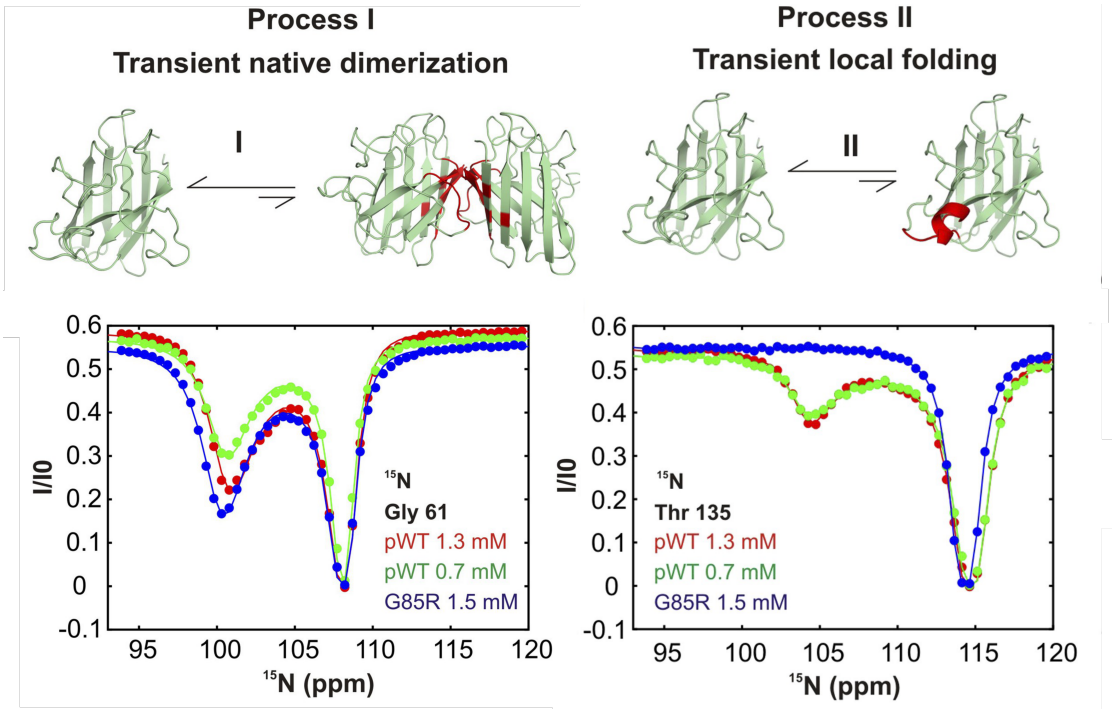


Fig. 1.10: CEST profiles differentiate between to independent SOD1 processes. The CEST profile for process I (native dimerisation) is concentration dependent, indicative of dimerisation. The CEST profiles for process II (transient local folding) are not concentration dependent. Both CEST profiles show a second minimum corresponding to the invisible state. Adapted from [90].

The immature SOD1 protein also samples non-native dimeric structures described in processes III and IV (Figure 1.9), which were characterised using CPMG-RD experiments [90]. CPMG-RD was used to detect two different exchange processes which both reported on different non-native dimer formation. These CPMG-RD results provide insight into the oligomerisation and aggregation pathway for immature SOD1, since one of the non-native dimers (process IV) is asymmetrical and able to form higher order oligomers with other dimers (Figure 1.11; based on *in silico* docking experiments) [90].

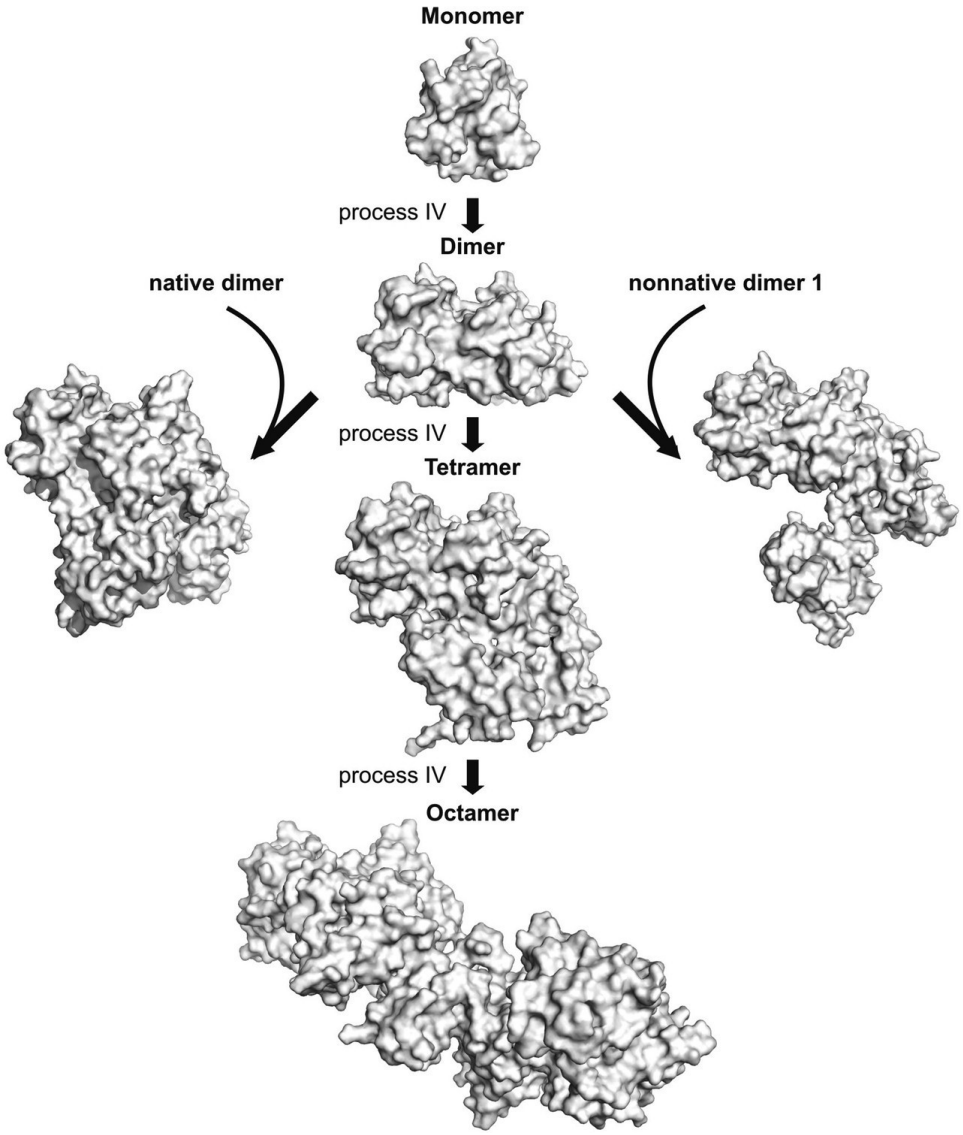


Fig. 1.11: SOD1 oligomerisation may proceed via a non-native dimer.
in silico docking experiments suggest that the non-native dimer from process IV may recruit other dimers for oligomerisation. Figure taken from [90].

Overall, the use of CEST and CPMG-RD studies show that the immature protein samples conformations which resemble the metal bound, mature SOD1 homodimer, even without the presence of metal ions. However, a consequence of the conformational flexibility required for metal binding and native dimer formation is the exploration of other potentially aggregation-prone species. Importantly, all of the 'invisible' species revealed by NMR are accessible through thermal fluctuation. These studies provide valuable insight into the conformational landscape of a folded protein and may provide structural

information required to prevent SOD1 aggregation. These studies also show that all of the low-populated 'invisible' states are important, whether it is for correct function or disease, and that the major population itself does not provide insight into function or dysfunction of the protein. It is also clear that in the case of SOD1, the binding of ligands (Cu and Zn ions and intermolecular disulphide bonds), can drastically alter the energy landscape of the protein, and pull the major state into a new energy minimum, which is highly stable and functional (such as the mature SOD1 enzyme).

1.3.2 Functionally important 'invisible' states in enzyme energy landscapes

'Invisible' states can also be important in protein function and CPMG-RD experiments on a number of enzymes have shown that enzymes often pre-sample functional structures in order to guide the chemical reaction cycle [9, 8] and have shown that this conformational change can be the rate limiting step in enzyme catalysis [9, 105].

CPMG-RD experiments carried out on a range of different enzymes [9, 8, 105] such as *E. coli* dihydrofolate reductase (DHFR [106]), adenylate kinase [107], HIV protease [108], cyclophilin A [109], RNAse A [110, 111] and others [112, 113], have shown that 'invisible' conformational states corresponding to functional structures are present. In all of these studies, two important features are common. The first feature is that even without the presence of the substrate, the ground state structure exchanges with functional states observed in the enzymatic cycle [9]. The second feature is that the rate limiting step of the enzymatic cycle is consistent with the rate of exchange between conformational states observed with CPMG-RD experiments, indicating that the conformation change itself is the rate-limiting step [9]. These observations reinforce the notion that low-populated functional states are present in the energy landscape and are accessible via thermal fluctuations, even without the ligand present.

The five-state cycle of DHFR [106], presented in Figure 1.12 is a case where CPMG-RD experiments have been used to detail the complex energy landscape for each step in the cycle which shed light on the observed rates of catalytic turnover and slower than expected rate of hydride transfer by the enzyme [106].

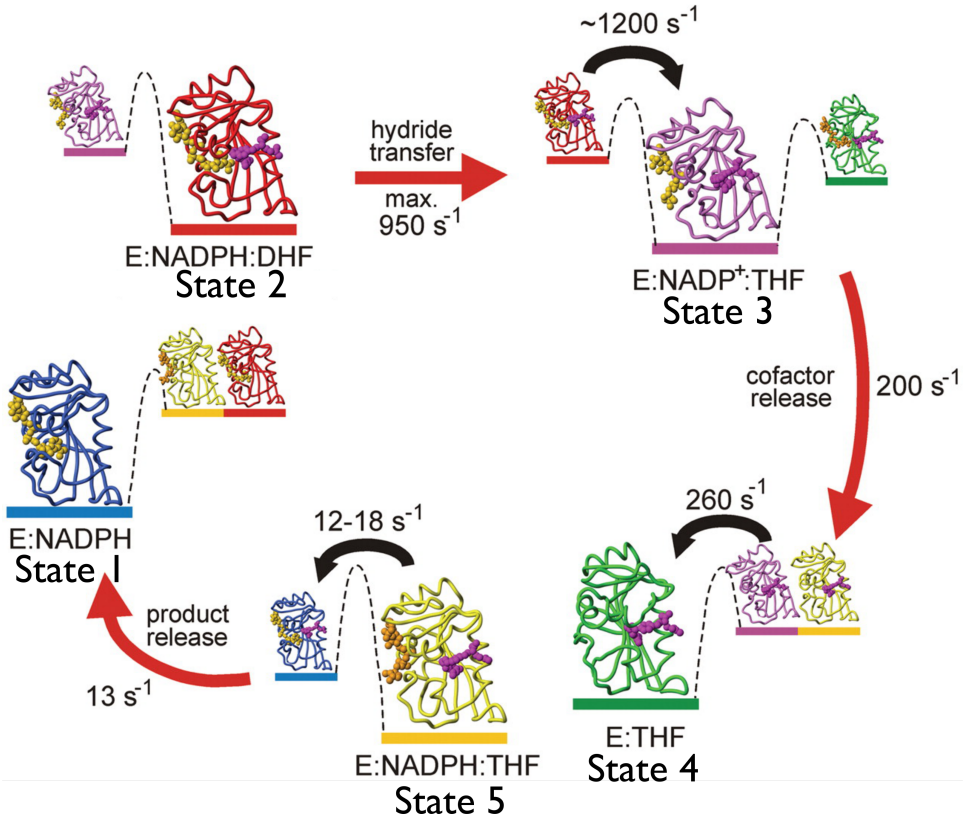


Fig. 1.12: The dynamic landscape of DHFR catalysis, revealed by CPMG-RD NMR. Ground states are shown as large structures and their associated excited states are shown as smaller structures adjacent and coloured to match the major states they resemble in other steps of the cycle. Measured rate of interconversion measured from CPMG-RD experiments are shown as black arrows. Measured rate constants from pre-steady-state enzyme kinetics are shown on red arrows. NADPH and NADP+ are shown in gold. Substrate and product analogues are shown in magenta. Figure adapted from [106].

The cycle in Figure 1.12 shows the five step cycle for the catalytic action of DHFR. For each state in the cycle, CPMG-RD experiments revealed the presence of excited states. The excited states were consistent with either the previous or next step in the cycle, or both. The rate of hydride transfer (State 2 to 3) was not limited by the fast hydride transfer, but rather by the exchange of State 3 with an excited state resembling State 2. The rate limiting step of the catalytic cycle was limited by the conversion of State 5 to State 1, which occurred on a a timescale consistent with the exchange of State 5 with an excited state resembling State 1, revealed by CPMG-RD.

The examples presented in this section highlight the importance of understanding invisible states and show how they can be studied in detail through CEST and CPMG-

RD NMR experiments. These experiments however are not effective when populations of the excited sates are less than 0.5% [12]. States populated below 0.5% can still be important in protein function and disease. Fortunately HDX-NMR is able to probe the effect of these states on hydrogen exchange behaviour of the native state.

1.3.3 Excited states of human lysozyme variants cause amyloidosis

The familial variants of human lysozyme are examples of point mutations which cause a population shift in the protein conformational ensemble which results in aggregation and disease [114, 115, 116, 117]. There are seven amyloidogenic lysozyme familial variants which cause systemic amyloidosis involving neuropathy and deposition of lysozyme amyloid fibrils in the kidneys and other tissues. The known mutations are D76H [114], I56T [114], W64R [115], F57I [116], F57I-T70N [116], T70N-W112R [117] and T70N [116] (T70N is the only non-amyloidogenic mutation) The positions of the amino acid substitutions are shown on the wild-type lysozyme structure in Figure 1.13.

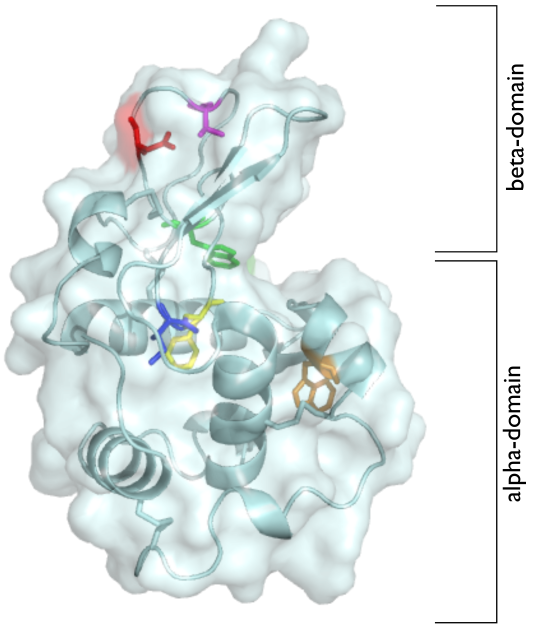


Fig. 1.13: Familial lysozyme mutation sites. The sites of all lysozyme familial mutations [114, 115, 116, 117] are shown on the structure of WT lysozyme (PDB:1REX [118]). Residues 67 (red), 64 (green), 56 (blue), 57 (yellow), 70, (magenta) and 112 (orange) are shown as sticks. Disulphide bonds are also shown as sticks. Produced with the PyMol molecular graphics system [119], which was used to produce all further protein structure images. Adapted from [120].

HDX-NMR experiments on the amyloidogenic D67H [58, 59] and I56T [59] variant proteins showed that some regions of the protein (β-domain and C-terminal region, residues 36-102) undergo local unfolding which results in HDX rates over 10-fold faster compared to WT in the region. The remainder of the variant protein (the α-domain) behave the same as the corresponding region in the WT protein, demonstrating that the observed effect is indeed local. Moreover, the observed local unfolding in the β-domain and C-terminal region of the variant proteins showed coherent exchange behaviour in HDX experiments monitored by mass spectrometry, which showed that the entire region unfolded cooperatively [58, 59]. As a consequence of the observed cooperativity of the local unfolding process, the entire region will periodically wholly unfold, meaning that even at very low populations, the partially unfolded state may well be sufficient to initiate the aggregation of the variants. The rate constant for 'opening' in the local unfolding

event occurs on a timescale 5-10 fold faster than the rate of global unfolding determined kinetically by stopped flow folding experiments [58, 59, 121].

The region responsible for the local unfolding event also corresponds to the proteolysis-resistant part of the fibril, suggesting that this area composes the fibril core [122]. Similar HDX studies on the WT and non-amyloidogenic T70N variant show that the partially unfolded state is present in the conformational landscape of the WT and T70N proteins. However, the partially unfolded state is far less populated in WT and T70N than in the disease variants and is only observable at higher temperatures which presumably make the energy barrier between the native fold and the partially unfolded state lower, making the partially unfolded state more accessible [123]. This proposed mechanism of lysozyme aggregation is depicted in Figure 1.14 and highlights the importance of lowly-populated states in aggregation.

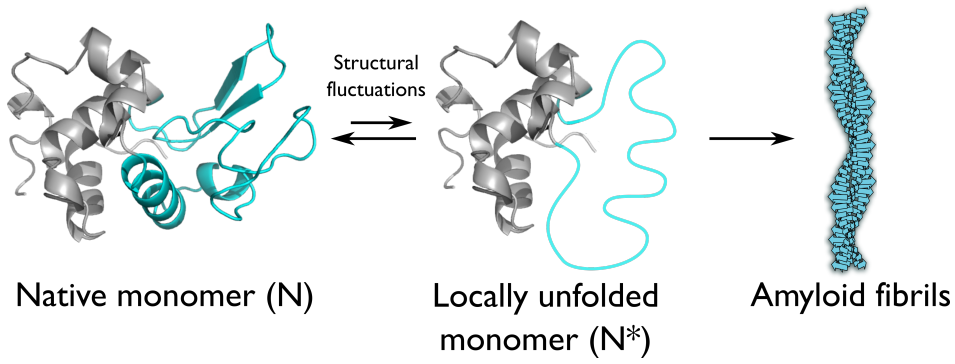


Fig. 1.14: The proposed mechanism of lysozyme amyloid formation.
The natively folded lysozyme (left) partially unfolds in β-domain and C-terminal region (residues 36-102) to form a partially unfolded N^* state. The locally unfolded portion of the N^* state, which is accessed from thermal fluctuations (especially in disease variants) then forms the lysozyme fibril core. The fibril structure is for illustrative purposes only. Figure adapted from [13].

There are a number of other proteins other than SOD1 and human lysozyme where aggregation occurs via an excited states accessible through thermal fluctuations from folded states, under physiological conditions, such as human transthyretin (TTR).

1.3.4 TTR aggregation proceeds via and excited state

TTR is a 55 kDa homotetrameric retinol binding protein (Figure 1.15) which is responsible for various amyloid diseases [124, 125] involving the wild-type protein and many familial mutants [126]. The homotetrameric protein must dissociate into a monomeric species in order to aggregate, and a partially unfolded monomeric species, which is invisible at physiological pH, is associated with aggregation [127, 128, 129]. Experiments at low-pH increase the population of the partially unfolded species and increase the aggregation propensity of the protein, strengthening the link between this species and aggregation [127, 128, 129]. This invisible, aggregation prone species has been investigated with CPMG-RD experiments [124]. These CPMG-RD experiments were carried out on a monomer-only variant and showed that there was an excited state present at approximately 3% at equilibrium, which had chemical shift differences localised in the D, A, G, and H β-strands [124] (Figure 1.15), which also undergoes chemical shift changes at aggregation promoting, low pH. This excited state was only present to 1% in a non-aggregation prone variant (T119M) [130, 131, 124]. These TTR studies once again highlight the importance of excited states in protein aggregation and disease and show how single point mutants can swing the balance of aggregation prone and non-aggregation prone conformations.

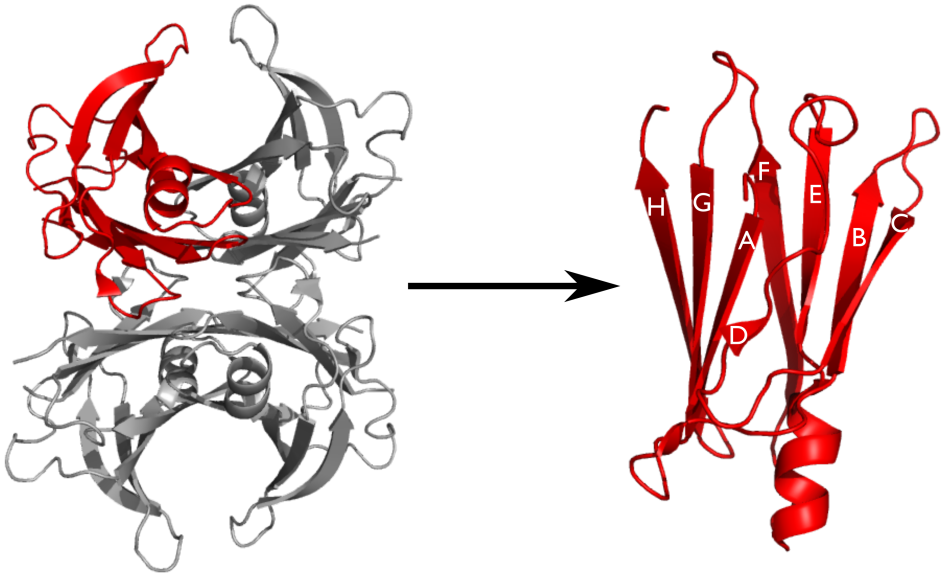


Fig. 1.15: The crystal structure of human TTR.
The tetrameric structure of human TTR (PDB:1F41 [132]) is shown in grey, with a single monomer in the complex coloured red. The monomer alone is also shows in red with the strands labelled.

The protein β_2-microglobulin and dialysis related amyloidosis [133] is another example of amyloid disease caused by a globular protein with a complex energy landscape which includes aggregation-prone states. The first familial variant of β_2-microglobin, D76N, which causes a hereditary amyloid disease [115] is the focus of this thesis.

1.4 Amyloid diseases share a common underlying architecture

As a consequence of the metastability and conformational flexibility of proteins previously discussed, proteins can misfold to form insoluble, higher-order aggregates. The most thermally stable, low energy form of protein aggregates is the highly ordered, cross-β amyloid fold [134, 135]. Amyloid deposition can occur both intracellularly and extracellularly and cause disease [134]. There are more than 40 proteins associated with amyloid disease, which can cause neurodegenerative and nonneuropathic systemic amyloid disease [135, 136]. Despite the sequence and structural diversity of amyloid precursor proteins, all amyloid fibrils share a common underlying architecture which produces a

characteristic X-ray diffraction pattern, characterised by 10Å and and 4.7Å reflections (Figure 1.16a) [137, 138, 134]. Amyloid fibres also share a common morphology when observed by electron microscopy (EM) (Figure 1.16b) [137, 134]. Amyloid fibrils can also be identified by their ability to bind the dye Congo Red, producing green birefringence under polarised light (Figure 1.16c,d) [139, 140, 134].

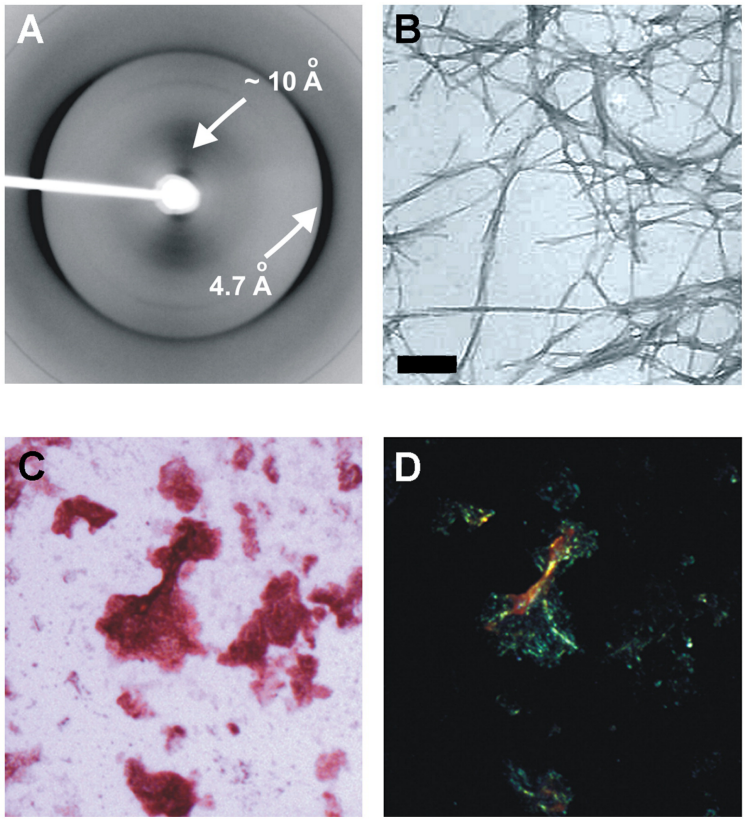


Fig. 1.16: Common hallmarks of amyloid.
The X-ray fibre diffraction patter in (a) shows the 10Å and 4.7Å reflections of the cross-β architecture. The negative stain EM image in (b) shows long straight amyloid fibrils. The images in (c) and (d) show Congo Red staining visualised with a bright-field light microscope and birefringence by cross-polarised light, respectively. The fibrils shows here are from wild-type β_2m formed at pH 2.5. Figure taken from [134].

Some amyloid fibrils are also able to bind Thioflavin-T (ThT), which is a benzathiole dye which is routinely used as a fluorescent probe for the presence of amyloid fibrils. The excitation maximum for ThT shifts from 385 nm to 450 nm when the ThT binds to amyloid fibres. This shift in excitation maximum is accompanied by a change in the

emission maximum from 445 nm to 482 nm [141, 142, 143, 144]. The chemical structure of ThT is presented in Figure 1.17.

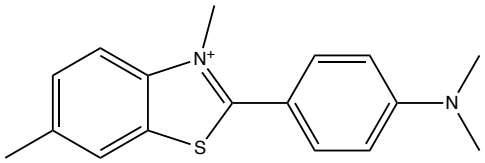


Fig. 1.17: The chemical structure of Thioflavin-T.

Monitoring amyloid formation *in vitro* with molecules such as ThT has led to the understanding that amyloid formation proceeds via a nucleation dependent mechanism (Figure 1.18) [134, 135, 145]. Under the nucleation dependent mechanism, spontaneous fibril growth (*de novo*) proceeds via an initial lag phase, corresponding to the formation of nucleation events, which are thermodynamically disfavoured [134]. Nucleation events are followed by the rapid elongation phase which is thermodynamically favoured. The addition of fibrillar seeds eliminates the lag phase as the seeds template the elongation phase [134].

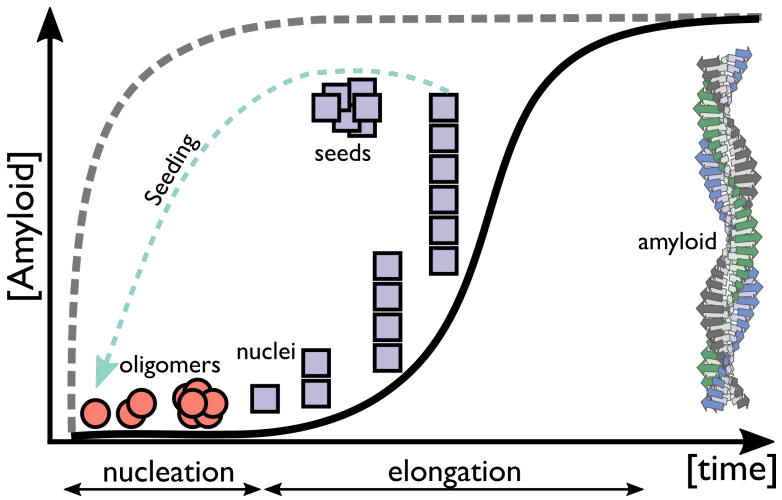


Fig. 1.18: Nucleation dependent mechanism of amyloid formation *in vitro*.
The black line shows *de novo*, spontaneous growth where the nucleus is formed in the lag phase, before rapid elongation towards the amyloid fibril. The grey dashed line depicts seeded growth where the lag phase is eliminated. Adapted from [134].

1.5 β_2-microglobin

1.5.1 β_2-microglobin and dialysis related amyloidosis

β_2 microglobulin (β_2m) is a component of the major histocompatibility complex class I (MHC-I) which plays an important role in antigen presentation [146, 147]. A MHC-I monomer consists of a heavy chain which is non-covalently assembled with a β_2m monomer which serves as the light chain in the MHC-I complex [148]. The physiological role of the MHC-I complex is to present a set of peptides which represent the cellular contents, on the surface of the cell [148]. These peptides are presented to T-lymphocytes, which can detect the presence of 'foreign' peptides, which may be indicative of a viral infection inside the presenting cell. If 'foreign' peptides are detected, the T-lymphocytes can recruit other immune cells, such as cytotoxic T-cells to kill the infected cell [149, 148]. The atomic structure of the MHC-I complex including β_2m [150] and the crystal structure of monomeric β_2m [151] are presented in Figure 1.19 and show that the structure of the 99 residue, $\sim$12 kDa β_2m is a typical of an immunoglobulin domain with a seven stranded β-sandwich structure with strands A, B, D and E on one sheet and strands C, F and G on the other sheet. The presence of a single disulphide bond between C25 and C80 stabilises the protein and reduction of this disulphide bond destabilises the native protein [152, 153].

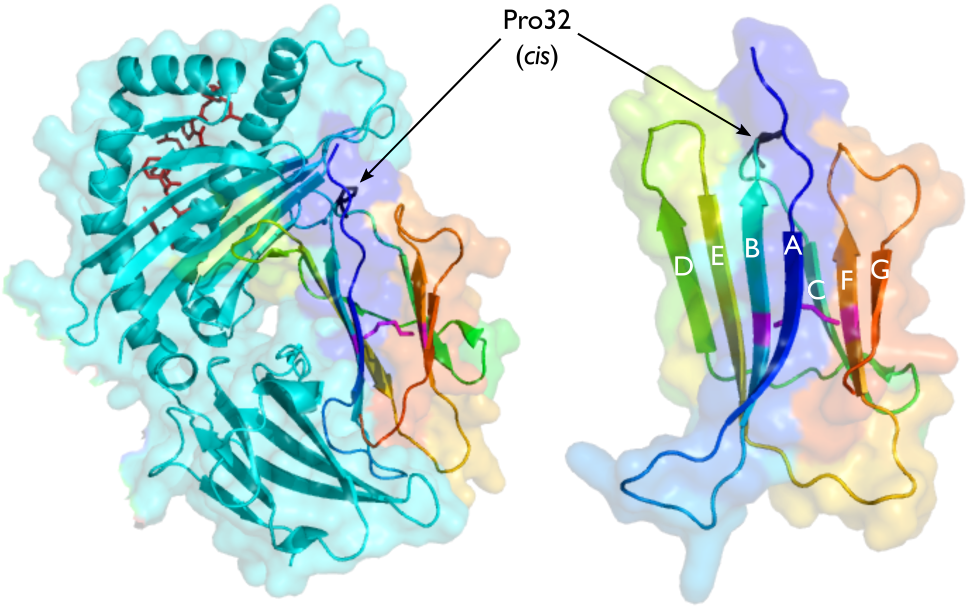


Fig. 1.19: The structure of the MHC-I complex and monomeric β_2m.
The crystal structure of the MHC-I complex (a) (PDB:3MYJ [150]), the heavy chain (cyan) and the β_2m (rainbow) are in complex with a Wilms Tumour 1 transcription factor variant (WT1-R1Y) peptide (red). The monomeric β_2m crystal structure (b) (PDB:1LDS [151]) shows the disulphide bond between C25 and C80 (magenta) and the proline at position 32 (black). The β_2m strands in (b) are labelled and are coloured in the same way in (a).

Since β_2m is non-covalently associated with the MHC-I complex, dissociation of the β_2m protein into the blood serum occurs [154]. In healthy individuals, β_2m is cleared from the serum via the kidneys [154]. However, individuals undergoing long-term hemodialysis for kidney failure are unable to effectively clear the β_2m from the blood serum, resulting in an increased serum β_2m concentration, from an average of 0.16 µM (5 heathy subjects) to 3.2 µM (11 patients) [154]. The increased serum β_2m levels result in the formation of β_2m amyloid fibrils, which deposit in bones and joints. The localisation of these amyloid deposits has been verified through tracking of intravenously administered iodine-131-labelled β_2m by whole body scintigraphy (Figure1.20) [155, 154, 156, 157, 158]. The deposition of such amyloid fibrils results in pathological bone and joint destruction in the disorder known as dialysis related amyloidosis (DRA) [154, 159, 160]. However, while increased serum β_2m levels are observed in DRA patients, the β_2m concentration alone is

not sufficient to cause DRA as increased β_2m levels are observed in many dialysis patients, who do not display symptoms of DRA [133]. The presence of additional factors must therefore be important in the formation of β_2m amyloid deposition in DRA.

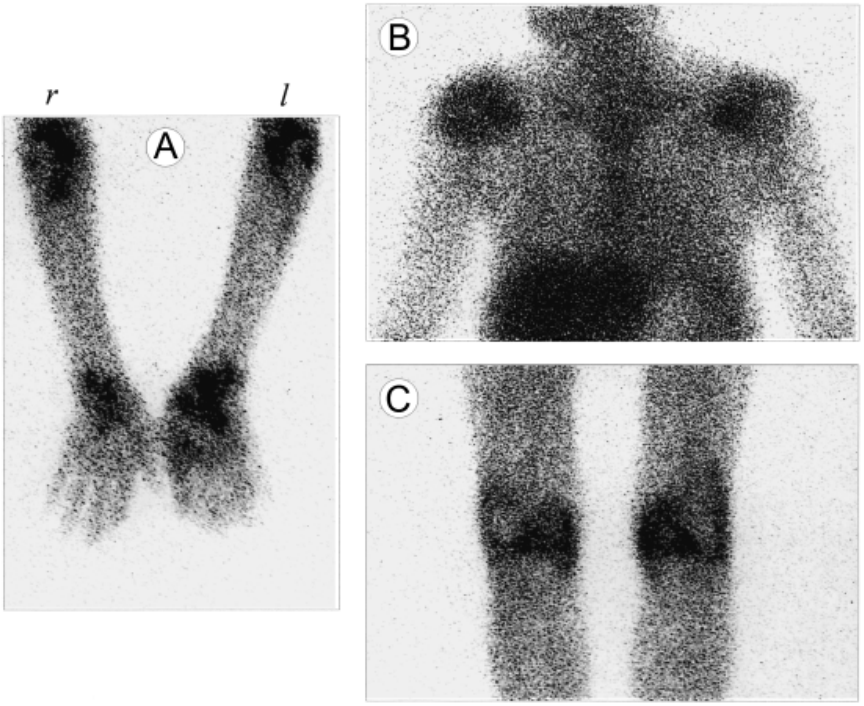


Fig. 1.20: Localisation of β_2m amyloid deposits in dialysis related amyloidosis. Whole body scintigraphy showing the localisation of β_2m amyloid deposits through tracking of intravenously administered iodine-131-labelled β_2m. Amyloid deposits can be detected in the hands and forearms (A), the shoulders (B) and knees (C). There is uptake in the liver in (B), which is unspecific and not due to amyloid. Figure taken from [155].

The insufficiency of full-length β_2m alone in the formation of β_2m amyloid fibres is mirrored *in vitro*. At physiological pH, full-length β_2m does not aggregate, however, other contributory factors [161, 162, 163, 164, 165], such as the presence of copper ions [166, 167], collagen [168], heparin [169] and post translational modifications (such as the cleavage of the N-terminal six residues of β_2m) [170, 171] result in the formation of β_2m amyloid fibrils.

These *in vitro* observations, together with extensive studies of the β_2m folding pathway provide insight into the mechanism of β_2m aggregation. [172]

Stopped-flow and double-jump refolding experiments [172] have shown that β_2m folds via a sequential process (Scheme 1). In the dead-time of the stopped flow refolding

experiment the unfolded species transitions into an intermediate known as the I_1-state, this step in the folding pathway is known as the 'burst phase'. The I_1-state next transitions into the I_2-state, this step is detectable in the stopped-flow experiments and is termed the 'fast phase'. The I_2-state converts into the Native state on a timescale of minutes, and this 'slow phase' was later found to be limited by the isomerisation of Proline-32 from a *trans*-to-*cis* configuration. The role of Pro32 in the 'slow phase' of β_2m refolding was identified with further double jump and NMR refolding experiments which showed that substitution of P32 with V32 eliminates the slow folding phase required to form the native fold [173, 174]. The I_2-state is therefore commonly referred to as the I_{trans}-state (I_T-state).

$$\mathrm{U} \xrightarrow[\text{burst}]{} \mathrm{I}_1 \xrightarrow[\text{fast}]{} \mathrm{I}_2 \xrightarrow[\text{slow}]{} \mathrm{N} \qquad \text{(Scheme 1)}$$

Where U corresponds to the unfolded state, I corresponds to folding intermediates and N corresponds to the native state.

The I_T-state was later shown to be an amyloid precursor in β_2m aggregation by its ability to elongate both *ex vivo* [175] and *in vitro* [176] derived β_2m fibrils. The study of mutants such as P5G and P32G, which have increased I_T-state populations, compared to the WT proteins (approximately 60%, 30% and 5% respectively) showed that the rate of formation of β_2m amyloid-like fibrils is dependent on the population of the I_T-state at equilibrium [177, 176].

The link between the I_T-state and β_2m aggregation was further developed by studies on the ΔN6 truncation variant, which is found in amyloid deposits of DRA patients [170, 171]. Refolding studies have shown that the ΔN6 variant does not exhibit a 'slow phase' [172] and the ^{1}H-^{15}N HSQC spectrum of ΔN6 closely resembles that of the I_T state observed during real-time refolding experiments [174, 176, 88]. Finally, the solution NMR structure of ΔN6 revealed that the Pro32 residue is indeed in the *trans* isomersiation state [88] and hence ΔN6 can serve as a structural mimic of the I_T state. ThT fluorescence assays have shown that ΔN6 readily forms amyloid-like fibrils under physi-

ological pH, and without the need for external factors *in vitro*, confirming the I_T state as an amyloidogenic precursor [177].

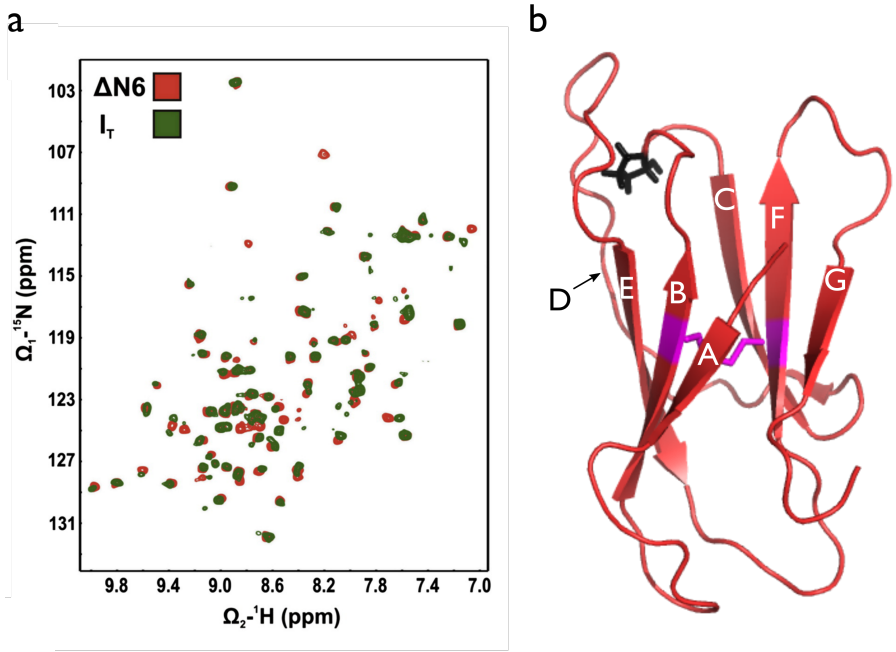


Fig. 1.21: ΔN6 as a structural mimic of the I_T-state.
The ^{1}H-^{15}N-SOFAST-HMQC of the I_T-state (2 minutes after refolding was initiated) and ΔN6 at equilibrium in (a) show how the I_T-state and ΔN6 are structurally similar (Figure taken from [88]). The NMR solution structure of ΔN6 (PDB:2XKU [88]) in (b) shows Pro32 (black) in the *trans* configuration, confirming that the structure of ΔN6 is indeed trapped in the I_T-state.

The presence of ΔN6 in DRA amyloid deposits begins to explain the importance of other contributing factors required for DRA to occur. Interestingly, the ability of copper ions to induce WT β_2m aggregation *in vitro* [166] may be due to the observation that the copper ions induce proline 32 isomerisation in WT β_2m [178, 179, 180]. The role of copper ions in β_2m aggregation is complemented by the correlation of Cu^{2+} levels in dialysis membranes and the incidence of DRA [166, 181, 182].

Further studies of the ΔN6 variant have also demonstrated the ability of ΔN6 to enhance the ability of WT Hβ_2m to form amyloid fibrils. Figure 1.22 shows how ΔN6 forms amyloid fibrils at pH 7.2 (37 °C, 200 rpm) while wild type Human β_2m (Hβ_2m) and the murine β_2m (mβ_2m) do not form fibrils under these conditions. However, ΔN6

promotes WT Hβ_2m aggregation as either a 1:1 ratio or in sub-stoichiometric amounts (1:99 ratio). In contrast, ΔN6 does not promote mβ_2m aggregation, in fact, mβ_2m inhibits ΔN6 aggregation (Figure 1.22).

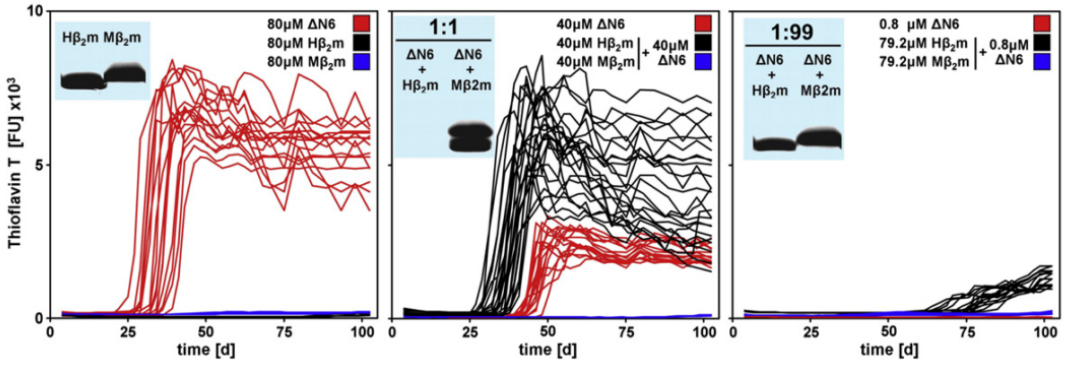


Fig. 1.22: ΔN6 enhances WT β_2m aggregation propensity *in vitro*.
The left panel shows the ThT fluorescence if 80 μM ΔN6 (red), Hβ_2m (black) and mβ_2m (blue) at pH 7.2, 37 °C, 200 rpm. The middle panel shows ThT fluorescence for 40 μM ΔN6 (red), 40 Hβ_2m with 40 μM ΔN6 (black) and 40 mβ_2m with 40 μM ΔN6. The blue inserts show SDS-PAGE analysis of the soluble fraction remaining after centrifugation of the samples at 14000 x *g*, 10 min. Figure obtained from [88].

The inhibition of ΔN6 aggregation by mβ_2m has been further investigated, resulting in the identification of a kinetically trapped inhibitory complex, revealed by NMR paramagnetic relaxation enhancement experiments [34]. Figure 1.23 shows the structure of this complex and highlights two key residues in mβ_2m which are crucial for effective inhibition of ΔN6 aggregation.

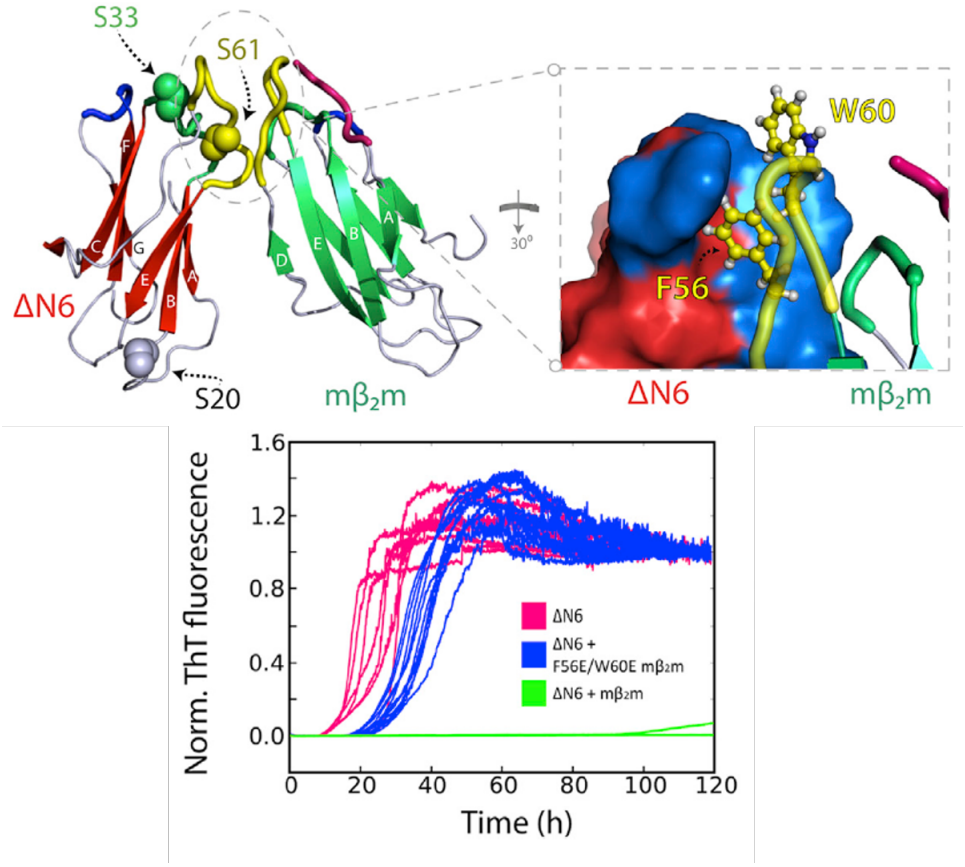


Fig. 1.23: The inhibitory ΔN6-mβ_2m complex.
(top) The lowest-energy calculated structure of the ΔN6 (red)-mβ_2m (green) complex highlighting F56 and W60 in the interface (space fill) (left panel). Interface residues are coloured blue on ΔN6 (right) as a space filling model, the key mβ_2m residues are coloured yellow. (Bottom) Aggregation kinetics of 20 μM ΔN6 alone pH 6.2 (red), ΔN6 mixed with a 2-fold molar excess of mβ_2m (green) or ΔN6 mixed with a 2-fold molar excess of F56E/W60E mβ_2m measured using ThT fluorescence. Taken from [34].

Real-time refolding experiments [60] have shown that mβ_2m does not form amyloid fibrils because it has a different energy landscape to WT Hβ_2m (Figure 1.24). NMR studies showed that a different, partially unfolded state known as the I_1 state is occupied to approximately 7%, thus reducing the population of the I_T state to a negligible level. These results reinforce the importance of the I_T state in fibril formation, and suggest that strategies to tip, or alter, the folding energy landscape may be used to stop amyloid formation.

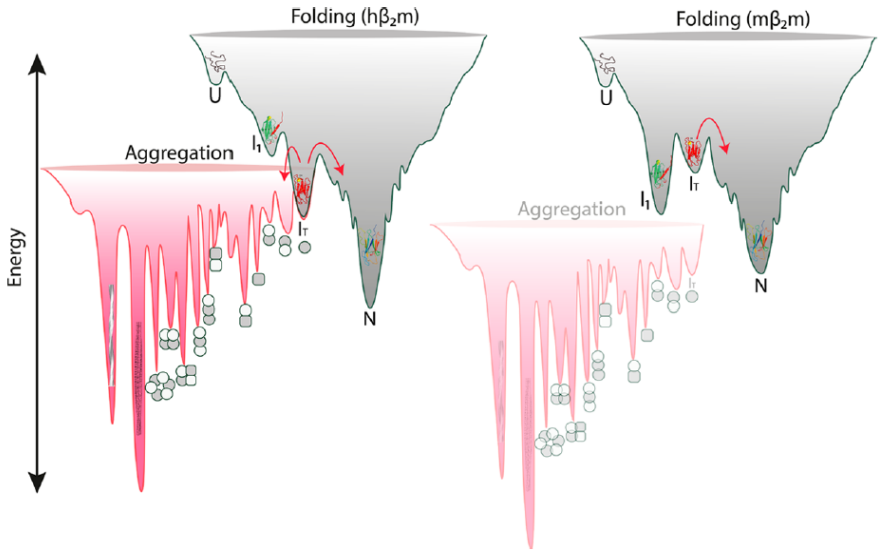


Fig. 1.24: Theoretical energy landscapes for β_2m folding and aggregation. The protein folding pathways (grey) for Hβ_2m and mβ_2m illustrating the increased population of the non-aggregation prone I_1 state in mβ_2m and the lower population if the aggregation prone I_T which, prevents mβ_2m from entering the aggregation landscape (red) (energy levels of aggregation landscape are for illustration purposes only). Taken from [60].

The studies of both mβ_2m and ΔN6 show the importance of alternate species and energy landscapes in β_2m aggregation and show how subtle changes in the conformation landscape of a protein can drastically change its aggregation propensity.

1.5.2 D76N β_2-microglobin associated systemic amyloid disease

In 2012 a novel, naturally occurring β_2m variant was identified as the causative agent of an hereditary, late-onset, fatal, systemic amyloid disease in a french family [183]. The patients suffered from "progressive gastrointestinal symptoms and autonomic neuropathy" [183] which was caused by deposition of D76N β_2m amyloid fibrils "in the spleen, liver, heart, salivary glands, and nerves" [183]. The amyloid deposits were shown to contain exclusively D76N variant β_2m and no other common amyloid proteins could be identified by immunohistochemical staining with monoclonal antibodies. Proteomic analysis of *ex vivo* fibrils by 2D gel electrophoresis, and tryptic peptide analysis by mass spectrometry could not detect any full length or ΔN6 WT Hβ_2m nor was any ΔN6 D76N β_2m present. The affected patients were heterozygous for the WT Hβ_2m and the D76N variant, they had

normal kidney function and thus had normal serum β_2m levels of 1.35-1.49 mg/L [183]. A comparison of D76N amyloidosis and DRA is shown in Figure 1.25.

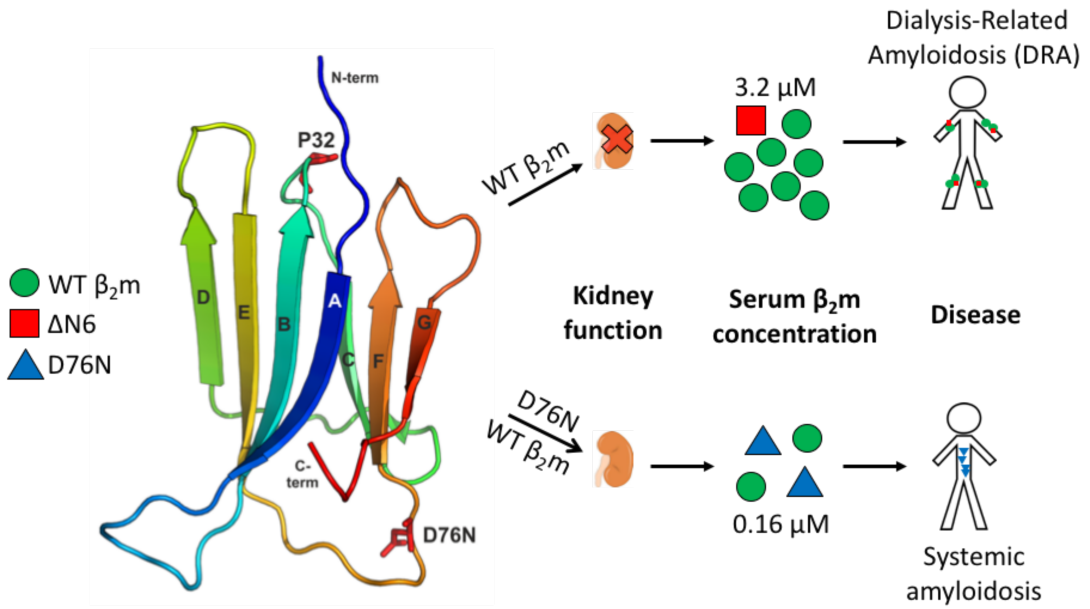


Fig. 1.25: WT vs D76N disease progression schematic.
A schematic illustrating the contrasting progression of DRA (top) and D76N amyloidosis (bottom). A comparison of the kidney function between the two diseases shows how DRA patients (top) have impaired kidney function, whilst WT-D76N β_2m heterozygotes (bottom) have normal kidney function. The impaired kidney function of DRA patients results in increased serum β_2m concentrations (green circles), from an average of 0.16 μM to 3.2 μM and a small amount of ΔN6 (β_2m which has had the N-terminal 6 residues proteolytically cleaved) (red squares) is enough to trigger aggregation, resulting in pathological bone destruction in the disorder DRA. There is no such increase in serum β_2m concentration in D76N heterozygotes who have both WT Hβ_2m and D76N (blue triangles) in the serum however no WT Hβ_2m or ΔN6 is observed in the D76N amyloid deposits present in the viscera of affected individuals suffering from systemic amyloidosis.

The D76N variant was shown to be less stable to GdnHCl denaturation by 11.3 kJ.mol^{-1} [184]. The calculated ΔG°_{un} (H_2O) of WT Hβ_2m was 23.8 kJ.mol^{-1} whereas the ΔG°_{un} (H_2O) of D76N was 12.6 kJ.mol^{-1}, at 30°C. This is remarkably more destabilised than ΔN6 (determined by urea denaturation) [177], which has a $\Delta\Delta G^\circ_{un}$ of 3.8 kJ.mol^{-1} compared to WT. However, it is known that thermostability does not correlate with aggregation propensity [152] as mβ_2m has a $\Delta\Delta G^\circ_{un}$ of 12.6 kJ.mol^{-1}, compared to human WT β_2m, measured by urea denaturation [60], and hence is less stable than the highly amyloido-

genic D76N, yet does not form amyloid. The instability of D76N has also been measured by differential scanning calorimetry which indicates that the D76N variant is less stable to temperature by a similar amount (ΔG°_{un} 11.7 kJ.mol^{-1} at 37 °C) [184] to the value obtained by GdnHCl denaturation experiments [184].

Further studies showed that the D76N variant formed amyloid fibrils *in vitro* at physiological pH (7.4) with 20% *(v/v)* Tetrafluoroethylene (TFE), in both the presence and absence of fibrillar seeds, whilst the WT Hβ_2m did not form amyloid fibrils under these conditions [183]. Importantly, the presence of heparin, which is a key modulator of WT Hβ_2m fibrillogenesis [176, 169] did not affect the rate of D76N fibrillation (the lag time remained the same) [183], suggestive of a different aggregation mechanism of the two proteins.

To determine the effect of the D76N substitution on the structure of β_2m the crystal structure was obtained for the D76N β_2m variant (Figure 1.26). Analysis of this structure suggests the presence of two new hydrogen bonds, N76-Y78 and Y78-T73 and analysis of the B factors for this region show that the EF-loop is more rigid in the D76N variant than in WT Hβ_2m. In general the structures of the WT and D76N β_2m variant are similar [183]. Further computational analysis has been undertaken to analyse the effect of the D76N amino acid change on the hydrogen bonding interactions with residue 76 and further networks dependent on the D76 residue [184, 185].

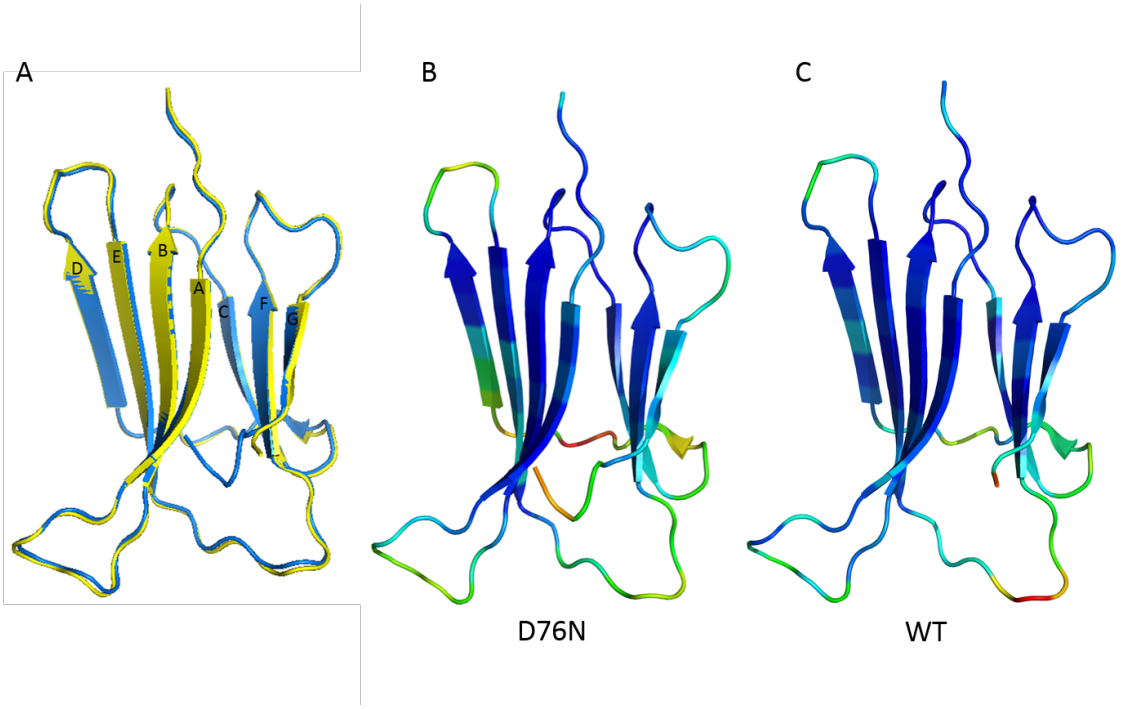


Fig. 1.26: Crystallographic B-factors for WT and D76N crystal structures. (A) Shows an overlay of the crystal structures for WT (yellow, PDB:1LDS [151]) and D76N (blue,PDB:4FXL [183]) β_2m. (B) D76N and (C) WT Hβ_2m coloured by B-factor, where blue corresponds to low B-factors and red corresponds to higher B-factors (less well ordered). D76N B-factors are in the range (8.3 to 32.0 Å^2) and WT in the range (1.0 to 45.5 Å^2).

Computational analysis of hydrogen bonds (H-bonds) involving position 76 in WT and D76N β_2m revealed that D76 in WT Hβ_2m forms a strong hydrogen bond to N42 which also forms a hydrogen bond with E77 and that there is a salt bridge between E77 and K94. There is a weaker set of interactions in WT Hβ_2m involving D76 and T73, a very weak, fluctuating H-bond between T73 and the side chain of R97 and a salt bridge between R97 and E16 [184]. According to Mangione et al. [184], all of these interaction appear to be present but they are all weaker in the D76N variant (except for 76-97 and 97-Asn17, which is the same as in WT). Chandrasekaran et al. [185] performed a similar analysis and postulated that the above interactions are all lost after a 27 ns simulation period apart from the interaction between N76 and T73. This analysis also suggests that loss of hydrogen bonding networks in the AB loop and DE loop lead to greater flexibility in these regions and that the loss of key interactions in the EF loop result in more frustration in the EF loop of the D76N variant [185].

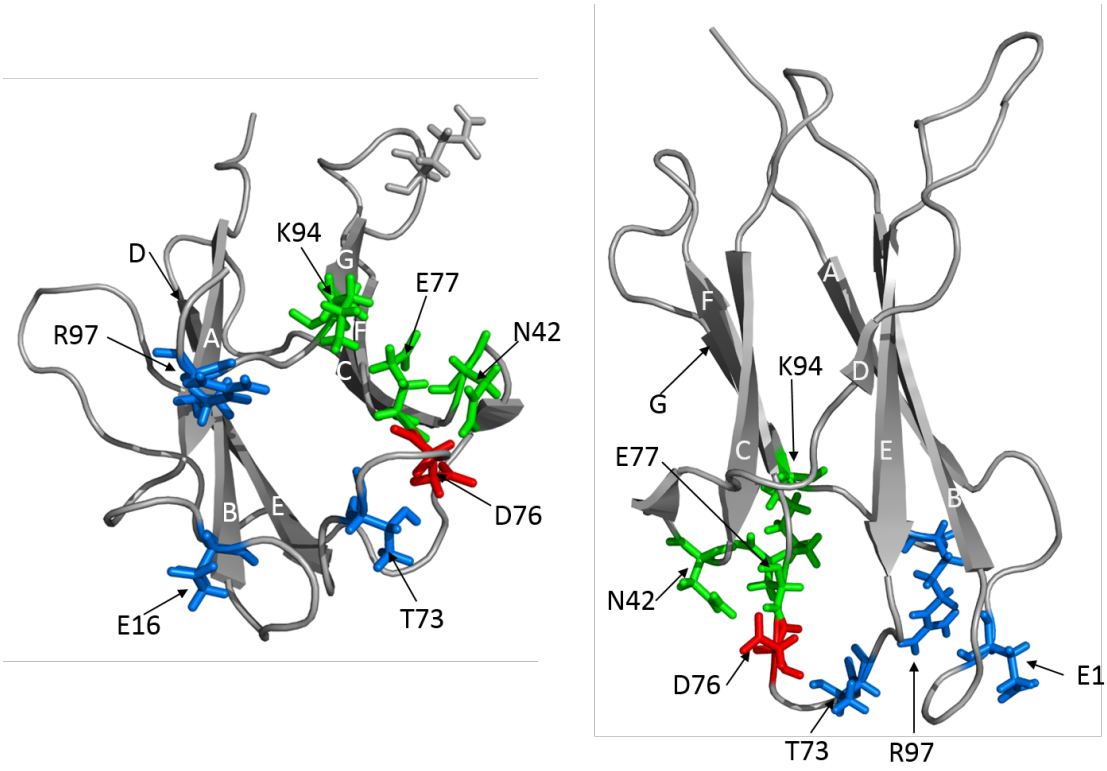


Fig. 1.27: Hydrogen bonding networks involving residue 76 in β_2m. The residues involved in the hydrogen bond networks with residue 76 (red) postulated by Mangione et al. [184] are shown in green for the network involving D76, N42, E77 and the salt bridge to K94. The blue residues are shown are T73, R97 and E16 (salt bridge between R97 and E16). (PDB:2XKS, WT Hβ_2m NMR structure [88]).

Since the I_T-state is known to be critically important for β_2m aggregation, the folding pathway of D76N was investigated with GdnHCl/refolding experiments [184]. These experiments showed that the D76N variant follows a similar folding pathways to WT Hβ_2m, with an initial fast phase followed by a slow phase corresponding to the isomerisation of Pro32 [184]. From these results, the authors [184] postulate that the D76N variant populates the partially folded I_T state to approximately 25% at equilibrium, in contrast to approximately 5% for WT Hβ_2m [184], possibly explaining the increased amyloidogenicity of the D76N variant.

Independent *in silico* studies of the D76N I_T-state [186, 187] have both suggested that the D76N I_T-state is structurally different to the WT I_T-state and that these differences drive the aggregation of D76N. Molecular dynamics simulations by Chong et al. [186] proposed that the D76N I_T-state has a larger solvent exposed surface area, a

more disordered D-strand and a larger solvation free energy, all of which contribute to the enhanced aggregation propensity of D76N, the structural model of the I_T-state used in these studies in presented in Figure 1.28b.

Rui et al. [187] suggested that D76N formed two different I_T-state structures, the first of which was the same as the WT I_T-state, however the second I_T-state was unique to D76N and had unfolded N-terminal and C-terminal regions. According to *in silico* docking experiments carried out with the I_T-state structures, the second D76N I_T-state was more prone to oligomerisation. The two D76N I_T-state models from these simulations are shown in Figure 1.28c,d.

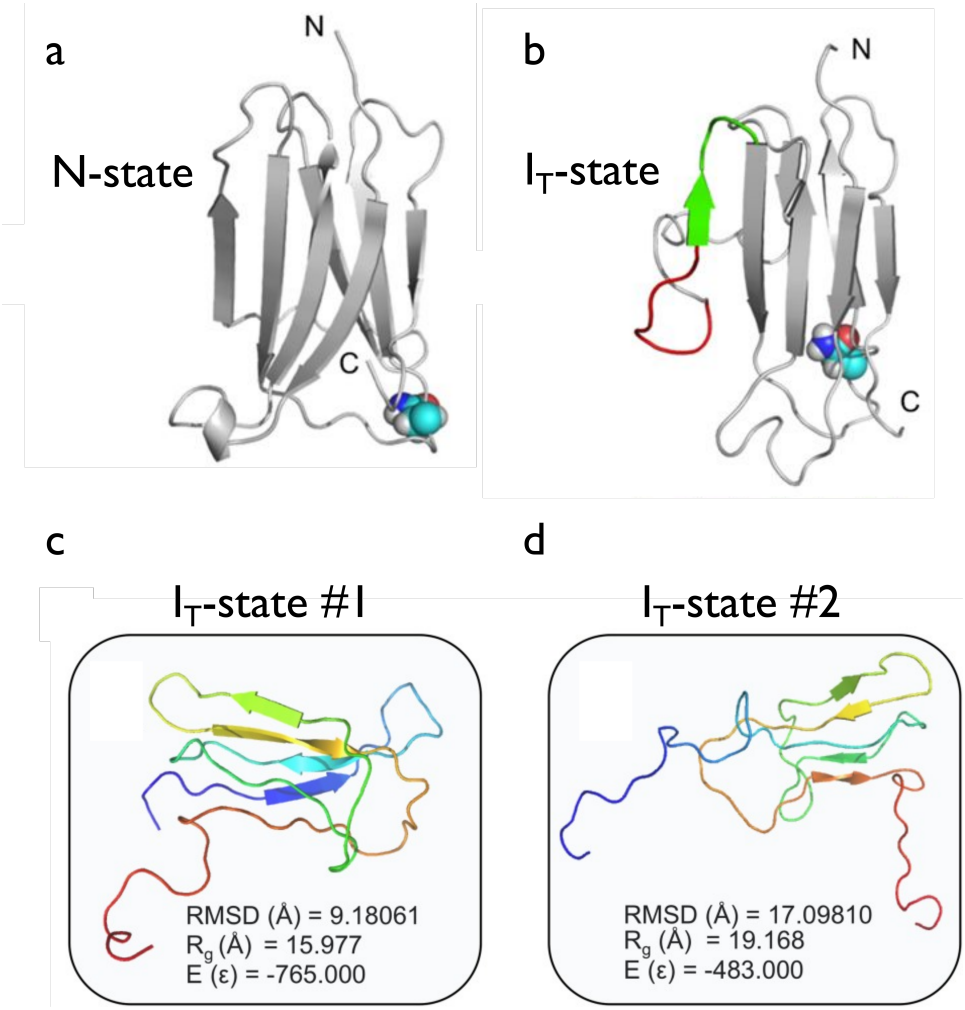


Fig. 1.28: Simulated structures of the D76N I_T-state.
The simulated D76N I_T-state structures from Chong et al. [186] in b and from Rui et al. [187] in c and d, compared to the D76N N-state structure in a. Adapted from [186, 187].

The observation that D76N and WT Hβ_2m did not appear to co-aggregate in patients [183] has been further investigated [184, 188]. Mangione et al. have shown that D76N and WT Hβ_2m co-aggregate *in vitro*, however, in the conditions used, D76N disappears from solution into the insoluble fibril fraction first, followed by addition of WT Hβ_2m to the D76N fibrils [184]. The observed addition of WT Hβ_2m to D76N fibrils can be inhibited by the presence of the chaperone α-crystallin [184, 188]. Further investigation of the co-aggregation of WT Hβ_2m and D76N, followed by FTIR spectroscopy showed that in an equimolar mixture of WT Hβ_2m and D76N, D76N becomes entirely fibrillar before WT Hβ_2m begins to form fibrils [188]. The chaperone α-crystallin prevents WT Hβ_2m elongation of D76N fibrils by directly binding to the D76N fibrils (visualised with AFM imaging) and thus blocking the binding of WT Hβ_2m to D76N fibrils [188]. Chaperones may therefore be responsible for the lack of co-aggregation observed in heterozygotic D76N amyloidosis patients.

Despite the reduced stability and increased amyloid propensity of the D76N β_2m variant, the protein evades cellular quality control mechanisms and is successfully expressed and secreted from cells. This can be explained by the ability of the MHC-I to stabilise D76N β_2m, protect it from cleavage by proteases and therefore facilitate its transport to the cell surface, where it is free to dissociate from the MHC-I complex and form extracellular aggregates [188].

In summary, while it is clear that D76N is less stable and indeed more amyloidogenic than WT Hβ_2m, there is not a molecular understanding about how a single amino acid substitution of a solvent exposed residue leads to this increased ability of the protein to form amyloid fibrils. It has been postulated that D76N occupies the I_T state, however there has yet to be a direct visualisation of this folded intermediate or any experimental comparison of how the structure of this intermediate compares with the I_T state of WT Hβ_2m. D76N pathology also progresses in a manner which is different to the pathology associated with the WT protein. The study of the D76N variant could give vital clues as to the origin of amyloidogenicity in folded proteins and why this particular mutation, which is the only β_2m variant found in humans, causes amyloid disease.

1.6 Aims and scope

The D76N β_2m variant is the first familial variant and at the time of writing, the only familial variant of β_2m identified. This single amino acid substitution, in a 100% solvent-exposed site results in a radically different disease progression in patients heterozygous for the mutation. The D76N related disease progression results in a late-onset, systemic and fatal amyloid disease with visceral amyloid deposits rather than the deposition of amyloid in the bones and joints, associated with the WT β_2m associated disease, DRA. The *ex vivo* D76N amyloid deposits also contained no WT β_2m or ΔN6 protein. This D76N variant was, therefore, an exciting opportunity to study the effect of the amino acid substitution on the energy landscape of the protein and to understand how it can cause aggregation and an alternate disease progression.

The D76N variant raises questions about the role of the I_T-state in D76N aggregation: Is there a greater population of I_T-state at equilibrium? Is the D76N I_T-state different?

There are further question about the D76N variant regarding the mutation site: How important is the identity of residue 76 and is there something specific about the N76 residue important for aggregation?

Further questions arose during the course of the studies presented here regarding the conformational dynamics of D76N: Is there an I_T-state independent excited state involved in aggregation? What is the nature of this state?

Studies of D76N are described in this thesis with the following chapters:

Chapter 1: Introduction.

An introduction to concepts surrounding energy landscapes and excited states. Techniques to study such states and how these states are important in function and aggregation.

Chapter 2: Materials and methods.

A presentation of all experimental details used in the experiments in Chapters 3-5

Chapter 3: Global properties and native structure of D76N β_2m.

The effect of the D76N substitution on the structure and stability of β_2m was investigated with NMR and biophysical methods. The aggregation propensity of the

D76N variant was compared to the I_T-state mimic (ΔN6) and the population of I_T-state at equilibrium for D76N was probed. The effect of saturation mutagenesis at position 76 in β_2m was also studied in terms of stability and aggregation. The results showed that while D76N is destabilised, the native structure remains unperturbed. The results also showed that D76N aggregated faster than ΔN6 and that an increased I_T-state population for D76N could not be detected. It was also found that position 76 is critically important in stability and aggregation of β_2m.

Chapter 4: Probing the D76N I_T-state with real time refolding experiments.

In this chapter the refolding of D76N and other position 76 variants was compared to WT and showed that the refolding pathway was largely unaffected by the mutation. The D76N I_T-state was visualised by NMR and remains the same as the WT I_T-state. The generation of a ΔN6-D76N variant showed that the D76N I_T-state aggregates at the same rate as the WT I_T-state and more slowly than full length D76N.

Chapter 5: Conformational dynamics of D76N.

In this chapter the conformational dynamics of D76N was probed through a variety of NMR methods and showed subtle differences with WT β_2m. Data-guided metadynamic-metainference molecular dynamics (MM-MD) simulations were performed and showed that D76N occupies a range of conformational substates, which resemble both the crystal structures of MHC-bound D76N and monomeric D76N which have some key differences in the AB loop and D-strand.

Chapter 6: Summary and conclusions.

This chapter summarises the results in Chapters 3-5 and contrasts the findings on D76N aggregation from this thesis with the current paradigms for β_2m aggregation.

Chapter 2

Materials and methods

2.1 Chemicals

All chemicals were purchased from Sigma-Aldrich (Dorset, UK), unless specified otherwise. ^{15}N labelled ammonium chloride and uniformly labelled ^{13}C glucose for isotopic labelling of proteins for NMR experiments was purchased from Cambridge Isotope Laboratories (Massachusetts, USA).

2.1.1 *E.coli* bacterial strains

The following strains of *Escherichia coli* were used:

- DH5α (Invitrogen), Genotype: F- ϕ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-
- BL21 [DE3], Genotype: E.coli B F^- dcm ompT hsdS(rB^- mB^-) gal λ(DE3)
- BL21 [DE3] pLysS, Genotype: E.coli B F^- dcm ompT hsdS(rB^- mB^-) gal λ(DE3) [pLysS Camr]

DH5α cells were used for molecular biology applications, BL21 [DE3] cells were used in the split-β-lactamase assay and BL21 [DE3] pLysS cells were used for protein expression.

2.2 Molecular biology

2.2.1 Site-directed mutagenesis

Site directed mutagenesis was carried out using a Q5 Site-Directed Mutagenesis Kit (New England Biolabs) to introduce point mutations into both the expression plasmid (Figure 2.1) and the tripartite fusion construct used for the split-β-lactamase assay (Figure 2.5). The primers used for the Q5 mutagenesis are listed in Table 2.1

Table 2.1: Primers used for site-directed mutagenesis of expression constructs and tripartite fusion constructs required for the split-β-lactamase assay. The same reverse primer was used for all reactions. Purchased from Eurofins Genomics. The codon for residue 76 is shown in lower case.

Mutant	Primer
D76E-Forward	CACTGAAAAAgaaGAGTATGCCTG
D76K-Forward	CACTGAAAAAaaaGAGTATGCCTG
D76R-Forward	CACTGAAAAAcgcGAGTATGCCTGCCGTG
D76H-Forward	CACTGAAAAAcatGAGTATGCCTG
D76S-Forward	CACTGAAAAAagcGAGTATGCCTGCC
D76T-Forward	CACTGAAAAAaccGAGTATGCCTGCC
D76C-Forward	CACTGAAAAAtgcGAGTATGCCTGCCG
D76Y-Forward	CACTGAAAAAtatGAGTATGCCTG
D76N-Forward	CACTGAAAAAaacGAGTATGCCTG
D76Q-Forward	CACTGAAAAAcagGAGTATGCCTGC
D76G-Forward	CACTGAAAAAggcGAGTATGCCTGC
D76A-Forward	CACTGAAAAAgcgGAGTATGCCTGC
D76V-Forward	CACTGAAAAAgtgGAGTATGCCTGC
D76L-Forward	CACTGAAAAAattGAGTATGCCTGCCGTG
D76I-Forward	CACTGAAAAAattGAGTATGCCTGC
D76M-Forward	CACTGAAAAAatgGAGTATGCCTGCCGTG
D76F-Forward	CACTGAAAAAtttGAGTATGCCTGC
D76W-Forward	CACTGAAAAAtggGAGTATGCCTGCG
D76P-Forward	CACTGAAAAAccgGAGTATGCCTGCCGTG
D76X-Reverse	GGGGTGAATTCAGTGTAG

DNA sequencing (GENEWIZ) was performed to verify the site-directed mutagenesis reaction was successful and that the correct mutation was introduced. The plasmid

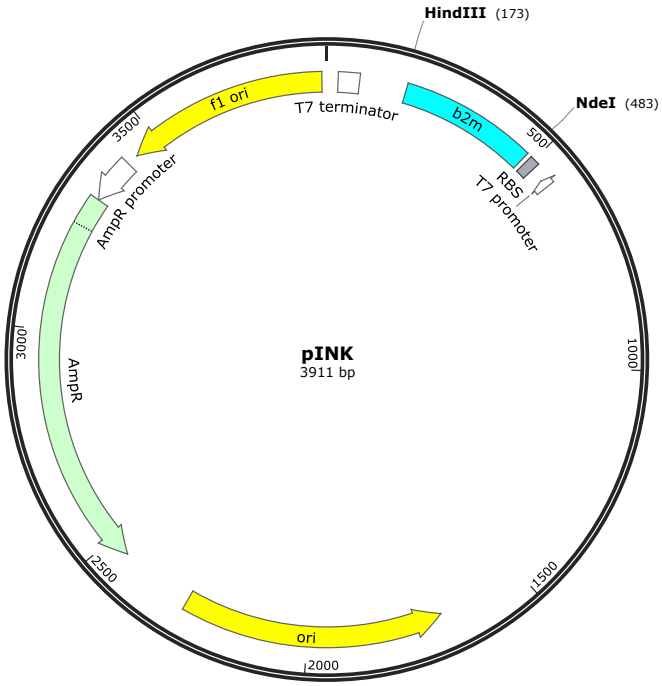
DNA was then isolated from fresh overnight cultures transformed with the correct plasmid using the Wizard Plus SV DNA miniprep kit, used according to the manufacturer's instructions. The plasmids were then stored at -80°C.

2.3 Protein preparation and general protein analysis

2.3.1 Expression plasmid

The D76N β_2m variant was over-expressed in BL21 [DE3] pLysS cells containing a pET-23a plasmid encoding the D76N β_2m gene, mutated from WT β_2m gene (Figure 2.1) with site directed mutagenesis (Methods 2.2.1). Expression plasmids for D76E and D76A were also made using site-directed mutagenesis (Methods 2.2.1). ΔN6 was overexpressed with a similar plasmid (Provided by Dr Theodoros Karamanos) containing the ΔN6 gene rather than the full length β_2m gene, this plasmid was used to generate the ΔN6-D76N expression plasmid, using site-directed mutagenesis (Methods 2.2.1). All primers used are presented in Table 2.1.

a



b



Fig. 2.1: The pINK plasmid, containing the WT β_2m gene.
The pINK plasmid [189] (a), containing the WT β_2m gene which has been cloned into the NdeI/HindIII restriction sites of a pET23a plasmid. The plasmid for both was used for site directed mutagenesis to make D76 mutants. The ΔN6 expression construct is the same, but contains the ΔN6 gene rather than the full length β_2m gene. The plasmid contains a Ampicillin resistance selection marker, and expression of β_2m is under the control of the T7 promoter. The sequence of WT β_2 is shown in (b).

2.3.2 Bacterial culture media

LB media (Melford) was used for growth of unlabelled (^{14}N, ^{12}C) protein. 25 g of granulated LB was used per litre of culture, each 25 g of LB contained 10 g bacto-tryptone, 5 g yeast extract and 10 g NaCl. HCDMI minimal media was used for the growth of isotopically labelled protein. HCDMI media was made by dissolving the components in Table 2.2 up to 1 L in H_2O, followed by the addition of the components in Table 2.3. After autoclaving at 120°C at 15 p.s.i. for 20 min, the appropriate filter sterilised antibiotics were then added after media had cooled to room temperature.

Table 2.2: Components made up to 1 L to make HCDMI media prior to autoclaving. (*: isotopically labelled as appropriate).

K_2HPO_4	10 g
KH_2PO_4	10 g
Na_2HPO_4	7.5 g
K_2SO_4	9 g
*NH_4Cl	1 g

Table 2.3: Components added to 1 L HCDMI media after autoclaving. (*: isotopically labelled as appropriate)

1 M $MgCl_2$	2 mL
1 M $CaCl_2$ (autoclaved)	100 µL
*20% *(w/v)* glucose	20 mL

2.3.3 Expression and purification

Starter cultures were inoculated from glycerol stocks of BL21 [DE3] plysS *E. coli* cells carrying the D76N, WT or ΔN6 β_2m plasmid into 100 ml LB supplemented with 50 µg/mL carbenicillin and 50 µg/mL chloramphenicol. The starter cultures were incubated overnight with aeration (200 rpm, 37°C).

Large scale cultures were set up (1 L in 2 L flasks for ^{14}N, 500 ml in 2 L for ^{15}N) with 50 µg/mL carbenicillin and inoculated with 5 mL starter culture. Cultures were incubated (200 rpm, 37°C) to an OD_{600} of 0.6, followed by induction of protein expression with

1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were incubated overnight (200 rpm, 37°C).

Cells were harvested the following morning (16 hrs post-induction) by centrifugation at 15000 rpm, 20 min (Heroaus continual action centrifuge). The lysis of the cell pellet was initiated by the addition of 200 mL lysis buffer (100 μg/ml hen egg white lysozyme, 20 μg/ml DNAse 1, 50 μg/ml phenylmethanesulfonyl fluoride (PMSF) and 1 mM EDTA in 25 mM Tris.HCl pH 8.0). After incubation at room temperature for 30 min, cells were further lysed with a constant cell disrupter system (Constant systems) at 30 kp.s.i.

The cell lysate was centrifuged at 15000 rpm for 30 min at 4 °C (Beckman JLA 16.250 rotor). The cell pellet containing the inclusion bodies was washed with 25 mM Tris.HCl pH 8.0, three or more times (by re-suspending, centrifuging, discarding supernatant and repeating as above), until the supernatant was clear and there was no dark 'slimy' material remaining in the pellet. The pellet was solubilised in 25 mM Tris.HCl pH 8.0 containing 8 M urea and incubated with gentle stirring overnight at room temperature, followed by centrifugation (as above) for 1 hr. The protein was then refolded by dialysis in 3500 Da molecular weight cut off (MWCO) dialysis tubing (Snakeskin Dialysis Tubing, Thermo Scientific) in 5 L 25 mM Tris.HCl pH 8.0, 4 °C which was changed 5 times after a 3 hour equilibration period per change.

Refolded protein was purified by ion exchange chromatography. Protein was loaded onto a 5 ml HP Q-Sepharose column (GE Healthcare), for small volumes, or a GE Healthcare XK 50 column packed with Q-Sepharose Fast Flow (GE Healthcare), for large volumes. The column was pre-equilibrated with 2 column volumes of 25 mM Tris.HCl pH 8.0, 25 °C, 5 ml/min. The bound protein was eluted with a gradient of 0-125 mM NaCl in 25 mM Tris.HCl pH 8.0. Eluted fractions were concentrated with a Sartorious vivaspin 2, 3000 Da MWCO, PES spin concentrator for small volumes, for large volumes of eluted protein, the sample was dialysed into water to remove buffer salts, followed by freeze drying, before resuspension into a smaller volume of 10 mM sodium phosphate buffer, pH 7.4, for gel filtration.

A size exclusion chromatography step was conducted with filtered (0.2 μM filter Fischer Scientific) ion exchange product, using a Superdex-75 26/600 Prep column (Amer-

sham Biosciences). The peak was collected, concentrated then samples were aliquoted and stored at -80°C. Alternatively, the monomeric material was dialysed into in 3500 Da MWCO dialysis tubing (Snakeskin Dialysis Tubing, Thermo Scientific) into H_2O, 4°C which was changed 5 times after a 3 hour equilibration period per change to remove buffer salts and subsequently freeze dried before storage at -80°C.

Molecular weight calibrants are shown on a gel filtration trace from a Superdex-75 26/600 Prep column (Amersham Biosciences) in Figure 2.2.

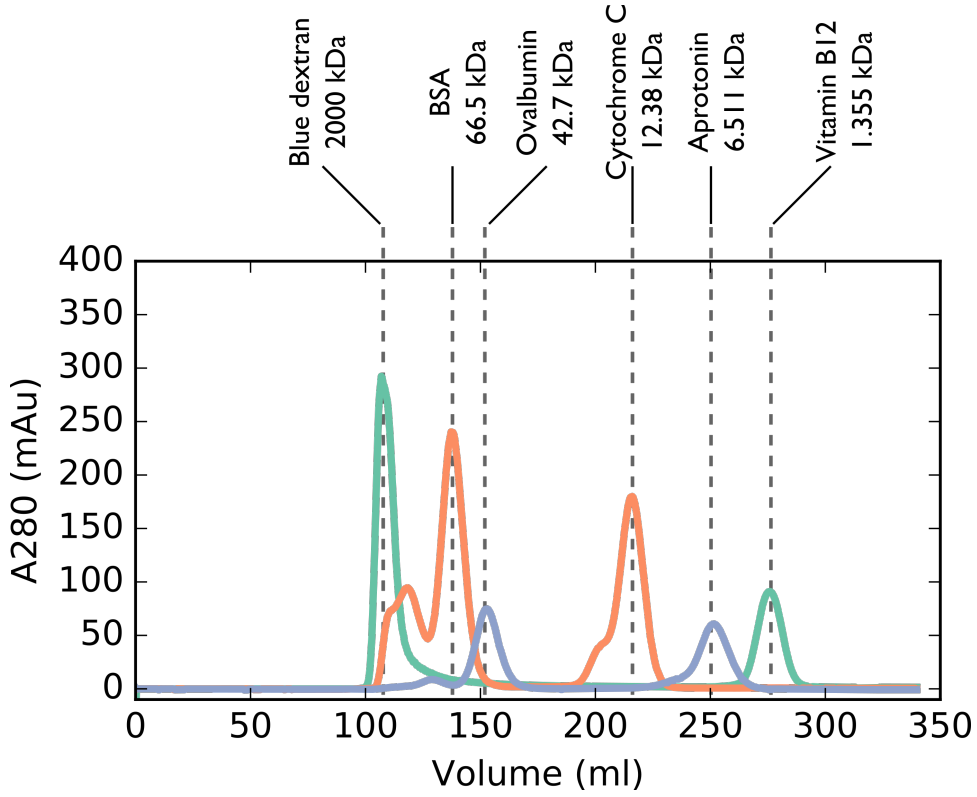


Fig. 2.2: Molecular weight calibrants gel filtration.
Size exclusion chromatography profiles for molecular weight calibrants, molecular weights are shown in the labels above each peak. Data kindly provided by Atenas Posada Borbon.

2.3.4 SDS-PAGE

A Tris-tricine buffered SDS-PAGE gel was used for resolution of approximately 12 kDa β_2m. The composition of the resolving gel and stacking gel are described in Table 2.4.

Table 2.4: 30% Tris-tricine buffered SDS-PAGE gel components

Solution component	Resolving gel: Volume to add (mL)	Stacking gel: Volume to add (mL)
30% Acrylamide:0.8% (w/v) bis-acrylamide	7.5	0.83
3 M Tris.HCl, 0.3% (w/v) SDS pH 8.45	5.0	1.55
H_2O	0.44	3.72
Glycerol	2.0	-
10% (w/v) ammonium persulphate	0.05	0.10
Tetramethylethylenediamine	0.005	0.10

Samples for SDS-PAGE analysis were diluted with an equal volume of 2x loading buffer (50 mM Tris.HCl pH 6.8, 100 mM dithiothreitol (DTT), 2% *(w/v)* SDS, 0.1% *(w/v)* bromophenol blue, 10% *(v/v)* glycerol. Samples were boiled for 2 minutes and loaded into the wells in the gel. The gel was run with a constant current of 60 mA until samples entered the resolving gel, then the current was increased to 80 mA. Gels were stained with Instant Blue (Expedeon protein solutions) overnight, prior to digital imaging (Syngene InGenius gel system).

2.3.5 Mass spectrometry

Electrospray ionisation mass spectrometry (ESI-MS) measurements were performed on a Synapt G1 Q-TOF HDMS (Waters corporation, Wilmslow, Manchester, UK) equipped with a Nanomate nESI autosampling device (Advion Biosciences, Ithica, New York, USA). Time of flight (TOF) calibration was performed using 5 mg/ml CsI. Samples were buffer exchanged using 7000 Da MWCO Zeba spin desalting columns (Thermo scientific, Hemel Hempstead, UK) before 30 minutes drop dialysis on a 0.05 µM nitrocellulose dialysis membrane (Millipore). 50 µL of sample was placed on the membrane, in a petri dish filled with 150 mM ammonium acetate pH 7.4. The final sample was diluted to a protein concentration of 10 µM. Data were analysed using Masslynx v4.1, Driftscope v3.0, and PULSAR v1.0.

Instrument conditions are summarised in Table 2.5. To generate the nESI, 1 kV potential was applied to the nanospray chip in the auto sampling device, assisted by a 0.7 psi nanoflow pressure.

Mass spectrometry was carried out by Owen Cornwell.

Table 2.5: Instrument conditions used for mass spectrometry analysis of purified β_2m

Instrument Parameter	**Conditions used**
Cone voltage (V)	70
Backing Pressure (mbar)	2.1
IMS T-wave velocity (ms^{-1})	300
IMS T-wave height (V)	4-10V ramp (100% cycle)
Trap release time (μs)	100
Quadrupole profile	AUTO
m/z acquisition range	500-8000
Trap CE (V)	5
Transfer T-wave velocity (ms^{-1})	248
Transfer T-wave height (V)	3

2.3.6 Estimation of protein concentration using UV absorption at 280 nm

Protein solutions were diluted to give absorbance values at 280 nm below 1. Beer-Lambert's law was used to calculate protein concentration based upon the extinction coefficient 20065 $M^{-1}cm^{-1}$ for all β_2m variants. Extinction coefficients were calculated with the ProtParam tool included on the ExPASy Server [190].

2.4 Circular dichroism spectroscopy

2.4.1 Background

Circular dichroism (CD) spectroscopy measures the difference in absorbance between left-handed and right-handed circularly polarised light components of plane polarised light. The difference in absorbance between the two components is typically reported as ellipticity (θ) in degrees [191]. The differential absorbance of the two components of circularly polarised light occurs in chiral chromophores. The chromophores in proteins include the peptide bond (for the spectral region below 240 nm) aromatic side chains (260-320 mm) and disulphide bonds (260 nm but weak and broad) [191].

In the peptide bond region, which is the region used in this work, absorption is due to two major transitions. The first transition is centred around 220 nm, and due to an $\mathrm{n} \rightarrow \pi*$ transition. The second, most intense transition is due to a $\pi \rightarrow \pi*$ transition. There can also be contribution to the signal in this range from aromatic side chains [191]. The resulting CD spectra are influenced by the secondary structure in the protein studied, characteristic CD spectra representing different secondary structures are shown in Figure 2.3.

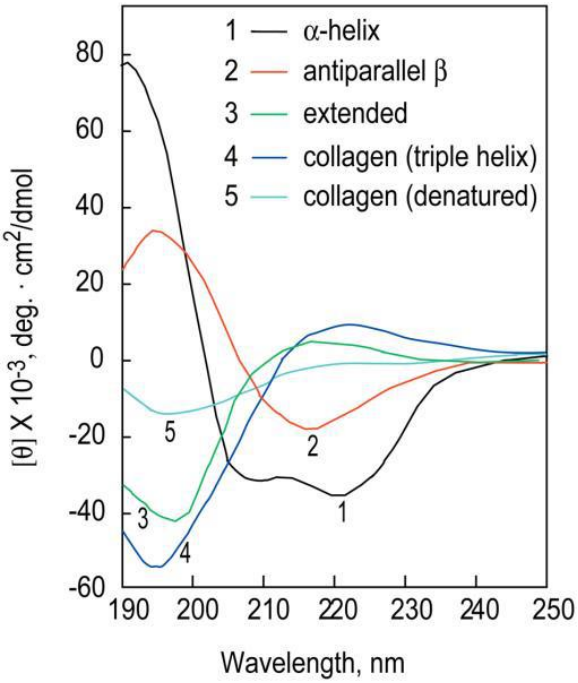


Fig. 2.3: Examples of different CD spectra produced by various secondary structure elements.
CD spectra of poly-L-lysine in α-helical (1, black) antiparallel β-sheet (2, red) and an extended conformation (3, green). Placental collagen is in its triple-helical (4, blue) and denatured (5, cyan) conformations. Figure taken from [192].

All CD experiments were performed on an Applied PhotoPhysics Chirascan spectropolarimeter.

2.4.2 Thermal denaturation, monitored by CD

For thermal denaturation experiments, an initial spectrum at 25°C was recorded, then the temperature was reduced to 20°C and increased in 5°C steps with a setting time of 120 s at each temperature up to 90°C. At the end of the temperature ramp experiment an additional spectrum was recorded 25°C. Each spectrum was acquired in the far-UV range (190-260 nm) with a with a step size of 1 nm with 1 s-per-point sampling. Two repeats were acquired for each temperature. The path length used was 1 mm on samples containing 20 μM protein in 25 mM sodium phosphate buffer at the required pH.

The data was converted into mean residue ellipticity (MRE) from measured ellipticity (m°) with Equation 2.1.

$$\mathrm{MRE} = \mathrm{m^{\circ}.M/10.L.C} \tag{2.1}$$

Where MRE is in deg.cm^2/dmol. C is concentration (g/L), M is average molecular weight (g/mol) and L is path length of cell (cm).

The MRE for each timepoint was plotted as a function of temperature and fit to a 2-state model (Equation 2.2) with the software package CDPal [193].

$$\mathrm{E} = \mathrm{e}^{-\frac{\Delta \mathrm{H_m}}{\mathrm{R}}\left(\frac{1}{\mathrm{T_m}}-\frac{1}{\mathrm{T}}\right)+-\frac{\Delta \mathrm{C_p}}{\mathrm{R}}\left(\frac{\mathrm{T_m}}{\mathrm{T}}-1+\ln\left(\frac{\mathrm{T}}{\mathrm{T_m}}\right)\right)} \tag{2.2}$$

Where $\Delta \mathrm{H_m}$ is the change in enthalpy at the denaturation midpoint $\mathrm{T_m}$ and $\Delta \mathrm{C_p}$ is the difference in heat capacity between the two states.

$\Delta \mathrm{C_p}$ is fixed at zero during the fit to a two-state model with the software package CDPal [193] as it is assumed that the difference in heat capacity between the two states is independent of temperature, while the fit is optimised for $\Delta \mathrm{H_m}$ and $\mathrm{T_m}$. Since the thermal denaturation process in not reversible, $\mathrm{T_{m,app}}$ values are quoted.

2.4.3 Real-time refolding, monitored by CD

Protein stocks were dialysed against the unfolding solution (0.8 M urea, pH 2.5) for 1 hr. The unfolded protein was then subjected to CD to confirm the protein was unfolded. To initiate the refolding reaction, refolding buffer (300 mM sodium phosphate, at the desired final pH) was added at a ratio of 2:1 (unfolded:refolding), so that the buffer was 100 mM sodium phosphate at the desired pH. These experiments were carried out at 20°C. The concentration of unfolded protein was 30 µM, which was reduced to 20 µM by the addition of refolding buffer. Data acquisition was initiated immediately after addition of unfolding buffer to the cuvette containing the unfolded protein. The CD experiments

were set up to acquire a spectrum per minute, ranging from 200-260 nm with a step size of 1 nm, with a sampling time of 0.5 s per point.

Refolding rate constants were calculated by fitting the far-UV signal (in MRE) at specific wavelengths as a function of time. The equations used are specified appropriately in the corresponding results chapter (Equations 4.2, 4.1 and 4.3, in Chapter 4).

2.5 Equilibrium unfolding experiments

250 mL urea stock solutions of 25 mM sodium phosphate buffer, pH 7.4 containing 0 M and 10.5 M urea were made in 250 mL volumetric flasks. The exact urea concentration was calculated using the Celti refractometer and equation 2.3.

$$117.66 \cdot (\Delta \mathrm{N}) + 29.753 \cdot (\Delta \mathrm{N})^2 + 185.56 \cdot (\Delta \mathrm{N})^3 \tag{2.3}$$

Where ΔN is the difference in refractive index between the 0 M buffer and the 10.5 M urea buffers.

Stock solutions containing 4 μM protein were made at urea concentrations of 0, 2, 4, 6, 8, and 10 M urea. The stock solutions were combined to make 1 mL solutions for every 0.2 M urea increment up to 10 M. The solutions were incubated overnight at 25°C before measurement of tryptophan fluorescence on a PTI Quantamaster C-61 spectrofluorimeter. The samples were excited at 280 nm and the emission monitored at 325 nm over 60 seconds. The signal over 60 seconds was then averaged and normalised to the signal of the 10 M sample and then corrected for the urea dependence of tryptophan fluorescence using Equation 2.4.

$$\mathrm{y} = \frac{(\mathrm{cx} + \mathrm{d}) + \mathrm{e}^{\frac{\mathrm{G-mx}}{\mathrm{RT}}} . (\mathrm{ax} + \mathrm{b})}{\mathrm{e}^{\frac{\mathrm{G-mx}}{\mathrm{RT}}} + 1} \tag{2.4}$$

Where G is ΔG°_{UN}, kJ.mol^{-1}, m is the m-value, kJ.mol^{-1}.M^{-1}. R is the gas constant, 8.314 J.K^{-1}mol^{-1}, T is temperature in Kelvin (298K). $(\mathrm{ax} + \mathrm{b})$ and $(\mathrm{cx} + \mathrm{d})$ correspond to the pre and post transition baselines, respectively. Midpoint urea values are calculated by solving the equation where $\Delta G^{\circ}_{UN} = 0$.

Equation 2.4 was then used to fit the normalised and corrected data to extract values for ΔG°_{UN} and the m-value.

2.6 Aggregation assays

2.6.1 *In vitro* fibrillation assays

For fibrillation assays protein was either stored as lyophilised powder or as a concentrated solution at -80˚C. Lyophilised protein was dissolved in sterile water, filtered (0.22 μm, Millipore) and diluted to the desired concentration in the appropriate buffer. Stock solutions of protein were centrifuged at 14000 xg for 10 min followed by filtration (0.22 μm, Millipore) before dilution into the appropriate buffer. NaCl was added from a 2 M stock to make the desired ionic strength for the experiment. 10 μM Thioflavin-T (ThT) was added to all experiments. Each condition was repeated five times in parallel from the same stock which was divided into 100 μL aliquots per well in a Corning 96-well polystyrene microtitre plate. Plates were sealed with clear polyolefin sealing film (STAR-LAB) and incubated at 37 ˚C for at least 48 hrs with constant shaking at 600 rpm. ThT fluorescence was monitored (excitation 440 nm, emission 480 nm) with a Fluorostar Optima, BMG Labtech plate reader.

T_{half} values were calculated by fitting normalised data (between 0-1) for each replicate to a generalised logistic function (Equation 2.5) and calculating the value of time at the midpoint of the curve. This analysis was carried out using python (Python Software Foundation, https://www.python.org/) scripts kindly provided by Emma Cawood.

$$\mathrm{Y(t)} = \mathrm{A} + \frac{\mathrm{K} - \mathrm{A}}{\left(1 + \mathrm{Q}\mathrm{e}^{-\mathrm{B}\left(\mathrm{t} - \mathrm{M}\right)\frac{1}{\mathrm{v}}}\right)} \tag{2.5}$$

Where A is the pre-transition baseline (lower asymptote), K is the post-transition baseline (upper asymptote), B is the growth rate and M is the time of maximal growth. Q and v are parameters which affect the transitions from and to the growth phase. Y is the normalised signal and t is time [194, 195].

2.6.2 Electron microscopy

Carbon coated copper EM grids were placed coated-side down onto sample drops containing undiluted material from the *in vitro* fibrillation assay for 30 seconds. The grids were then blotted with filter paper to remove excess solvent and the sample. Grids were then placed onto drops of 4% *(w/v)* uranyl acetate for 30 sec to stain. Grids were then blotted again and air-dried before analysis.

2.6.3 Fibril yield assay

Fibril yield assays were carried out by resuspending all material from an *in vitro* fibrillation assay condition and setting aside a 'whole' sample from the experiment. Next the rest of the sample was centrifuged on a bench top centrifuge for 30 minutes at 14000 rpm, the supernatant was then extracted for the 'soluble' fraction. Finally, the 'whole' and 'soluble' samples were analysed with SDS-PAGE (Methods 2.3.4) to assess how much protein sample remains soluble, compared to the 'whole' fraction which contains both soluble and insoluble material.

2.7 Nuclear Magnetic Resonance spectroscopy

2.7.1 Data processing

All raw NMR data were processed with the NMRPipe software package before further analysis [196]. Further analysis was carried out with CCPNmr Analysis [197], PINT [198], nmrglue [199] and the Python programming language (Python Software Foundation, https://www.python.org/).

2.7.2 Backbone Assignment

The sample for 3D experiments contained: 325 μM D76N (^{1}H, ^{15}N, ^{13}C), 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3 in 10 mM sodium phosphate buffer pH 7.4. HSQC, BEST-HNCA, BEST-HNCOCA, BEST-HNCO, BEST-HNCOCACB [200] and HBHACONH [201] experiments were carried out at 750 MHz with an HCN triple resonance cryoprobe and VARIAN INOVA console.

Data were processed with NMRPipe [196] and sequential assignments were determined through analysis of the BEST-HNCA, BEST-HNCOCA experiments, using the software package CCPNmr Analysis [197].

2.7.3 Chemical shift perturbation (CSP) analysis

To calculate amide chemical shift perturbations, peaks were picked with CCPNmr Analysis software [197]. The difference in peak position between two different conditions were calculated with Equation 2.6 [64].

$$\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2} \tag{2.6}$$

Where δ^{1}H and δ^{15}N are the observed change in the proton and nitrogen dimensions, respectively.

2.7.4 Real-time refolding experiments monitored using NMR

Protein stocks were dialysed against the unfolding solution (1.5 M urea, pH 2.5, 10% *(v/v)* D_2O) for 1 hr. To initiate the refolding reaction, 150 µL refolding buffer (500 mM sodium phosphate, pH 7.4) was added to 350 µL of unfolded protein so that the final buffer was 166.6 mM sodium phosphate at pH 7.4. These experiments were carried out at 20°C. The concentration of unfolded protein was 450 µM, which was reduced to 300 µM by the addition of refolding buffer.

Data acquisition began after a period of 30 s due to the time required to insert the NMR tube into the spectrometer after the manual addition of refolding buffer. To follow the folding reaction by NMR, ^{1}H-^{15}N SOFAST-HMQC spectra were recorded every 60 s, with 100 points in f1 (^{15}N) and 956 in f2 (^{1}H), and 2 scans per spectrum. With a spectral width of 15.97 ppm in f2 and 22.00 ppm in f1 and recorded with a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

Spectra were processed in NMRPipe [196] and all spectra were analysed with the software package PINT [198]. All spectra were loaded into PINT and for each spectrum, the peak volumes were calculated by fitting peaks to a lorentzian lineshape. For each residue, the total peak volume was plotted as a function of time and was fitted to the equations (Equations 4.2 and 4.1) for exponential decay, detailed in Chapter 4, Section 4.2.2.

To generate the subtracted I_T-state only spectrum (Chapter 4, Section 4.2.2), the N-state spectrum was scaled and subtracted from the refolding spectrum using NMRPipe [196], 'addNMR' functionality.

2.7.5 HDX-NMR

Hydrogen exchange experiments were conducted by freeze drying a sample containing 150 µM protein, 25 mM sodium phosphate, at pH 7.4. This sample was then resuspended in 100% D_2O with 38.3 mM NaCl and 0.1 mM DSS. Data acquisition was initiated immediately after resuspension with a dead time of approximately 3 min. ^{1}H-^{15}N-SOFAST-HMQC spectra were taken every 5 minutes for 500 spectra (42 hours total) then taken approximately every 24 hours until there was no observable signal. The peak

intensity for each observable peak was plotted as a function of time, then fitted to an exponential decay function (Equation 2.7) to calculate rates of decay. Peak intensities were calculated with CCPNmr Analysis [197]

Spectra were acquired with 242 points in the indirect dimension, 956 in the direct dimension and with 4 scans, depending on sample concentration. With a spectral width of 15.97 ppm in f2 (^{1}H) and 25.00 ppm in f1 (^{15}N). Data were acquired with a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

$$\mathrm{y} = \mathrm{a} \cdot \mathrm{e}^{(-\mathrm{bx})} + \mathrm{c} \tag{2.7}$$

Where y is the peak intensity/volume, a is the initial peak intensity/volume, b is the rate of decay, and c is the final peak intensity/volume.

Protection factors were calculated with Equation 2.8.

$$\mathrm{P_i} = \frac{\mathrm{k_{int}}}{\mathrm{k_{ex,(obs)}}} \tag{2.8}$$

Where P_i is the protection factor, k_{int} is the intrinsic rate of hydrogen exchange (calculated with SPHERE [202]) and $k_{ex,(obs)}$ is the observed rate of hydrogen exchange.

2.7.6 ^{15}N transverse relaxation rate determination

T_2 relaxation experiments were carried out at 600 MHz and at 750 MHz, with relaxation delays ranging from 16 ms to 160 ms over 11 increments (with repeated points for error estimation). All experiments were carried out with 256 points in the indirect dimension, 1024 in the direct dimension and with 8 or 16 scans, depending on sample concentration.

For each residue, peak volumes were calculated with the PINT [198] software package and plotted as a function of relaxation delay, which were then fitted to an exponential decay function to extract rates of decay. Errors bars for the R_2 relaxation rates are the uncertainty in the fit for that residue.

Experiments at pH 7.4 were carried out in 25 mM sodium phospate buffer pH 7.4, 90.8 mM NaCl, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3. Experiments at pH 6.2 were carried out at in 25 mM sodium phospate buffer pH 6.2, 120 mM NaCl, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3. NaCl was added to make equal ionic strengths of 155 mM.

2.7.7 ^{15}N relaxation dispersion experiments

For CPMG-RD experiments, a fixed relaxation delay of 40 ms (at 600 MHz and at 750 MHz) was used with varying refocussing frequencies ranging from 100 to 2000 Hz (with repeated frequencies for error estimation). All experiments were carried out with 200 points in the indirect dimension (^{15}N), 2048 points in the direct dimension (^{1}H) and with 32 scans.

Experiments at 600 MHz were carried out using a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console. Experiments at 750 MHz were carried out using a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and Bruker Avance III HD console.

Experiments at pH 7.4 were carried out in 25 mM sodium phospate buffer pH 7.4, 90.8 mM NaCl, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3. Experiments at pH 6.2 were carried out at in 25 mM sodium phospate buffer pH 6.2, 120 mM NaCl, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3. NaCl was added to make equal ionic strengths of 155 mM.

$R_{2,eff}$ values were calculated with Equation 2.9. The $R_{2,eff}$ values were then plotted against the frequency of the CPMG pulse train for each increment and fitted to the Bloch-McConnell equations for two-site exchange using the software package PINT [198]. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test [198].

$$\mathrm{R_{2,\,eff}} = \frac{1}{\mathrm{T}} \ln \frac{\mathrm{I_0}}{\mathrm{I_x}} \tag{2.9}$$

Where I_x is the peak volume for a given CPMG frequency (ν_{CPMG}), I_0 is the peak volume in a reference spectrum where no CPMG pulse train is applied. T is the relaxation delay of 40 ms, used for all incrememnts apart from the reference spectrum where T = 0 [203].

2.7.8 Measurement of residual dipolar couplings

RDCs were measured using sensitivity-enhanced in-phase and anti-phase (IPAP) J-coupled ^{1}H-^{15}N-HSQC spectra [69, 204]. Alignment was achieved by addition of 15mg/ml Pf1 filamentous phage (ASLA Biotech.) to 300 μM D76N, 90.3 mM NaCl, 5 % D2O, 0.02% NaN_3 samples. Identical spectra were also recorded for samples without any alignment media added.

IPAP spectra were recorded with 1280 points in the indirect dimension (^{15}N) to ensure high resolution peak positions could be measured with 1024 points acquired in the direct dimension (^{1}H) and 32 scans. Experiments at 750 MHz were carried out using a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and Bruker Avance III HD console.

Peaks were fitted with the software package PINT to obtain the peak positions and the difference in splitting (dipolar coupling) between the aligned and unaligned spectra for each residue were measured. This difference is the RDC.

2.8 Split-β-lactamase assay

The Split-β-lactamase assay assay was carried out as previously described [205]. The appropriate plasmid was transformed into *E. coli* BL21 [DE3] cells which were incubated overnight at 37°C on a tetracycline (10 μg/mL) selection plate (LB-Agar). 100 mL of sterile LB (10 μg/mL tetracycline) was inoculated from a single colony and incubated at 37°C, 200 rpm overnight. 1 mL of the overnight culture was used to inoculate 100 mL sterile LB (10 μg/ml tetracycline) and incubated at 37°C, 200 rpm to an OD_{600} of 0.6. Arabinose (final concentration 0.02% *(w/v)*) was then used to induce the expression of the construct, followed by a further 1 hr incubation (37°C, 200 rpm). The culture was then serially diluted in tenfold increments into 170 mM NaCl solution. The diluted cultures were then applied to 48 well agar plates containing 10 μg/ml tetracycline and 0.02% *(w/v)* arabinose and increasing concentrations of ampicillin (0-140 μg/ml). Plates were then incubated overnight (18 hrs) and scored based on visual identification of colonies (more than 10 individual colonies or lawn required to be scored). Where the maximal cell dilution at which colonies could be observed MCD_{GROWTH} was recorded for each ampicillin concentration . The layout of the plate used for the assay and the scoring procedure are shown in Figure 2.4.

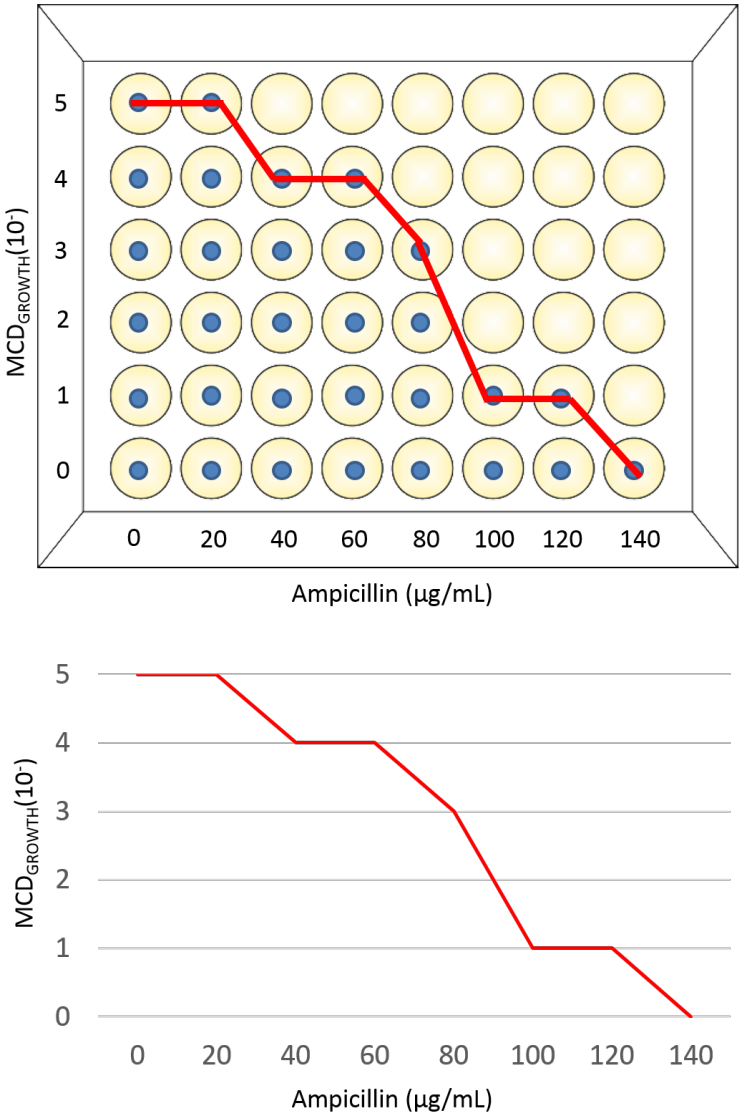


Fig. 2.4: Example schematic of the 48 well plate used for MIC assay.
Blue circles correspond to lawns of bacterial growth. Serial dilutions are go from undiluted culture $MCD_{GROWTH}(10^{-0})$ up to $MCD_{GROWTH}(10^{-5})$. The serial dilutions are incubated at ampicillin concentrations ranging from 0 µg/mL to 140 µg/mL. The results are plotted as an antibiotic survival curve of the MCD_{GROWTH} after 18 h incubation at 37°C. Blue dot represents bacterial growth.

The area under curve value (AUC) was calculated with Equation 2.10

$$\mathrm{AUC} = \sum_{i+1}^{7} \frac{y_i + Y_{i+1}}{2} \times (x_{i+1} - x_i) \tag{2.10}$$

Where x_i and y_i are the x and y axis values at each ampicillin concentration

The constructs used for saturation mutagenesis were generated by site-directed mutagenesis (Methods 2.2.1). The tripartite fusion construct plasmid map is shown in Figure 2.5

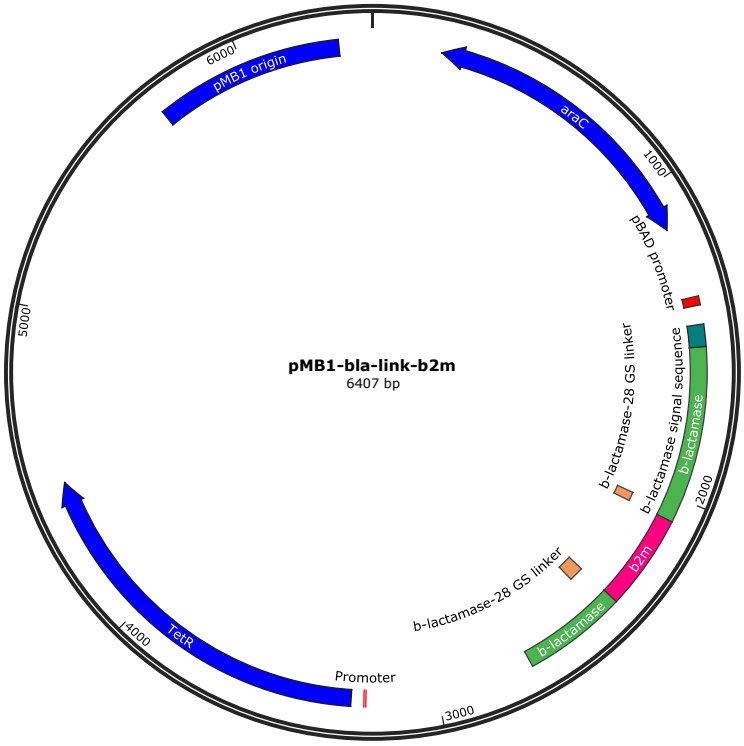


Fig. 2.5: Plasmid map of the tripartite fusion construct used in the split-β-lactamase assay.
The β_2m sequence is inserted inbetween the two β-lactamase domains, connected by 28 residue glycine-serine linkers. Under control of the pBAD promoter, inducible with arabinose and under tetracycline antibiotic selection. The WT plasmid [205] was used to make all D76 mutants using site-directed mutagenesis.

2.9 Metadynamic-metainference molecular dynamics simulations

2.9.1 Experimental setup

All simulations were carried out using GROMACS [206] version 2016.4 patched with plumed [207, 208] version 2.5-dev. The gromacs 'pdb2gmx' tool was used to convert the D76N crystal structure (PDB:4FXL, [183]) into a coordinate file, add hydrogens and impose both the AMBER99SB*-ILDN forcefield [209] and TIP3P water model [210]. Next, energy minimisation *in vacuo* was carried out. The system was then placed into a dodecahedral box with the gromacs 'editconf' tool and solvated using the gromacs 'solvate' tool with 7713 water molecules, After energy minimisation in water an NPT equilibration was carried out, followed by an NVT equilibration. The NVT ensemble was used hereafter with a time-step of 2 fs. LINCS constraints [211] were used with lincs-iter=2 and lincs-order=6 settings (these settings are important due to the increased strain induced by restraints such as backbone chemical shifts). The Verlet cutoff-scheme was used for neighbour searching, electrostatics were treated with a Particle-Mesh Ewald method, with a 0.9 nm cut-off and Van der Waals interactions were also given a 0.9 nm cut-off. Temperature coupling was carried out with the v-rescale method at 300 K.

The equilibrated system was then split into 24 replicas for the metadynamic-metainference run.

Before any metadynamics was applied, the system was pre-equilibrated with the RDC data in order to calculate the scaling factor required to simplify the restraining potential for the RDC restraint. The scaling factor is the slope of the linear fit between the experimental data and the predicted RDCs [67] over the ensemble when the correlation between the experimental and predicted RDC is close to 1. This was achieved by first calculating the average RDC across the replicas for each residue. Next the correlation between the experimental RDCs and the calculated RDCs from the simulation are calculated. Then, a restraint was applied to to the correlation to bring it close to 1. Finally, once the correlation was close to 1, the slope of the linear fit between the experimental and calculated RDCs was calculated, this is the scaling factor. In this case, the correlation

plateaued at approximately 0.9, and the scaling factor at this point was approximately 0.007. The 24 replicas, now pre-equilibrated with the RDC data were used for the full metainference, metadynamic simulations.

Metainference [32, 33] was applied using backbone chemical shifts (CS)(Cα, Cβ, CO, Hα, HN, NH) and N-H RDCs using Camshift [212] to model backbone chemicals shifts and the θ method [67] to model RDCs. Metainference was carried out using an 'outliers' noise type, averaged over 500 steps. The averaged calculated CS and RDC values were then compared to the experimental values (taking errors from the metainference into account) to generate the energy penalties based on agreement with the experimental data, which are added to the parallel-bias metadynamics potential [213].

Parallel-bias metadynamics [213] was applied to the 24 replica simulation based on the Collective Variables (CVs) presented in Table 5.2. Gaussians of height 1.2 were deposited at a pace of 500 with a biasfactor of 30, using multiple walkers [214]. The sigma values (widths) of the metadynamics gaussians, and the grid max and min values are quoted in Table 2.6. The simulation ran for 175 ns per replica.

Table 2.6: Metadynamics settings used for each CV

Collective variable	sigma (width)	Grid min	Grid max
N-term	1.0	0	10
Pro32	0.15	π	-π
AB-loop	1.5	0	19
Distance76-42	0.03	0	5
Distance76-73	0.03	0	5
D-strand-APBR	0.25	0	20
Pro32:N-term	0.5	0	30
Pro32:DE-loop	0.5	0	30

Each of the 24 replicas in the simulation ran for 175 ns

2.9.2 Generation of free energy surfaces

Predicted free energy surfaces were generated using the plumed [207, 208] sum_hills tool and python scripts provided by Dr Massimiliano Bonomi (Cambridge, UK). The free

energy surfaces are generated by converting the parallel bias metadynamics energy potential for each timestep into a population weight using Equation 2.11. Next a grid is made which splits up the CV space into 100 equal parts. Then for each value of a particular CV, the population weight is added to the corresponding grid space (for example proline 32 isomerisation can be split into 100 π/100 section between -π and π.). This can be done on one CV only, or can be scaled up to 2 CVs, making a 100 by 100 grid. Finally the weights for each histogram (space in the grid) can be converted back into free energy with the relationship in Equation 2.12. In all cases the minimum value of energy is set to zero and all other energies are quoted in reference to the minimum value.

$$\mathrm{w} = \mathrm{e}^{((\mathrm{B}-\mathrm{B}_{\mathrm{max}})/\mathrm{kBT})} \tag{2.11}$$

Where w is the population weight for a particular CV value, B is the parallel bias metadynamics potential, kBt is a product of the Boltzmann constant and temperature (300K) given in kJ.mol^{-1} (2.494339)

$$\mathrm{E} = -\mathrm{kBT} \cdot \ln(\mathrm{w}) \tag{2.12}$$

Where w is the population weight for a particular CV value, B is the parallel bias metadynamics potential, kBt is a product of the Boltzmann constant and temperature (300K) given in kJ.mol^{-1} (2.494339)

The trajectories used for making the monodimensional landscapes were first slit up into 11 parts, to make 11 different free energy surfaces which could be compared as a qualitative check of convergence. For instance, if the energy landscape stops changing in later segments, then the metadynamics for that particular CV is converged since the energy potentials are no longer having an effect on the landscape. The full trajectories were used for making multidimensional landscapes.

2.9.3 Clustering analysis

T carry out the clustering analysis, the simulation trajectory was concatenated into 50000 frames. Next the MDAnalysis python library [215, 216] was used to calculate the backbone RMSD between all pairs of frames in the trajectory. Next the frames were separated into clusters based on a 2Å backbone RMSD cut-off through an iterative process where the frame with the most neighbours (within cut-off distance) was identified, then all neighbours were added to this cluster and removed from the pool. The next cluster was made by again finding the structure with the most neighbours and removing that cluster from the pool and so on. The populations of the individual clusters were generated by adding up the weights calculated in the analysis above, for each frame in the cluster. This clustering analysis was carried out using scripts kindly provided by Dr Massimiliano Bonomi (Cambridge, UK).

Chapter 3

Global properties and native structure of D76N β_2m

3.1 Introduction

The D76N variant of β_2m, which causes a hereditary, late-onset and fatal system amyloid disease in individuals heterozygous for the mutation [183], has an almost identical crystal structure (PDB:4FXL [183]) to the WT β_2m crystal structure (PDB:1LDS [151]), with a root-mean-square deviation (RMSD) of 0.339 Å (Figure 3.1). The D76N β_2m variant is less stable to GdnHCl denaturation than WT β_2m by 11.3 kJ.mol^{-1} [184]. However, for β_2m thermostability does not correlate with aggregation propensity [217]. Murine β_2m (mβ_2m) has a ΔG°_{un} of 12.6 kJ.mol^{-1}, compared to WT Human β_2m, calculated by urea denaturation [60], which is as unstable as the highly amyloidogenic D76N [184], yet mβ_2m does not form amyloid [60]. There are other destabilised Human β_2m variants such as V37A which has a $\Delta\Delta G^{\circ}_{UN}$ of 8.5 kJ.mol^{-1} and does not form amyloid at pH 7.5 and 37°C [177, 217]. In fact, a comprehensive analysis of destabilising β_2m mutations showed that the thermal stability of a β_2m mutant does not correlate with its aggregation propensity [217]. Instead, the equilibrium population of the folding intermediate (known as the I_T-state) correlates with β_2m aggregation in various proline variants of β_2m (such as P5G and P32G) which modulate the equilibrium population of the I_T-state [177].

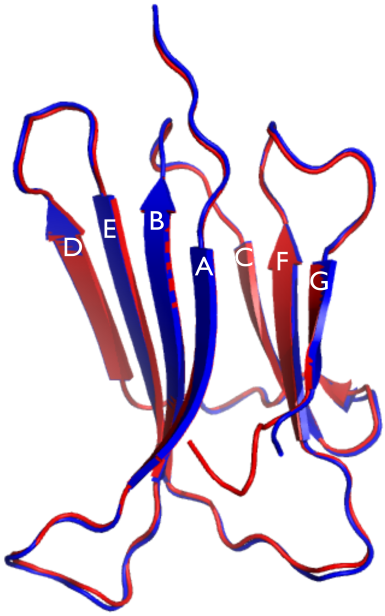


Fig. 3.1: WT and D76N crystal structures.
The WT (blue, PDB:1LDS [151]) and D76N (red, PDB:4FXL [183]) crystal structures share an RMSD of 0.339 Å. The structural alignment was carried out with the PyMol molecular graphics system [119].

Since the decreased thermal stability of D76N is not sufficient to explain its increased aggregation propensity, the folding pathway was investigated with GdnHCl denaturation/refolding experiments by Mangione et al. [184]. These experiments showed that the D76N variant follows a similar folding pathway to WT β_2m, and the authors postulate that the D76N variant populates the partially folded I_T-state at approximately 25% at equilibrium [184]. In contrast, WT β_2m populates the I_T-state to just approximately 5% possibly explaining the increased aggregation propensity of D76N, compared with WT β_2m [184, 177].

In this chapter, D76N β_2m was purified under a range of isotopic labelling regimes to facilitate the study of D76N β_2m by NMR. The stability of the D76N variant was also studied by equilibrium urea denaturation experiments, and through thermal denaturation. The effect of the mutation on the structure, overall fold and the equilibrium I_T-state population of the D76N variant was also studied by CD and at single residue resolution

by NMR. The aggregation of D76N was also examined, in comparison to the I_T-state mimic ΔN6 to further investigate the role of I_T-state in D76N aggregation.

Finally, the specific importance of the residue at position 76 in β_2m was investigated by saturation mutagenesis of position 76, in a split-β-lactamase assay [205, 218], and two new mutants with substitutions at position 76 were purified, and their structure, stability and aggregation were characterised.

3.2 Results

3.2.1 Stability and aggregation of D76N β_2m

Recombinant WT and D76N β_2m were purified under a range of isotopic labelling regimes (^{1}H-^{14}N-^{12}C, ^{1}H-^{15}N-^{12}C and ^{1}H-^{15}N-^{13}C) by the protocol detailed in Methods 2.3. The size exclusion chromatography elution profiles for WT and D76N (Figure 3.2a,b) show a single peak at approximately 200-220 ml. The blue regions in the elution profiles (Figure 3.2a,b) were further analysed by SDS-PAGE (Figure 3.2c, Methods 2.3.4) which shows that the proteins are each a single, pure species of approximately 12 kDa. The exact masses of the purified proteins were measured by denaturing ESI-MS (Methods 2.3.5). The electrospray ionisation mass spectra (ESI-MS) in Figure 3.3 show that the pure proteins are single species with an observed molecular weight of 11860.30 (± 0.19) Da and 11859.38 (± 0.65) Da for WT and D76N, respectively, which is in agreement with the expected masses of 11860.38 Da and 11859.39 Da. The observed molecular masses confirm that for each protein the intramolecular disulphide bond is formed.

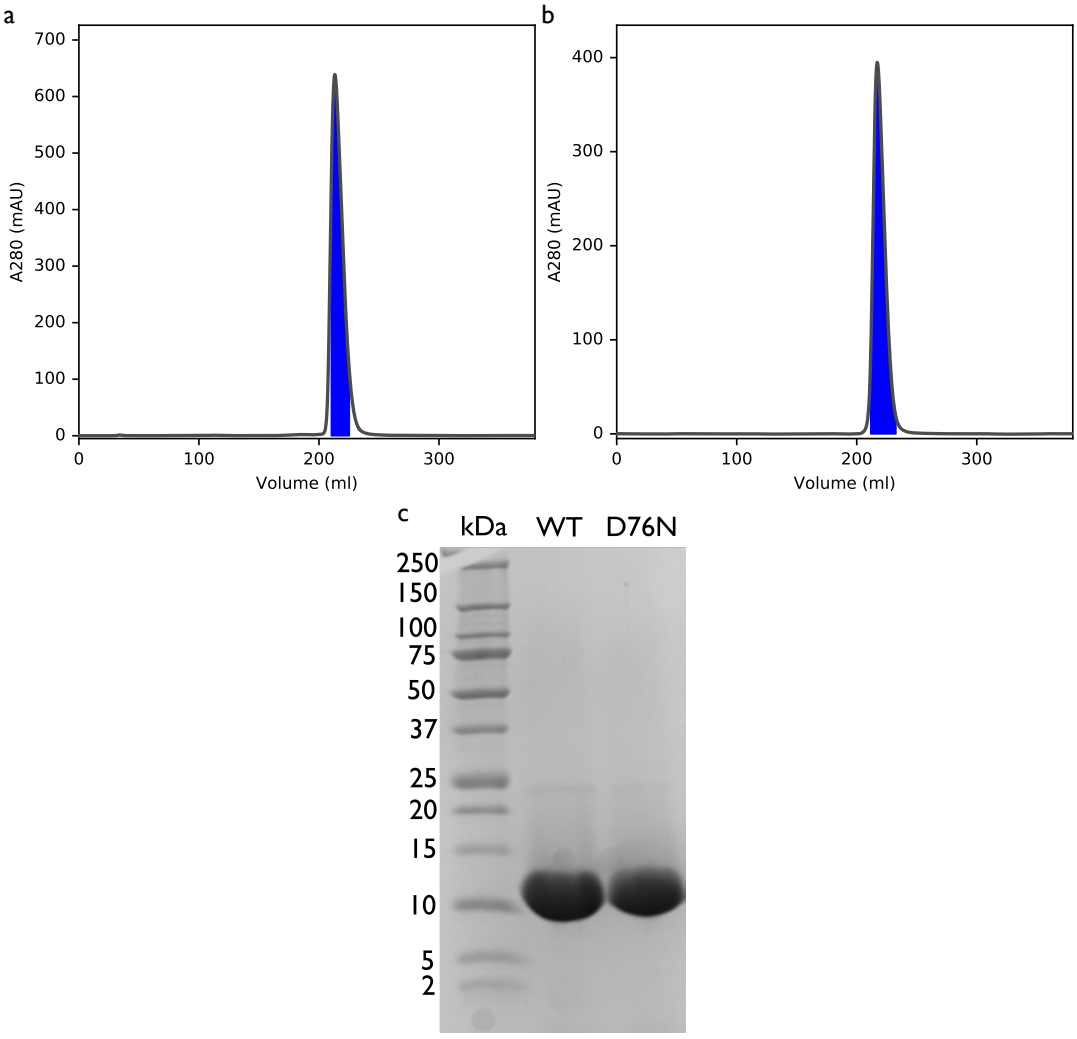


Fig. 3.2: Purification of recombinant D76N and WT protein.
Size exclusion chromatography profiles for WT (a) and D76N β_2m (b) The blue areas correspond to the material taken forward for further experiments and the gel in c, shows the pure samples from a and b as single bands compared to the molecular weight marker (kDa).

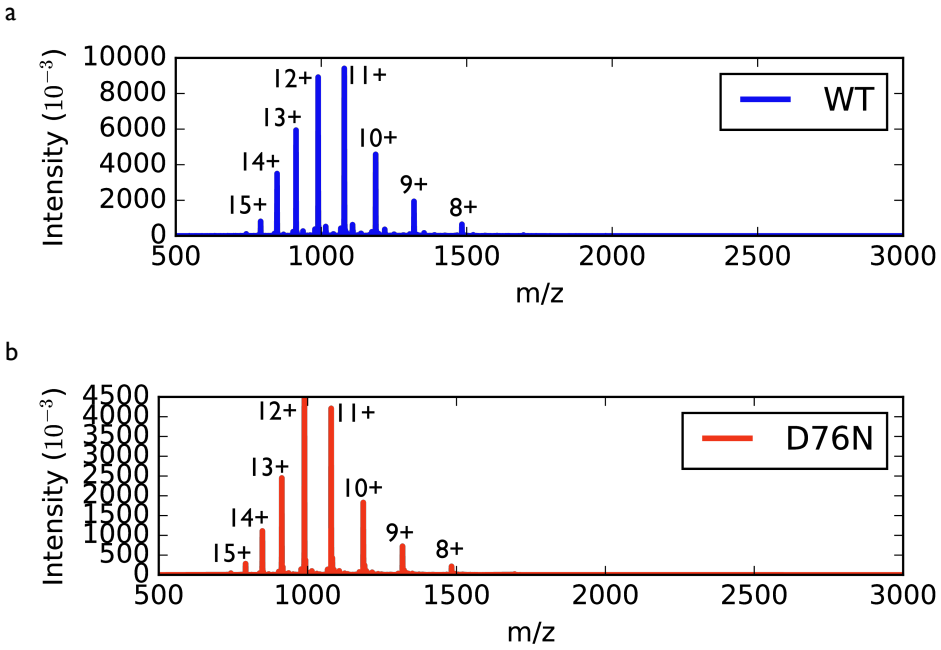


Fig. 3.3: WT and D76N Mass spectra.
Denatured ESI-MS mass spectra of purified WT (a) and D76N (b) for exact mass calculation. The observed masses are 11860.30 (± 0.19) Da and 11859.38 (± 0.65) Da for WT and D76N, respectively.

Far-UV CD spectroscopy was carried out to assess the secondary structure of the purified D76N protein and to verify that the D76N protein was correctly folded (Methods 2.4.2). In Figure 3.4a the far-UV CD spectra (200-260 nm) of WT and D76N show that the proteins share a similar global fold, which is distinct from the ΔN6 spectrum, also shown in Figure 3.4a.

WT β_2m and ΔN6 have similar β-sheet content (36.4 and 36.2%, respectively) and contain no α-helical content, as calculated from the NMR structures (PDB: 2XKS and 2XKU [88], respectively) with DSSP [88, 219, 220]. In fact, WT β_2m has a remarkably small far-UV signal for a protein with 36.4% β-sheet content; this is due to a positive contribution to the far-UV signal from aromatic residues, which does not occur in the ΔN6 structure, due to an alternative packing arrangement of aromatic side chains in ΔN6 [177, 179].

The differences between the WT β_2m and ΔN6 CD spectra can be used as a probe for equilibrium I_T-state population [177]. In Figure 3.4a a simulated 25% I_T-state spectrum (calculated by addition of 75% WT spectrum + 25% ΔN6 spectrum) is compared to

the spectra of WT, D76N and ΔN6. Comparison of the spectra in Figure 3.4a shows that there is no detectable increase in the I_T-state population in D76N compared to WT and that an equilibrium I_T-state population of 25% would be clearly distinguishable from the WT spectrum by the magnitude of the CD minimum at 216 nm. The lack of a distinguishable difference in I_T-state population in the WT and D76N CD spectra point to a disagreement with the results of Mangione et al. [184] who suggested that D76N had an increased I_T-state population of 25% compared to the much lower I_T-state population of 5% observed for WT β_2m.

Far-UV CD was also utilised to probe the thermal stability of D76N, compared to WT, through the execution of a temperature ramp (Figure 3.4b-d, Methods 2.4.2). The CD-signal at 216 nm as a function of temperature was used to calculate $T_{m,app}$ values for WT and D76N, by fitting the data to a 2-state model with the software package CDPal [193]. WT ($T_{m,app}$ = 62.4°C ± 0.4) is approximately 9°C more stable than D76N ($T_{m,app}$ = 53.5°C ± 0.6), consistent with other studies [221, 222, 223]. The thermal denaturation process is non-reversible, presumably due to aggregation, which is visible in the cuvette after the assay has finished.

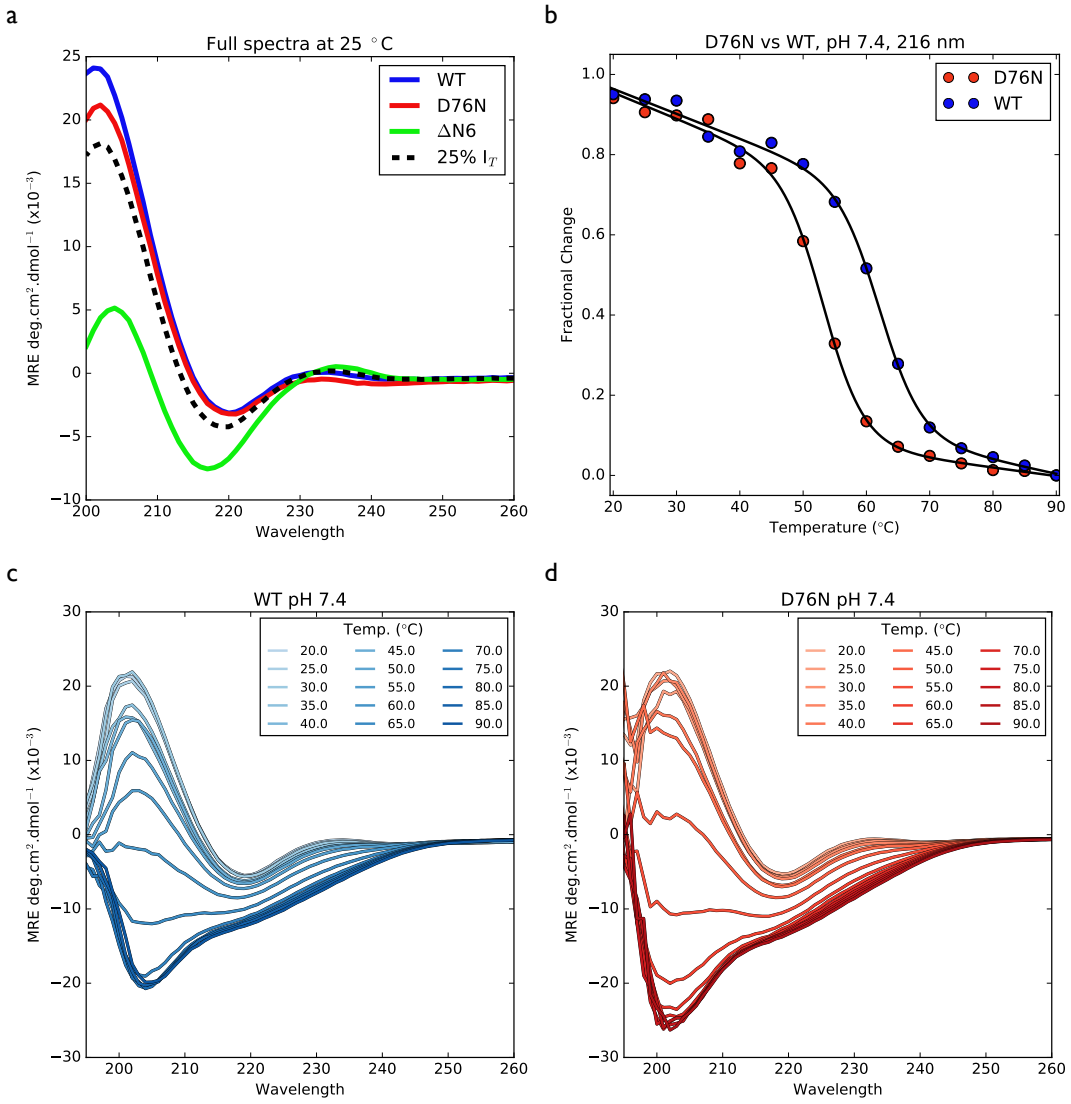


Fig. 3.4: D76N and WT structure and thermal stability, monitored by far-UV CD spectroscopy at pH 7.4.

The far-UV CD spectra for D76N (red), WT (blue), ΔN6 (green), and a simulated 25% I_T-state/75%I_T-state spectrum (dashed, black) at 25 °C are shown in a. In b, the temperature ramp data (c,d) are shown as a single wavelength (216 nm), fitted to the equation for two-state exchange model using the CDPal software package [193] for calculation of $T_{m,app}$ values (Methods 2.4.2). The temperature ramp experiments were carried out at pH 7.4 in 25 mM sodium phosphate buffer in the range 20-90 °C in 5 °C steps, spectra at these temperatures are shown in (c) for WT and (d) for D76N.

Since the thermal denaturation experiments were not reversible, the effect of the D76N amino acid substitution on the global stability of the β_2m fold was measured by equilibrium urea denaturation experiments, monitored by tryptophan fluorescence

(Methods 2.5), which are reversible for β_2m. The unfolding curves (Figure 3.5) show that the D76N variant is less stable than WT β_2m. Fitting the data to two-state equilibrium model (Methods 2.5, Equation 2.4) allows ΔG°_{UN} values of 18.1 kJ.mol^{-1} and 27.6 kJ.mol^{-1} to be extracted for D76N and WT, respectively ($\Delta\Delta G^\circ_{UN}$ of 9.5 kJ.mol^{-1}). The m value, which is an indicator of the surface area of the protein which is buried in the folded state is, 4.78 kJ.mol^{-1}.M^{-1} for WT and 4.70 kJ.mol^{-1}.M^{-1} for D76N, further supporting the conclusion that WT and D76N β_2m share the same native fold and unfold cooperatively under the conditions employed.

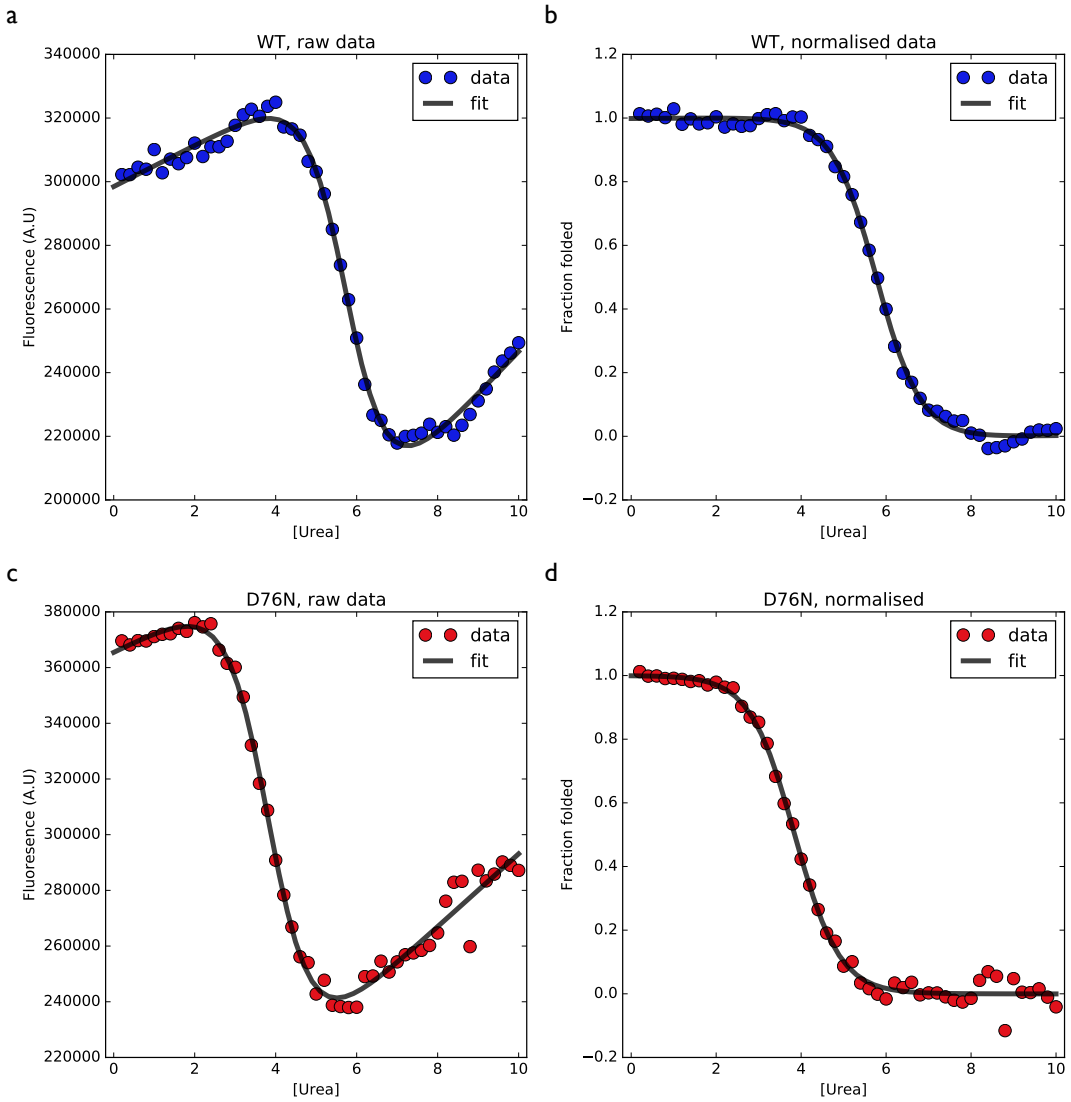


Fig. 3.5: Equilibrium urea denaturation of D76N and WT β_2m, followed by tryptophan fluorescence.

Equilibrium urea denaturation curves for 4 µM WT (blue) and D76N (red) in 25 mM sodium phosphate buffer, pH 7.4, 25 °C. The raw tryptophan fluorescence data (280 nm excitation, 325 nm emission) is shown on the left (a, c) and the normalised data (fraction folded) is shown on the right (b, d). The individual data points taken for every 0.2 M urea increment are shown as dots and the line of best fit to a two-state equilibrium model is shown as a black line (Methods 2.5, Equation 2.4).

The aggregation kinetics of D76N was next monitored by ThT fluorescence (Methods 2.6.1) and show that D76N readily forms amyloid fibrils at pH 7.4. This result is in stark contrast to WT β_2m which requires a range of extrinsic factors or harsh conditions

to form amyloid *in vitro*, such as low pH [224, 217, 225], the presence of Cu^{2+} ions [166, 178, 167] or TFE [226, 227].

To assess the role of the I_T-state in D76N aggregation the aggregation kinetics of D76N were also compared to ΔN6 (an I_T-state structural mimic). However, most ΔN6 studies to date were conducted at pH 6.2, where the protein aggregates more quickly than at higher pH (pH 7.2 and 8.0) [88, 64]. Experiments detailed so far in this chapter were conducted at pH 7.4; therefore, it was essential to test the effect of pH on the aggregation, stability and structure of D76N. The aggregation kinetics in Figure 3.6 show that D76N aggregates faster at pH 6.2 than at pH 7.4, yet resulted in fibrils with similar overall morphology, consistent with long, straight and unbranched amyloid fibres (Figure 3.6c).

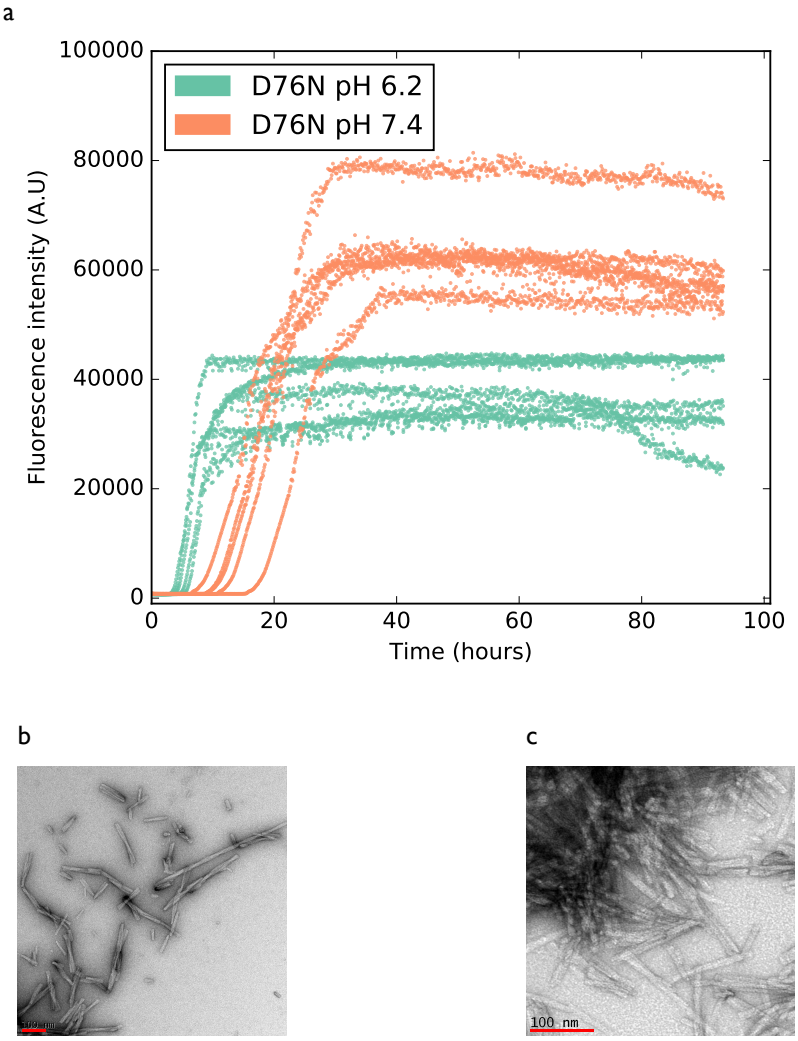


Fig. 3.6: D76N aggregation kinetics at pH 7.4 and pH 6.2.
(a) Kinetics of D76N aggregation monitored by ThT fluorescence, 60 µM D76N, 25 mM sodium phosphate buffer, pH 7.4 (orange) and pH 6.2 (green) with a final ionic strength of 200 mM (made by addition of NaCl), 10 µM ThT, 0.02% *(w/v)* NaN_3, 37°C, 600 rpm. (b,c) Negative stain EM micrographs of amyloid fibrils (Methods 2.6.2) from reaction endpoint material at pH 7.4 (b) and pH 6.2 (c), scale bar (red) = 100 nm). Five replicates for each condition are shown.

In Figure 3.7 the aggregation kinetics of D76N and ΔN6 are compared at pH 6.2 (Methods 2.6.1), a pH known to enhance ΔN6 aggregation [88, 64]. Remarkably, given the very different equilibrium population of I_T-state between D76N (similar to WT β_2m) and ΔN6 (an I_T-state mimic [88]), the aggregation of D76N (T_{half} 6.567 ±0.679) proceeds more quickly than that of ΔN6 (T_{half} 18.017 ±1.883 hours). The faster aggregation rate observed for D76N, compared to ΔN6 , demonstrates that the aggregation of D76N

cannot be purely I_T-state driven. Even an equilibrium population of 25% I_T-state, as proposed by Mangione et al. [184] would not be sufficient to explain the faster aggregation of D76N compared with ΔN6. These results suggest that D76N must aggregate via a mechanism distinct from WT and ΔN6, consistent with the different disease phenotype associated with D76N [183].

However, there must, in some part, be a common basis to the aggregation of D76N and ΔN6, since murine-β_2m (mβ_2m), which inhibits ΔN6 aggregation via a specific bi-molecular interaction [64], also retards the aggregation of D76N (Figure 3.8). The inhibitory effect of mβ_2m on D76N aggregation was tested through the aggregation assay in Figure 3.8, where the addition of mβ_2m to D76N at an equimolar ratio (40 μM D76N + 40 μM mβ_2m) vastly increased the lag time for aggregation. This inhibitory effect is specific to mβ_2m and not solely due to molecular crowding since the mixture of 40 μM D76N + 40 μM mβ_2m aggregated slower than both 40 μM D76N and 80 μM D76N alone (Figure 3.8).

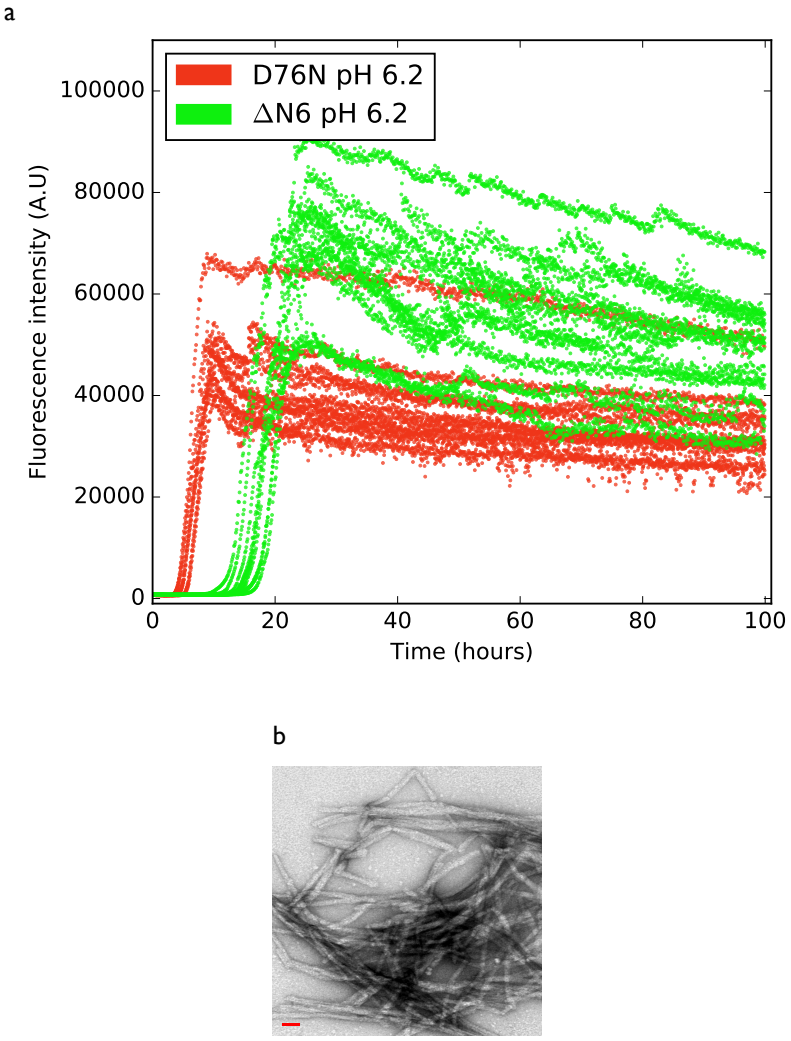


Fig. 3.7: Comparison of the aggregation rates of D76N and ΔN6.
(a) Kinetics of aggregation of D76N (red) and ΔN6 (green) monitored by ThT fluorescence, 30 μM protein, 25 mM sodium phosphate buffer, pH 6.2 with a final ionic strength of 180 mM (made by addition of NaCl), 10 μM ThT, 0.02% *(w/v)* NaN_3, 37 °C, 600 rpm. (b) Negative stain EM micrographs of amyloid fibrils (Methods 2.6.2) from reaction endpoint material for ΔN6, scale bar (red) = 100 nm. 10 replicates for each condition are shown.

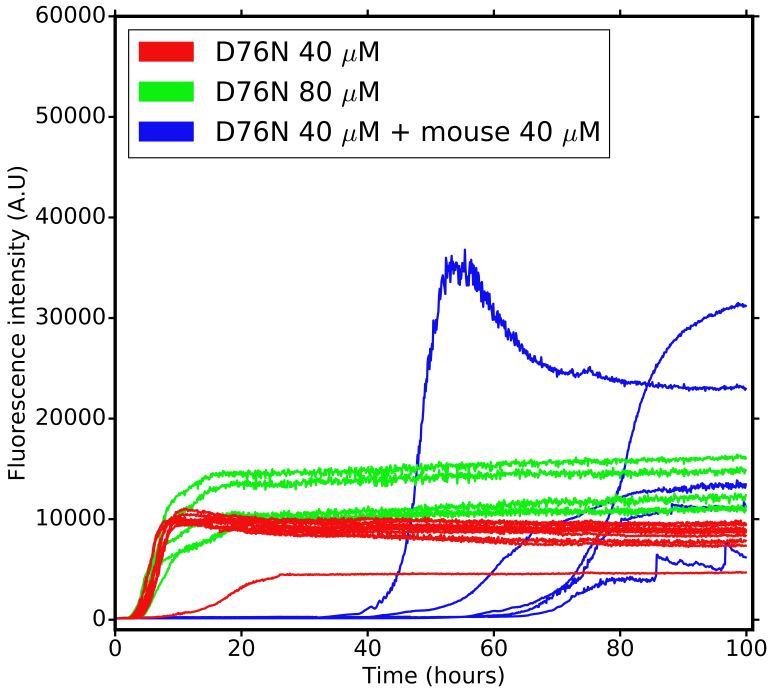


Fig. 3.8: Inhibition of D76N aggregation by murine-β_2m.
(a) Aggregation kinetics of 40 µM D76N (red) (10 replicates), 80 µM D76N (green) and an equimolar mixture of 40 µM D76N and 40 µM mβ_2m (blue) (both 5 replicates) in 25 mM sodium phosphate buffer, pH 6.2 with a final ionic strength of 180 mM (made by addition of NaCl), 10 µM ThT, 0.02% *(w/v)* NaN_3, 37 °C, 600 rpm.

In order to verify the accelerated aggregation of D76N at pH 6.2 was not due to either unfolding of native D76N or an increased equilibrium I_T-state population, the far-UV CD spectra of WT and D76N at pH 6.2 were compared (Figure 3.9a, Methods 2.4.2). The far-UV CD spectra show that there is no global change in the protein fold or increase in I_T-state population for D76N caused by the change in pH. The D76N and WT far-UV spectra in Figure 3.9a do have some small differences; however, this is likely due to differences in concentration, given that both the minima and maxima are affected equally.

The effect of the pH change from pH 7.4 to pH 6.2 on the thermal stability of the WT and D76N proteins was also examined through a temperature ramp monitored by far-UV CD (Figure 3.9b-d, Methods 2.4.2). The $T_{m,app}$ values at pH 6.2 were calculated as a function of the far-UV CD signal at 216 nm in the same manner as the $T_{m,app}$ values at pH 7.4 (Figure 3.4). At pH 6.2 the $T_{m,app}$ for WT β_2m is 65.5°C ± 0.5, and the D76N $T_{m,app}$ is 54.6°C ± 0.1, which results in a difference of approximately 11°C between WT

and D76N at pH 6.2, compared to the 9°C difference observed at pH 7.4. Interestingly, the D76N $T_{m,app}$ is unchanged by the change in pH, whereas the WT $T_{m,app}$ increases from 62.4°C ± 0.4 to 65.5°C ± 0.47. All $T_{m,app}$ values are summarised in Table 3.1.

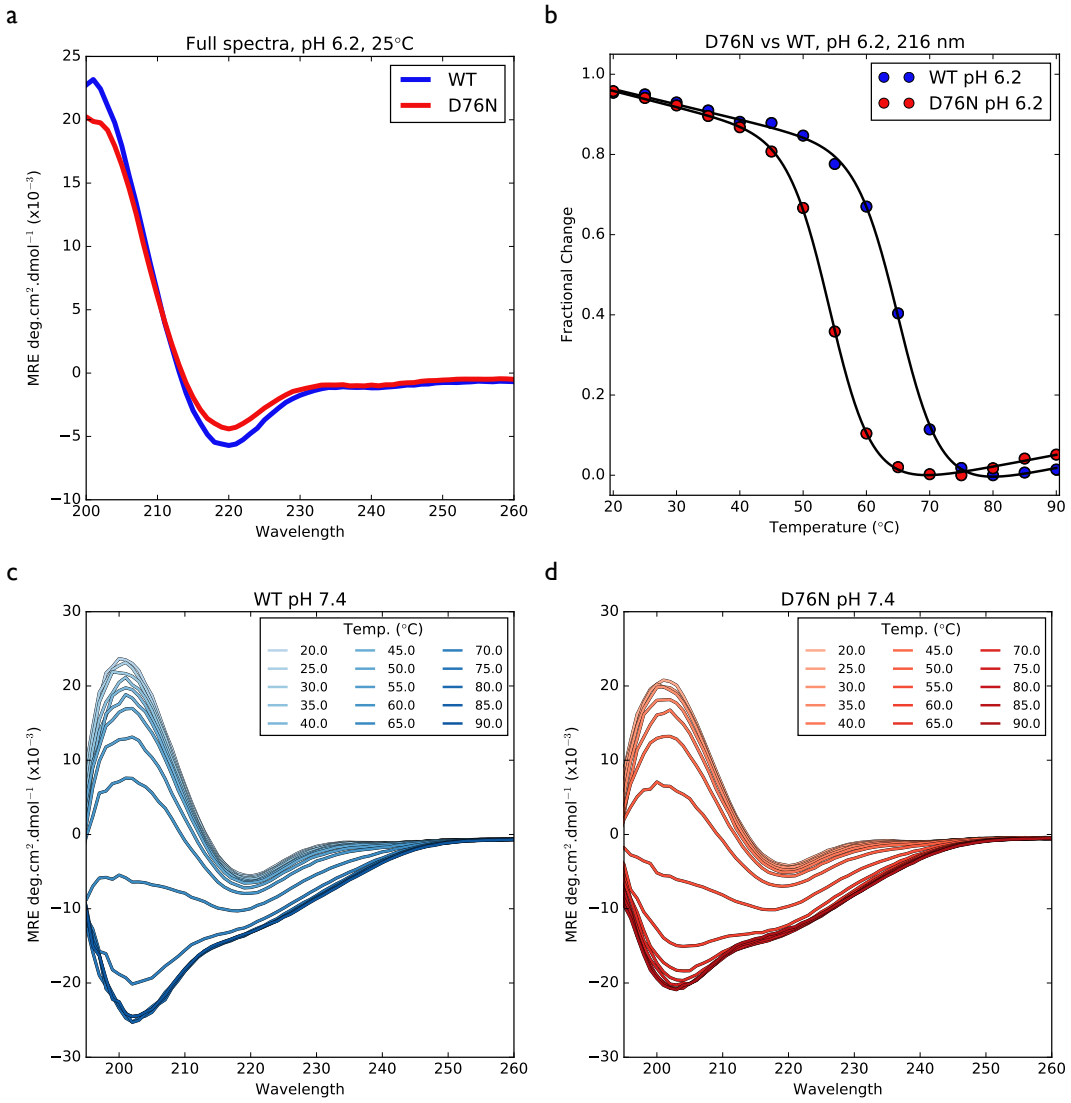


Fig. 3.9: Thermal stability of D76N and WT β_2m, monitored by far-UV CD spectroscopy at pH 6.2.
The far-UV CD spectra for D76N (red) and WT (blue) at 25 °C are shown in a. In b, the temperature ramp data (c,d) are shown as a function of temperature at a single wavelength (216 nm) fitted to the equation for a two-state exchange model using the CDPal software package [193] for calculation of $T_{m,app}$ values (Methods 2.4.2). The temperature ramp experiment were carried out at pH 6.2 in 25 mM sodium phosphate buffer in the range 20-90 °C in 5 °C increments.

Table 3.1: Apparent Tm values, for WT and D76N β_2m in 25 mM sodium phosphate buffer at pH 7.4 and pH 6.2

Variant	$T_{m,app}$ (°C)
WT pH 7.4	62.4 ± 0.4
D76N pH 7.4	53.5 ± 0.6
WT pH 6.2	65.5 ± 0.5
D76N pH 6.2	54.6 ± 0.1

3.2.2 Assessing the residue level effect of the D76N substitution using NMR

In order to study the residue level differences between D76N and WT by NMR, backbone residue assignments were required for D76N. D76N backbone assignments were obtained through the use of the BEST-HNCA, BEST-HNCOCA assignment experiments [200] (Figure 3.10) in combination with a comparison of the D76N ^{1}H-^{15}N HSQC spectrum with the assigned ^{1}H-^{15}N HSQC spectrum of WT β_2m (BMRB:17165, [88]). The remaining backbone chemical shifts were obtained from BEST-HNCO [200], BEST-HNCOCACB [200], and HBHACONH [201] experiments (Methods 2.7.2). In total, 86 HN, 86 N, 75 CO, 85 CA, 65 CB and 54 HA assignments were obtained, and are presented in Table 3.2. The data used for pH 7.4 assignments of D76N were not sufficient to assign residues 74 and 75 which have low intensity (perhaps indicative of exchange broadening). Residues 74 and 75 were assigned through another dataset collected at pH 6.2, otherwise under the same conditions, which provided a better dataset which was used to assign residues 74 and 75. Residues 74 and 75 in D76N can be observed in later experiments carried out at high resolution but the peaks are still broad.

Although all of the peaks in the D76N ^{1}H-^{15}N HSQC spectrum (Figure 3.11) are assigned (apart from side chain peaks), there are a further 14 residues, which do not give rise to visible resonances in the ^{1}H-^{15}N HSQC spectrum. Five of the missing residues are proline (5, 14, 32, 72, 90), which do not produce peaks in ^{1}H-^{15}N HSQC spectra. However, nine further residues are not present presumably because of exchange broadening or degeneracy with other peaks. The missing non-proline residues are I1, G29, E36, S57, K58, W60, S61, F62, Y63, and S88; the majority of these residues are found in, or adjacent to, the DE-loop (residues 57-61), which is known to be dynamic in WT β_2m [88]. The backbone assignments acquired for D76N were used to calculate chemical shift perturbations (CSPs) in Figures 3.12 and 3.13.

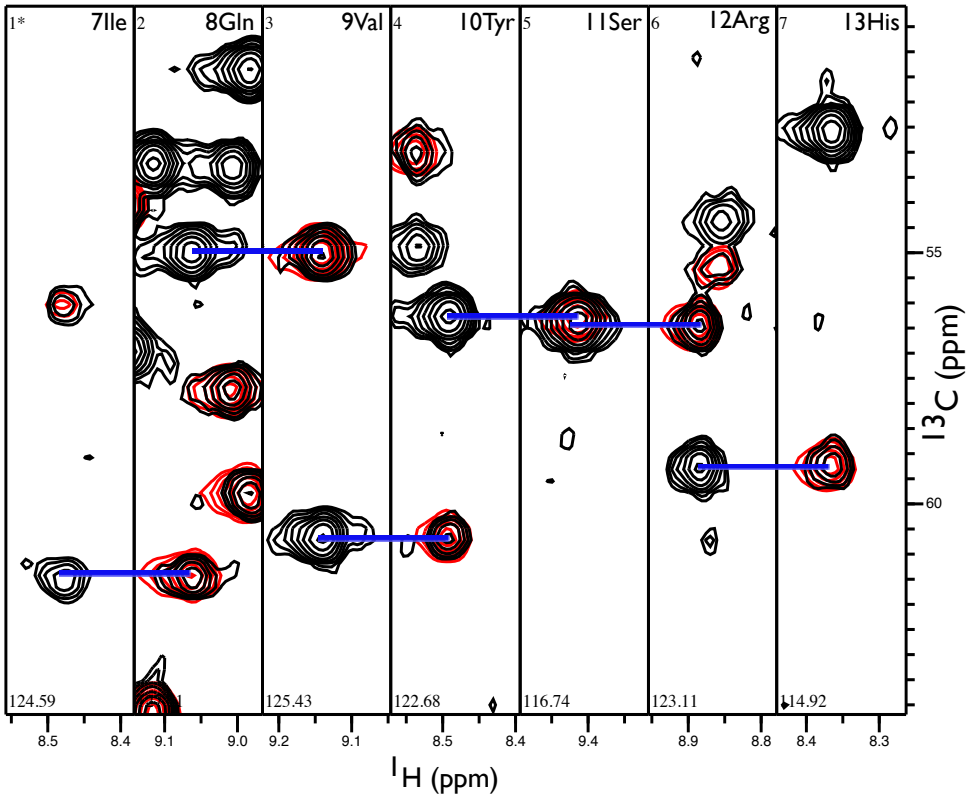


Fig. 3.10: Backbone resonance assignment of D76N.
Example slices from 750 MHz BEST-HNCA (black) and BEST-HNCOCA (red) [200] showing sequential assignment for residues 7 to 13. The sample for 3D experiments contained: 325 µM D76N (^{1}H, ^{15}N, ^{13}C), 0.1 mM DSS, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3 in 10 mM sodium phosphate buffer pH 7.4.

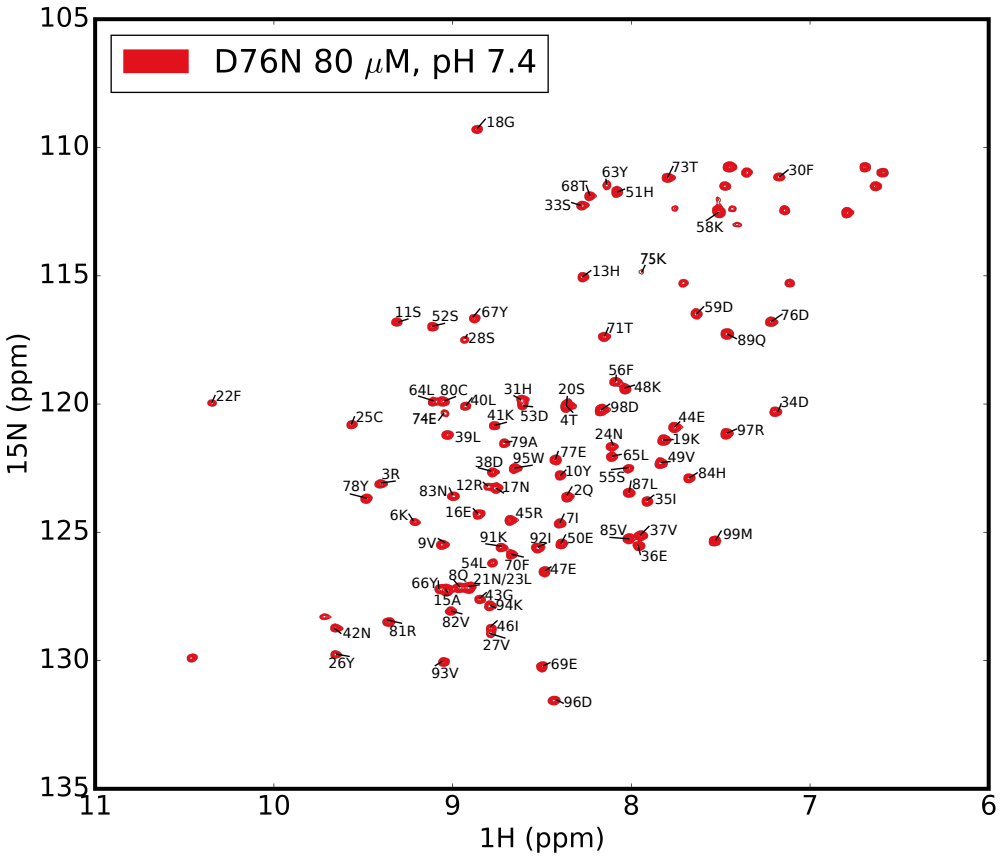


Fig. 3.11: ^{1}H-^{15}N HSQC spectrum of D76N, showing resonance assignments. ^{1}H-^{15}N HSQC spectrum of 80 µM D76N, in 25 mM sodium phosphate buffer pH 7.4, 0.02% (w/v) NaN_3, 10% *(v/v)* D_2O, acquired with a 600 MHz Varian INOVA spectrometer (Varian Inc., California, USA) with a room temperature probe, at 25°C, labelled with assignments.

Table 3.2: Backbone chemical shift assignments for D76N β_2m (ppm), ('-', denotes, no assignment). The sample for 3D experiments contained: 325 µM D76N (^{1}H, ^{15}N, ^{13}C), 0.1 mM DSS, 5% *(v/v)* D_2O, 0.02% *(w/v)* NaN_3 in 10 mM sodium phosphate buffer pH 7.4

Residue	HN	N	CO	CA	CB	HA
1	-	-	176.201	-	-	-
2	8.424	123.469	176.414	55.931	29.906	-
3	9.466	122.995	175.891	55.643	-	-
4	8.459	119.917	-	59.456	-	-
5	-	-	176.489	-	32.061	4.399
6	9.276	124.533	176.708	55.997	32.819	4.408
7	8.469	124.586	175.475	61.499	41.879	4.738
8	9.054	127.107	174.348	54.996	33.254	4.955
9	9.134	125.428	175.774	60.676	34.109	5.35
10	8.482	122.679	173.961	56.268	39.362	5.338
11	9.405	116.743	175.205	56.413	65.066	-
12	8.867	123.107	175.829	59.272	33.371	-
13	8.352	114.923	-	52.575	-	-
14	-	-	177.416	64.167	31.55	4.535
15	9.116	127.165	177.442	53.234	19.987	4.105
16	8.924	124.209	175.743	55.285	32.745	4.417
17	8.844	123.118	177.987	54.376	37.272	-
18	8.931	109.202	173.884	45.519	-	3.524
19	7.898	121.29	-	54.336	33.772	4.723
20	8.441	119.918	173.985	59.769	63.006	4.268
21	8.982	127.089	174.364	51.367	42.112	4.921
22	10.421	119.809	-	57.715	-	-
23	8.989	127.091	173.762	53.287	41.618	-
24	8.194	121.562	173.997	51.648	41.668	-
25	9.656	120.774	172.082	53.954	41.799	-

Residue	HN	N	CO	CA	CB	HA
26	9.729	129.661	176.917	56.246	34.104	-
27	8.877	128.903	174.898	59.857	33.552	-
28	9.017	117.457	-	57.214	-	-
29	-	-	174.002	-	-	-
30	7.241	111.103	174.941	53.972	-	-
31	8.693	119.793	-	57.964	42.031	-
32	-	-	177.293	62.698	35.166	-
33	8.359	112.244	174.263	60.867	62.315	-
34	7.261	120.173	174.61	55.095	40.473	4.386
35	8.004	123.675	-	60.966	-	-
36	-	-	-	-	-	4.59
37	8.042	125.426	173.987	60.833	34.15	5.306
38	8.861	122.567	174.901	52.692	46.108	4.987
39	9.119	121.117	175.172	53.678	43.732	5.015
40	9.006	120.001	176.109	52.933	45.831	5.028
41	8.855	120.766	176.827	54.35	34.123	4.521
42	9.743	128.644	176.108	54.353	-	4.357
43	8.919	102.53	173.552	45.523	-	4.164
44	7.851	120.868	176.307	54.36	31.818	4.573
45	8.759	124.51	-	57.816	30.626	4.266
46	8.867	128.664	176.53	62.014	38.859	3.968
47	8.559	126.358	177.611	58.65	30.587	4.173
48	8.109	119.289	175.154	55.904	30.841	4.502
49	7.937	122.186	175.303	61.062	-	4.398
50	8.472	125.261	174.586	54.424	32.464	4.474

Residue	HN	N	CO	CA	CB	HA
51	8.147	111.779	175.723	53.732	30.582	-
52	9.176	116.857	-	57.592	65.609	-
53	8.673	119.948	177.446	54.779	-	-
54	8.851	126.109	176.911	56.651	-	-
55	8.121	122.316	171.367	57.545	-	-
56	8.174	119.239	-	56.116	-	-
57	-	-	-	-	-	-
58	-	-	-	-	-	-
59	8.174	116.553	-	-	-	-
60	-	-	-	-	-	-
61	-	-	-	-	-	-
62	-	-	-	-	-	-
63	-	-	174.694	56.731	-	-
64	9.198	119.903	-	55.782	41.599	3.652
65	8.2	122.086	175.697	54.042	46.73	-
66	9.153	127.162	175.596	57.013	41.986	5.401
67	8.968	116.626	174.173	55.982	41.101	5.989
68	8.317	111.778	172.007	60.203	70.175	4.918
69	8.575	130.152	176.33	56.779	-	4.257
70	8.766	125.763	172.318	55.307	41.042	4.802
71	8.222	117.131	-	58.498	-	-
72	-	-	-	-	-	-
73	7.904	111.342	-	-	-	-
74	9.147	120.433	-	-	-	-
75	8.073	114.878	176.806	56.582	32.961	4.343

Residue	HN	N	CO	CA	CB	HA
76	7.322	116.724	174.645	52.995	40.875	5.192
77	8.526	122.193	175.506	54.898	32.812	4.788
78	9.57	123.763	175.732	56.817	42.682	-
79	8.803	121.461	174.156	51.115	24.358	5.08
80	9.133	119.813	171.779	52.982	43.472	5.132
81	9.451	128.463	174.544	53.852	33.881	-
82	9.077	127.893	173.894	60.423	35.793	4.961
83	9.07	123.578	173.424	51.253	41.61	5.206
84	7.723	122.86	175.761	57.087	36.182	4.524
85	8.087	124.985	175.873	64.535	31.261	3.982
86	7.524	110.694	175.17	63.268	69.891	4.533
87	8.12	123.36	-	53.462	-	-
88	-	-	-	-	-	-
89	7.54	117.183	-	-	-	-
90	-	-	176.416	63.598	31.892	4.508
91	8.799	125.481	175.193	55.104	35.356	4.53
92	8.59	125.458	176.37	60.484	38.861	4.805
93	9.14	129.977	175.715	61.514	34.05	4.305
94	8.866	127.8	176.76	56.672	-	4.434
95	8.731	122.395	175.039	56.818	28.291	4.658
96	8.496	131.459	175.525	53.104	41.755	4.477
97	7.531	120.89	176.194	56.66	29.948	3.429
98	8.253	120.124	175.765	54.703	41.249	-
99	7.633	125.201	-	57.356	-	-

Figure 3.12a shows a comparison of the ^{1}H-^{15}N HSQC spectra for both D76N and WT β_2m. The D76N spectrum in Figure 3.12a is well dispersed, indicating the protein is folded at pH 7.4, 25°C, consistent with the results shown in Figure 3.4 . The spectrum also resembles the ^{1}H-^{15}N HSQC spectrum of WT β_2m, also shown in Figure 3.12a. While the ^{1}H-^{15}N HSQC spectra for both D76N and WT β_2m are similar, there are some differences local to the amino acid substitution (Figure 3.12b,c). These differences could be because there is a different amino acid present, which confers a different chemical environment to the surrounding residues. The D76N amino acid substitution does not cause any long-range chemical shift perturbations. The chemical shift perturbations caused by the D76N substitution are shown in Figure 3.12b (Methods 2.7.3) and are mapped onto the crystal structure of D76N (PDB:4FXL, [183]) in Figure 3.12c.

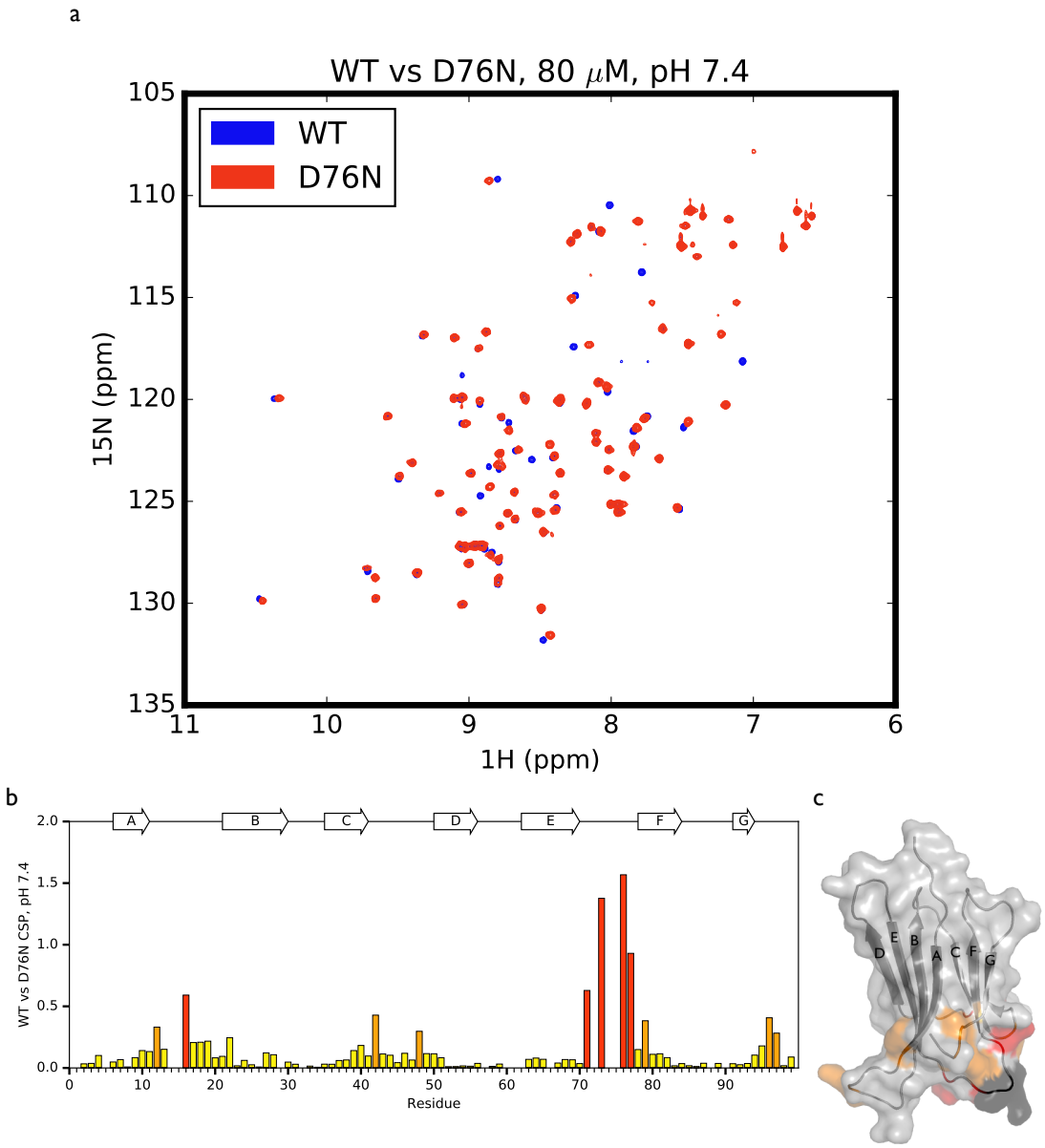


Fig. 3.12: Comparison of the ^{1}H-^{15}N HSQC spectra of WT and D76N at pH 7.4. (a) ^{1}H-^{15}N HSQC spectrum of 80 μM D76N (red) and of 80 μM WT β_2m (blue), in 10mM sodium phosphate buffer pH 7.4, 0.02% (w/v) NaN_3, 10% *(v/v)* D_2O, acquired with a 600 MHz Varian INOVA spectrometer (Varian Inc., California, USA) at 25°C with a room temperature probe. The CSPs between the two spectra in (a) are shown in (b) and mapped onto the structure of D76N (PDB:4FXL, [183]) in (c). CSPs between the WT and D76N spectra were calculated with the equation $\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2}$. Residues with CSPs less than one standard deviation from the mean average of all CSPs are coloured yellow, orange corresponds to residues with CSPs between 1-2 standard deviations and red corresponds to residues with CSPs greater than 2 standard deviations from the mean (standard deviation = 0.26). The secondary structure is shown as ribbons in (b).

Due to the faster aggregation observed for D76N at pH 6.2 compared to pH 7.4 (Figure 3.6), CSP analysis between D76N and WT β_2m was also carried out at pH 6.2 (Figure 3.13, Methods 2.7.3). The ^{1}H-^{15}N HSQC spectra for WT and D76N are compared in Figure 3.13a, and the CSP analysis at pH 6.2 shows a similar pattern to the CSPs observed at pH 7.4 (Figure 3.12), again localised to the residues surrounding position 76 in the structure. While the ^{1}H-^{15}N HSQC spectra change between pH 7.4 and pH 6.2, the changes are the same for WT and D76N, suggesting that the effect of the change in pH from pH 7.4 to pH 6.2 on D76N aggregation is not due to a large conformational change in the native structure.

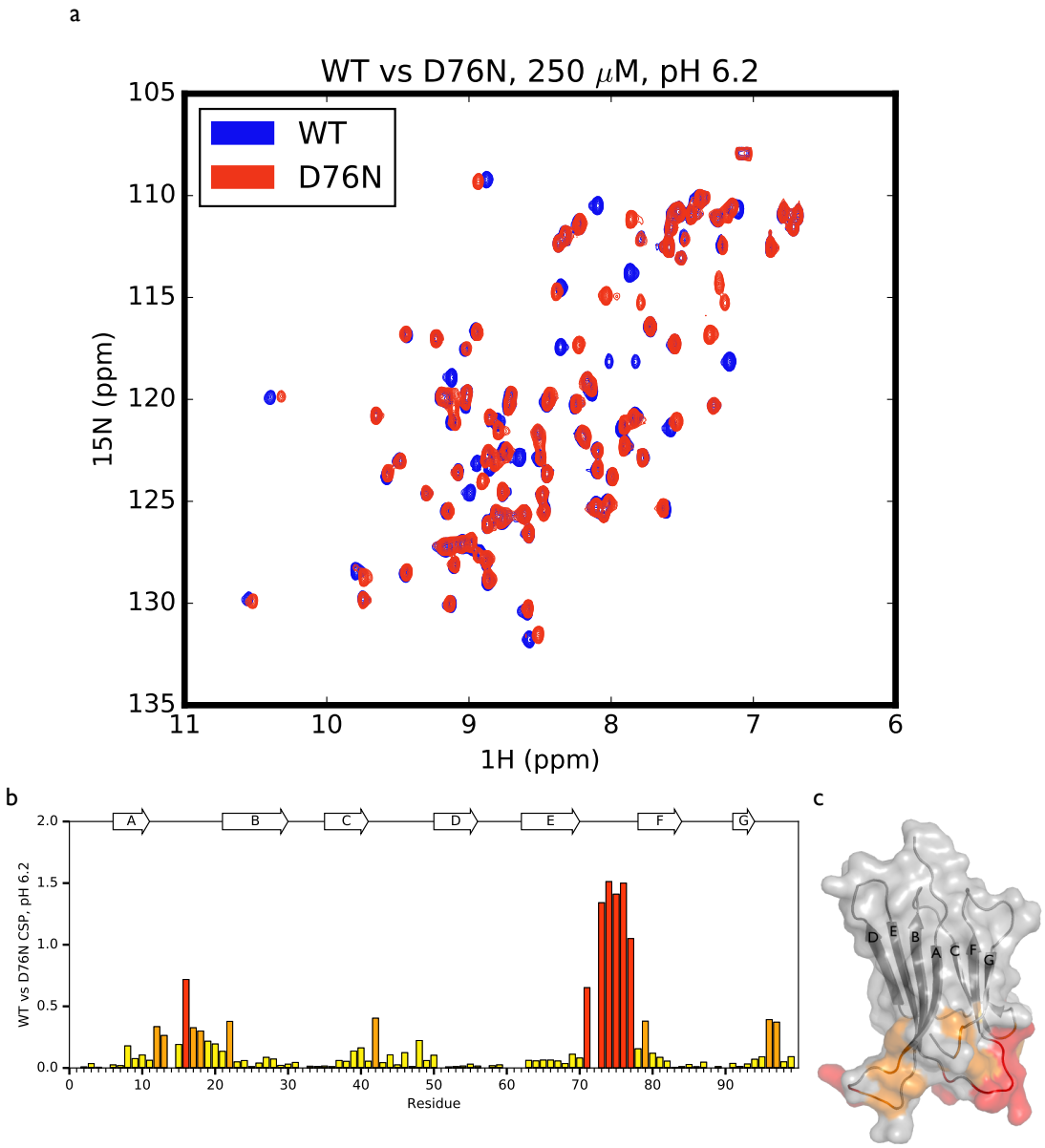


Fig. 3.13: Comparison of WT and D76N ^{1}H-^{15}N HSQC spectra at pH 6.2. (a) ^{1}H-^{15}N HSQC spectrum of 250 µM D76N (red) and of 250 µM WT β_2m (blue), in 25mM sodium phosphate buffer pH 6.2, 0.02% (w/v) NaN_3, 10% *(v/v)* D_2O, recorded on a 600 MHz Oxford NMR magnet with QCI-P-cryoprobe; proton-optimised quadruple resonance NMR ínverse' probe. at 25°C. The CSPs between the two spectra in (a) are shown in (b) and mapped onto the structure of D76N (PDB:4FXL) [183] (c). CSPs between the WT and D76N spectra were calculated with the equation $\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2}$. The standard deviation from the pH 7.4 analysis (0.26) was used as a cut-off. Residues with CSPs less than one standard deviation from the cut-off are coloured yellow, orange corresponds to residues with CSPs between 1-2 times the cut-off and red corresponds to residues with CSPs two times greater than the cut-off. The secondary structure is shown as ribbons in (b).

3.2.3 The role of residue type at position 76 in structure stability and aggregation of β_2m

In this section, a split-β-lactamase assay was utilised to determine whether other amino acid substitutions could be tolerated at position 76 in β_2m, and to understand if this would cause aggregation in the new variant.

The split-β-lactamase assay (Figure 3.14) is an *in vivo* system which has been used to detect aggregation of intrinsically disordered proteins and to assay inhibition of their aggregation with small molecules [205]. The split-β-lactamase assay has also been utilised to measure the effect of stabilising and destabilising mutations on the *in vivo* stability of Im7 protein variants [218]. The split-β-lactamase system consists of a tripartite fusion construct, where the test protein is inserted in between two domains of the β-lactamase enzyme (TEM-1) via two glycine/serine linkers, 28 residues in length (Methods 2.8). Upon expression of the β-lactamase construct in the bacterial periplasm, the two β-lactamase domains interact to form an active enzyme, conferring resistance to ampicillin. If the test protein aggregates, the two β-lactamase domains are unable to form an active complex, resulting in sensitivity to ampicillin and a lower MCD_{GROWTH} score in the assay (Methods 2.8). Importantly, the split β-lactamase assay requires no prior knowledge of the structure of the protein, and no protein purification is required, which facilitates the rapid assessment of many protein sequences.

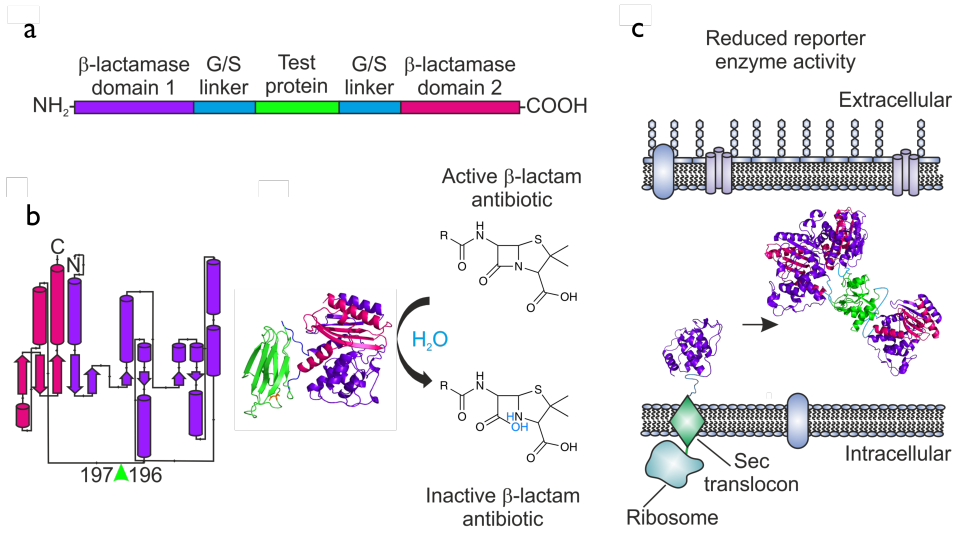


Fig. 3.14: Split β-lactamase assay schematic.
(a) The tripartite fusion construct composed of the two β-lactamase domains (purple and red) with the test protein (green, β_2m (PDB:4FXL, [183])) separated by glycine/serine linkers (blue). (b) The complete β-lactamase enzyme including the location of the test protein and linkers (PDB:1BTL, [228]), coloured as same (a). (c) If the test protein aggregates (green) the β-lactamase enzyme becomes less effective, leading to lower MCD_{GROWTH} scores. Figure adapted from [205].

Tripartite fusion constructs for WT and D76N (βla-β_2m-WT and βla-β_2m-D76N, respectively) were generated and the split-β-lactamase assay was carried out in the ampicillin range 0-140 μg/mL (Methods 2.8). The MCD_{GROWTH} curves are presented in Figure 3.15 and show increased antibiotic sensitivity for the D76N construct, consistent with its lower stability and higher aggregation propensity. The data in Figure 3.15 establishes that the split-β-lactamase assay can distinguish between WT and D76N (consistent with previous results [205]); however, it is not clear if the difference in stability or aggregation between the two proteins, or both, is responsible for the observed effect.

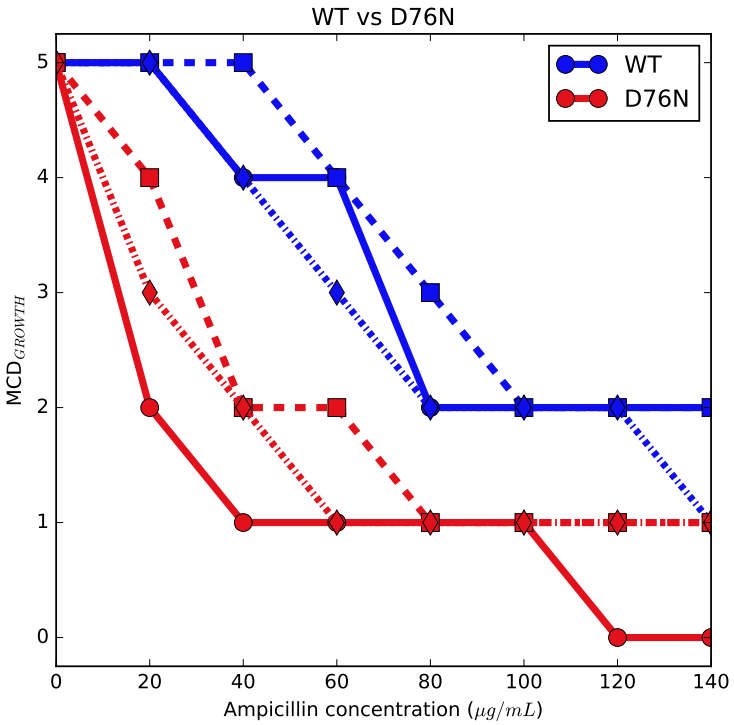


Fig. 3.15: Split-β-lactamase assay for D76N and WT β_2m.
Antibiotic survival curve of the maximal cell dilution allowing growth (MCD_{GROWTH}) after 18 h incubation at 37°C in *E. coli* BL21 [DE3] cells, over 0-140 µg/mL ampicillin concentrations for βla-β_2m-WT (blue) and βla-β_2m-D76N (red) constructs. The three replicates are represented as different line styles.

Saturation mutagenesis at position 76 was carried out to understand the significance of the residue identity at position 76. All variants (D76X variants) were assayed with the split-β-lactamase assay in triplicate, (Methods 2.8) and the corresponding traces are shown in Figures 3.16 and 3.17.

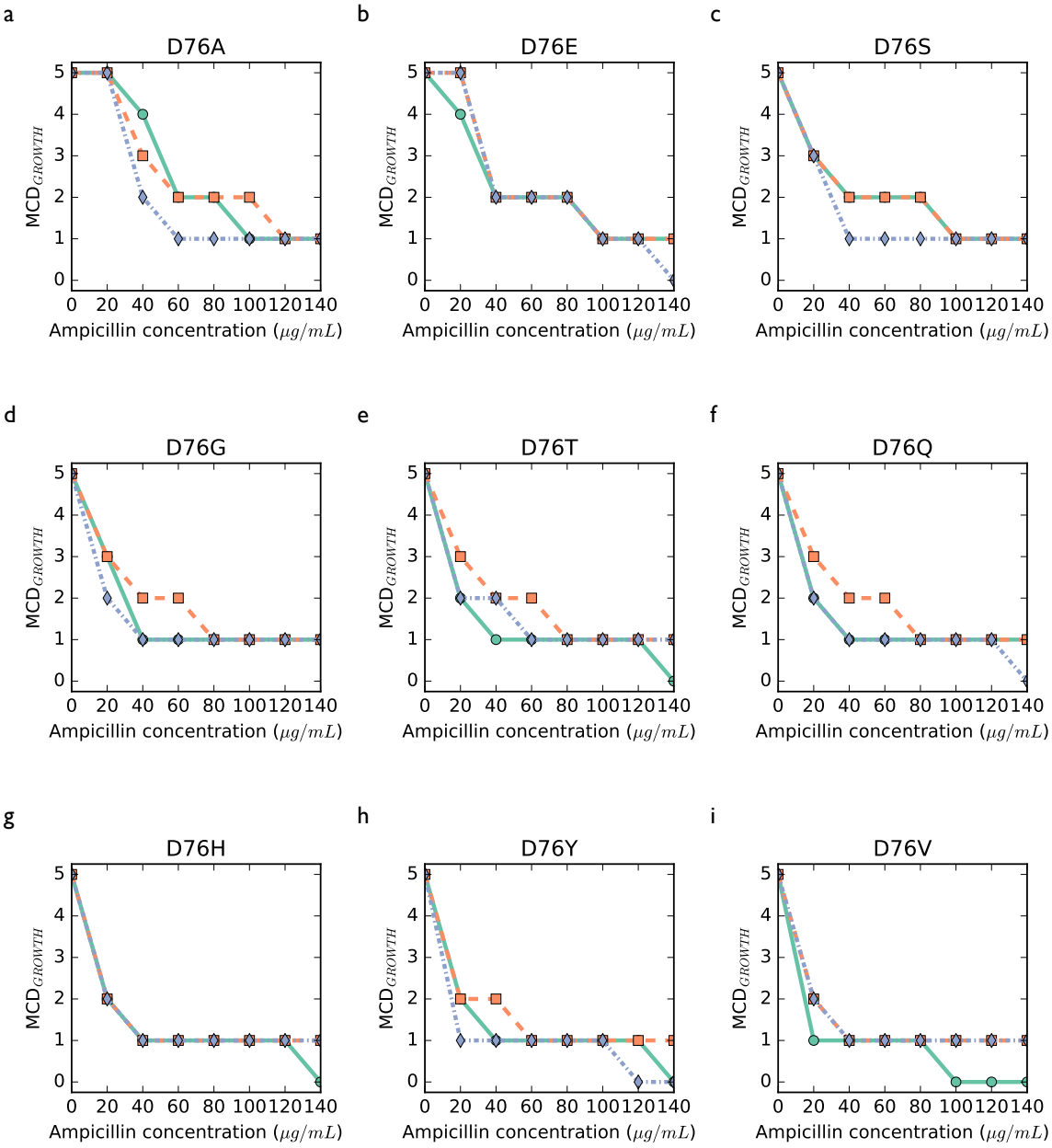


Fig. 3.16: Split-β-lactamase assay for D76X constructs.
Antibiotic survival curve of the maximal cell dilution allowing growth MCD_{GROWTH} after 18 h incubation at 37°C in *E. coli* BL21 [DE3] cells, over 0-140 μg/mL ampicillin concentrations for D76X constructs. The three replicates are represented as different line styles and colours.

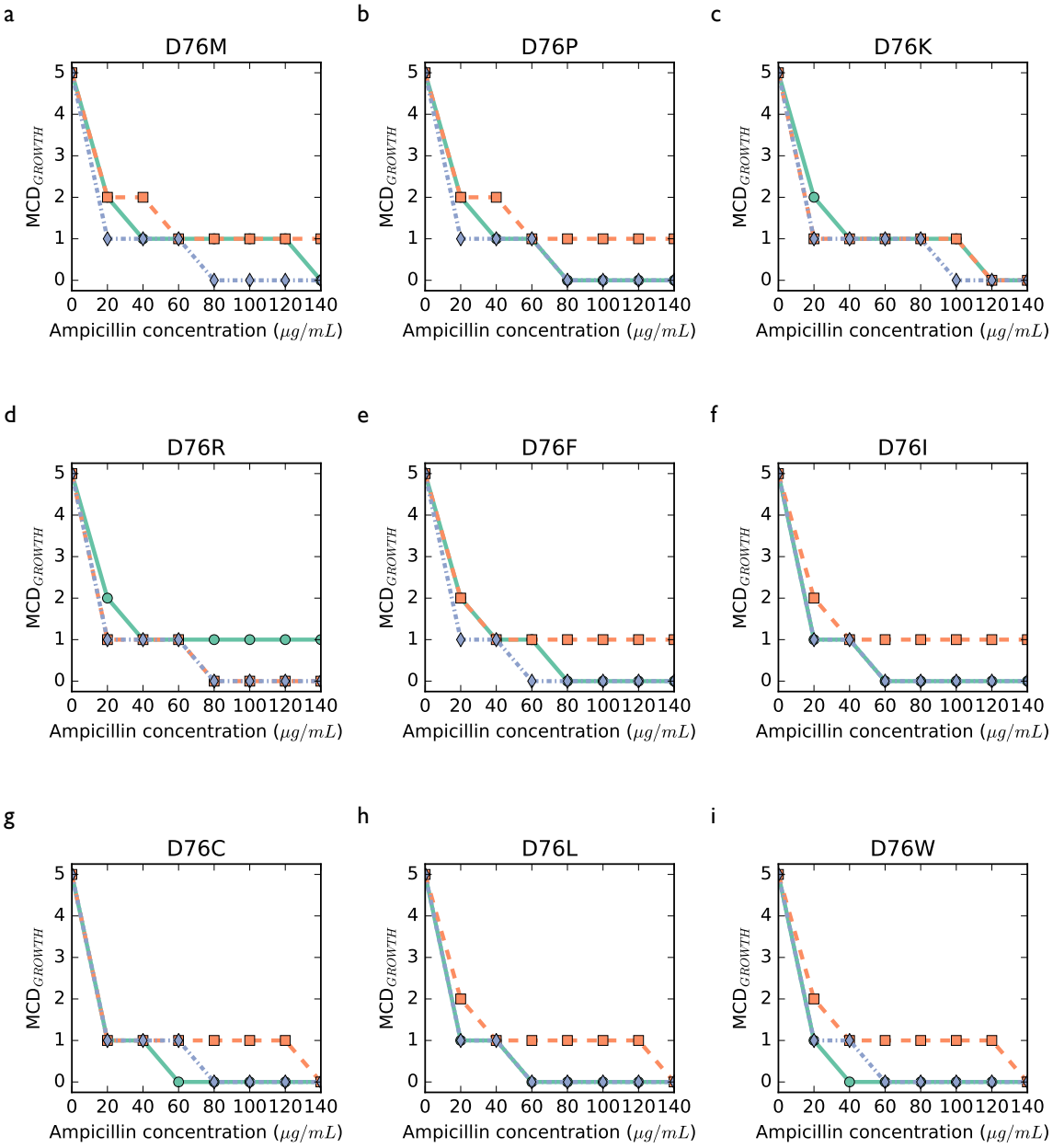


Fig. 3.17: Split-β-lactamase assay for D76X constructs.
Antibiotic survival curve of the maximal cell dilution allowing growth MCD_{GROWTH} after 18 h incubation at 37°C in *E. coli* BL21 [DE3] cells, over 0-140 µg/mL ampicillin concentrations for D76X constructs. The three replicates are represented as different line styles and colours.

A quantitative comparison of all mutants (D76X mutants) was carried by calculating the area under the curve (AUC) for each replicate and the mean average AUC for each replicate is presented in Figure 3.18. The AUC values in Figure 3.18 show that any amino acid substitution at position 76, other than the WT Asp, results in lower MCD_{GROWTH} scores in the split-β-lactamase assay. The data in Figure 3.18 can be arbitrarily split into four different groups based on the AUC values. The first group contains only the WT sequence with an AUC score over 400. The second group contains E and A constructs, with scores 300-400. The third group contains D76N with S, G, T and Q. The final group (group 4) contains residues H, Y, V, M, P, K, R, F, I, C, L, and W.

The low scores for many of the residues in group 4 can be rationalised by the introduction of hydrophobic residues into a solvent-exposed site (W, L, I, F, V, Y, M), the introduction of an opposite charge to the natural D residue found at position 76 (R, K) and the formation of intermolecular disulphide bonds (C). The remaining residues in group 4 (H, P) may also be explained by charge (H) and conformational restriction (P). The group 3 residues (excluding G), which share similar scores to D76N (S, T, Q) are polar residues, which may be better tolerated within the surrounding environment. Perhaps the most intriguing group is group 2 (A, E) with scores between WT and D76N. E shares the same charge as D (-1) so may be better tolerated than N at position 76; however, it is not entirely clear why A would have a better score than N, based on criteria such as charge, size or hydrophobicity.

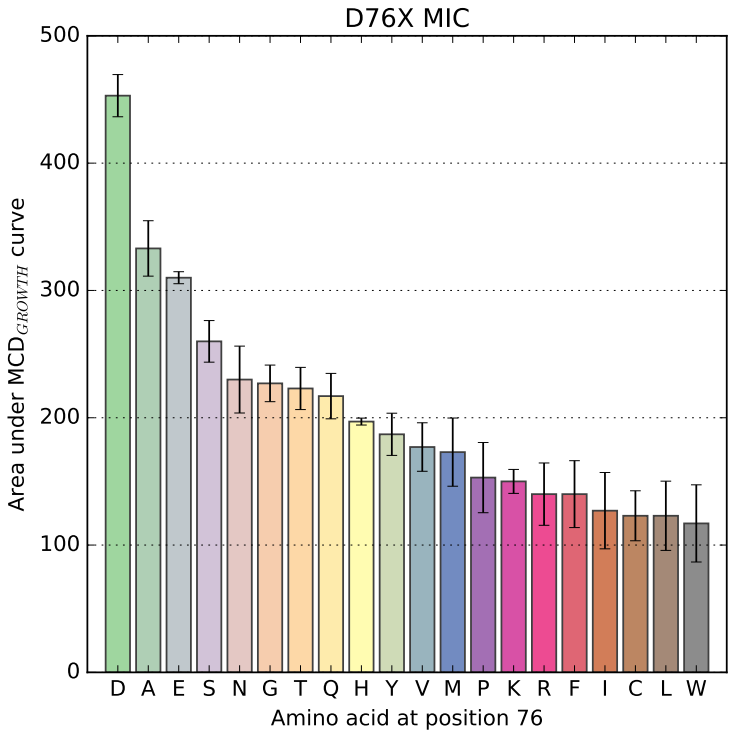


Fig. 3.18: Comparison of all the MCD_{GROWTH} value for D76X variants measured using the split-β-lactamase scores.
MCD_{GROWTH} for all D76X constructs represented as the mean average area under the curve (AUC) for each MCD_{GROWTH} curve (3 replicates) with S.E.M error bars. AUC values were calculated with Equation 2.10 as detailed in Methods 2.8.

The Pearson's Correlation between the hydrophobicity and residue size (van der Waals volume), with the observed AUC scores, were analysed and show that the correlation between hydrophobicity index at pH 7 [229] is -0.55, and the correlation with residue size [230] is -0.48 (Figure 3.19a,b). The correlations presented show that there is not significant relationship over all 20 amino acids between residue size or hydrophobicity and the split-β-lactamase assay scores. However, when the hydrophobic residues alone are studied (A, Y, V, M, F, I, L, W) the correlation is -0.97 (Figure 3.19c), indicating that increased hydrophobicity does correlate with a reduced split-β-lactamase assay score. This result does not account for the readout of the split-β-lactamase assay entirely because non-hydrophobic residues can have very low scores, resulting in the correlation of -0.55 observed in Figure 3.19a.

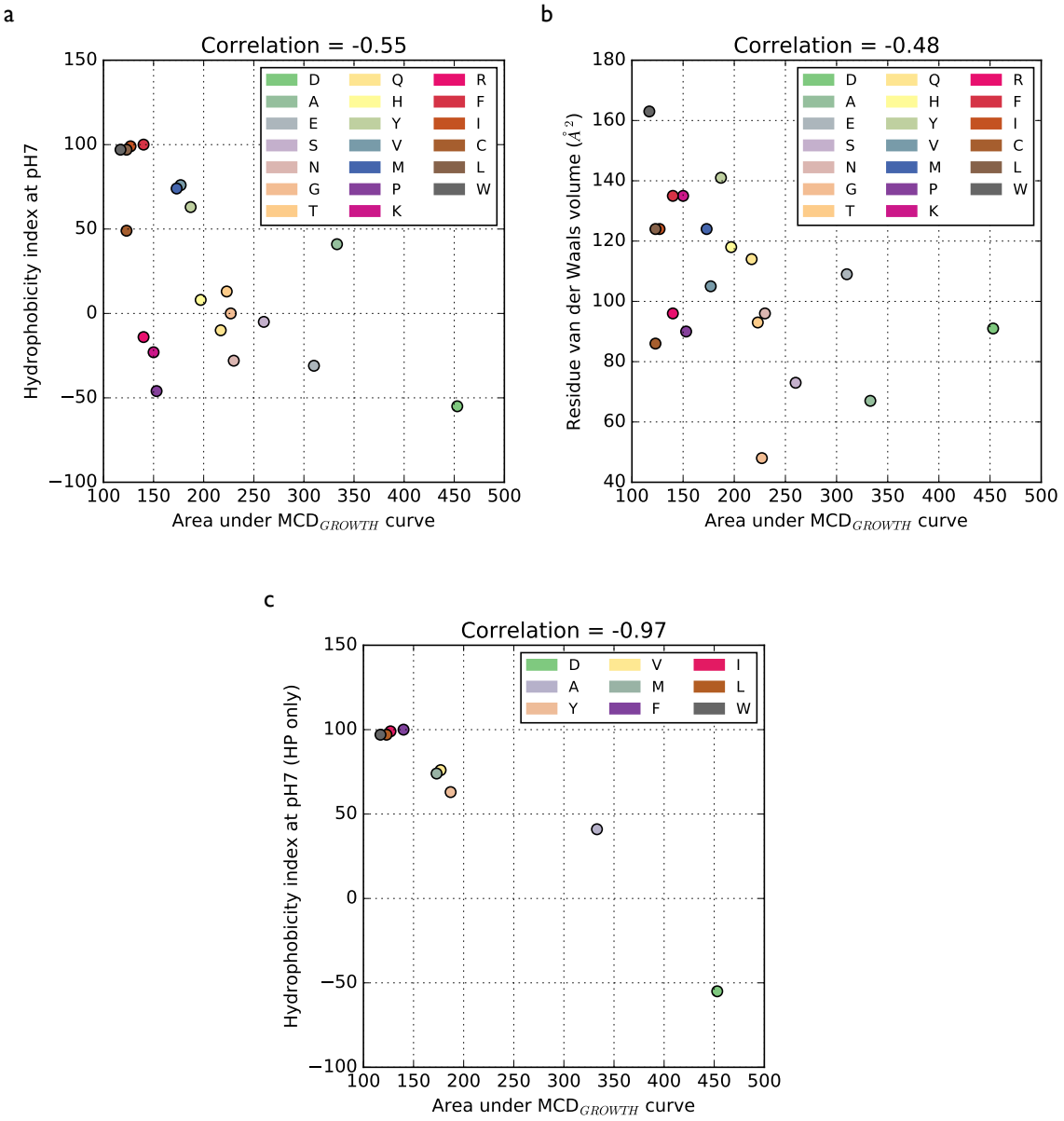


Fig. 3.19: Correlation between split-β-lactamase assay score with hydrophobicity and residue size.
Pearson's Correlation between the hydrophobicity (a) and residue size (b) with the observed AUC scores, for all variants. The correlation with only the hydrophobic residues (and WT) is shown in (c). Hydrophobicity index scores at pH 7 are taken from [229] and residue van der Waals volumes (Å) from [230].

D76E and D76A were selected for purification to study the effect of these mutations on stability, aggregation and structure. D76E and D76A were chosen due their position

between WT and D76N in the AUC values in Figure 3.18. Protein purification was carried out with the same protocol used for WT and D76N purification (Methods 2.3).

In Figure 3.20a,b, the size exclusion chromatography elution profiles for D76E and D76A show a single peak at approximately 200-220 ml. The blue regions in the elution profiles (Figure 3.20a,b) were carried forward for further experiments, including an SDS-PAGE experiment (Figure 3.20c), which shows the protein is a single, pure species of approximately 12 kDa. The exact masses of the purified proteins were measured by denaturing ESI-MS. The ESI-MS spectra in Figure 3.21 show that the pure protein is a single species with an observed molecular weight of 11874.26 (± 0.12) Da and 11816.16 (± 0.16) Da for D76E and D76A respectively. Both of these values agree with the expected masses of 11874.42 Da and 11816.38 Da, respectively, for the oxidised (disulphide-bond-containing) protein.

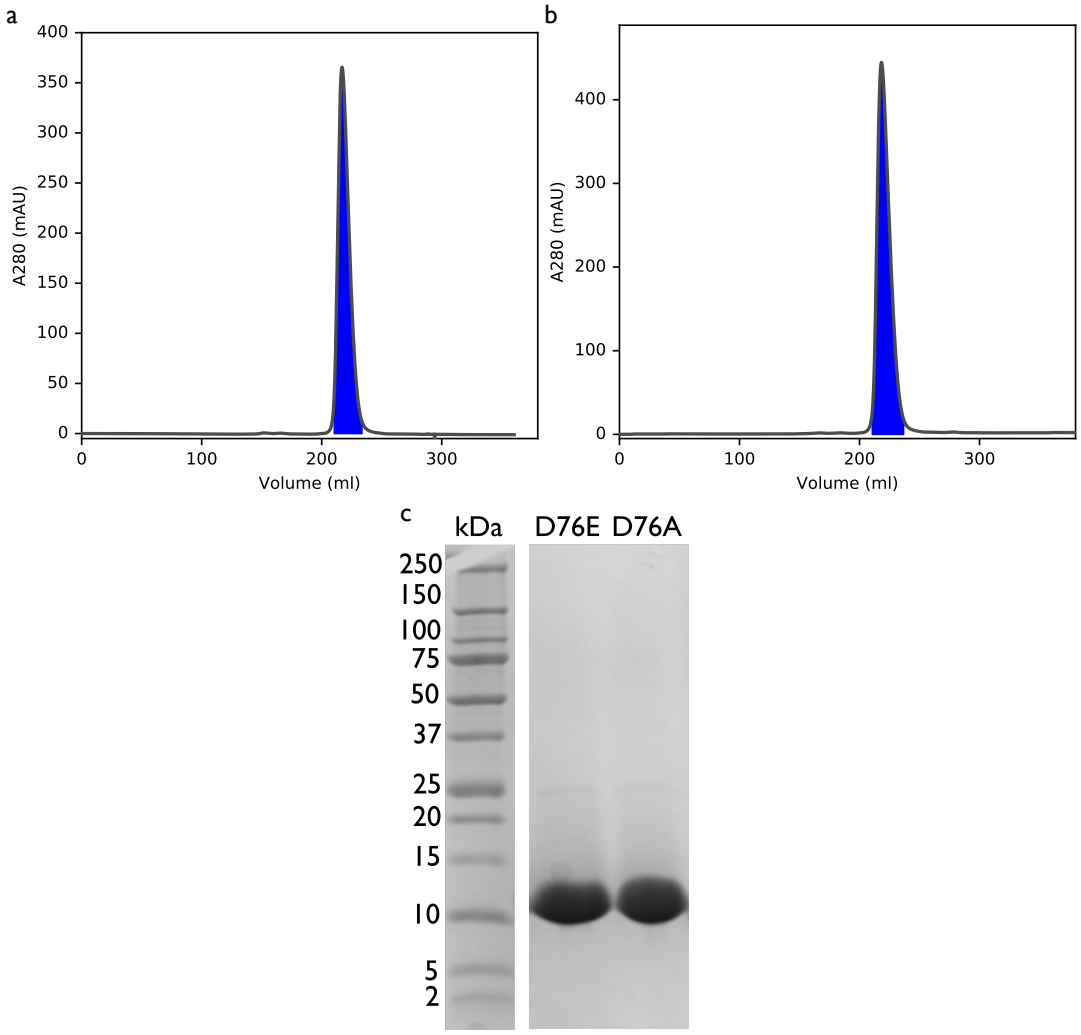


Fig. 3.20: Purification of D76E and D76A variants.
Size exclusion chromatography profile for D76E (a) and D76A β_2m (b), the blue area corresponds to the material taken forward for further experiments. The gel in c, shows the pure samples from a and b as single bands.

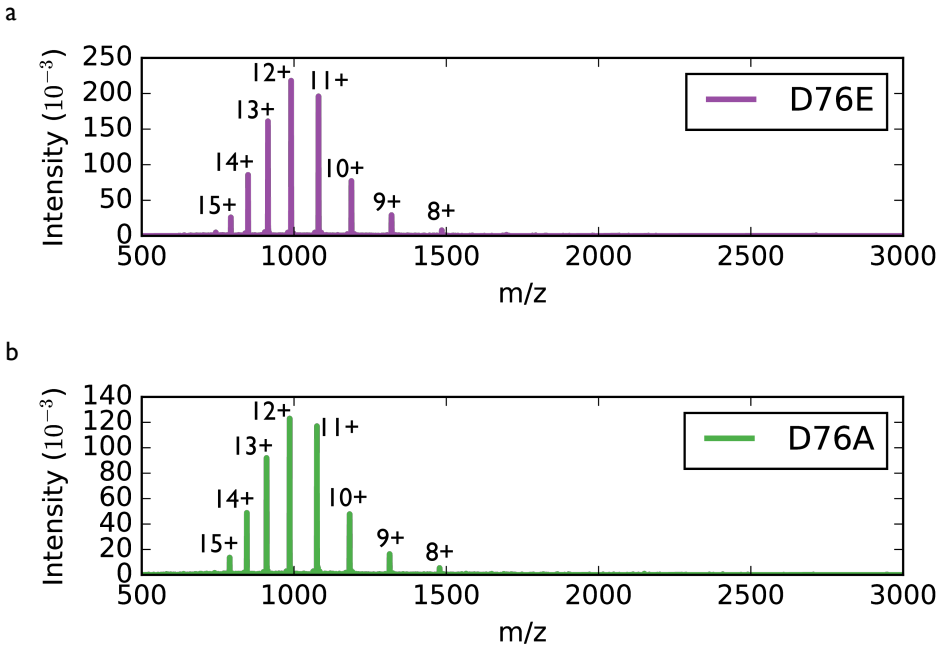


Fig. 3.21: D76E and D76A Mass spectra.
Denatured ESI mass spectra of purified D76E (a) and D76A (b) for exact mass calculation. The observed molecular weights are 11874.26 (± 0.12) Da and 11816.16 (± 0.16) Da for D76E and D76A, respectively.

The global fold of the D76E and D76A variants was compared to WT and D76N using far-UV CD spectroscopy. The resulting spectra (Figure 3.22a) show that all variants are folded, and have a spectrum typical of β_2m, however, there are some small differences in the spectra in the magnitudes of the 205 nm maximum and the 220 nm minimum.

The thermal stability of the variants was also probed using a temperature ramp, monitored by far-UV CD (Figure 3.22b-d, Methods 2.4.2). The $T_{m,app}$ values for D76E and D76A observed through following the 216 nm signal as a function of the temperature, are compared to the WT and D76N data from Figure 3.9. WT β_2m is the most stable with a $T_{m,app}$ of 65.5°C ± 0.5, D76E and D76A have the same $T_{m,app}$ value of 57.6°C ± 0.3 and 57.6°C ± 0.2 respectively, and D76N is the least stable with a $T_{m,app}$ of 54.6°C ± 0.1 (summary in Table 3.3). This order matches the order observed in the split-β-lactamase data in Figure 3.18.

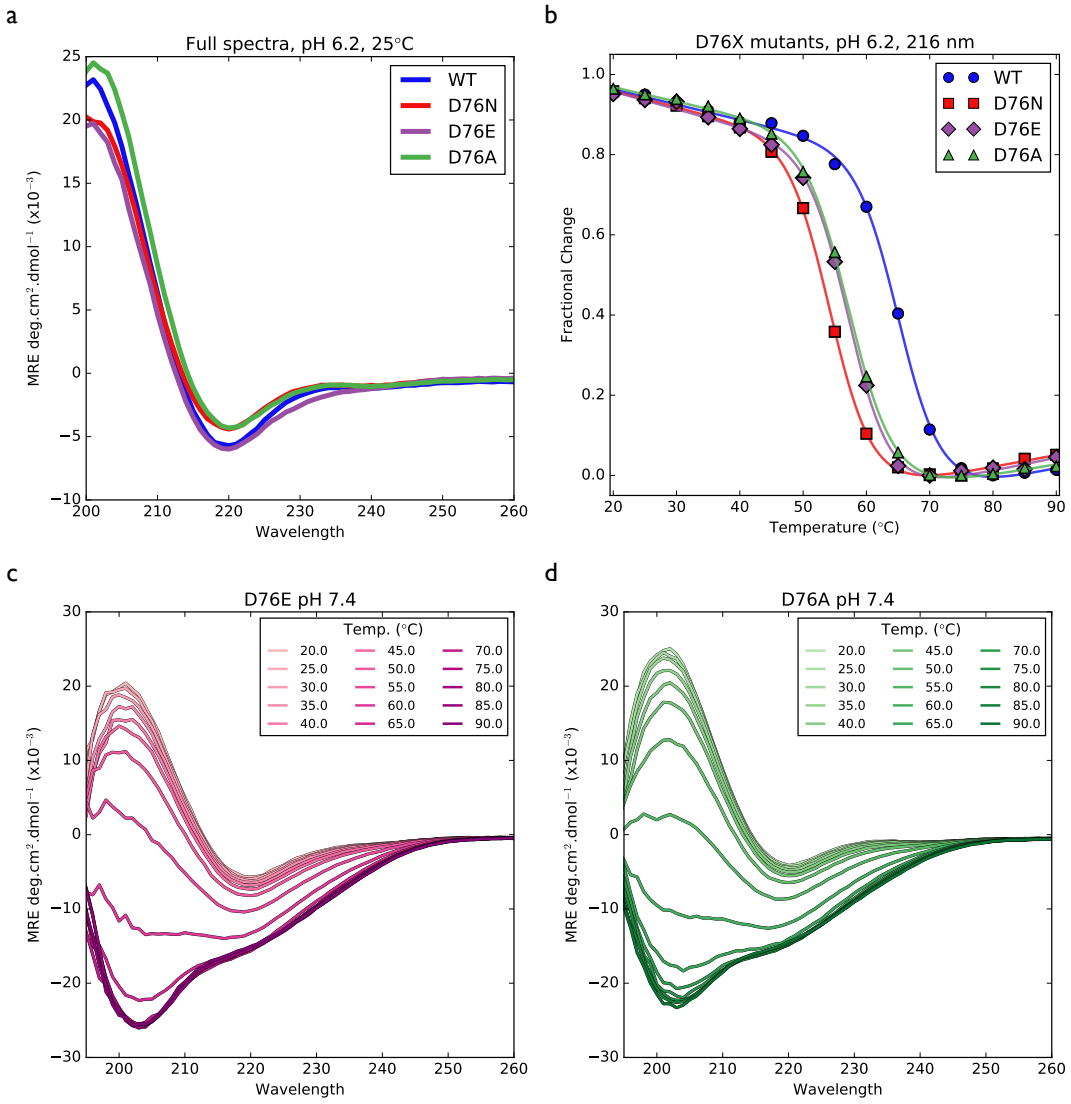


Fig. 3.22: D76E and D76A temperature ramp, monitored by far-UV CD spectroscopy at pH 6.2.

Far-UV CD spectra for D76N (red), WT (blue), D76E (pink) and D76A (green) at 25 °C are shown in a. In b, the temperature ramp data (c,d) are shown as a function of a single wavelength (216 nm) fit to the equation two-state exchange model using the CDPal software package [193] for calculation of $T_{m,app}$ values. The temperature ramp experiment were carried out at pH 6.2 in 25 mM sodium phosphate buffer in the range 20-90 °C in 5 °C steps. (c) and (d) show all spectra recorded in the temperature ramp experiment for D76E and D76A, respectively.

Table 3.3: Apparent Tm values, for WT, D76N, D76E and D76A β_2m in 25 mM sodium phosphate buffer at pH 7.4 and pH 6.2

Variant	$T_{m,app}$ (°C)
WT pH 7.4	62.4 ± 0.4
D76N pH 7.4	53.5 ± 0.6
WT pH 6.2	65.5 ± 0.5
D76N pH 6.2	54.6 ± 0.1
D76E pH 6.2	57.6 ± 0.3
D76A pH 6.2	57.6 ± 0.2

In addition to the stability of the proteins, the ability of the D76E and D76A variants to aggregate to form amyloid fibrils was carried out at pH 6.2 and compared with D76N (Figure 3.23), interestingly, D76E and D76A both aggregate readily, yet more slowly than D76N. D76N aggregated the fastest, followed by D76A and D76E in conditions in which WT β_2m is not able to aggregate. T_{half} values are presented in Table 3.4.

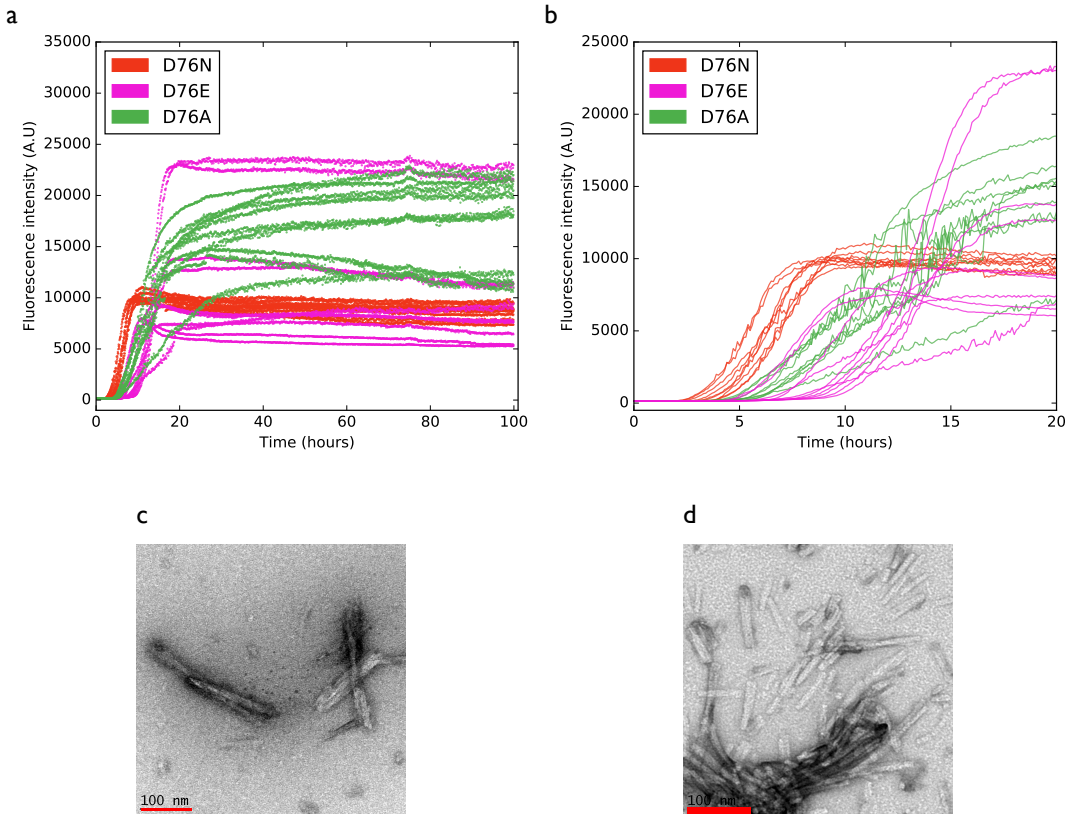


Fig. 3.23: Comparison of the aggregation rates for D76E and D76A with D76N. (a,b) Aggregation kinetics of D76N (red), D76E (pink) and D76A (green) monitored by ThT fluorescence. 40 μM protein, 25 mM sodium phosphate buffer, pH 6.2 with a final ionic strength of 180 mM (made by addition of NaCl), 10 μM ThT, 0.02%, NaN_3, 37 °C, 600 rpm (b, zoomed in view of lag phase). (c,d) Negative stain EM micrographs of amyloid fibrils from reaction endpoint material for D76E (c) and D76A (d), scale bar (red) = 100 nm.

Table 3.4: Aggregation T_{half} values, for D76N, D76E and D76A β_2m in 25 mM sodium phosphate buffer pH 6.2

Variant	T_{half} (hours)
D76N	6.5 ± 0.3
D76E	13.6 ± 2.5
D76A	11.6 ± 2.8

The results from the CD temperature melt (Figure 3.22) and the aggregation assay (Figure 3.23) both match the observed order from the split-β-lactamase assay. The effect of the D76E and D76A mutations on the structure of the β_2m fold was investigated through the measurement of ^{1}H-^{15}N HSQC spectra for D76E and D76A, in comparison

to WT and D76N (Figure 3.13). The spectra in Figure 3.24 show that all variants share a similar overall fold, at least as indicated by the small differences in chemical shift. However, there is a subset of peaks with different chemical shifts, which are presented in the CSP analysis in Figure 3.25.

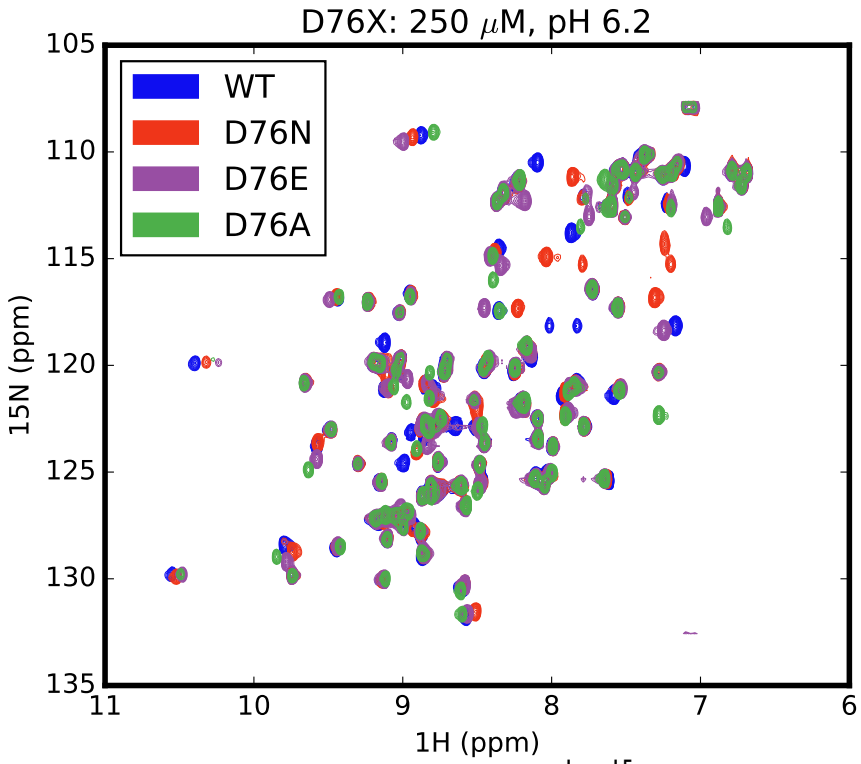


Fig. 3.24: Comparison of D76E and D76A variants ^{1}H-^{15}N HSQC, with WT and D76N β_2m.
^{1}H-^{15}N HSQC spectra of WT (blue), D76N (red), D76E (pink) and D76A (green). 250 µM protein in 25mM sodium phosphate buffer pH 6.2, 0.02% (w/v) NaN_3, 10% *(v/v)* D_2O, recorded on a 600 MHz Oxford NMR magnet with QCI-P-cryoprobe; proton-optimised quadruple resonance NMR ínverse' probe. at 25°C.

The CSP analysis of the WT protein compared to D76E and D76A (Figure 3.25) shows a similar pattern of CSPs to the differences observed between WT and D76N under the same conditions in Figure 3.13. These chemical shift differences are again clustered around the area surrounding the mutation site in the protein structure; although the observed magnitude of the CSP between D76E and D76A with WT values are greater than the CSPs between D76N and WT, in the areas surrounding position 76.

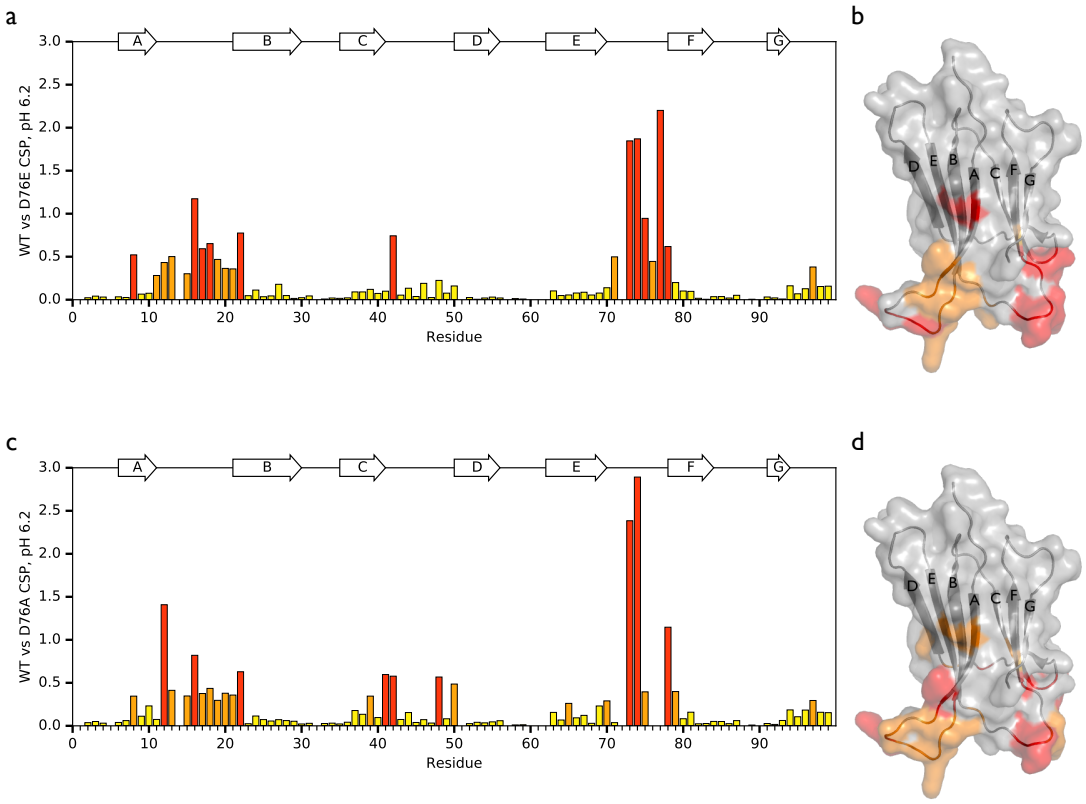


Fig. 3.25: CSP analysis of D76E and D76A, in comparison to WT β_2m at pH 6.2. CSPs between the WT and D76E (a) or D76A (d) were calculated with the equation $\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2}$ (Methods 2.7.3). The standard deviation from the pH 7.4 analysis (0.26) was used as a cut-off. Residues with CSPs less than the cut-off are coloured yellow, 1-2 times the cut-off are orange and over 2 times the cut-off are red. The CSPs are mapped onto the structures in b and c for D76E and D76A, respectively (mapped onto WT structure PDB: 1LDS [151]). The secondary structure elements are shown as ribbons on (a) and (c).

3.3 Discussion

In this chapter the structure, stability and the aggregation of D76N β_2m were examined. The native β_2m fold was shown to be unaffected by the D76N amino acid substitution, this is demonstrated firstly by comparison of the existing crystal structures (Figure 3.1) and in this chapter by comparison of the far-UV CD spectra (Figure 3.4), and ^{1}H-^{15}N HSQC spectra of D76N compared to WT β_2m (Figure 3.12). The D76N amino acid substitution destabilises the β_2m fold, which translates to a $\Delta\Delta G^{\circ}_{UN}$ of 9.5 kJ.mol^{-1}, measured by equilibrium urea denaturation at pH 7.4 and 25 °C (Figure 3.5) and a difference in $T_{m,app}$ of approximately 9°C (Figure 3.4. Table 3.3), consistent with other studies [221, 222, 223]. The lower global stability of D76N is not sufficient however to explain the aggregation propensity of the variant, as it is well established that the stability of β_2m variants does not correlate with their aggregation propensity [217, 177] (Table 3.5).

Table 3.5: Summary of $\Delta\Delta G^{\circ}_{UN}$ and aggregation propensity of β_2m variants

$\Delta\Delta G^{\circ}_{UN}$ values (Data taken from [177], apart from D76N data)

β_2m variant	$\Delta\Delta G^{\circ}_{UN}$ (kJ.mol^{-1})	Fibrils?
WT	-	x
D76N	9.5	✓
ΔN6	3.8	✓
V37A	8.5	x
P5G	1.5	✓
P32G	1.8	Elongation only

Instead, the mechanism of β_2m aggregation is understood to proceed via a folding intermediate with a *trans* proline at position 32, (the I_T-state), and the equilibrium population of this state correlates with β_2m aggregation propensity [176, 177]. One such variant (P32G) which populates the I_T-state at 25% at equilibrium, is not able to aggregate *in vitro* without the presence of fibrillar seeds [176, 177]. This is an important observation when considering the proposal that D76N occupies the I_T-state at 25% at equilibrium [184], because 25% I_T-state is not sufficient to initiate aggregation in P32G. The published conclusion by Mangione et al. [184] that D76N populates the I_T-state at 25% at equilibrium is also not reflected in the CD spectra in Figure 3.4, as a 25% popu-

lation would be distinguishable from the WT spectrum, which is indeed the case for the P32G variant [176, 177], but not for D76N. Furthermore, a 25% population of I_T-state would be detectable in the D76N ^{1}H-^{15}N HSQC spectrum, as shown for P32G [176], as the exchange between the two states occurs on a very slow timescale, limited by proline isomerisation with a half-time in the order of minutes [172, 175, 176] which would present a separate set of peaks belonging to the I_T-state.

The role of the I_T-state in D76N aggregation is further probed in Figure 3.7, where the *in vitro* aggregation of D76N is compared to ΔN6, an I_T-state structural mimic [177, 88]. D76N aggregated more rapidly than ΔN6, in conditions favourable for ΔN6 aggregation [177, 88], which shows that either the D76N I_T-state is exceptionally amyloidogenic, so much so that even at populations akin to WT β_2m, can aggregate faster than the I_T-state structural mimic or, D76N aggregation proceeds in an I_T-state independent manner. The observation that mβ_2m inhibits D76N aggregation, however, suggests that there must be some common basis to D76N and ΔN6 aggregation as mβ_2m inhibits ΔN6 via a specific bimolecular interaction [64]. Also, acceleration of D76N aggregation at pH 6.2, compared to pH 7.4, mirrors the effect observed for ΔN6 [88], again suggesting that there is a common basis for aggregation. The D76N I_T-state is the subject in Chapter 4, where a detailed analysis of the D76N I_T-state is carried out.

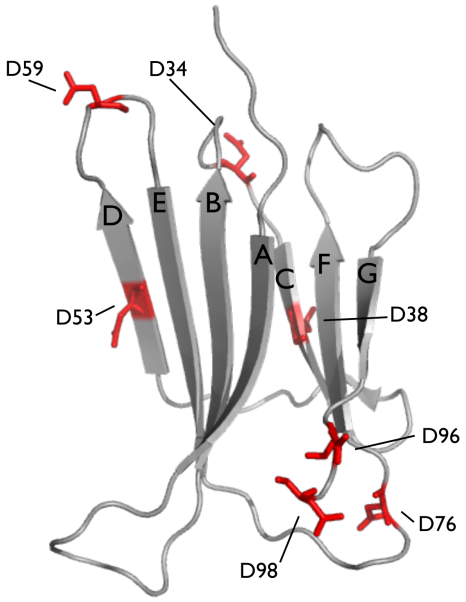


Fig. 3.26: Positions of the seven aspartate resides in β_2m.
The positions of all seven aspartate residues are shown on the WT β_2m structure (PDB:2YXF [231]).

There are seven aspartate (D) residues in β_2m (34, 38, 53, 59, 76, 96, 98) (Figure 3.26, all of which are highly conserved in mammalian β_2m sequences [222]. Systematic substitution of the six aspartate residues to asparagine (N) (with the exception of D76N) did not produce aggregation-prone variants or perturb the β_2m structure or stability of any of the variants [222]. This study showed therefore, that the change in charge resulting from a D-N substitution does not in itself drive the aggregation of β_2m. Instead, the saturation mutagenesis of position 76, carried out in Section 3.2.3, shows the importance of the aspartate residue at position 76, and any change in amino acid either decreases the stability, or increases the aggregation propensity, as measured by the split-β-lactamase assay (Section 3.2.3). The interactions between the D76 side chain must, therefore, be significant in the context of the β_2m structure and its ability to form amyloid. The exquisite specificity for an aspartate at residue 76 in the β_2m structure is perhaps best exemplified by the D76E variant which can form amyloid and is less stable than WT β_2m by 7.9°C, when a D-E substitution retains the same charge and the sidechain only differs by single CH_2. Both the D76E and the D76A variants studied in this chapter were able to form amyloid

and were destabilised compared to WT, however, they were not as unstable as D76N and did not aggregate as fast as D76N either, this suggests that there is a specific interaction formed by N76, which may preclude stabilising interactions still possible in D76E and D76A. The interactions of the N76 residue in D76N are investigated in Chapter 5.

The split-β-lactamase assay used in Section 3.2.3 successfully predicted the aggregation propensity of the D76E and D76A variants. However, the exact physical properties the split-β-lactamase assay is sensitive to remain unclear, as there appears to be some link between hydrophobicity and residue size; however, outliers such as the D76K and D76R substitutions show that the assay is not merely probing a property such as hydrophobicity. The split-β-lactamase assay is powerful, therefore, in the case of position 76 when fundamental properties such as hydrophobicity are not sufficient to explain the effect of mutations. The structural context of a mutation must also be considered, and this assay can glean insight into the role of residues in the structure of a protein by examining outliers such as D76K and D76R. Further work, to understand the split-β-lactamase assay further may include saturation mutagenesis of other positions such as D53 and D98, which are areas unaffected by the D-N substitution [222], and the examination of both destabilised, non aggregating mutations such as V37A [177, 217] or murine β_2m [60] and stable, aggregation-prone variants such as P5G [177].

Chapter 4

Characterisation of the D76N I_T-state using real-time NMR

4.1 Introduction

The folding pathway of β_2m proceeds via a long-lived, structured, folding intermediate [172, 175, 176]. The slow rate of conversion from the structured intermediate to the native folded state of β_2m is hindered by the necessary conversion of the prolyl-peptidyl bond between His31 and Pro32 from a *trans* conformation to an energetically unfavourable *cis* configuration [176, 174, 173]. The folding intermediate with a trans-Pro32 is known as the I_T-state [176, 174, 173]. Mutagenesis of proline residues in β_2m has shown that the equilibrium population of the I_T-state to the N-state is directly proportional to the aggregation rate of the β_2m variant [176, 177]. In fact, a truncated form of β_2m found in DRA patients [232, 171] (known as ΔN6), in which the N-terminal six residues have been cleaved proteolytically becomes trapped in the I_T-state and therefore aggregates more quickly than any β_2m variant studied prior the discovery of D76N, presumably because the ΔN6 β_2m variant cannot escape the I_T-state configuration [88]. This ΔN6 variant is a good structural mimic of the WT β_2m I_T-state and matches the ^{1}H-^{15}N HSQC and far-UV CD spectrum attributed to the I_T-state generated during real-time refolding experiments which directly visualise the I_T-state for WT β_2m [88, 174].

In Chapter 3, no increase in D76N I_T-state population could be observed, additionally *in vitro* aggregation experiments demonstrated that D76N aggregated more quickly than

the ΔN6 variant of β_2m which is trapped in the I_T-state. This prompted the further study, carried out in this chapter to understand how D76N is more aggregation prone than ΔN6.

Through real-time refolding experiments the D76N I_T-state was studied to understand if it was structurally different to the WT I_T-state, leading to a more aggregation prone I_T-state. The rate of refolding for some D76X mutants (Chapter 3) was also probed to investigate the presence of some alternative folding species, which could have an effect on the folding rate. Finally, a ΔN6-D76N mutant was produced to test if this mutant was also an I_T-state mimic, which would allow the aggregation rate of the D76N I_T-state to be investigated.

4.2 Results

4.2.1 Real-time refolding of β_2m variants, monitored using far-UV CD

It is possible to distinguish between ΔN6 and WT β_2m using far-UV CD spectroscopy (200-260 nm) with the former showing a deeper minimum at 216 nm, presumably due to the rearrangement of aromatic side chains in the I_T-N transition [177, 179] (Figure 4.1). These sidechain rearrangements lead to a positive aromatic contribution to the far-UV CD spectrum of natively folded WT β_2m and the resulting spectrum appears to have lower intensity in the far-UV than expected based on its β-strand content. The difference between natively folded β_2m and the spectrum of the I_T-state mimic, ΔN6, is illustrated in Figure 3.4 in Chapter 3. Additionally, various proline variants of WT β_2m studied [177] show that as the population of the I_T-state at equilibrium, measured through size exclusion chromatography and NMR, is reflected in the equilibrium far-UV CD spectrum.

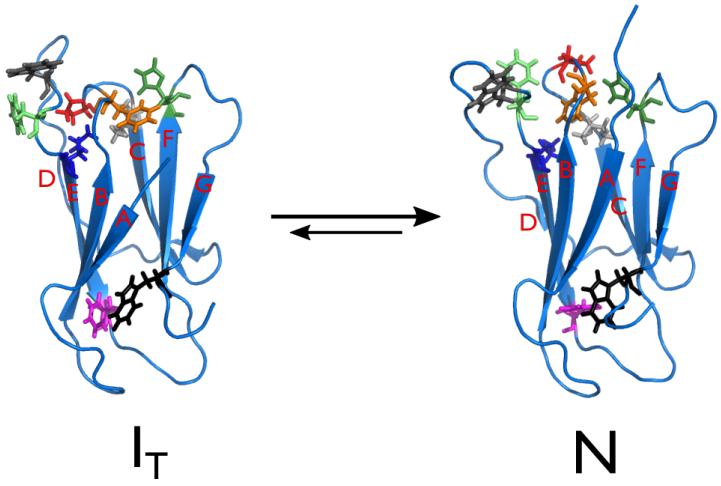


Fig. 4.1: Structural rearrangements involved in the I_T-state to native transition. The ΔN6 (left, PDB:2XKU) and WT β_2m (right, PDB:2XKS) NMR structures [88] reveal the structural transition in the I_T-state to Native state transition. The rearrangement of aromatic sidechains for residues F30 (orange), P32 (red), I35 (light grey), W60 (Dark grey), F62 (light green), L64 (blue), F70 (pink), H84 (dark green) and W95 (black) are shown as sticks.

The difference between the β_2m N-state, which has a *cis*-proline at position 32, and the I_T-state, where the proline at position 32 is *trans*, was utilised to probe the transi-

tion from the I_T-state to the native state in real-time refolding experiments, monitored by far-UV CD. In these refolding experiments described in Methods 2.4.3 (schematic in Figure 4.2), the refolding of acid-unfolded β_2m variants was initiated by the addition of a refolding buffer (pH 7.4 and pH 6.2). Upon the initiation of refolding, the I_T-state is rapidly formed and proceeds to slowly convert into the N-state, over the course of minutes. The isomerisation of proline 32 thus limits the rate of refolding to a time-scale amenable to real-time studies.

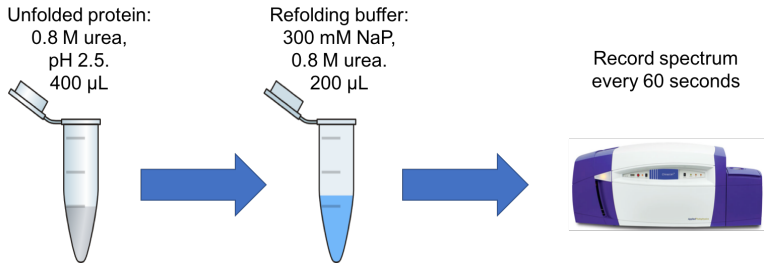


Fig. 4.2: Schematic describing the protocol used for real time refolding reaction studies, monitored by far-UV CD.
In the real-time refolding reaction depicted above, refolding buffer (300 mM sodium phosphate) at the desired pH for refolding is added to the acid unfolded β_2m (buffer: 0.8 M urea, pH 2.5). Once refolding has been initiated, far-UV CD spectra (200-260 nm) are recorded every 60 s for the duration of the time-course (Methods 2.4.3).

In Figure 4.3 the refolding of WT, D76N and the two additional β_2m D76X variants (D76E and D76A) were followed in real time using far-UV CD spectroscopy. In these real-time refolding experiments, far-UV CD spectra in the range 200-260 nm were acquired every 60 seconds, following the initiation of the refolding reaction upon the addition of refolding buffer to the acid unfolded protein. In Figure 4.3 a selection of full spectra are shown, at five-minute intervals and show that the transition from I_T-N can be observed in all variants, characterised by changes in the spectra at 216 nm and 235 nm, confirming that each variant folds via an I_T-like intermediate to form a folded N-state.

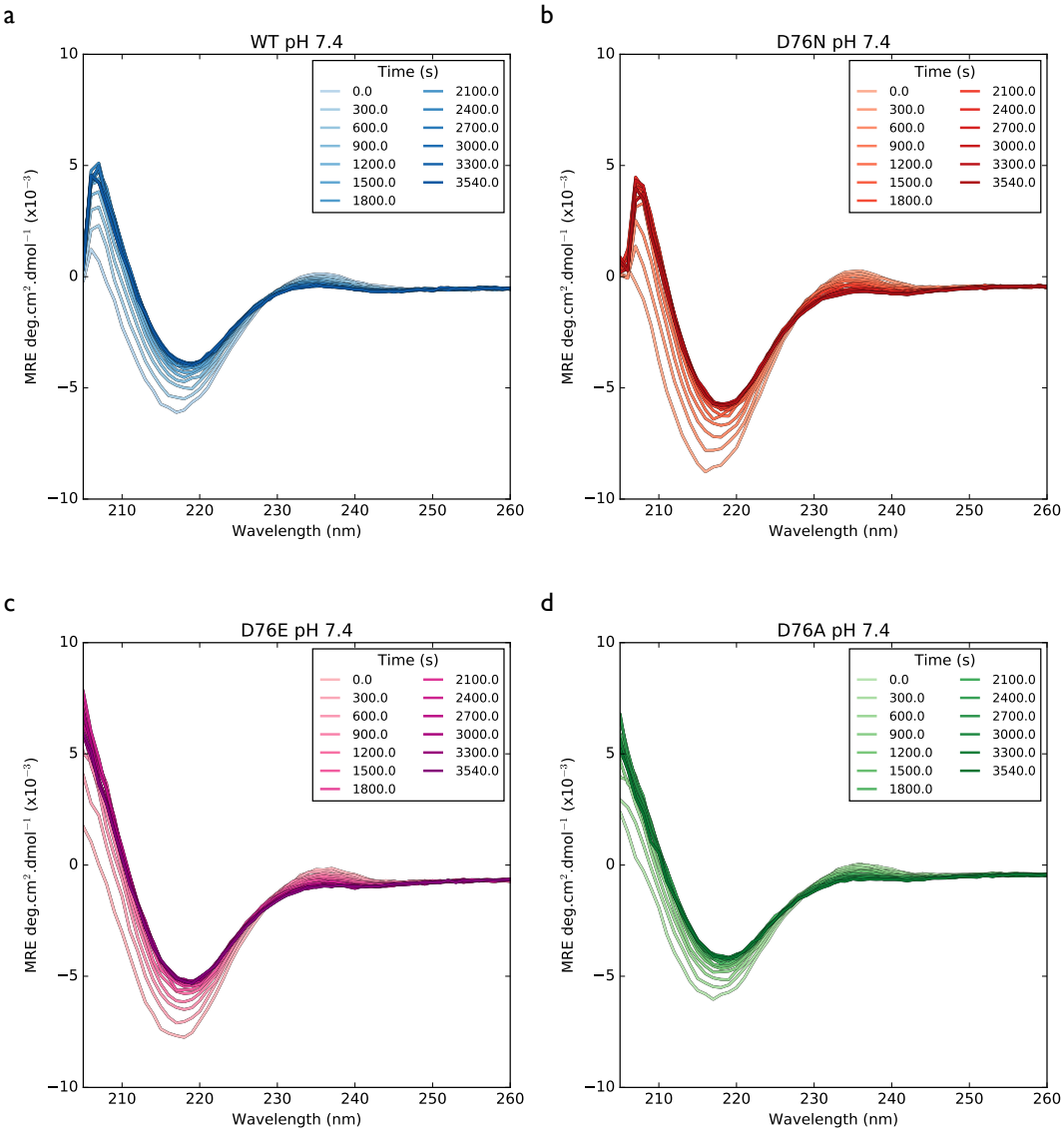


Fig. 4.3: Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy. Far-UV CD spectra recorded for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green. Spectra were recorded every 60 s over the refolding time-course in the spectral range 200-260 nm, however spectra corresponding to 5 minute intervals are shown, for clarity. In all plots, spectra get darker as the time-course progresses, showing the change in the spectrum over time. These experiments were carried out at 20 °C at a final protein concentration of 20 μM in 100 mM sodium phosphate pH 7.4 (Methods 2.4.3).

The changes in the spectra observed during the refolding time course in Figure 4.3 are studied in more detail in Figures 4.4 and 4.5, where the CD intensity at 216 nm and 235 nm during the refolding reaction is plotted versus time. The change in signal at

216 nm can be fitted to a monoexponential equation (Equation 4.2), the rate constants (b) for each variant are presented in Table 4.1.

$$y = ae^{-bx} + c \tag{4.1}$$

Where y is the CD intensity at the chosen wavelength, x is time, c is the value of y at infinite time and a is the starting value and b is the rate constant.

Fitting the data at 235 nm (Figure 4.5) to a monoexponential equation (Equation 4.1) results in rate constants 2-3 times slower than those obtained at 216 nm (Table 4.1). This discrepancy in refolding rate constants at 216 nm and 235 nm may reflect the faster formation of secondary structure (216 nm), followed by the reordering of side chain packing (235 nm) required to form a fully native fold. Regardless, the difference in rate constants at different wavelengths is the same for all variants and is therefore a general effect not related to the aggregation of D76N.

$$y = -ae^{-bx} + c \tag{4.2}$$

Where y is the CD intensity at the chosen wavelength, x is time, c is the value of y at infinite time and a is the starting value and b is the rate constant.

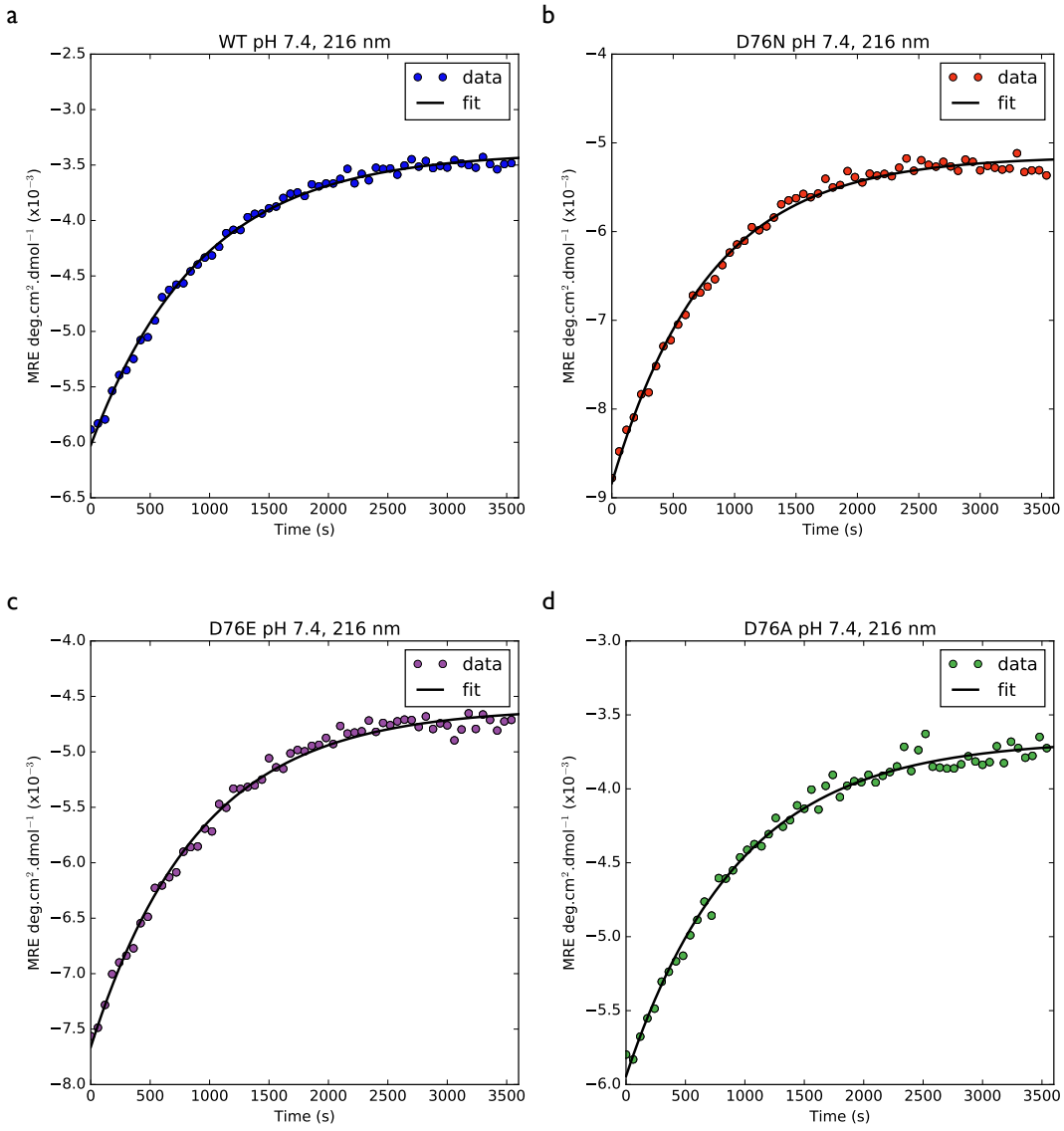


Fig. 4.4: Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy as a function of time at 216 nm.
MRE at 216 nm as a function of time for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green are plotted as circles and fitted to the equation $y = -ae^{-bx} + c$ (Equation 4.2), shown as a black line. These experiments were carried out at 20 °C at a final protein concentration of 20 μM in 100 mM sodium phosphate buffer at pH 7.4, with spectra acquired at 60 s intervals.

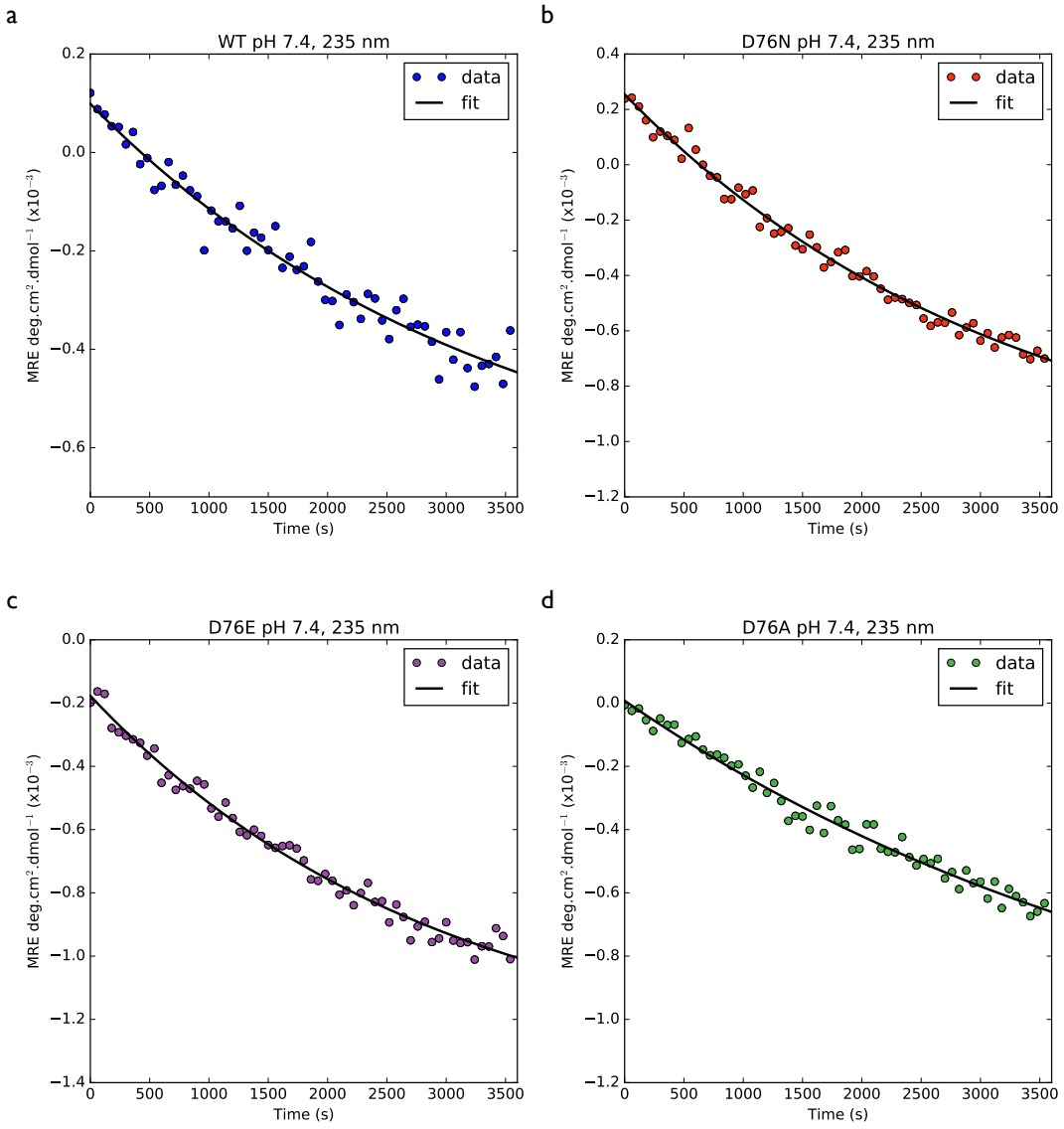


Fig. 4.5: Real-time refolding of β_2m variants at pH 7.4, followed by CD spectroscopy as a function of time at 235 nm.
MRE at 235 nm as a function of time for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green are plotted as circles and fitted to the equation $y = ae^{-bx} + c$ (Equation 4.1), shown as a black line. These experiments were carried out at 20 °C at a final protein concentration of 20 μM in 100 mM sodium phosphate buffer at pH 7.4, with spectra acquired at 60 s intervals.

Table 4.1: Refolding rate constants, measured by far-UV CD at 216 nm and at 235 nm for all β_2m variants at pH 7.4. Observed rate constants for β_2m variants were followed at 216 nm and 235 nm and the data fitted using equations $y = -ae^{-bx} + c$ (Equation 4.2) and $y = ae^{-bx} + c$ (Equation 4.1), respectively. The rate constants are the values for the coefficient, b, in the fitting equations 4.2 and 4.1. Errors are one standard deviation on the fitting parameters. All values are quoted in s^{-1}.

Variant	WT	D76N	D76E	D76A
216 nm	$1.03x10^{-3}$	$1.27x10^{-3}$	$1.10x10^{-3}$	$1.06x10^{-3}$
216 nm error	$2.63x10^{-5}$	$3.23x10^{-5}$	$3.34x10^{-5}$	$3.90x10^{-5}$
235 nm	$2.95x10^{-4}$	$3.1x10^{-4}$	$3.42x10^{-4}$	$1.97x10^{-4}$
235 nm error	$6.11x10^{-5}$	$3.30x10^{-5}$	$3.71x10^{-5}$	$4.34x10^{-5}$

Based on the enhanced aggregation rate of D76N at pH 6.2 compared to pH 7.4 (observed in Chapter 3, Figure 3.6) the refolding reactions were carried out at pH 6.2. In Figure 4.6 the full far-UV CD spectra in the range 200-260 nm are shown; over the same time course as the pH 7.4 studies in Figure 4.3. The spectra under these conditions do not undergo the same dramatic change observed at pH 7.4 despite the fact that the proteins are natively folded at this pH at equilibrium (Chapter 3, Figures 3.9 and 3.13). Instead it appears that the minimum at 216 nm first becomes smaller, then after approximately 20 minutes the spectrum begins to revert back to the original state present at the start of the refolding reaction, this process can be seen in Figure 4.7. In Figure 4.7 the refolding kinetics followed at 216 nm are fit to Equation 4.3 and show the presence of another process, for which there are some small hallmarks of in the pH 7.4 data at 216 nm in Figure 4.3.

$$y = -ae^{-bx} + mx + c \tag{4.3}$$

Where y is the CD intensity at the chosen wavelength, x is time, the constant c represents both the value of y at infinite time and the y-intercept of the straight-line component of the equation. a is the starting value and b is the rate constant. m is the gradient of the straight line component of the curve.

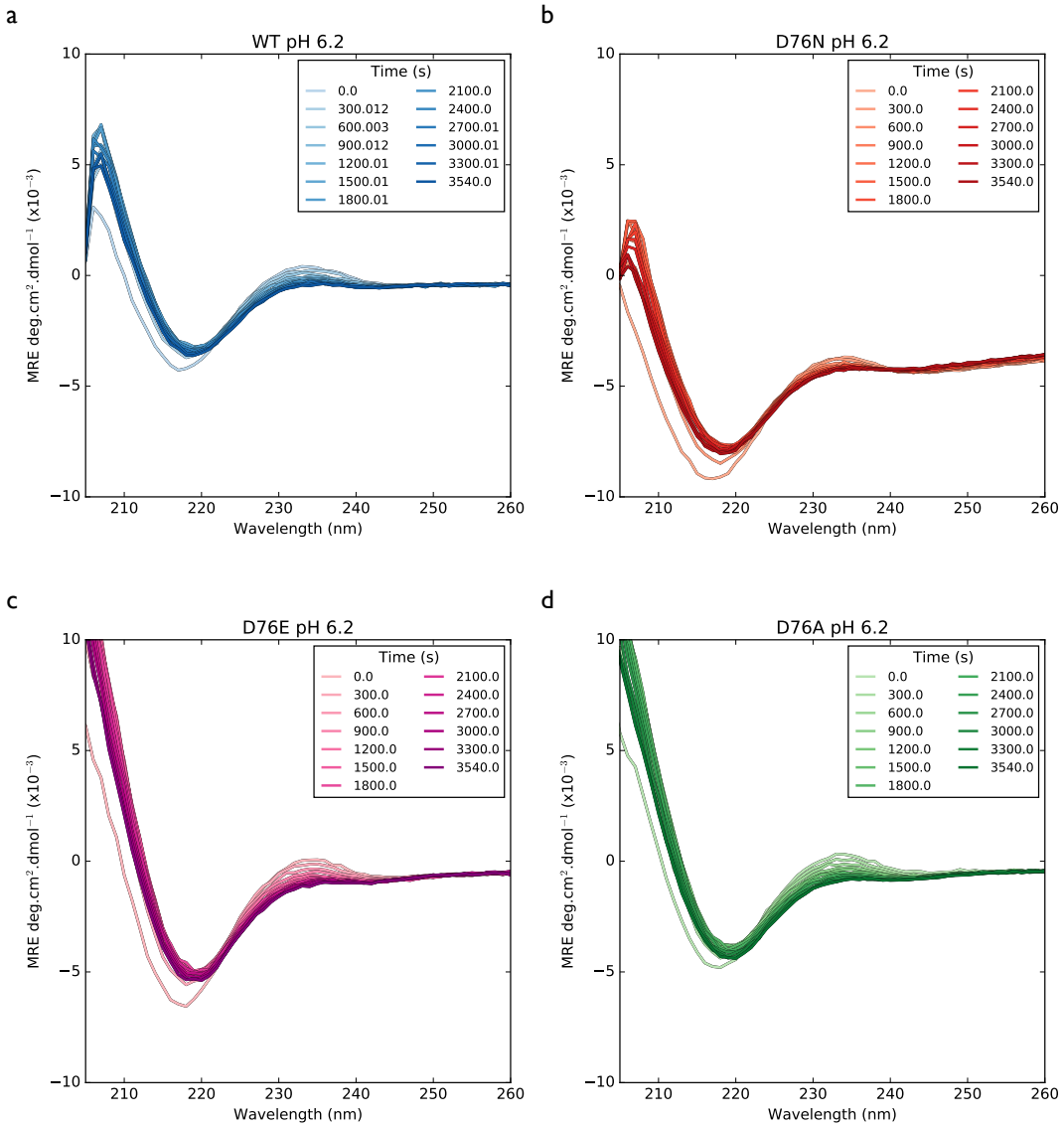


Fig. 4.6: Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy. Far-UV spectra recorded for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green. Spectra were recorded every 60 s over the refolding time-course in the spectral range 200-260 nm, however spectra corresponding to 5 minute intervals are shown, for clarity. In all plots, spectra get darker as the time-course progresses, showing the change in the spectrum over time. These experiments were carried out at 20 °C at a final protein concentration of 20 μM in 100 mM sodium phosphate buffer at pH 6.2.

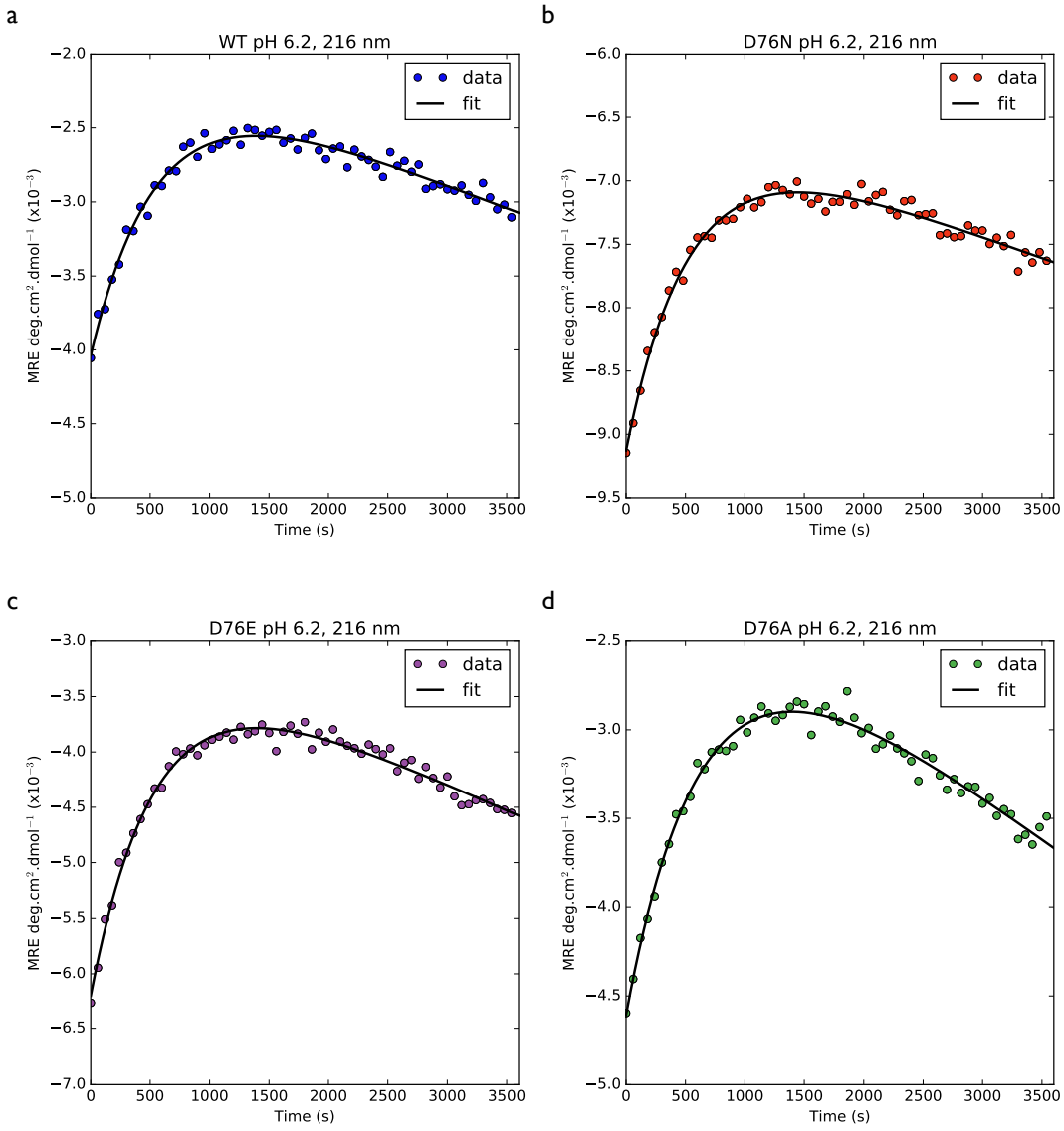


Fig. 4.7: Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy as a function of time at 216 nm.
MRE at 216 nm as a function of time for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green are plotted as circles and fit to the equation $y = -ae^{-bx} + mx + c$ (Equation 4.3), shown as a black line. These experiments were carried out at 20 °C at a final protein concentration of 20 µM in 100 mM sodium phosphate buffer at pH 6.2, with spectra acquired at 60 s intervals.

The refolding kinetics as a function time at 235 nm at pH 6.2 (Figure 4.8) do not follow the same kinetics observed at 216 nm (Figure 4.7) at pH 6.2, where the signal begins to decay after approximately 20 minutes. Instead the data at 235 nm appear to follow a single process and can be fitted to the monoexponential equation (Equation 4.1). The rate constants of refolding at pH 6.2 are summarised in Table 4.2.

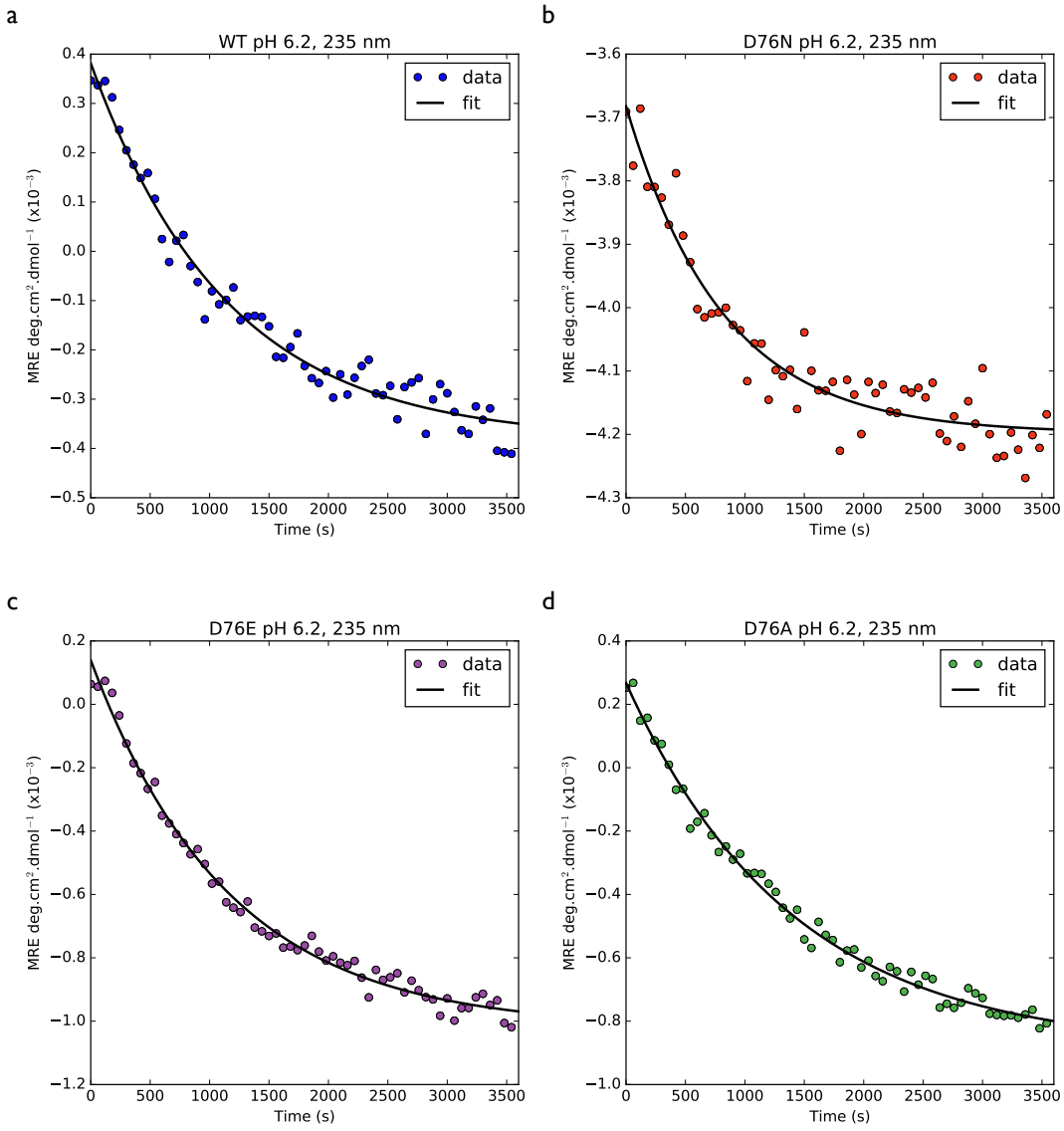


Fig. 4.8: Real-time refolding of β_2m variants at pH 6.2, followed by CD spectroscopy as a function of time at 235 nm.

MRE at 235 nm as a function of time for WT β_2m (a) in blue, D76N β_2m (b) in red, D76E β_2m (c) in purple and D76A β_2m (d) in green are plotted as circles and fit to the equation $y = ae^{-bx} + c$ (Equation 4.1) , shown as a black line. These experiments were carried out at 20 °C at a final protein concentration of 20 μM in 100 mM sodium phosphate buffer at pH 6.2, with spectra acquired at 60 s intervals.

Table 4.2: Refolding rate constants, measured by far-UV CD at 216 nm and at 235 nm for all β_2m variants at pH 6.2. Observed rate constants for β_2m variants monitored at 216 nm and 235 nm using equations $y = -ae^{-bx} + mx + c$ (Equation 4.3) and $y = ae^{-bx} + c$ (Equation 4.1), respectively. The rate constants are the values for the coefficient, b, in the fitting equations 4.3 and 4.1. Errors are one standard deviation on the fitting parameters. All values are in s^{-1}.

Variant	WT	D76N	D76E	D76A
216 nm	$1.67x10^{-3}$	$1.87x10^{-3}$	$1.84x10^{-3}$	$1.52x10^{-4}$
216 nm error	$1.06x10^{-4}$	$1.02x10^{-4}$	$8.66x10^{-4}$	$3.55x10^{-5}$
235 nm	$8.76x10^{-4}$	$1.23x10^{-3}$	$8.66x10^{-4}$	$7.19x10^{-4}$
235 nm error	$6.01x10^{-5}$	$1.2x10^{-4}$	$3.58x10^{-5}$	$3.28x10^{-5}$

The real-time refolding experiments carried out at pH 7.4 show that WT and D76N both fold via an I_T-state and that the refolding proceeds at a similar rate. The addition of the D76A and D76E variants to this analysis indicates that residue 76 does not appear to have a critical role in the refolding process, as both of these variants also fold at a similar rate. Interestingly, for all variants studied here, the rate constants observed by following the change at 216 nm is different from the refolding rate constants observed at 235 nm (Tables 4.1 and 4.2). These different rate constants at 216 nm and 235 nm indicate that the transition from the I_T-state to a native fold is a complex process with more than one process involved, perhaps due to packing of aromatic side chains as a result of the transition from *trans*-to-*cis* in proline 32 (Figure 4.1).

Under conditions in which D76N aggregates more rapidly *in vitro*, the refolding process appears to become more complex, still. It should be noted that visible precipitation was observed when the refolding reaction was carried out at pH 6.2, and the aggregation processes may be complicating the analysis further. Since the processes attributed to refolding and aggregation cannot be differentiated, no further refolding experiments at pH 6.2 were carried out. However, pH 6.2 remains a useful condition for studying other processes such as *in vitro* aggregation kinetics.

4.2.2 Real-time refolding of WT and D76N β_2m, followed by NMR

Previous work has shown that the I_T-state, can be visualised by NMR during the refolding process and that this intermediate has a distinct ^{1}H-^{15}N HSQC spectrum from that of the natively folded protein [88, 233, 234, 235].

In the following work, the I_T-states of both WT and D76N were visualised using ^{1}H-^{15}N NMR, to identify differences in chemical shift and to monitor refolding at the single residue level of resolution. These studies aimed to unravel the role of the I_T-state in D76N aggregation, given the fact that there does not appear to be any detectable increase in I_T-state population for D76N at equilibrium. Together with the fact that D76N aggregates more rapidly than the WT I_T-state-mimic ΔN6, thus poses the question of whether the D76N I_T-state adopts some highly aggregation-prone state, distinct form the I_T-state of the WT protein.

To begin to answer these questions, real-time refolding reactions as described in Methods 2.7.4 and Figure 4.9 were carried out, and ^{1}H-^{15}N HMQC spectra were recorded every 60 seconds for 180 minutes.

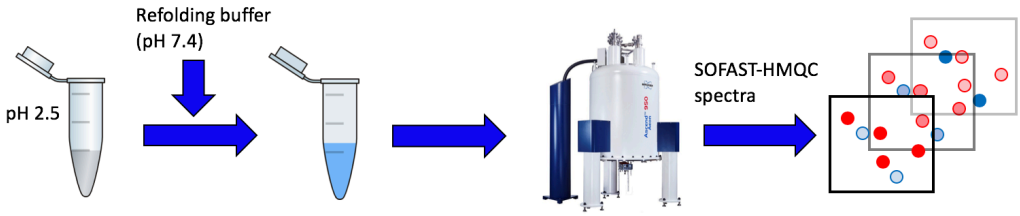


Fig. 4.9: Schematic of the real-time refolding reaction, followed by NMR. The sample protein is acid unfolded in unfolding buffer (1.5M urea, pH 2.5). Refolding is initiated upon the addition of refolding buffer (0.5M sodium phosphate, pH 7.4) raising the pH of the sample to pH 7.4. The sample is loaded into the NMR spectrometer and a SOFAST-HMQC experiments are recorded every 60 seconds for 180 minutes. The dead time from initiation of refolding to the start of data acquisition is 30 s. The change in NMR spectrum is monitored with time as peaks corresponding to the I_T-state disappear and peaks belonging to the natively folded state appear with time, as the *trans-cis* proline 32 isomerisation occurs and the protein transitions from the I_T-state to the native state.

The first spectrum acquired in the refolding reaction, after a dead time of 30 s and an acquisition time of 60 s is compared to the final spectrum acquired after 180 minutes

for WT in Figure 4.10 and D76N in Figure 4.11. For both WT (Figure 4.10) and D76N (Figure 4.11), the first spectrum (90 s, a) is well dispersed, indicating that the species present in the spectrum are at least partially folded. The spectrum of the species present at 90 s has clear differences to the spectrum of the natively folded species present after 180 min (b). In the spectrum recorded after 90 s, a distinct subset of peaks are observed, which disappear with time; these peaks correspond to the loss of I_T-state and gradually lose intensity as the refolding reaction proceeds and the population of I_T-state declines as it converts to the native folded state. On the other hand, the first spectrum also has another subset peaks which share a common chemical shift with peaks observed in the final, native spectrum, which can be observed in the overlay of the spectra 90 s and 180 min (Figure 4.10c and Figure 4.11c). The set of peaks that are common to both the I_T-state and the N-state represent the structural similarity between the two states, which can be observed in the structures of ΔN6 and WT β_2m shown in Figure 4.1. In addition to the I_T-state specific peaks, and the common peaks, there is also third a subset of peaks which are not present in the early phases of the time course, which grow in intensity over time, belonging to the N-state.

Peaks present in the early (90 s) spectra for both WT (Figure 4.10) and D76N (Figure 4.11), which belong to the I_T-state, are not detectable at equilibrium, further supporting the observation that an increased I_T-state population is not apparent in D76N at equilibrium.

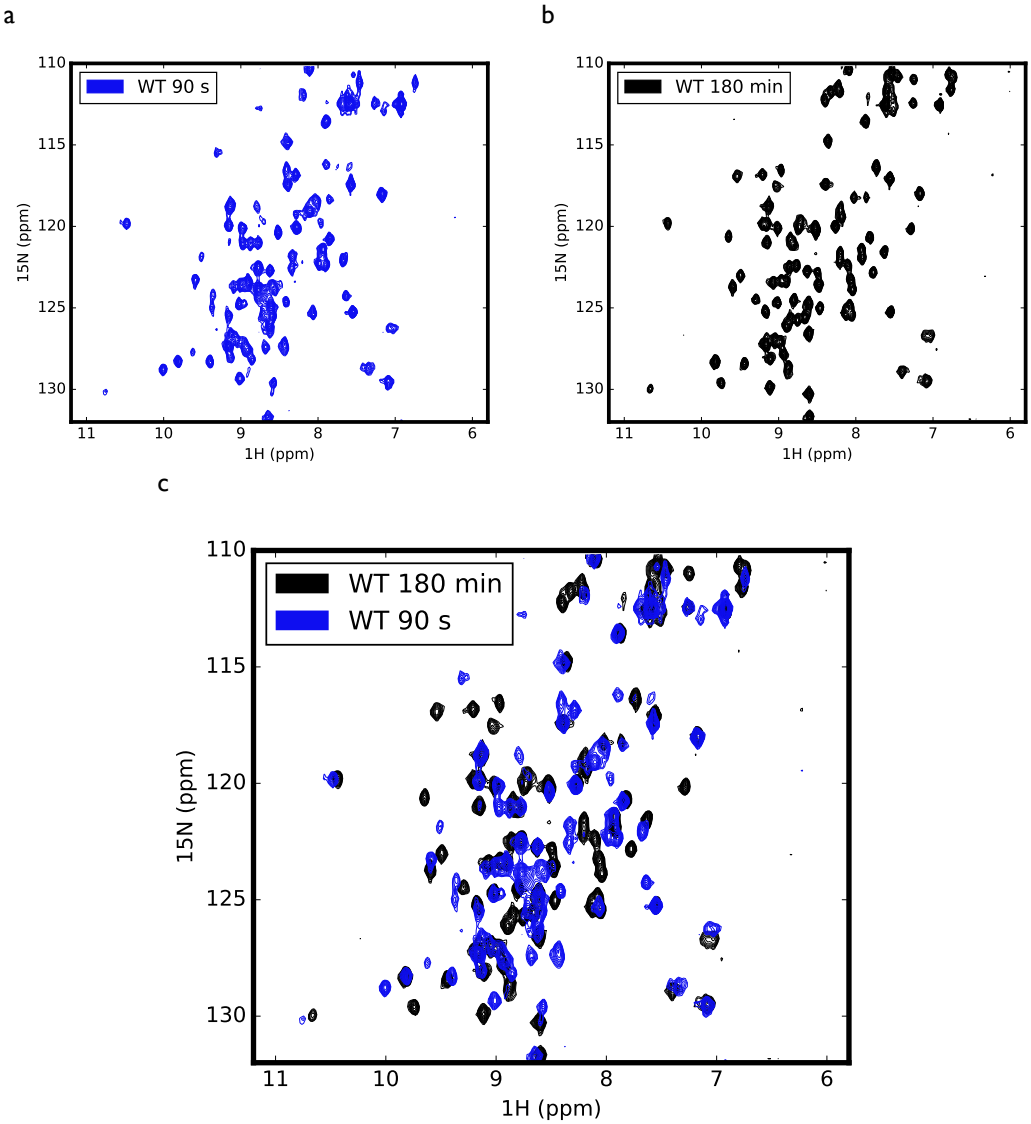


Fig. 4.10: Real time refolding of WT β_2m, monitored by NMR. (a) ^{1}H-^{15}N-SOFAST-HMQC spectrum recorded over a duration of 60 s after a dead time of 30 s from the initiation of the real-time refolding timecourse. (b) ^{1}H-^{15}N-SOFAST-HMQC spectrum 180 min after refolding was initiated, where no further changes in preceding ^{1}H-^{15}N-SOFAST-HMQC spectra are observed. (c) An overlay of the spectra from (a) and (b). These experiments were carried out at 20 °C at a final protein concentration of 300 µM at pH 7.4 in 1.0 M urea, 166.6 mM sodium phosphate buffer. Spectra were acquired at 60 s intervals, throughout the refolding process.

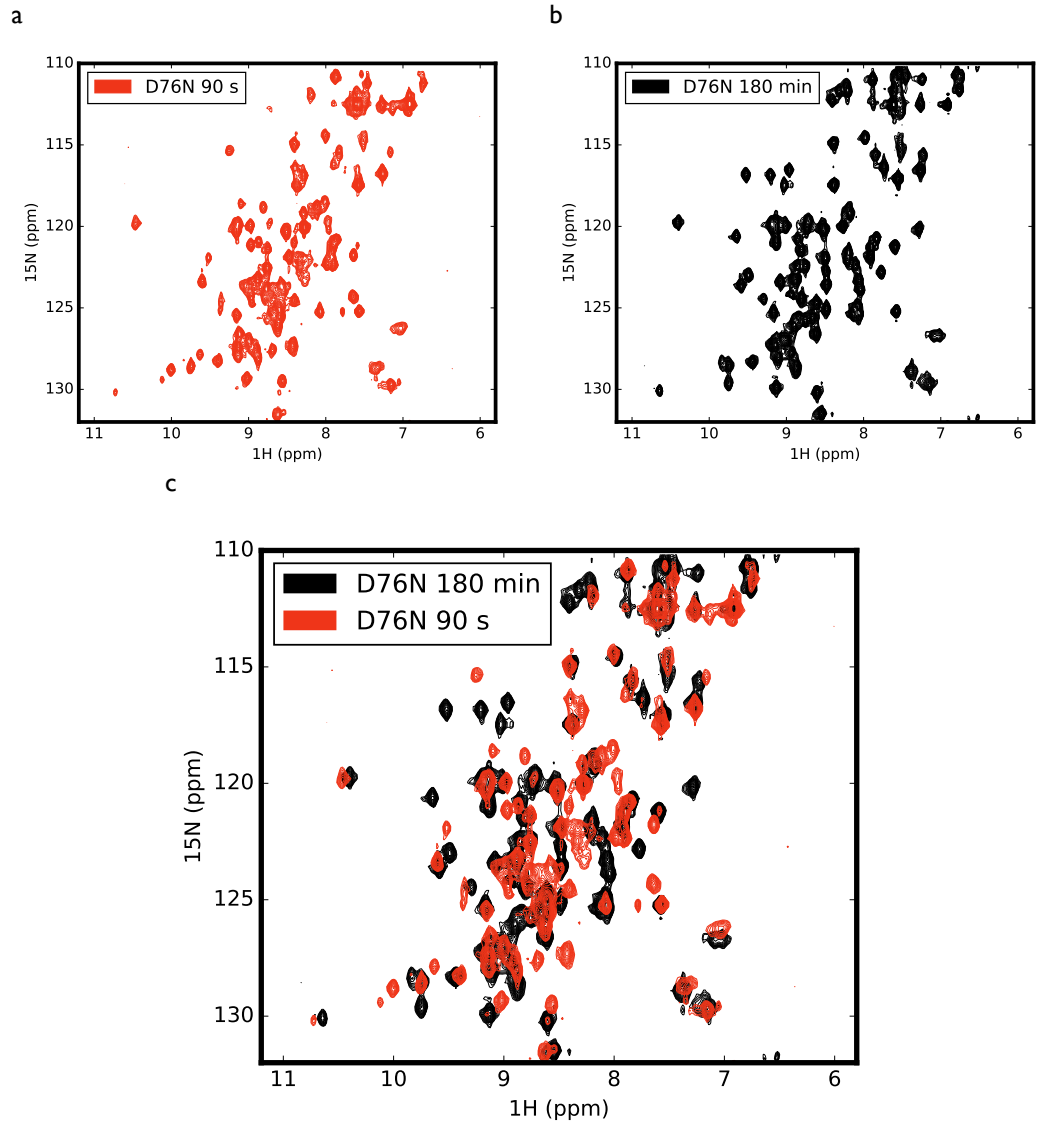


Fig. 4.11: Real time refolding of D76N β_2m, monitored by NMR. (a) ^{1}H-^{15}N-SOFAST-HMQC spectrum recorded over a duration of 60 s after a dead time of 30 s from the initiation of the real-time refolding timecourse. (b) ^{1}H-^{15}N-SOFAST-HMQC spectrum 180 min after refolding was initiated, where no further changes in preceding ^{1}H-^{15}N-SOFAST-HMQC spectra are observed. (c) An overlay of the spectra from (a) and (b). These experiments were carried out at 20 °C at a final protein concentration of 300 µM at pH 7.4 in 1.0 M urea, 166.6 mM sodium phosphate buffer. Spectra were acquired at 60 s intervals, throughout the refolding process.

Before in-depth chemical shift perturbation (CSP) analysis (Methods 2.7.3), and single residue refolding kinetic studies (Methods 2.7.4) could be carried out, assignments for the I_T-state peaks in the early spectra for WT and D76N were required. While it is possible

to carry out triple resonance experiments required for backbone assignment within the time course of these experiments [235, 236, 237], the structural similarity between ΔN6 and the WT I_T-state was utilised to transfer assignments for ΔN6 to the WT I_T-state. WT I_T-state assignments were transferred to the D76N I_T-state where possible. While direct backbone assignment of the D76N I_T-state would be ideal, to confirm these assignments, this method was sufficient to carry out CSP and kinetic analysis of the D76N I_T-state. The assignments used in the following analyses are shown in Figures 4.12 and 4.13.

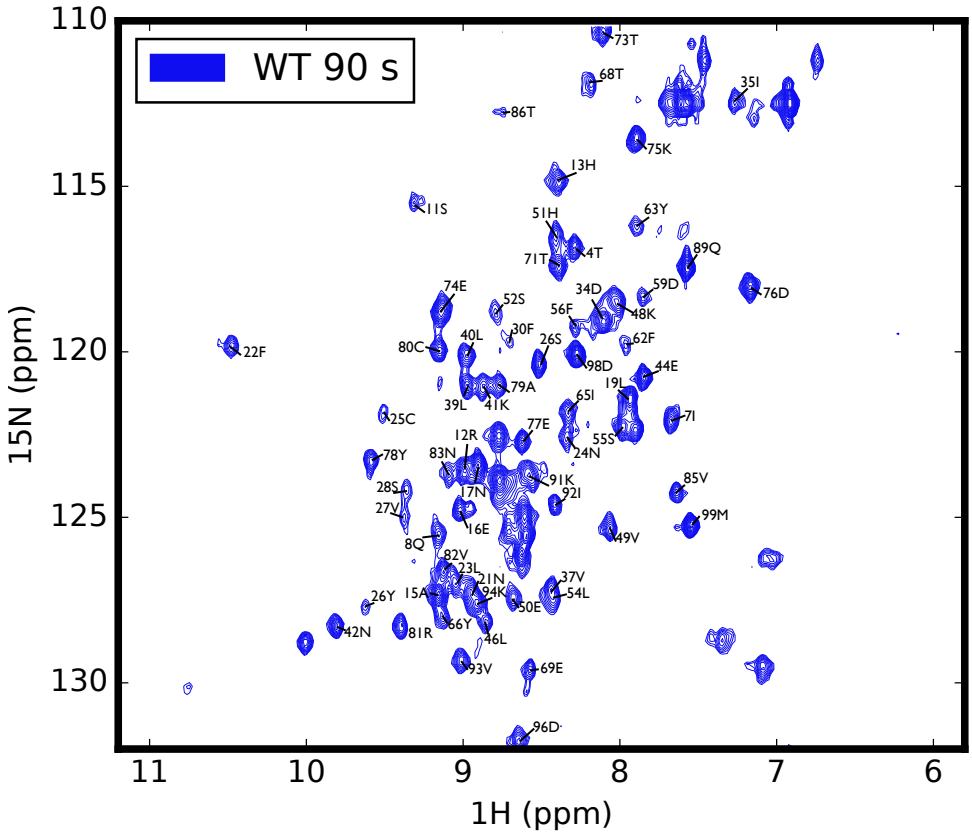


Fig. 4.12: ^{1}H-^{15}N assignments for the WT β_2m 90 s refolding spectrum, used for CSP and refolding kinetic analysis.
The assignments used here were transferred from the assignment of the ΔN6 spectrum (BMRB:17166 [88]). Some residues are not assigned due to spectral crowding.

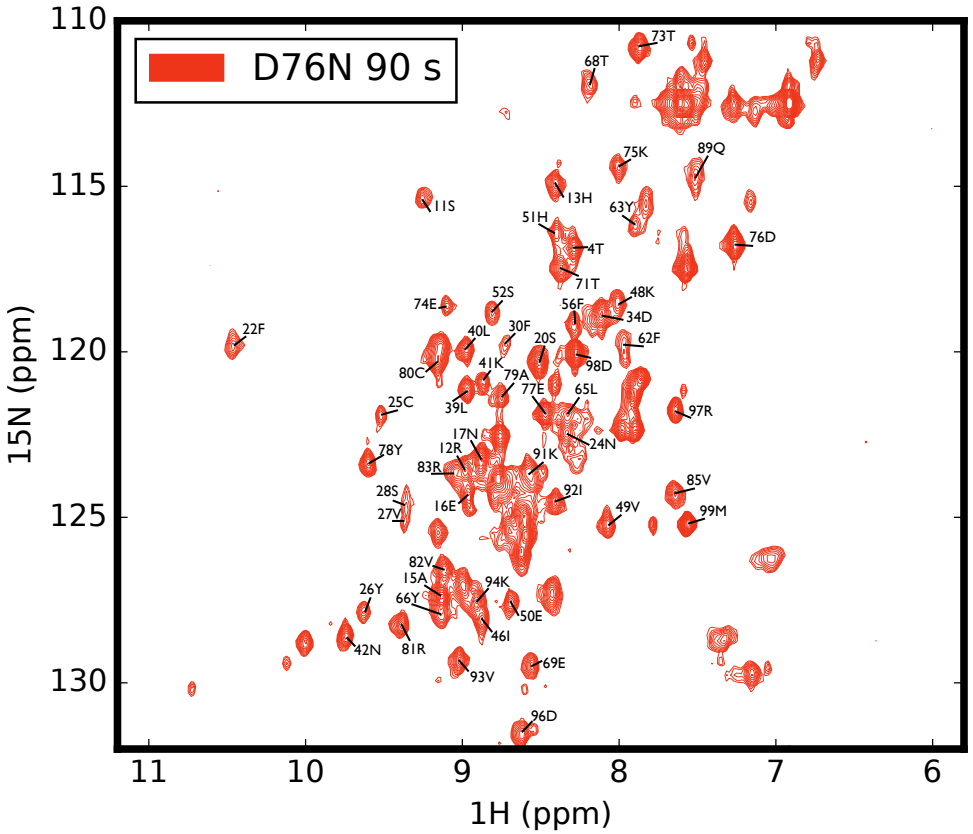


Fig. 4.13: ^{1}H-^{15}N assignments for the D76N β_2m 90 s refolding spectrum, used for CSP and refolding kinetic analysis.
The assignments used here were transferred from the a combination of the assignments of the WT ΔN6 spectrum (BMRB:17166 [88]) and the assignments for the D76N N-state spectrum assigned in Chapter 3. Some residues are not assigned due to spectral crowding.

To further investigate the differences between the I_T-state and N-state spectra observed in Figure 4.10 and 4.11 (WT and D76N, respectively), chemical shift perturbation analysis was carried out (Methods 2.7.3). The chemical shift perturbation (CSP) analysis in Figure 4.14 compares the first and last spectra of the refolding time-course for WT and D76N, using the assignments shown in Figures 4.12 and 4.13. In this analysis, the change in chemical shift is calculated with Equation 2.6.

Plots in Figure 4.14, show the magnitude of CSP between the N and I_T-states for WT (a) and D76N (b), coloured based upon the standard deviation from the mean average CSP. The CSP analysis in Figure 4.14 shows that the overall pattern of CSPs between the N and I_T-states are remarkably similar between WT and D76N, the most substantial differences for both proteins are found in the BC loop, in which the proline 32 residue responsible for the I_T-N transition is found. However, there are other areas of significant CSP, found in the D strand, which is known to be conformationally dynamic in β_2m [88] and in the FG loop, adjacent to the protein 32-containing BC loop. There is also a lack of resonances corresponding to the DE loop, this could be due to lack of assignment, as there are a number of unassigned peaks in the spectra in Figures 4.12 and 4.13, or could be due to exchange broadening due to the dynamics of the DE-loop. The CSPs are mapped onto the structures of WT and D76N in Figure 4.14c and d, respectively. This information indicates that D76N undergoes a very similar structural rearrangement in the I_T-N transition to WT β_2m.

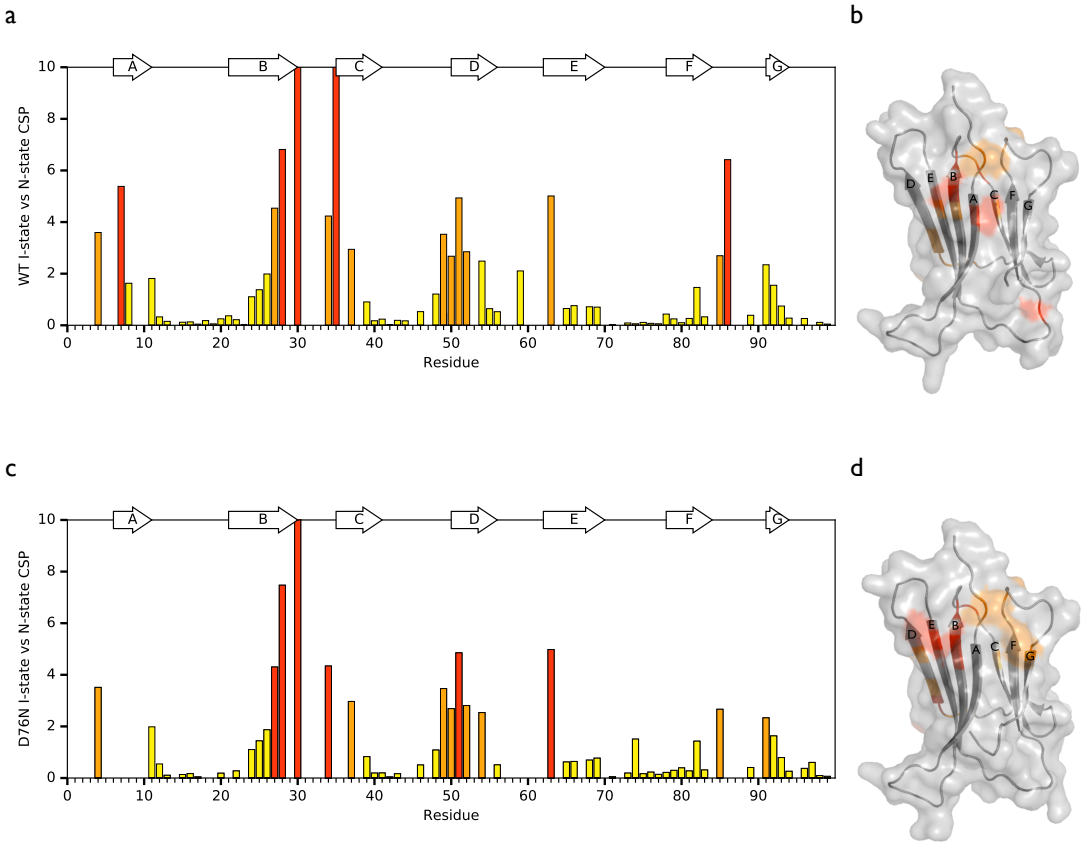


Fig. 4.14: Chemical shift differences between spectra recorded at 90 s and 180 min, during the real-time refolding reaction at pH 7.4.
CSPs between the 90 s spectrum and the 180 min spectrum, calculated with the equation $\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2}$ (Equation 2.6). CSPs less than one standard deviation from the mean average of all CSPs are coloured yellow, orange corresponds to 1-2 standard deviations and red corresponds to over 2 standard deviations from the mean. The CSPs from WT (a) and D76N (b) are mapped onto the structures in (c) and (d), respectively (PDB: 1LDS (WT) [151] and 4FXL (D76N) [183]), using the same colour code.

The analysis in Figure 4.14 shows that the differences between the N-state and I_T-state for WT are similar in location and magnitude to the differences observed between the D76N N-state and I_T-state. In Figure 4.15, the differences in ^{1}H-^{15}N HMQC spectra between the I_T-states for D76N and WT are compared. In Figure 4.15a the first spectrum in the time course for WT and D76N are compared and show a remarkably similar dispersion pattern, indicating that the two intermediate species share the same overall fold. The differences in chemical shift are quantified in Figure 4.15b and are mapped

onto the structure of ΔN6 in Figure 4.15c. The CSPs observed here are consistent with those observed between the native folds for WT and D76N studied in Chapter 3 (Figures 3.12 and 3.13), where the most significant chemical shift differences are observed in the residues spatially adjacent to the mutation site at position 76, in the EF-loop. The CSP analysis in Figure 4.15 provides further indication that the WT and D76N I_T-states are structurally very similar.

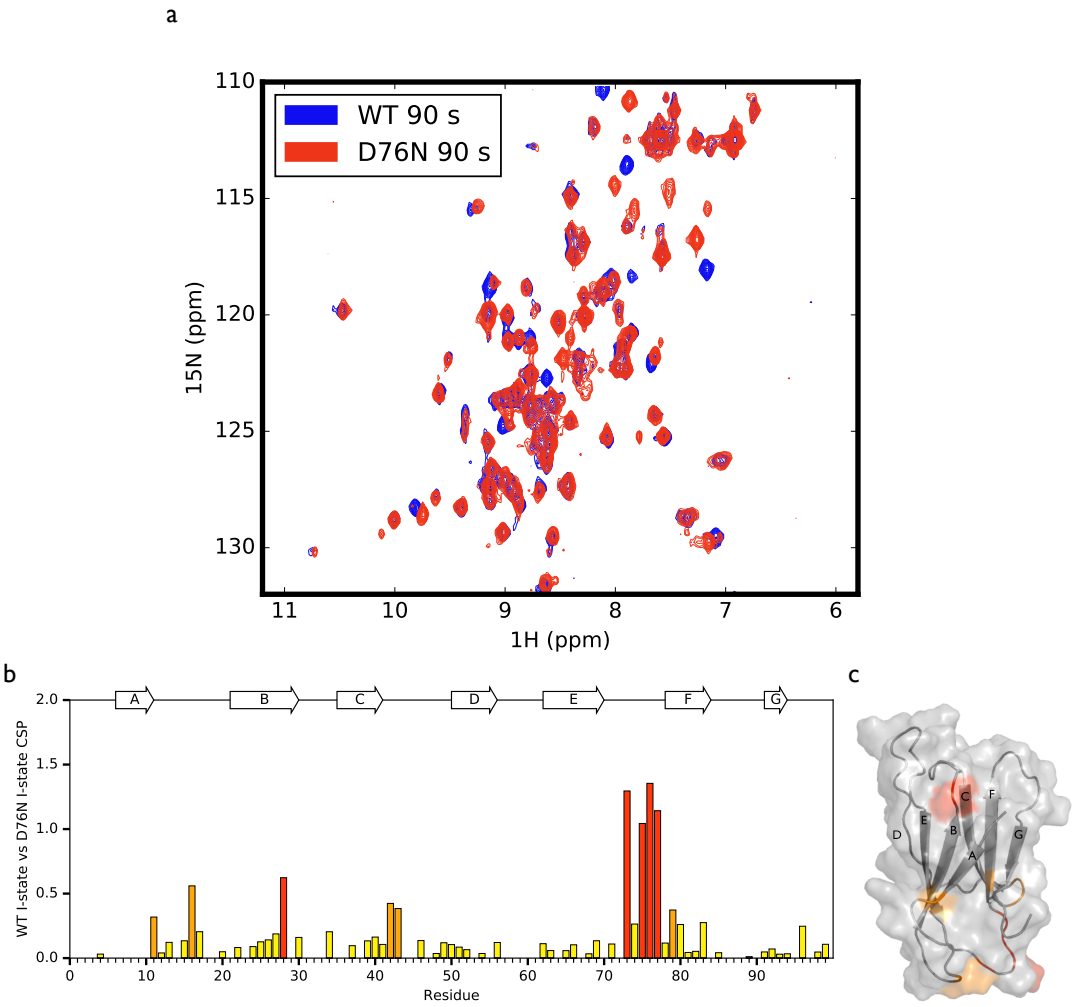


Fig. 4.15: Comparison of WT and D76N ^{1}H-^{15}N HMQC spectra recorded at 90 s, during the real-time refolding reaction.
The WT and D76N ^{1}H-^{15}N HMQC spectra recorded at 90 s in the real time refolding reaction at pH 7.4, 20° are shown in (a), the CSPs between the two spectra in (a) are shown in (b) and mapped onto the structure of ΔN6 (PDB:2XKU) in (c). CSPs between the WT and D76N spectra at 90 s were calculated with the equation $\Delta\delta\omega = \sqrt{\left(5 \cdot \delta^{1}\mathrm{H}\right)^2 + \left(\delta^{15}\mathrm{N}\right)^2}$. CSP less than one standard deviation from the mean average of all CSPs are coloured yellow, orange corresponds to 1-2 standard deviations and red corresponds to over 2 standard deviations from the mean.

As mentioned previously, the real-time refolding NMR experiments discussed here were carried out as a time-course with a spectrum recorded every 60 s for the full 180 min duration of the refolding time-course. Therefore it is possible to carry out kinetic analysis at the single residue level for the entire folding transition between I_T-N, particularly for resonances which are unique to either the I_T-state or the N-state. Such

an analysis was carried out for peaks corresponding to the N-state in Figure 4.16 and peaks corresponding to the I_T-state in Figure 4.17 for WT and D76N.

The kinetic analysis in Figure 4.16 was carried out by following the change in volume for peaks corresponding to the native state as a function of time and fitting the volumes to the monoexponential equation (Equation 4.2) used previously (Methods 2.7.4). Within this data set, there were peaks which shared a chemical shift with peaks in the I_T-state, which could not be fitted to the monoexponential process sufficiently. To filter out the peaks which did not fit to a single process, due to overlapping peaks, fits which had a standard deviation of more than 10% of the value for the rate constant (b in Equation 4.2) were removed from this analysis, (plots without this cut-off applied are shown in Appendix A) In Figure 4.16a-d, representative fits for the increase in volume of native resonances with time are shown. In Figure 4.16e,f the per-residue observed rate constants for WT (e) and D76N (f) are shown. The results show that the majority of peaks share the same rate constant of approximately 6.0×10^{-4} s^{-1}, consistent with previously observed rate constants for WT refolding by NMR [233]. The mean average rate constant of folding for the residues within the 10% cut off is $6.1 \times 10^{-4} \pm 6.2 \times 10^{-5}$ s^{-1} for WT and $6.3 \times 10^{-4} \pm 4.7 \times 10^{-5}$ s^{-1} for D76N. The results also show that WT and D76N fold at a similar rate, consistent with the observations made by monitoring the refolding process by far-UV CD (approx. 3×10^{-4} at 235 nm, Table 4.1). In Figure 4.16f there are rates for residues around the mutation site in the EF loop which appear to refold more slowly, this is due the peaks in question overlapping with I_T-state peaks, which makes the apparent rate appear slower, the same peaks are not shown in the WT data (Figure 4.16e) as the fits are not within the cut-off applied, however a similar pattern is observed (Appendix A).

A similar analysis of I_T-state decay was carried out in Figure 4.17 which shows the rate constant for the decay of I_T-state, where volumes as a function of time are fit to Equation 4.1. While this data set is sparse in comparison with the data in Figure 4.16, the results show that the rate constants for the reduction in population of I_T-state between WT and D76N are similar, but slightly faster in D76N ($5.6 \times 10^{-4} \pm 3 \times 10^{-5}$ s^{-1} for WT and $4.6 \times 10^{-4} \pm 9.75 \times 10^{-7}$ s^{-1} for D76N). Again, fits which had a standard deviation of more

than 10% of the value for the rate constant were removed from this analysis and the plots without this cut-off applied are shown in Appendix A).

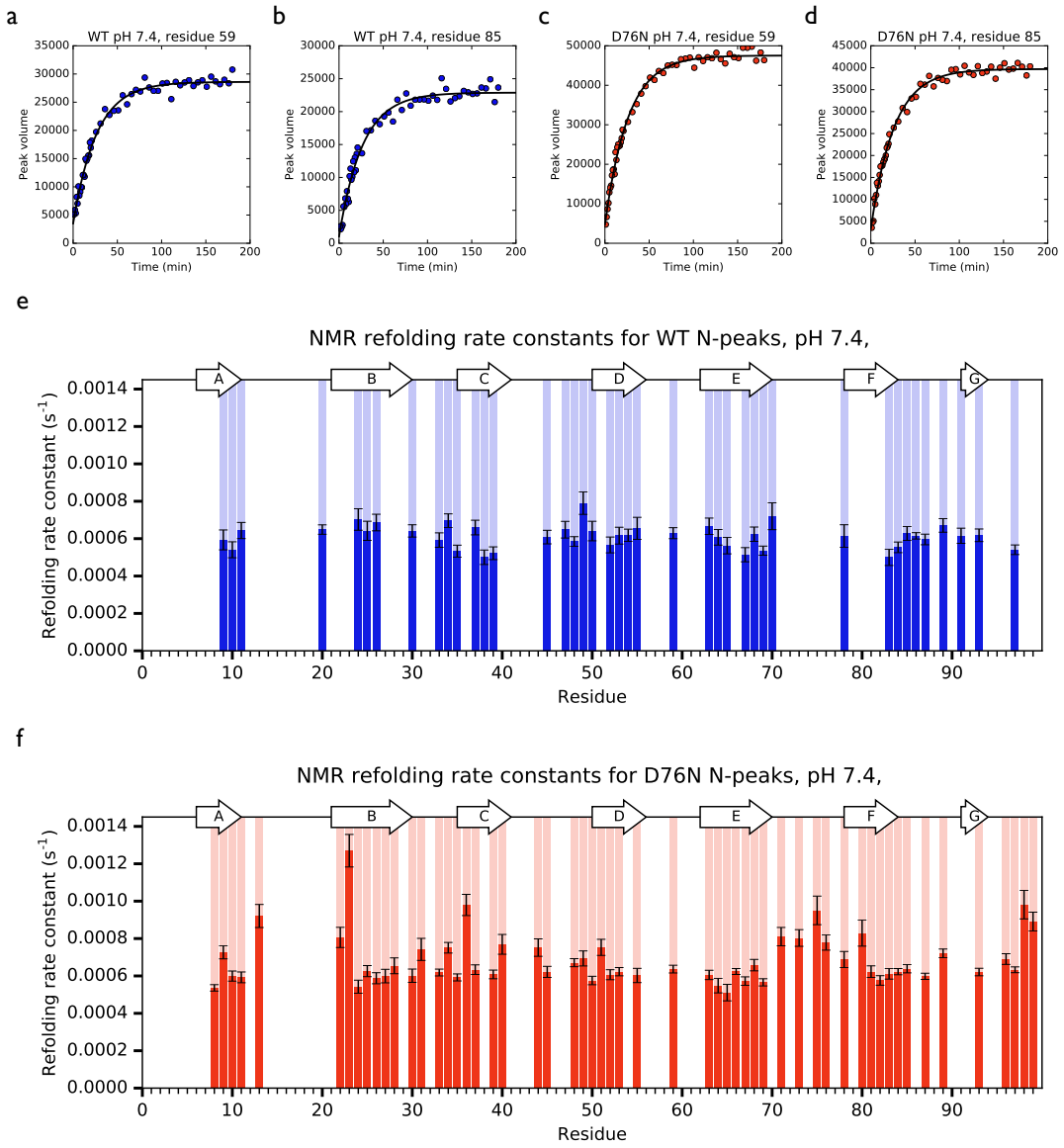


Fig. 4.16: Single residue refolding rates for N-state peaks.
(a-d) show representative data (WT in blue, D76N in red) and fits (black) for single residue refolding rates fitted to the equation $y = -ae^{-bx} + c$ (Equation 4.2). The rate constants for residues where the standard deviation on the fit is no more than 10% of the rate constant value are shown in (e) and (f) for WT and D76N, respectively.

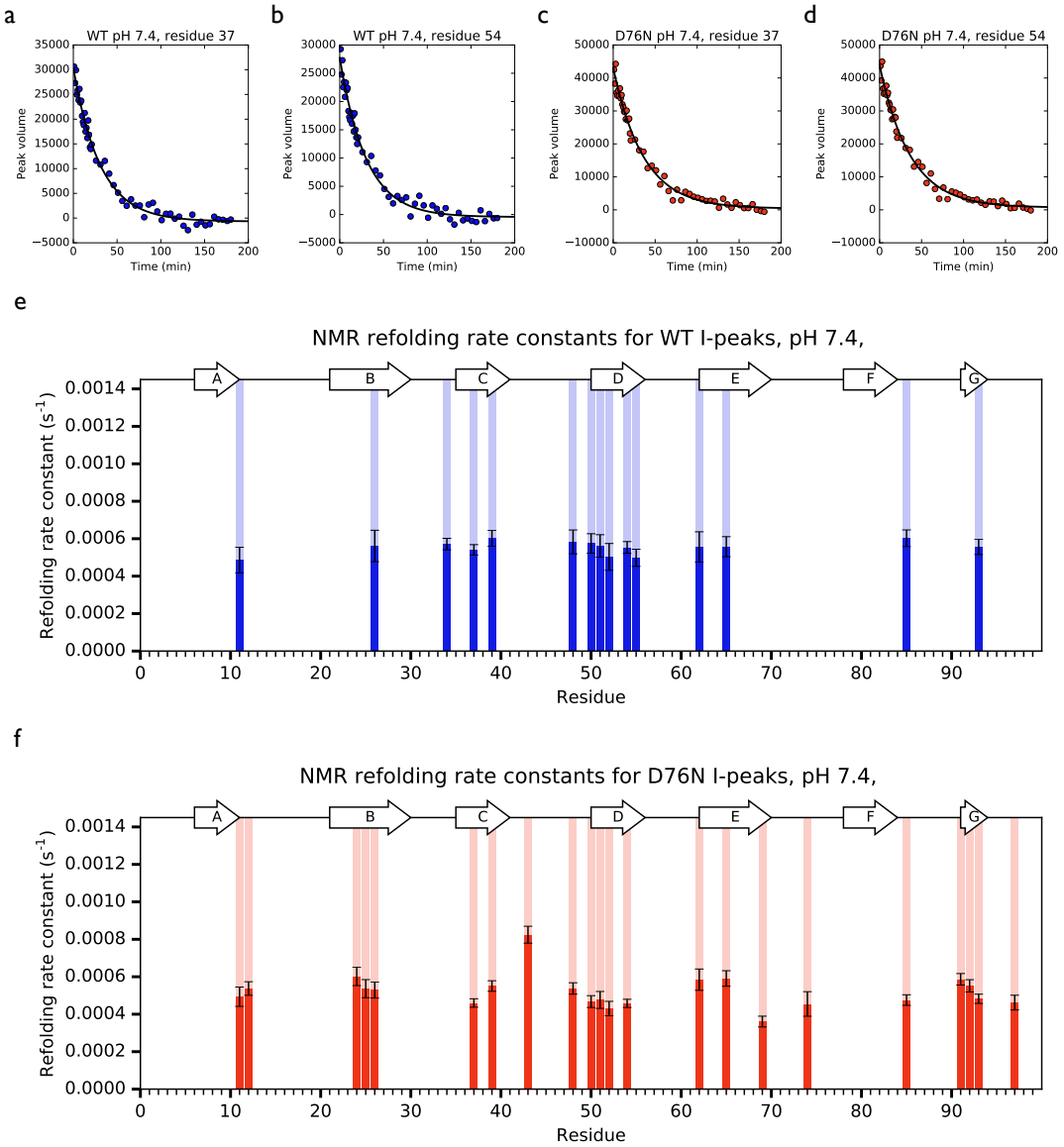


Fig. 4.17: Single residue refolding rates for I-state peaks.
(a-d) show representative data (WT in blue, D76N in red) and fits (black) for single residue refolding rates fitted to the equation $y = ae^{-bx} + c$ (Equation 4.1). The rate constants for residues where the standard deviation on the fit is no more than 10% of the rate constant value are shown in (e) and (f) for WT and D76N, respectively.

The first spectrum acquired in the refolding reaction time-course, shown in Figures 4.12 and 4.13, can only be obtained after a dead-time of 30 s. Within this dead-time some molecules will already have folded to the N-state. Therefore, in order to determine if the first spectrum acquired in the refolding reaction is a good representation of the I_T-state, the population of N-state at a given time point was calculated and the signal intensity of

the N-state resonances subtracted from the refolding spectra to yield a spectrum solely containing I_T-state resonances (Figure 4.18). To calculate the population of N-state at any time point; Equation 4.2 was rearranged to yield the % N-state for a given time point using Equation 4.4. Using the fitting coefficients for all residues where the standard deviation on the fit was less than 5%, and the median average rate constant of folding was used to scale an N-state spectrum (180 m spectrum) and then subtract it from the refolding spectrum to yield an I_T-state only spectrum. The resulting I_T-state only spectra are compared for WT and D76N using the refolding spectrum after 90 s and also the refolding spectrum after 300 s in Figure 4.18.

$$\%\mathrm{N} = \frac{100}{\mathrm{c}}\left(-\mathrm{a}\mathrm{e}^{-\mathrm{b}\mathrm{t}} + \mathrm{c}\right) \tag{4.4}$$

Where y is the CD intensity at the chosen wavelength, x is time, c is the value of y at infinite time and a is the starting value and b is the rate constant.

The results (Figure 4.18) show that the species observed at 90 s in the refolding reaction is a good representation of the I_T-state, even when the N-state contribution is 13.8% for WT and 12.8% for D76N. Interestingly, even the refolding spectrum after 300 s which contains visible N-state resonances (the N-state population for WT is 25.8% and D76N is 25.4%) can be used to reproduce the I_T-state spectrum once the native contribution is subtracted. This approach represents a simple and straightforward way to visualise folding intermediates, using prior kinetic information, either from NMR or other techniques. This approach could be utilised for situations where a longer dead time is used, if spectra need to be acquired over longer time period, or if the process in question proceeds more quickly than β_2m refolding and a spectrum containing resonances from another species is unavoidable.

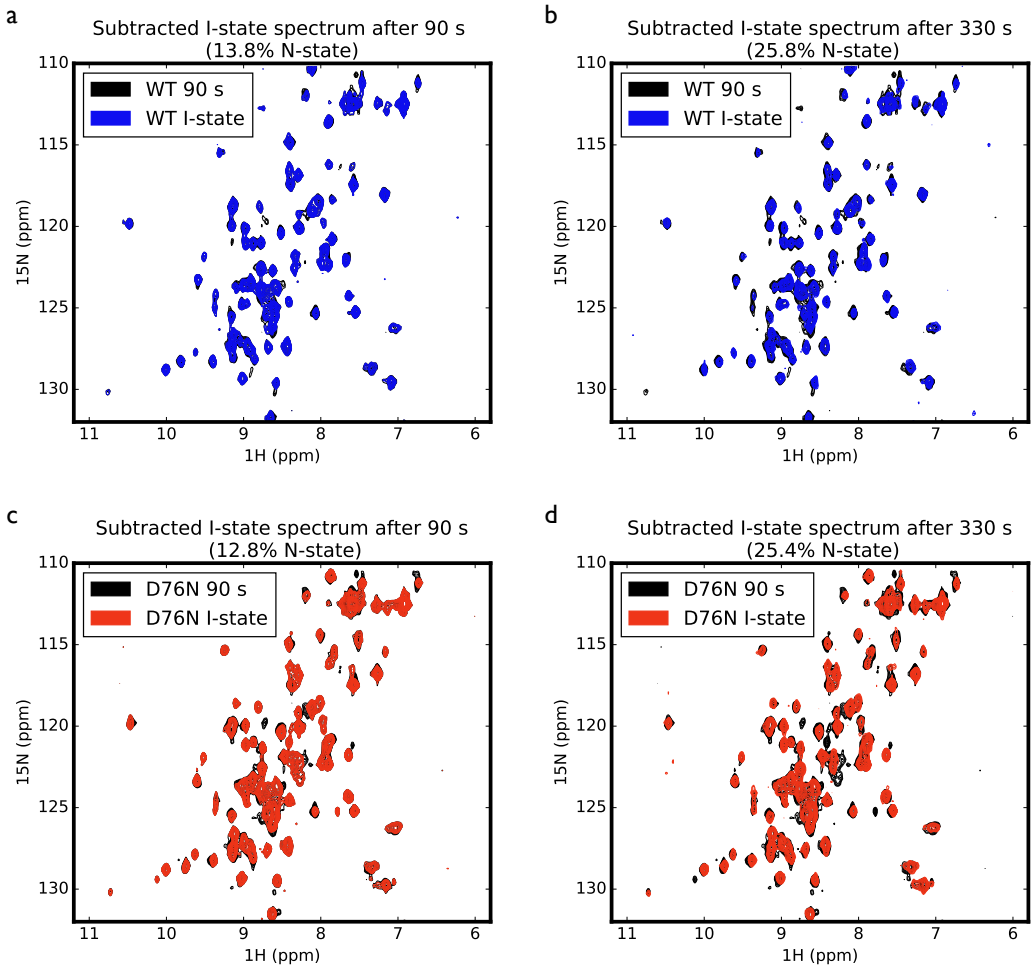


Fig. 4.18: Comparison of pure-I_T-state spectra and refolding spectra.
The spectra in blue (WT) and red (D76N) are the product of subtracting the native state contribution to the spectrum after 90 s (a,c) and after 330 s (b,d) and are compared to the refolding spectrum acquired after 90 s.

4.2.3 Generation of a trapped D76N I_T-state mimic (ΔN6-D76N)

The results presented so far indicate that the I_T-states of WT and D76N are remarkably similar regarding both the refolding kinetics and structure of the N and I_T-states. The critical question, however, is whether the D76N I_T-state is more aggregation prone than the WT I_T-state state. This is an important question since D76N aggregates faster than ΔN6, suggesting either the D76N I_T-state is more aggregation prone than the WT I_T-state, or that D76N aggregation proceeds via a route independent of the I_T-state. Given the similarities between the WT and D76N I_T-state discovered above, a ΔN6-D76N variant was made, as a potentially kinetically-trapped D76N I_T-state mimic, akin to the ΔN6-WT variant [88]. It should be noted, however, that while ΔN6 makes up approximately 30% of the fibrils responsible for DRA [232, 171] (with the other 70% full-length WT β_2m), there is no ΔN6-D76N in patients with D76N amyloidosis [183].

The ΔN6-D76N mutant was cloned and the variant protein purified as described in Methods 2.2.1 and 2.3. Figure 4.19a shows that the ΔN6-D76N eluted as a single peak in gel filtration at 200-230 ml elution volume consistent with the ΔN6-WT monomer species. The purified ΔN6-D76N produced a single band on an SDS-PAGE gel (Figure 4.19b). The purified protein had a molecular weight of 11.249 kDa, by mass spectrometry, consistent with ^{15}N isotopic labelling and was predominantly a single species with a small population at 11.13 kDa (Figure 4.19c). The yield of ΔN6-D76N was remarkably low (0.5 mg/ml) with much of the protein precipitating during the refolding step of the purification protocol. Typical yields for full-length D76N are approximately 15 mg/ml.

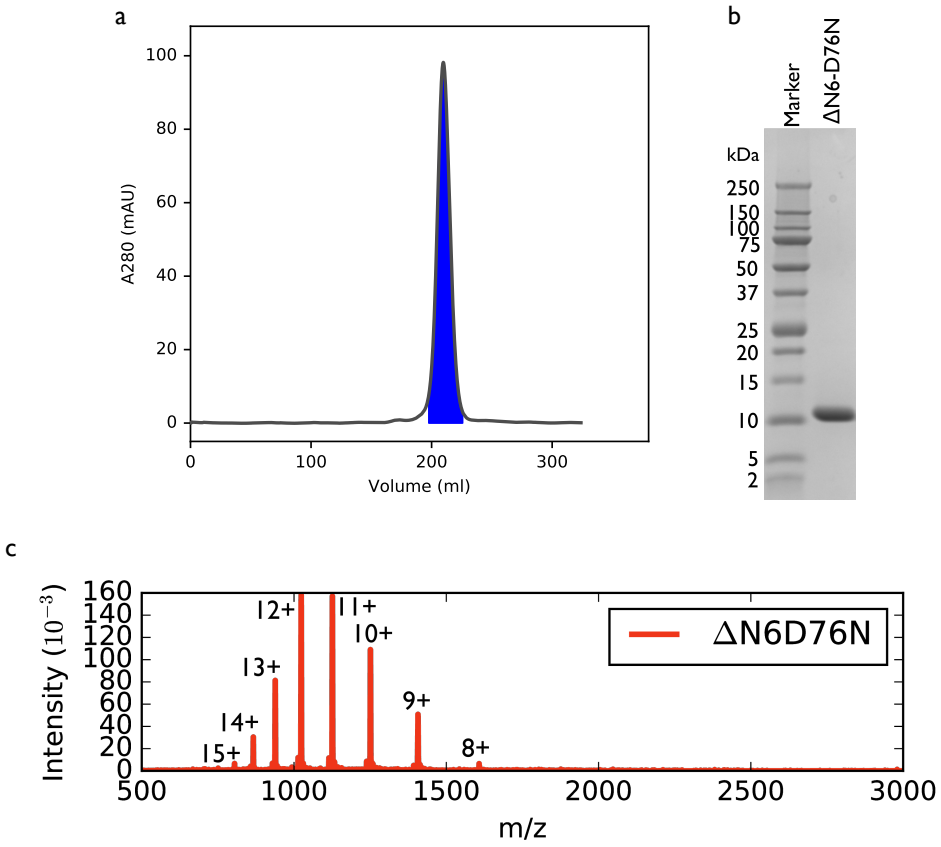


Fig. 4.19: Purification of ΔN6-D76N.
(a) size exclusion chromatography profile for ΔN6-D76N, the blue area corresponds to the material taken forward for further experiments. The gel in b, shows the pure sample from (a) as a single band. (c) Denatured ESI mass spectra of purified ΔN6-D76N for exact mass calculation. With an observed molecular weight of 11.249 ± 0.1 kDa. The expected mass is 11.269 kDa, if 100% 15N labelled.

The fold of ΔN6-D76N was first interrogated using far-UV CD spectroscopy and compared to the spectrum of ΔN6 (Figure 4.20a). The far-UV CD spectrum of ΔN6-D76N shows the same characteristics observed for ΔN6 and the folding intermediates of WT and D76N observed in Figure 4.3, although the signal for the ΔN6-D76N is lower intensity to that of ΔN6 due to problems with ΔN6-D76N sample preparation resulting in lower protein concentration, the final concentration however is unknown, so MRE calculation was done assuming a concentration of 20 μM. The thermal stability of ΔN6-D76N was also assessed and compared to ΔN6 (Figure 4.20b). ΔN6-D76N has a $T_{m,app}$ of 42.8°C compared to the ΔN6 $T_{m,app}$ which is 55.2°C. The difference in $T_{m,app}$ between

ΔN6 and ΔN6-D76N is 12.4°C, while the $T_{m,app}$ difference between WT and D76N N-states is 10.9°C (Chapter 3). The $T_{m,app}$ values for WT and D76N (full length and ΔN6 variants for both) are summarised in Table 4.3, and show that the combined effects of the D76N substitution and the ΔN6 truncation appear to be additive, resulting in a 22.7°C decrease in stability between WT and ΔN6-D76N.

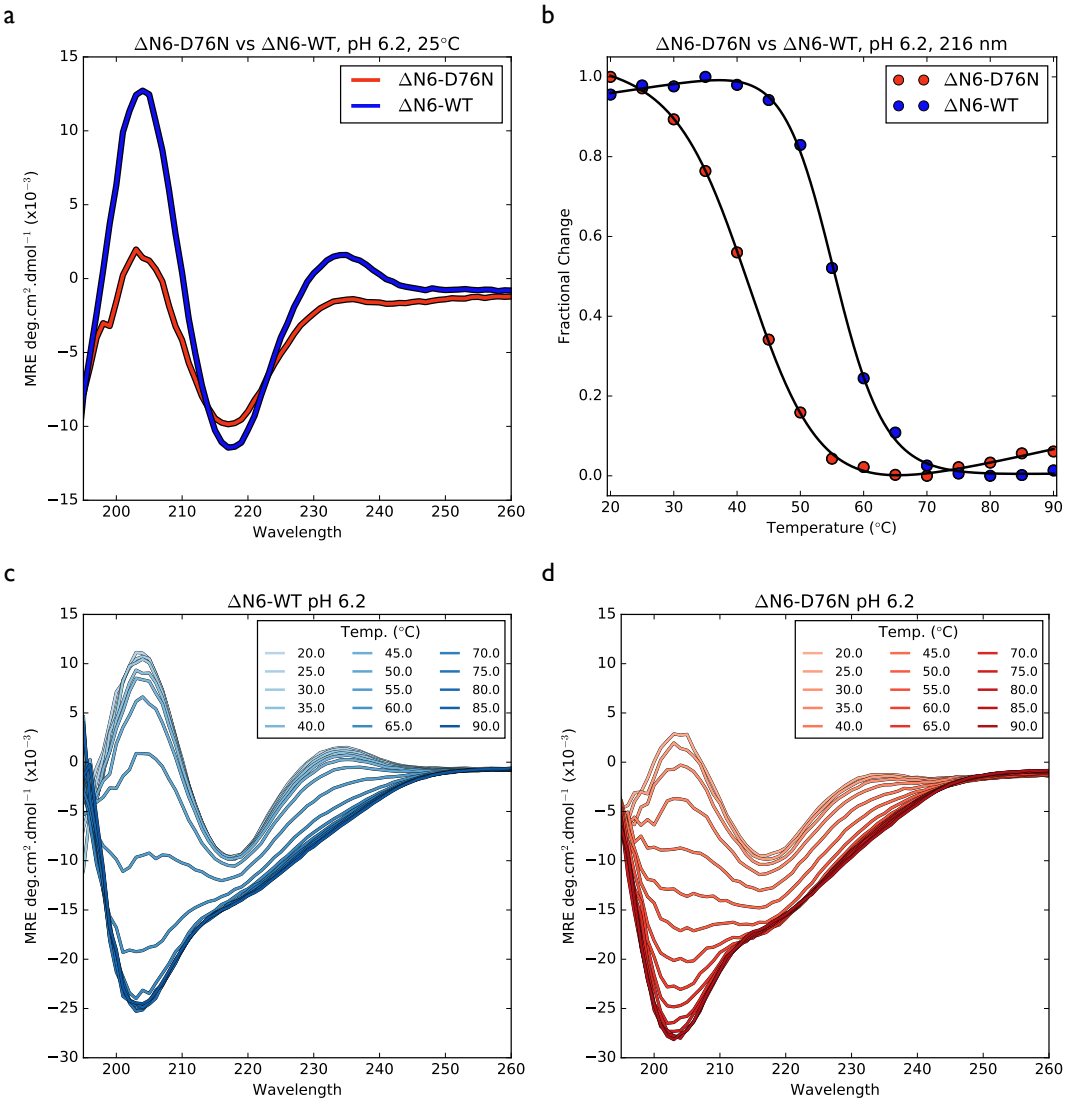


Fig. 4.20: ΔN6-D76N and ΔN6-WT temperature ramp, monitored by far-UV CD spectroscopy.
The far-UV CD spectra for ΔN6-D76N (red) and ΔN6-WT (blue) at 25 °C are shown in a. In b, the temperature ramp data (c,d) are shown as a function of temperature fitted to the equation two-state exchange model using the CDPal software package [193] for calculation of $T_{m,app}$ values. The temperature ramp experiment were carried out at pH 6.2 in 25 mM sodium phosphate buffer in the range 20-90 °C in 5 °C steps.

Table 4.3: $T_{m,app}$ values for WT and D76N (full length and ΔN6) at pH 6.2 in 25 mM sodium phosphate buffer, fitted to the equation two-state exchange model using the CDPal software package [193].

Variant	WT	D76N	ΔN6-WT	ΔN6-D76N
$T_{m,app}$ (°C)	65.5 ± 0.5	54.6 ± 0.1	55.2 ± 0.5	42.8 ± 0.9

The similarity of the ΔN6-D76N and D76N variants was further studied by NMR. The ^{1}H-^{15}N HMQC spectrum of ΔN6-D76N (Figure 4.21) shows a remarkable similarity to to that of the kinetically populated D76N I_T-state spectrum obtained using real time NMR. Small differences can be observed, yet may be a consequence of the different buffer and protein concentrations used in the refolding experiment in Figure 4.11 and the buffer used to study ΔN6-D76N. These results show that the ΔN6-D76N variant is a suitable mimic for the D76N I_T-state, akin to ΔN6-WT and the WT I_T-state.

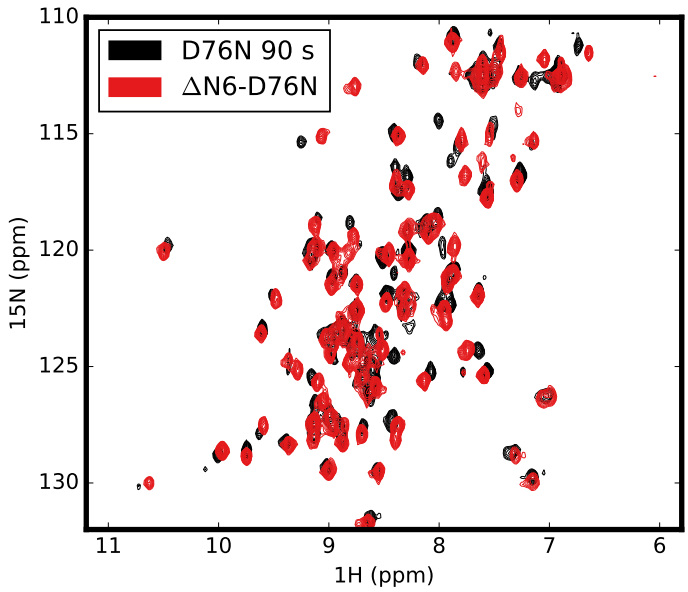


Fig. 4.21: Comparison of the ^{1}H-^{15}N HMQC spectrum of ΔN6-D76N and the D76N I_T-state at 293K.
^{1}H-^{15}N SOFAST HMQC spectra of ΔN6-D76N (red, 80μM protein, 25 mM sodium phosphate, pH 7.4) and the D76N 90 s refolding spectrum (black, 300 μM protein at pH 7.4 in 1.0 M urea, 166.6 mM sodium phosphate buffer).

Finally, the aggregation kinetics of the ΔN6-D76N variant was measured with ThT fluorescence (Figure 4.22, Methods 2.6.1) and show that ΔN6-D76N aggregates with a

rate comparable to the rate observed for ΔN6 (Chapter 3), which is slower than the rate observed for full-length D76N. The ΔN6-D76N half-time is 26.855 (±6.125) hours, while the ΔN6-WT half-time is 18.017 (±1.883) hours, both of which aggregate more slowly than the D76N variant which aggregates with a half-time of 6.567 (±0.679) hours (Chapter 3). This is a significant result, which shows how the I_T-states between D76N and WT are structurally similar and indeed aggregate with similar kinetics. Hence, differences between the I_T-states for WT and D76N cannot explain the enhanced aggregation rate of D76N. Since WT also appears to populate the same I_T-state as D76N, and WT does not aggregate under these conditions; the aggregation observed for D76N must proceed via an alternate species, independent of the I_T-state.

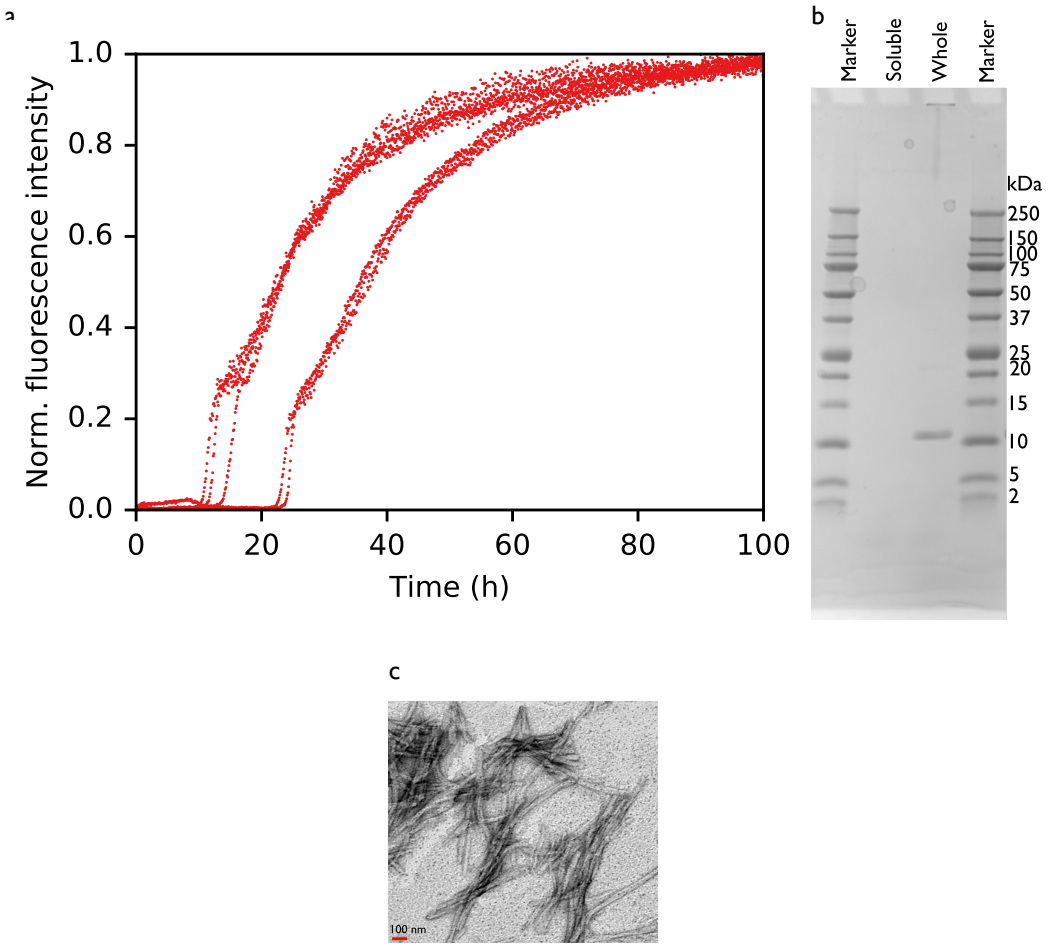


Fig. 4.22: ΔN6-D76N aggregation assay.
(a) ThT kinetics of D76N aggregation, 30 μM ΔN6-D76N, 25 mM sodium phosphate buffer, pH 6.2 with a final ionic strength of 180 mM (made by addition of NaCl), 10 μM ThT, 0.02%, NaN_3 *(w/v)*, 37 °C, 600 rpm. (b) Fibril yield (Methods 2.6.3), assessed through SDS-PAGE comparison of endpoint soluble material after centrifugation and the whole sample of soluble and insoluble material. (c) Negative stain EM micrographs of amyloid fibrils (Methods 2.6.2) from reaction endpoint material, scale bar (red) = 100 nm). Five replicates are shown.

4.3 Discussion

The real-time refolding experiments described in Section 4.2.1, show that WT β_2m, D76N, D76E and D76A all fold via an I_T-state, with the typical CD spectrum associated with this species [177] and with refolding rate constants on a timescale consistent with the *trans-cis* isomerisation of proline 32 [176, 174, 173]. The refolding rates observed by far-UV CD in Section 4.2.1 are similar for all variants and do not correlate with the observed aggregation rates for the four proteins studied in Chapter 3; in which D76N aggregates the fastest, followed by D76A and D76E, while WT β_2m does not aggregate at all under the same conditions. These observations show that while the residue at position 76 in β_2m is critical in enhancing aggregation, it does not appear to have such a role in the refolding pathway.

The real-time refolding of WT and D76N, followed by NMR was described in Section 4.2.2. In these studies, the I_T-states of WT and D76N were directly visualised by ^{1}H-^{15}N NMR show that the I_T-states are structurally similar with areas of significant chemical shift perturbation are focussed around the mutation site at position 76. This result is consistent with the differences observed between the native-state spectra of WT and D76N which also show differences predominantly around the mutation site. Further similarities between the D76N and WT I_T-state are observed in the analysis of refolding rate constants for both the increase in intensity of N-state peaks and the decay of I_T-state peaks during the time course. In both cases, the rates for WT and D76N are comparable, providing further evidence suggesting the role of residue 76 in refolding is minimal.

The production of a ΔN6-D76N variant in Section 4.2.3 successfully produced a D76N I_T-state state mimic with a similar NMR spectrum to the I_T-state spectrum observed during the D76N refolding timecourse. This ΔN6-D76N variant is destabilised by 22.7°C compared with WT β_2m, consistent with the additive effects of the D76N substitution and the ΔN6 truncation, suggesting a lack of cooperativity between the sequence changes. The ΔN6-D76N aggregates on a similar timescale to ΔN6, and more slowly than D76N.

The results described in this chapter provide strong evidence that the I_T-states between D76N and WT are similar. The aggregation rate also shows that the D76N I_T-state is not responsible for the aggregation propensity of D76N. Overall, therefore, the results presented suggest strongly that D76N aggregation must proceed via an I_T-state independent species, derived from the native state. The presence of such a species is investigated in the next chapter.

Chapter 5

Conformational dynamics of D76N

5.1 Introduction

The experiments carried out in Chapters 3 and 4 revealed that the native fold of D76N is structurally unperturbed by the amino acid substitution at position 76 and that the generic mechanism of β_2m aggregation, where the population of the folding intermediate known as the I_T-state correlates with aggregation propensity does not explain D76N aggregation. Instead, the native D76N protein was shown to be more aggregation prone than the I_T-state itself, suggesting that the aggregation of D76N proceeds via an I_T-state independent route.

The low intensity, broad peaks corresponding to residues 74 and 75 in D76N observed in the D76N ^{1}H-^{15}N HSQC spectrum in Chapter 3 (Figure 3.11), suggest that there may be exchange with other conformational states occurring in D76N. In this chapter, NMR experiments, including hydrogen-deuterium exchange, T_2 relaxation and relaxation dispersion experiments were carried out to investigate the dynamics of D76N over a range of timescales. In addition to the NMR experiments carried out to probe the dynamics of D76N, metadynamic-metainference molecular dynamics simulations were carried out to explore the predicted conformational free energy landscape of D76N. These simulations combined backbone chemical shifts and residual dipolar couplings as restraints with the prior knowledge gleaned from the NMR experiments to inform the

design of collective variables, required for metadynamics. The aim is to use this approach to identify rare conformational states in equilibrium with the native protein which may rationalise its amyloid potential.

5.2 Results

5.2.1 Exploring the role of protein dynamics in D76N aggregation with NMR

Hydrogen-deuterium exchange NMR spectroscopy as a probe for D76N dynamics

Hydrogen-deuterium exchange (HDX) NMR experiments are particularly sensitive to the presence of rarely populated species on a wide range of timescales [238, 39, 40]. HDX experiments were carried out on both D76N and WT β_2m to investigate whether there are any differences in D76N which may be responsible for its enhanced aggregation potential compared to WT β_2m.

HDX-NMR experiments were carried out (Methods 2.7.5) and SOFAST-^{1}H-^{15}N HMQC experiments [239] were utilised to follow the loss of amide signals due to exchange with the deuterium solvent with time. The SOFAST-^{1}H-^{15}N HMQC experiments were recorded every 5 minutes after a 3 minute dead time. The spectra in Figure 5.1 show the first spectrum acquired in each experiment (D76N and WT at pH 7.4); in these spectra, peaks corresponding to solvent-exposed loop residues exchange too quickly to be observed. Many residues in the core of the protein can be observed and the decay in their intensity was followed over time. The residues still visible after exchange for 8 minutes are shown in Figure 5.1 and mapped onto the structure of D76N in Figure 5.2.

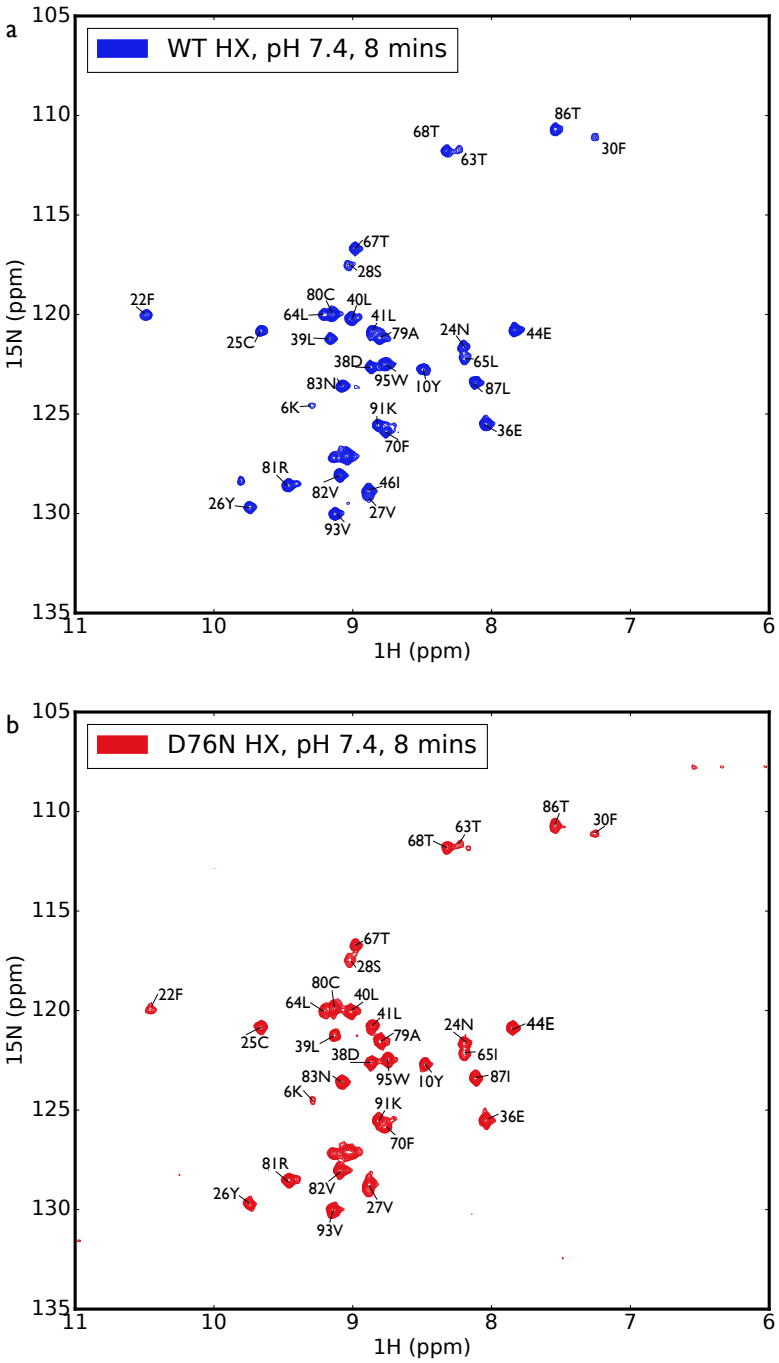


Fig. 5.1: WT and D76N spectra after initiation of hydrogen-deuterium exchange. SOFAST-^{1}H-^{15}N HMQC experiments for the first time points in hydrogen-deuterium exchange experiments for WT (a) and D76N (b) β_2m at pH 7.4, after 8 minutes of exchange (3 min dead-time, 5 min acquisition time).

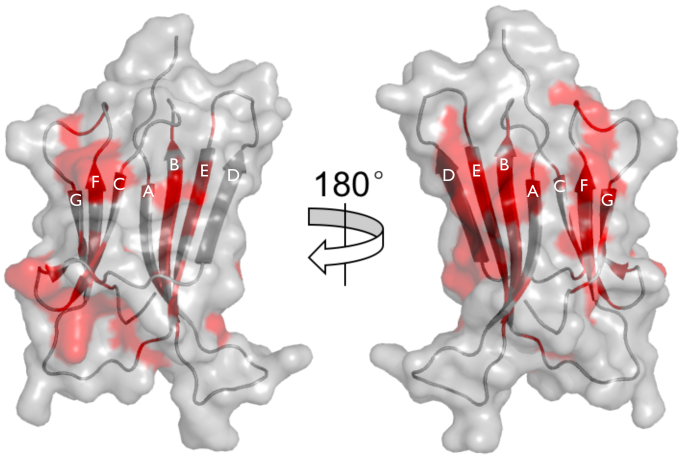


Fig. 5.2: Protected core residues for D76N and WT β_2m mapped onto the protein structure.
Core residues, still present in the initial SOFAST-^{1}H-^{15}N HMQC spectra acquired after 8 minutes of exchange (red) on the D76N crystal structure (PDB:4FXL, [183]).

The loss of intensity over time for each residue was measured and fitted to a monoexponential decay function (Equation 2.7). The rates of decay were extracted for each residue, then converted into protection factors by the method described in Methods 2.7.5. Examples of the intensity plots and fits are shown in Figure 5.3.

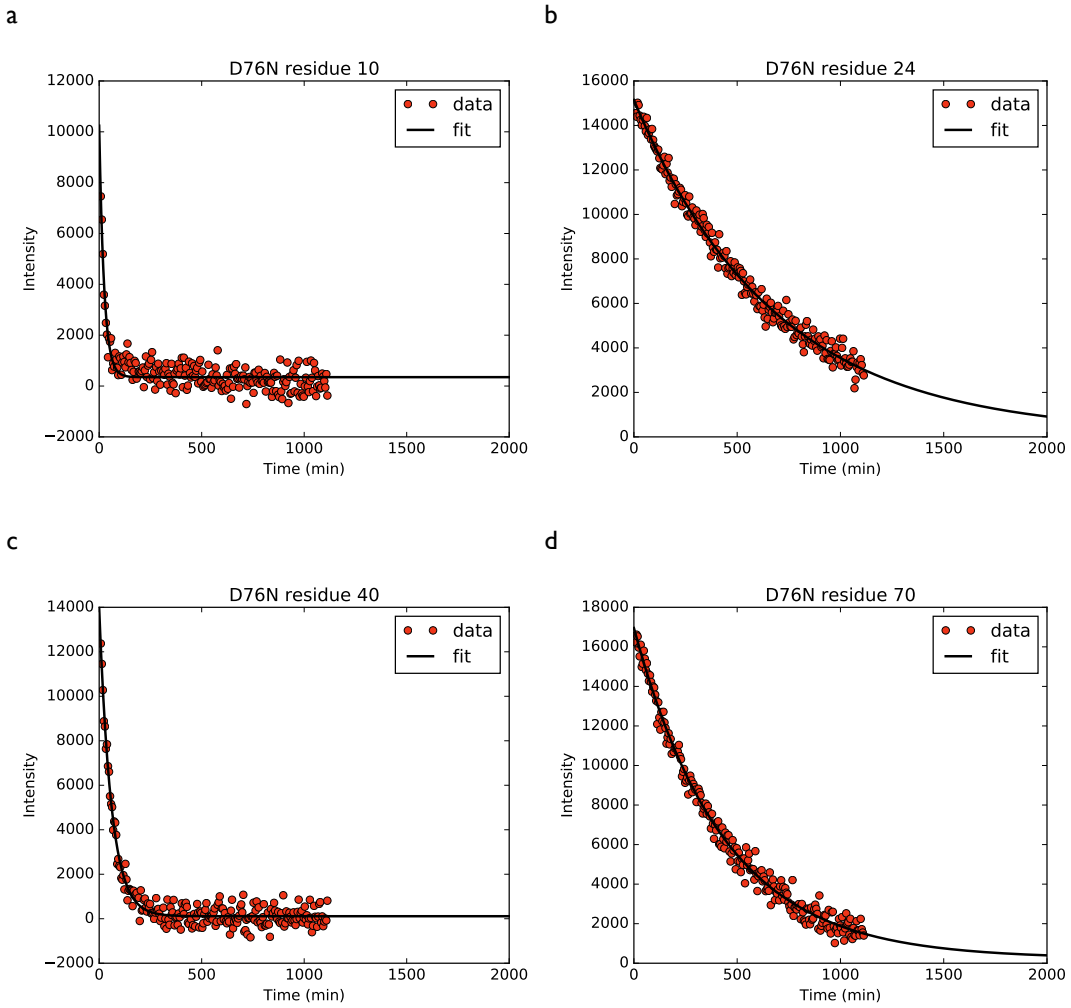


Fig. 5.3: Representative HD-exchange plots.
Representative plots for intensity decay over time in HX-NMR experiments for D76N at pH 7.4, the black line corresponds to the fit of the red data points to a monoexponential decay function (Equation 4.2).

The protection factors for residues in D76N compared to WT are shown in Figure 5.4, which shows the fold difference between the WT protection factor compared to the much lower protection factors for D76N (Figure 5.4c). Given that WT β_2m is 1.5 times more stable than D76N (based on ΔG°_{UN} values of 18.1 kJ.mol^{-1} and 27.6 kJ.mol^{-1} for D76N and WT, respectively, from Chapter 3), under exclusively EX1 exchange, the expected protection factor differences should be 1.5 times also. The results in Figure 5.4 show that D76N is far less protected than it should be if exchange were only possible from a global unfolding event, suggesting that D76N populates a less protected, yet at

least partially structured species at equilibrium which is responsible for the reduced protection factors observed in D76N compared to WT β_2m. Alternatively, HX may occur by an EX2 mechanism or a mixed EX1/EX2 mechanism, which is observed in WT β_2m [240] which is presumably due to the presence of the I_T-state.

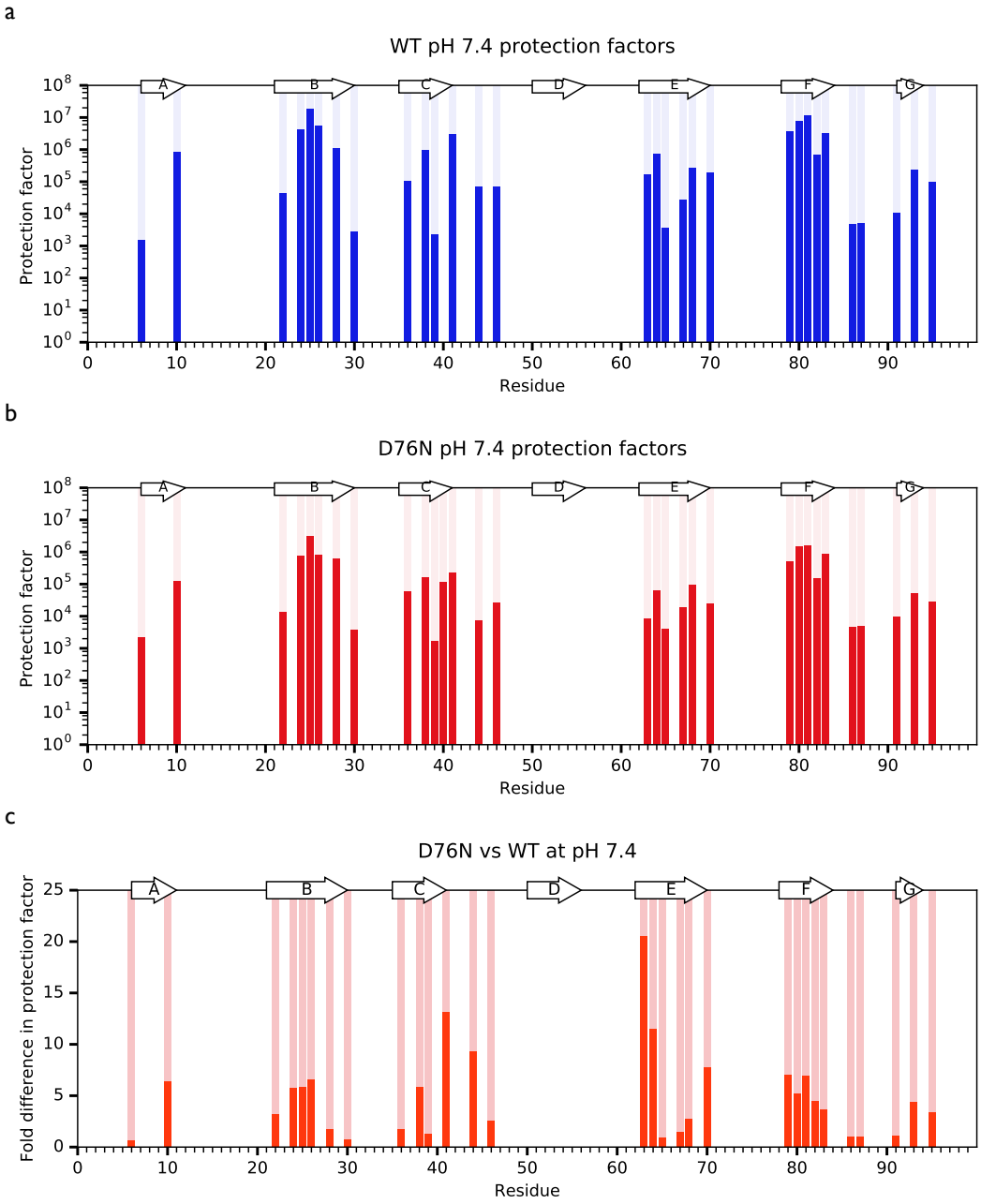


Fig. 5.4: Comparison of WT and D76N protection factors at pH 7.4.
The protection factors calculated for WT β_2m (a) and D76N (b) at pH 7.4. (c) The fold difference in protection factors between for residues in WT β_2m compared to D76N, calculated from HX-NMR experiments conducted at pH 7.4. The fold change in protection factor is calculated by dividing the WT/D76N. The pale bars indicate that there is data for a residue, while the lack of a transparent bar means that no data is available.

Based on the results of the HDX NMR experiments in this chapter, it is not clear whether the less protected species is a rarely populated state in equilibrium with the na-

tive state or if the protection factor differences between D76N and WT β_2m are instead a reflection of the lower stability of the D76N I_T-state compared to the WT I_T-state (Chapter 4, Figure 4.20), since, ΔN6 (the WT I_T-state structural mimic) is poorly protected from hydrogen exchange [60]. Therefore further experiments were carried out to probe the dynamics of D76N.

Measuring conformational dynamics with T_2 relaxation NMR experiments

^{15}N transverse relaxation (T_2) NMR experiments were carried out to further investigate the dynamics of D76N β_2m. T_2 relaxation experiments provide information about ps-ns dynamics as well as being sensitive to μs-ms timescale dynamics of each resonance observed in an ^{1}H-^{15}N HSQC spectrum [84]. T_2 relaxation is affected by the protein dynamics and also by interactions between proteins, so can provide information about interfaces formed between proteins [84]. Therefore, T_2 relaxation experiments were carried out to assess whether D76N had a different dynamic profile to WT β_2m and to investigate whether any additional interfaces could be identified between D76N monomers at increasing concentrations. Figure 5.5 shows the R_2 ($1/T_2$) relaxation rates for D76N and WT β_2m at 300 μM, pH 7.4. Some example fits are shown in Figure 5.6.

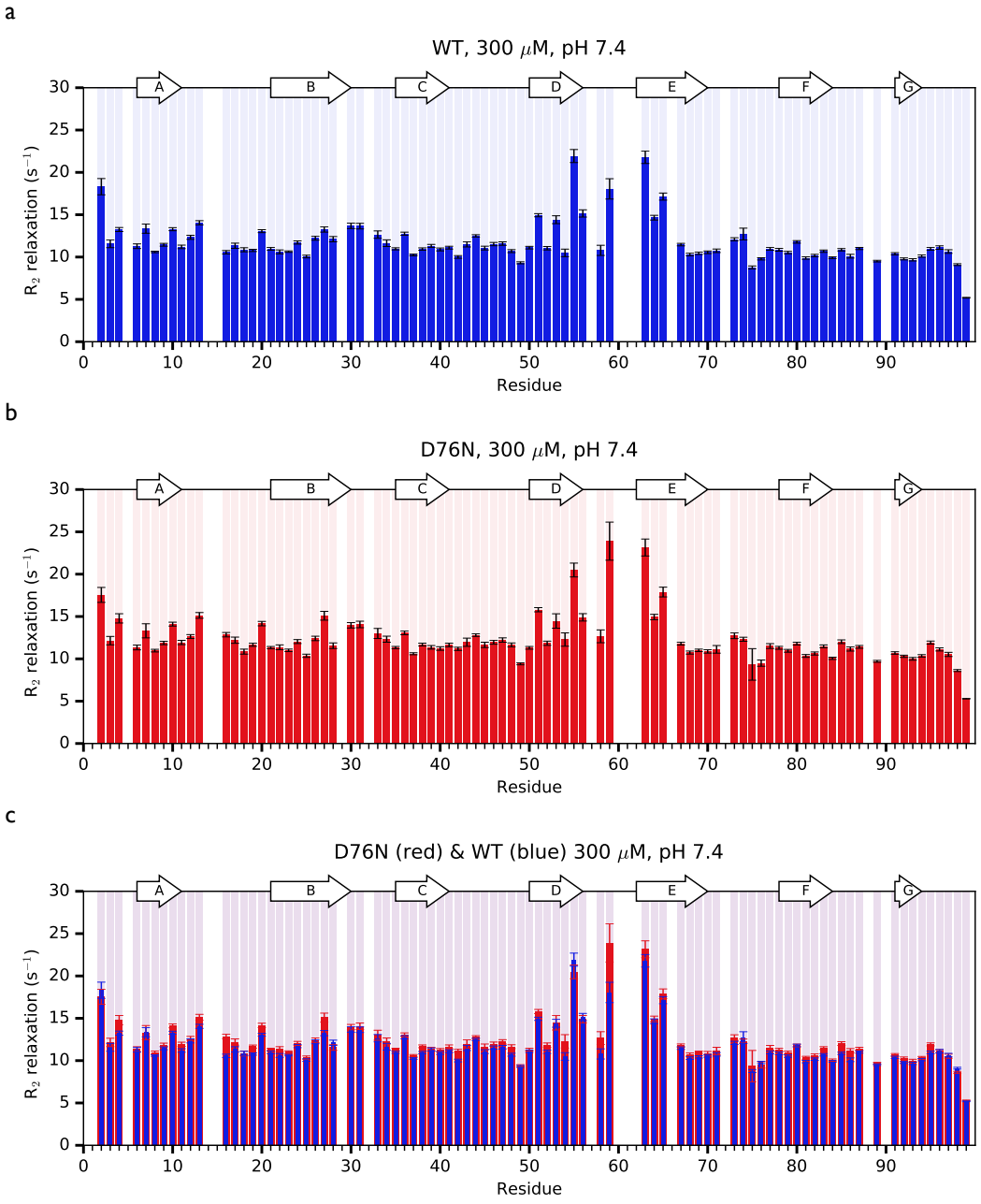


Fig. 5.5: Transverse relaxation rates for D76N and WT β_2m at pH 7.4. ^{15}N transverse relaxation measurements (R_2=1/T_2) of 300 μM D76N (red) 300 μM WT (blue) at pH 7.4, 25 °C. Experiments were carried out in 10 mM sodium phosphate buffer at pH 7.4 containing 0.02 % NaN_3 (w/v) and 10% (v/v) D_2O. Acquired at 25 °C with a 600 MHz room temperature probe. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue (D76N is red, WT is blue). When data for a residue is present in both data sets, the bar becomes purple, if no data is available for a residue, the background is coloured white.

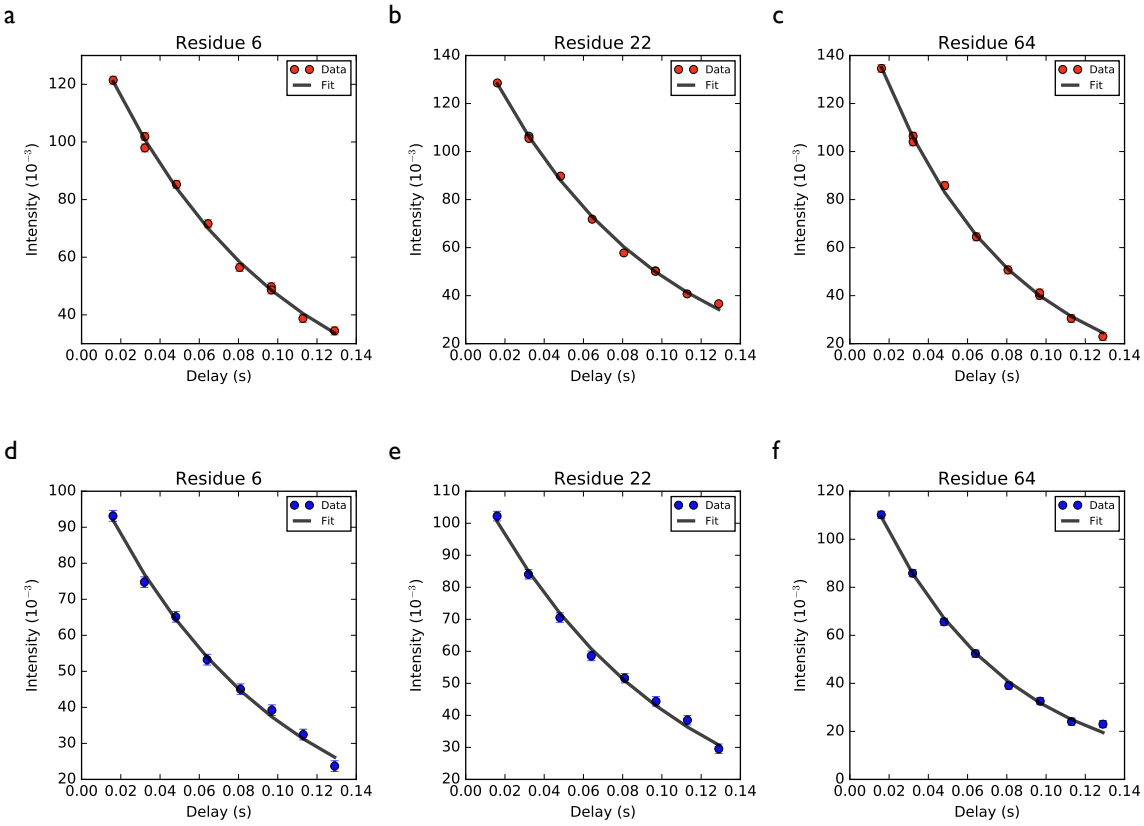


Fig. 5.6: Representative transverse relaxation decay curves for D76N and WT β_2m. Transverse relaxation traces for single residues, corresponding to the data shown in Figure 5.5 for D76N (a-c, red) and WT (d-f, blue) showing peak intensity against relaxation delay, fit to a monoexponential decay function (Equation 4.2).

For both WT and D76N there are areas with faster R_2 relaxation rates in the D-Strand, DE-loop and E-strand in comparison to the rest of the protein. There are also a number of residues in the DE-loop which are not detected, presumably due to exchange broadening (residues 57, 60, 61 and 62), which show that this area of the protein is highly dynamic. There are differences in this dynamic area in D76N, compared to WT, for residues 51, 52, 54, 58 and 59 in the D-strand and DE-loop. There are other smaller differences in R_2 relaxation rates in D76N compared with WT throughout the protein; these areas are shown on the 3D structure of D76N in Figure 5.7.

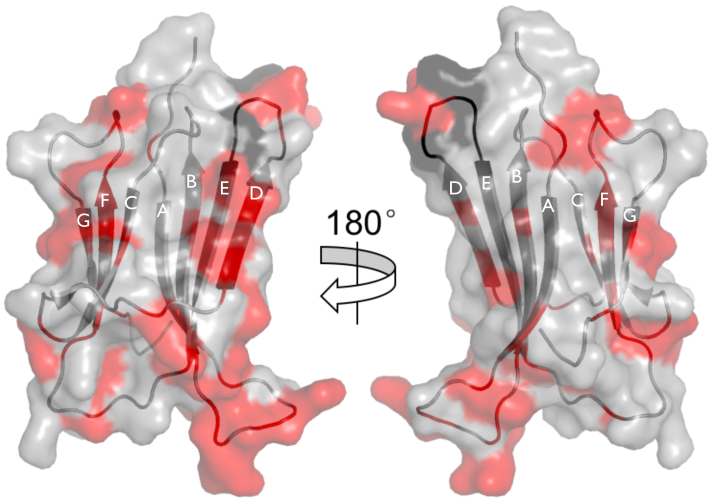


Fig. 5.7: Areas of perturbed dynamic behaviour in D76N compared to WT at pH 7.4. D76N residues with different transverse relaxation rates (where the difference between the fitted rates is greater than the sum of the errors for the fitted rates) at pH 7.4, compared to WT β_2m (red) on the D76N crystal structure (PDB: 4FXL, [183]).

In order to identify any additional interaction interfaces in D76N, R_2 relaxation rates were also measured at two different D76N concentrations (80 μM and 300 μM, Figure 5.8). At higher concentrations changes in the T_2 relaxation profile are likely to be due to protein-protein interactions rather than dynamic changes in monomers. Increasing the D76N concentration globally increased R_2 rates although the overall pattern per-residue remains the same. This global increase in R_2 relaxation was also observed in previous experiments on WT β_2m [88] and is consistent with the increased frequency of interactions at a higher concentration. The T_2 experiments show that WT and D76N β_2m share a similar dynamic profile on a μs-ms and/or ps-ns timescale observed in T_2 relaxation experiments at pH 7.4 which is suggestive of similar local dynamics, with a similar concentration dependence reflective of non-specific inter-molecular interactions.

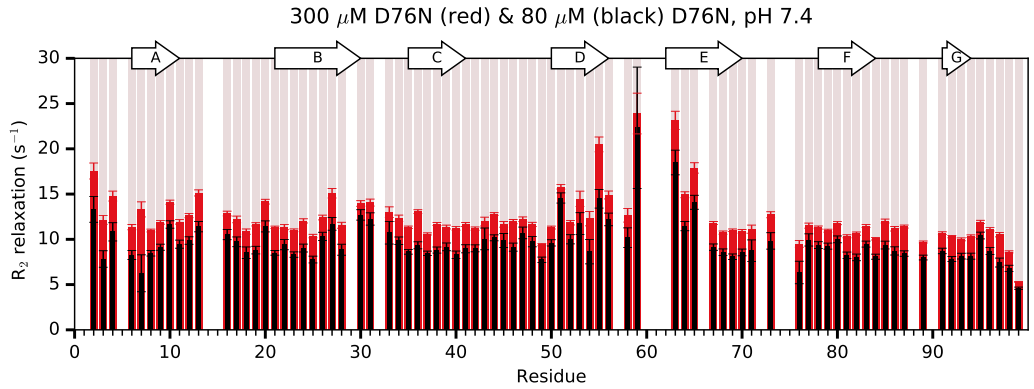


Fig. 5.8: Concentration dependence of D76N transverse relaxation rates. ^{15}N transverse relaxation measurements ($R_2=1/T_2$) of 300 μM D76N (red) 80 μM D76N (black) at pH 7.4, 25 °C. Experiments were carried out in 10 mM sodium phosphate buffer at pH 7.4 containing 0.02 % NaN_3 (w/v) and 10% (v/v) D_2O. Acquired at 25 °C with a 600 MHz room temperature probe. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue. When data for a residue is present in both data sets, the bar becomes dark red, if no data is available for a residue, the background is coloured white.

The aggregation assays in Chapter 3 show that D76N aggregates faster at pH 6.2 than at pH 7.4. Therefore, the R_2 relaxation rates for WT and D76N at pH 6.2 were compared (Figure 5.9). At pH 6.2, there is a marked difference in R_2 relaxation rates in the AB-loop, where D76N has faster R_2 relaxation rates than WT β_2m in this region. Interestingly there are observable differences in residues 75 and 76, which are too weak to measure at pH 7.4. Again there are additional small differences throughout the protein, which are shown on the structure of D76N in Figure 5.10.

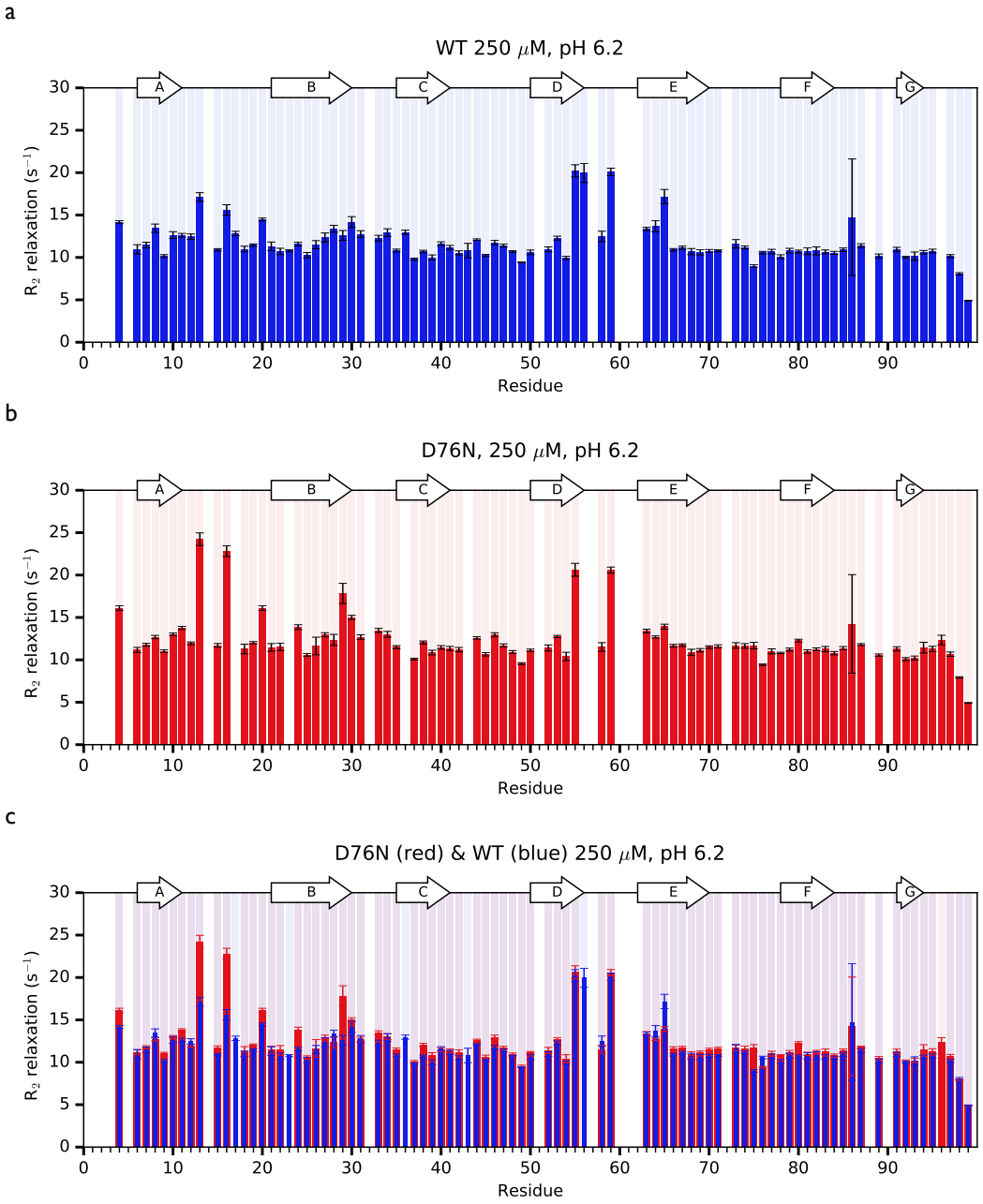


Fig. 5.9: Measuring D76N dynamics, compared to WT β_2m at pH 6.2. ^{15}N transverse relaxation measurements (R_2=1/T_2) of 250 μM WT (blue, a) and 250 μM D76N (red, b) and compared in (c). Data were acquired on samples in 25 mM sodium phosphate, at pH 6.2, 25 °C with a 750 MHz TCI cryoprobe, ^{1}H optimised triple resonance NMR inverse probe and Bruker Avance III HD console. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue (D76N is red, WT is blue). When data for a residue is present in both data sets, the bar becomes purple, if no data is available for a residue, the background is coloured white.

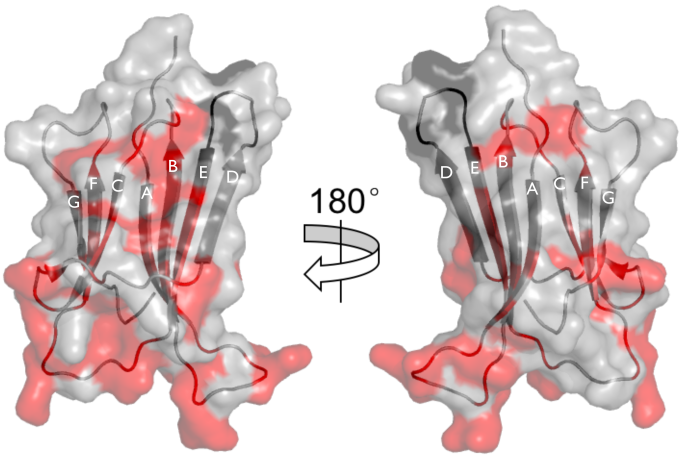


Fig. 5.10: Areas of perturbed dynamic behaviour in D76N compared to WT at pH 6.2. D76N residues with different transverse relaxation rates (where the difference between the fitted rates is greater than the sum of the errors for the fitted rates) compared to WT β_2m at pH 6.2 (red) on the D76N crystal structure (PDB: 4FXL, [183]).

The dynamics of the D76E and D76A mutants investigated in Chapter 3 and Chapter 4, which are also able to aggregate at pH 6.2, although not as quickly as D76N, were examined in Figure 5.11. Both variants show similar differences in relaxation rates at pH 6.2 to WT β_2m, with significant differences in the AB-loop and around the mutation site in the EF loop, with further differences in the D-strand and DE-loop. Interestingly, the D76A variant shows universally faster R_2 relaxation rates throughout the entire protein which suggests that interactions formed between the residue 76 side chain and other residues are important in the global dynamics of the protein.

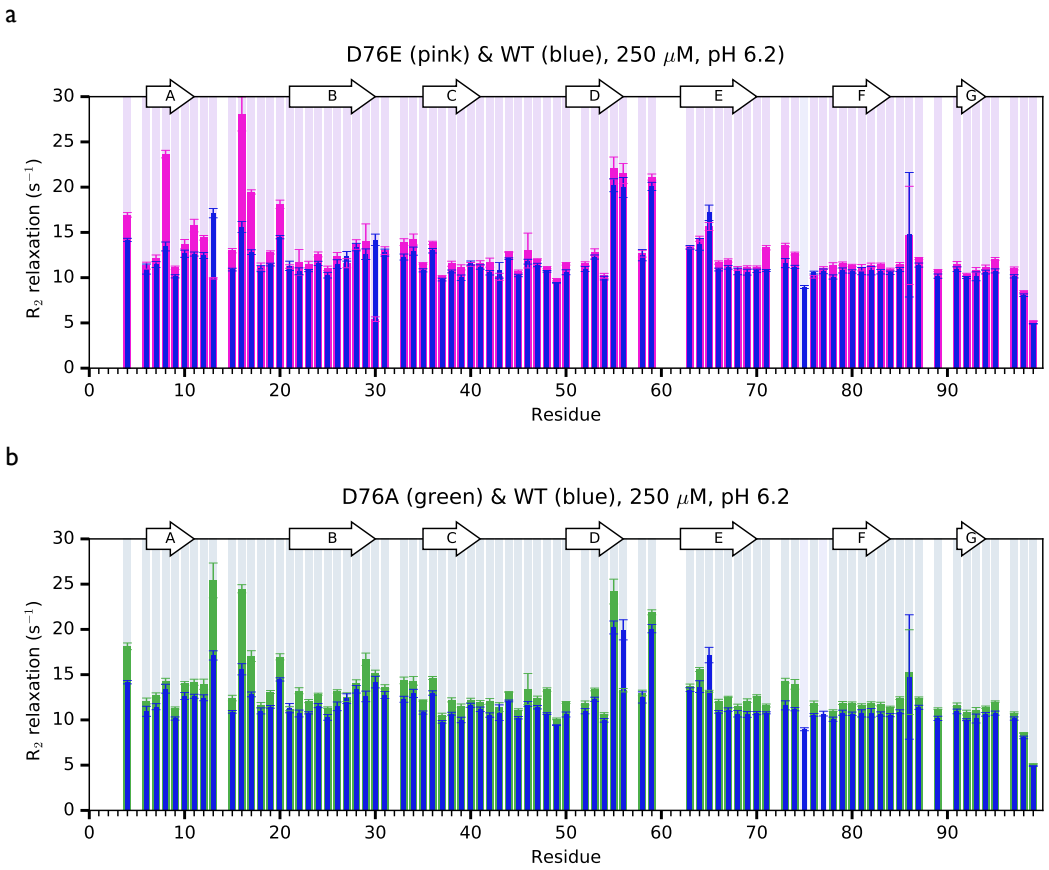


Fig. 5.11: Transverse relaxation rates of the D76E and D76A variants, compared to WT β_2m at pH 6.2.

Comparison of ^{15}N transverse relaxation measurements (R_2=1/T_2) of 250 μM D76E (a) and 250 μM D76A (b) with WT β_2m. Data was acquired on samples in 25 mM sodium phosphate, at pH 6.2, 25 °C with a 750 MHz TCI cryoprobe, ^{1}H optimised triple resonance NMR inverse probe and Bruker Avance III HD console. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue (D76E is purple, D76A in green and WT is blue). When data for a residue is present in both data sets, the bar becomes a mixture of the corresponding colours and if no data is available for a residue, the background is coloured white.

Probing invisible states with CPMG-RD NMR experiments

The results in Chapter 3 showed that the native β_2m fold is unperturbed by the amino acid substitutions at position 76, although the global stability and the aggregation propensity of β_2m was affected in D76N, D76E and D76A. In Chapter 4 the refolding of the D76 variants was studied, and the I_T-state of D76N was characterised and showed similar structural, aggregation and refolding properties to the WT I_T-state. The observations from Chapters 3 and 4, together with the results in this chapter so far, suggest that there could be an alternative conformational state, derived from the native β_2m fold, and independent from the I_T-state, which is involved in D76N aggregation. Therefore, Carr-Purcell-Meiboom-Gill sequence relaxation-dispersion (CPMG-RD) experiments [34, 9, 10] were carried out to investigate the presence of an alternative conformational state in D76N.

CPMG-RD experiments were carried out (Methods 2.7.7) on WT and D76N, at pH 7.4 (500 μM protein, 600 MHz field strength), for both WT and D76N. Per-residue relaxation dispersion curves were fitted to the Bloch-McConnell equations for two-site exchange using the software package PINT [198] (Methods 2.7.7). The results suggest there is indeed exchange with alternative conformational species. Representative dispersion curves are shown in Figure 5.12 for both WT and D76N. All residues which appear to be undergoing exchange with another species are mapped onto the WT (PDB:2YXF [231]) and D76N (PDB:4FXL [183]) crystal structures in Figure 5.14. The estimated R_{ex} values are shown in Figure 5.13. Examination of the exchanging residues in D76N and WT, on the protein structures shows that the D-strand, DE-loop and E-strand appear to be the most active parts of the structures in terms of exchange, and the number of residues experiencing exchange is greater in D76N than in WT β_2m. Residues are deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange is preferred over the fit to a constant model (no exchange) after the performance of an F-test. All dispersion curves and lists of residues undergoing exchange are presented in Appendix B.

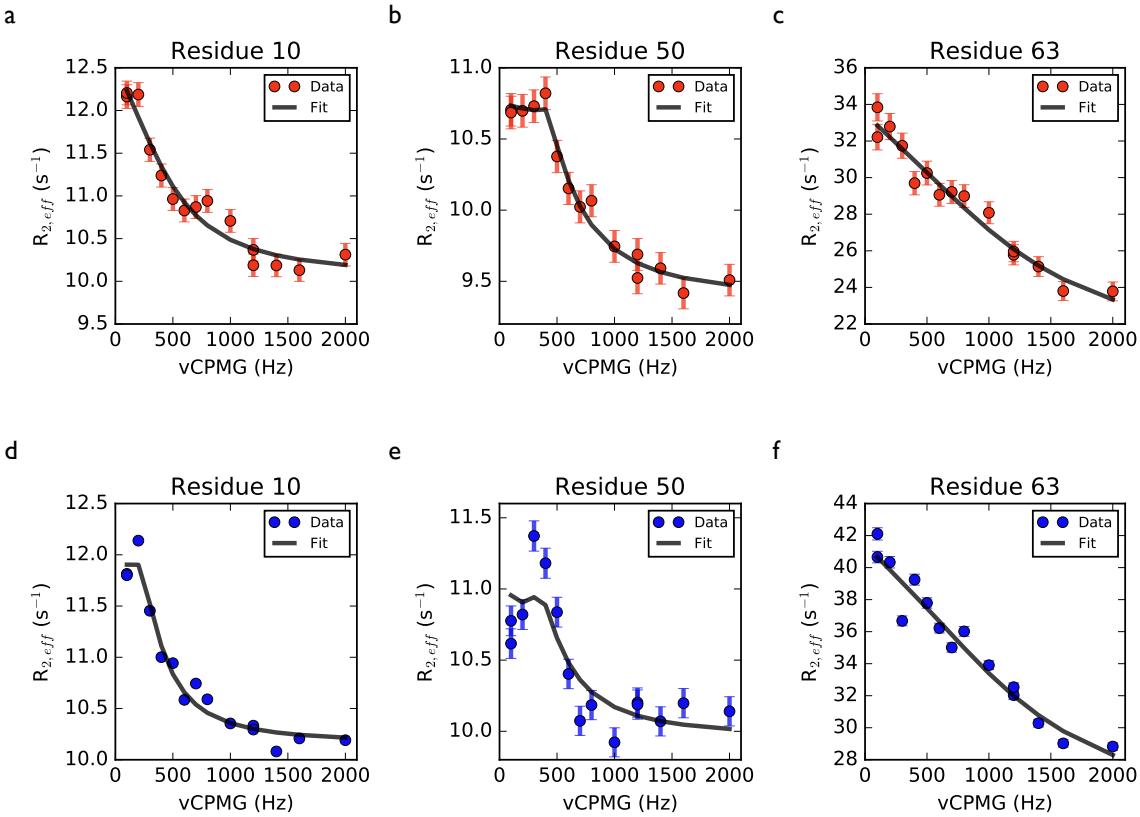


Fig. 5.12: Representative relaxation dispersion profiles for exchanging residues in D76N and WT β_2m at pH 7.4.
Selected representative CPMG relaxation dispersion curves for residues where the fit to the Bloch-McConnell equations for two-site exchange is preferred over the fit to a constant model (no exchange) after the performance of an F-test. The $R_{2,eff}$ values are plotted as dots (D76N red (a-c), WT blue (d-f)) and the fit is plotted as a black line. Error bars correspond to the uncertainty in lineshape fitting of the peak. Data were acquired with 500 μM samples at pH 7.4, recorded with a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

a

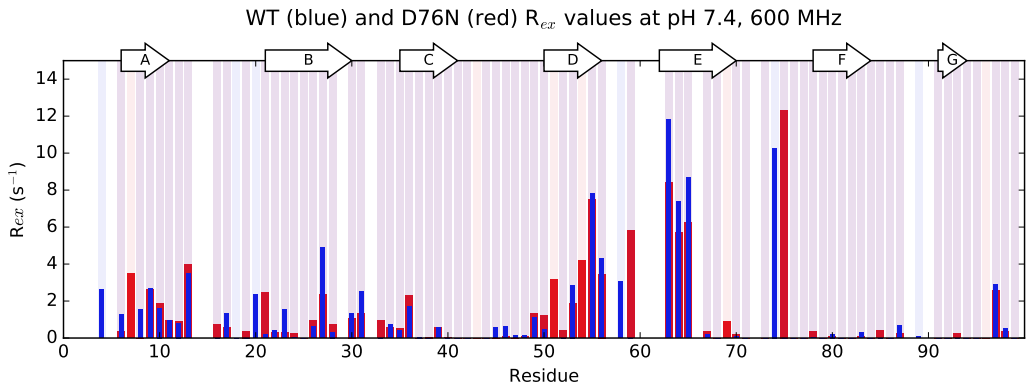


Fig. 5.13: Estimated R_{ex} values for WT and D76N, at pH 7.4 and 600 MHz Comparison of R_{ex} values for 500 µM WT with 500 µM D76N. Estimated R_{ex} values were calculated by subtracting the $R_{2,eff}$ values from the highest frequency νCPMG from the lowest frequency νCPMG $R_{2,eff}$ values (R_{ex} = $R_{2,eff}$(νCPMG=100) - $R_{2,eff}$(νCPMG=2000)). Data were acquired on samples in 25 mM sodium phosphate, at pH 7.4, 25 °C with a 600 MHz QCI-P cryoprobe, ^{1}H optimised triple resonance NMR inverse probe and Bruker Avance III HD console. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue (WT is blue and D76N in red a). When data for a residue is present in both data sets, the bar becomes a mixture of the corresponding colours and if no data is available for a residue, the background is coloured white.

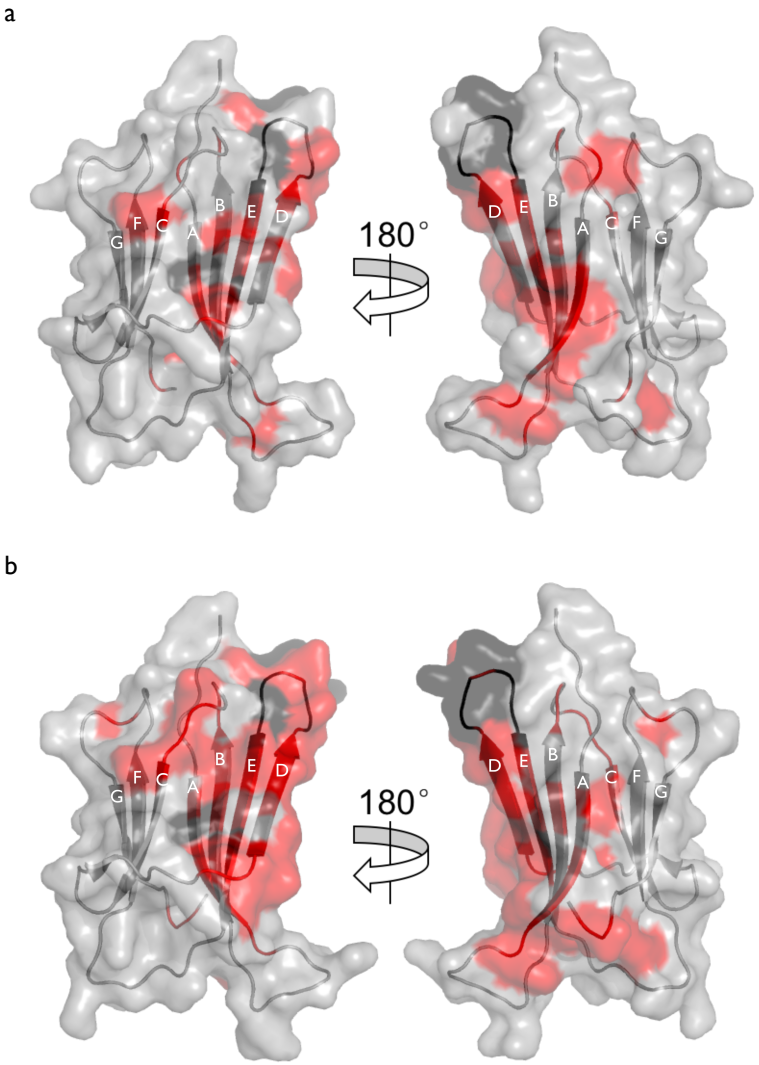


Fig. 5.14: Exchanging residues at pH 7.4, mapped on to WT and D76N crystal structures.
WT (PDB:2YXF [231]) (a) and D76N (PDB:4FXL [183]) (b) crystal structures with exchanging residues from CPMG experiment at pH 7.4, (500uM protein concentration, 600 MHz) and residues not visible due to exchange broadening coloured in black. Residues are deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange is preferred over the fit to a constant model (no exchange) after the performance of an F-test.

Next, the behaviour of D76N at pH 6.2 compared to pH 7.4 was examined (500 µM protein, 750 MHz field strength) and the dispersion curves (Figure 5.15) indicate that the exchange regimes at pH 6.2 are similar to those at pH 7.4. The structures in Figure 5.17 show that more residues undergoing exchange can be detected at 750 MHz. At both

pH 7.4 and pH 6.2, residues in the AB-loop exhibit exchange while at pH 6.2 residues in the EF-loop which are too broad to measure at pH 7.4 are also exchanging with another state. The similarity of the exchange patterns at pH 7.4 and pH 6.2 suggest that the species present at pH 6.2 are the same as the species present at pH 7.4 aggregation, which is possibly populated to a higher degree at pH 6.2. The estimated R_{ex} values are shown in Figure 5.16 and all dispersion curves and lists of residues undergoing exchange are presented in Appendix B.

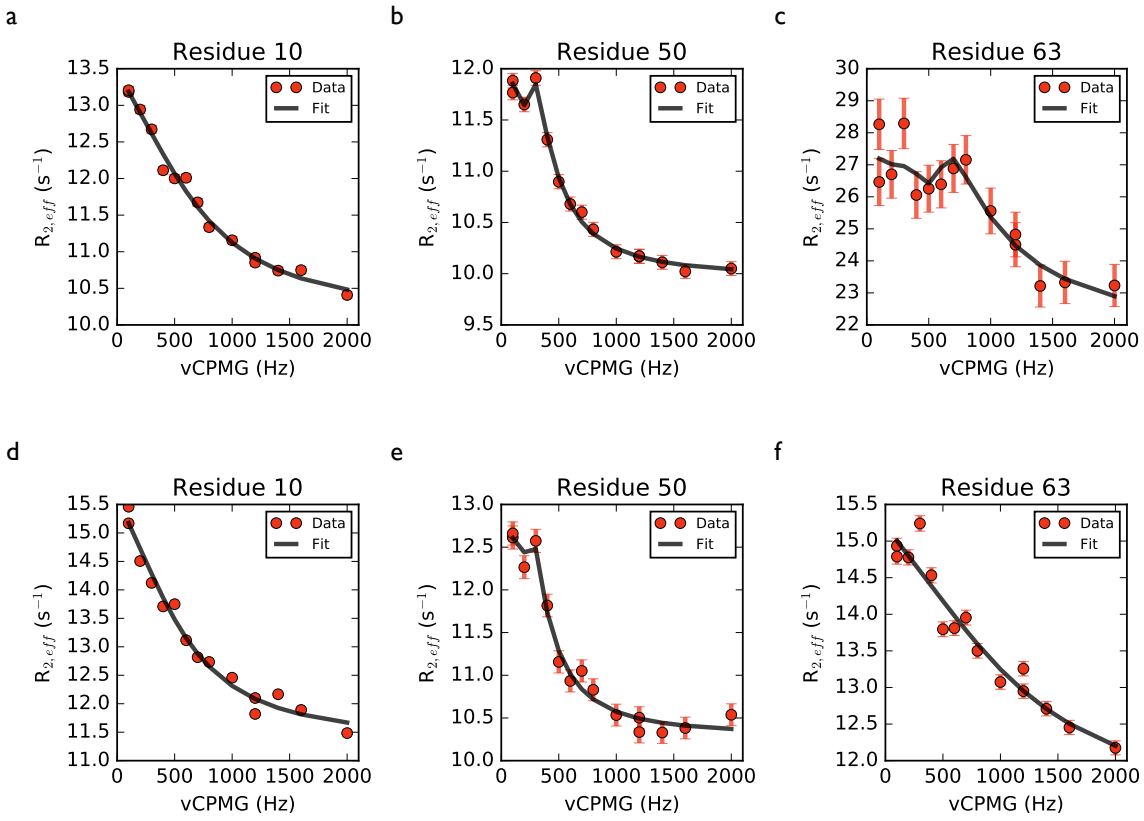


Fig. 5.15: Representative relaxation dispersion profiles for exchanging residues in D76N at pH 7.4 and pH 6.2.
Selected representative CPMG relaxation dispersion curves for residues where the fit to the Bloch-McConnell equations for two-site exchange is preferred over the fit to a constant model (no exchange) after the performance of an F-test. The $R_{2,eff}$ values are plotted as dots (D76N, pH7.4 (a-c) and D76N, pH6.2(d-f)) and the fit is plotted as a black line. Error bars correspond to the uncertainty in lineshape fitting of the peak. Data was acquired with 500 µM samples using a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and Bruker Avance III HD console.

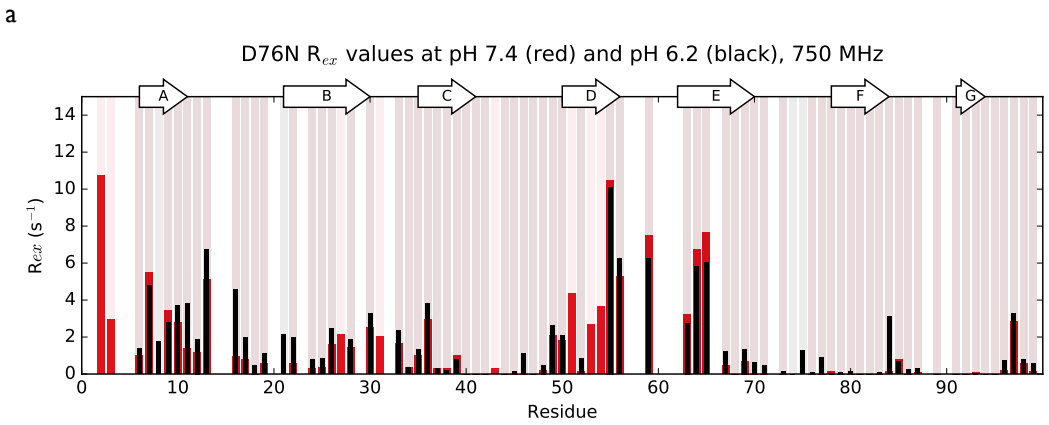


Fig. 5.16: Estimated R_{ex} values for D76N at pH 7.4 and pH 6.2 at 750 MHz Comparison of R_{ex} values for 500 µM D76N at pH 7.4 with pH 6.2. Estimated R_{ex} values were calculated by subtracting the $R_{2,eff}$ values from the highest frequency νCPMG from the lowest frequency νCPMG $R_{2,eff}$ values (R_{ex} = $R_{2,eff}$(νCPMG=100) - $R_{2,eff}$(νCPMG=2000)). Data were acquired on samples in 25 mM sodium phosphate, at pH 7.4, 25 °C with a 600 MHz QCI-P cryoprobe, ^{1}H optimised triple resonance NMR inverse probe and Bruker Avance III HD console. Secondary structure is denoted by the arrows above each plot. Transparent bars in the respective colours indicate that data are available for the particular residue (WT is blue and D76N is red). When data for a residue is present in both data sets, the bar becomes a mixture of the corresponding colours and if no data is available for a residue, the background is coloured white.

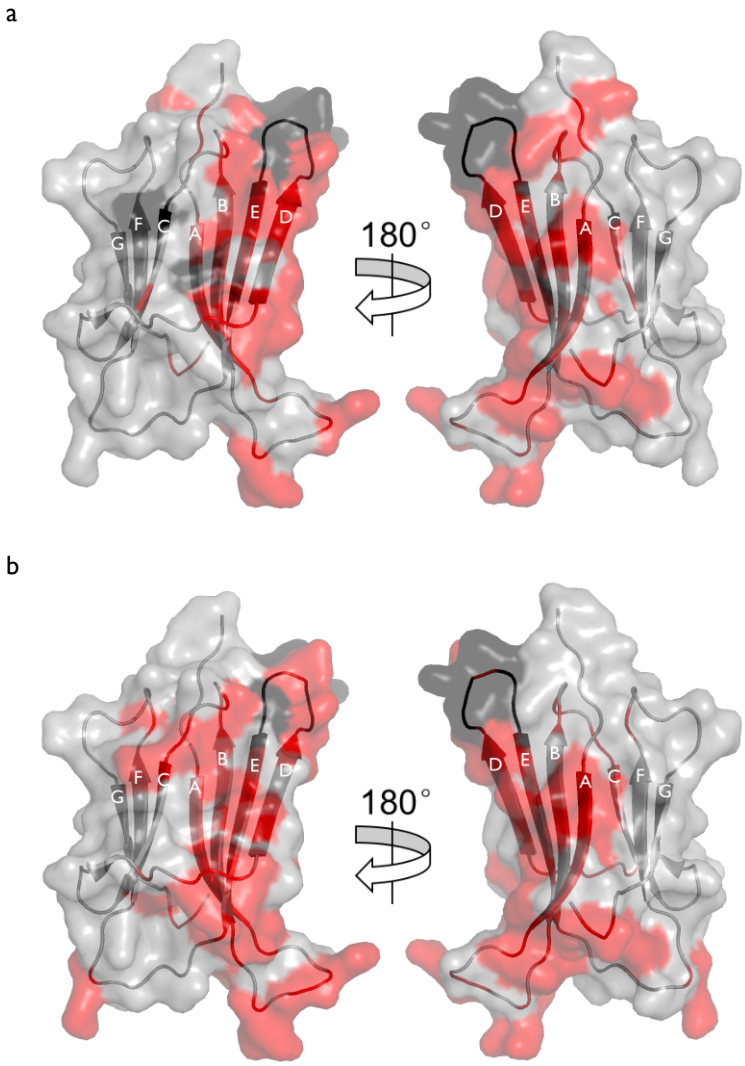


Fig. 5.17: The effect of pH on the D76N conformational exchange.
Exchanging residues identified at pH 7.4 (a) and pH 6.2 (b) for D76N, mapped onto the D76N crystal structure (PDB:4FXL [183]) and residues too broad to measure due to exchange broadening coloured in black. Data was acquired with 500 μM samples using a 750 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe. and Bruker Avance III HD console. Residues are deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange is preferred over the fit to a constant model (no exchange) after the performance of an F-test. Produced with PyMol molecular graphics system [119].

The CPMG-RD experiments show that both WT and D76N exchange with a hidden species. There may be some contribution to the exchange observed in WT and D76N from the equilibrium between the N-state and I_T-state or could be due to dimerisation,

since the protein concentration in the CPMG-RD experiments is 500 μM. The NMR data here provide an independent cross-validation of the simulations described in the next section. The simulations in the next section aimed to investigate the structures and populations of species present at equilibrium in D76N, including the I_T-state and other N-state derived species using replica-averaged, metadynamic-metainference molecular dynamics simulations.

5.2.2 Metadynamic-metainference molecular dynamics simulations as a tool to explore the conformational energy landscape of D76N

This section describes the use of metadynamic-metainference molecular dynamics simulations (MM-MD) [33, 32] to explore the conformational landscape of D76N and to provide physical insights into the observed experimental results described in this chapter. The MM-MD simulations are replica averaged simulations which incorporate experimental data using a Bayesian inference method to generate accurate structural ensembles, by estimating random, systematic errors, as well as errors in forward model prediction and error introduced by heterogeneous sample mixing [33, 32]. The ensemble averaged experimental data used in these simulations were both backbone chemical shifts (Chapter 3, Table 3.2) and ^{1}H-^{15}N residual dipolar couplings (Section 5.2.2), all collected for D76N at pH 7.4. The conformational ensemble generated through metainference was encouraged to explore conformational space through the use of parallel-bias metadynamics where a history-dependent energy potential was applied to a set of collective variables (CVs) (Section 5.2.2) in parallel [33, 241, 31]. The CVs are measurable properties to which the parallel bias metadynamics potential will be applied. CVs should discriminate between the various states of the system, so that modulating the values for each CV with the bias potential facilitates exploration of the energy landscape of the protein [31]. It is important therefore to include CVs which help explore 'slow' conformational motions [31].

Obtaining ^{1}H-^{15}N RDCs for D76N at pH 7.4

^{1}H-^{15}N residual dipolar couplings (RDCs) were measured for D76N at pH 7.4 as described in Methods 2.7.8 through the use of sensitivity-enhanced in-phase and antiphase (IPAP) J-coupled ^{1}H-^{15}N-HSQC spectra [69, 204]. In Figure 5.18 the upfield and the downfield components for both the unaligned (isotropic) and aligned (anisotropic) samples are shown, in these spectra the residual dipolar coupling is simply the difference in splitting (dipolar coupling) between the aligned and unaligned spectra for each residue [69].

The per-residue RDC values are shown in Figure 5.18c. The sample was aligned with Pf1 filamentous phage.

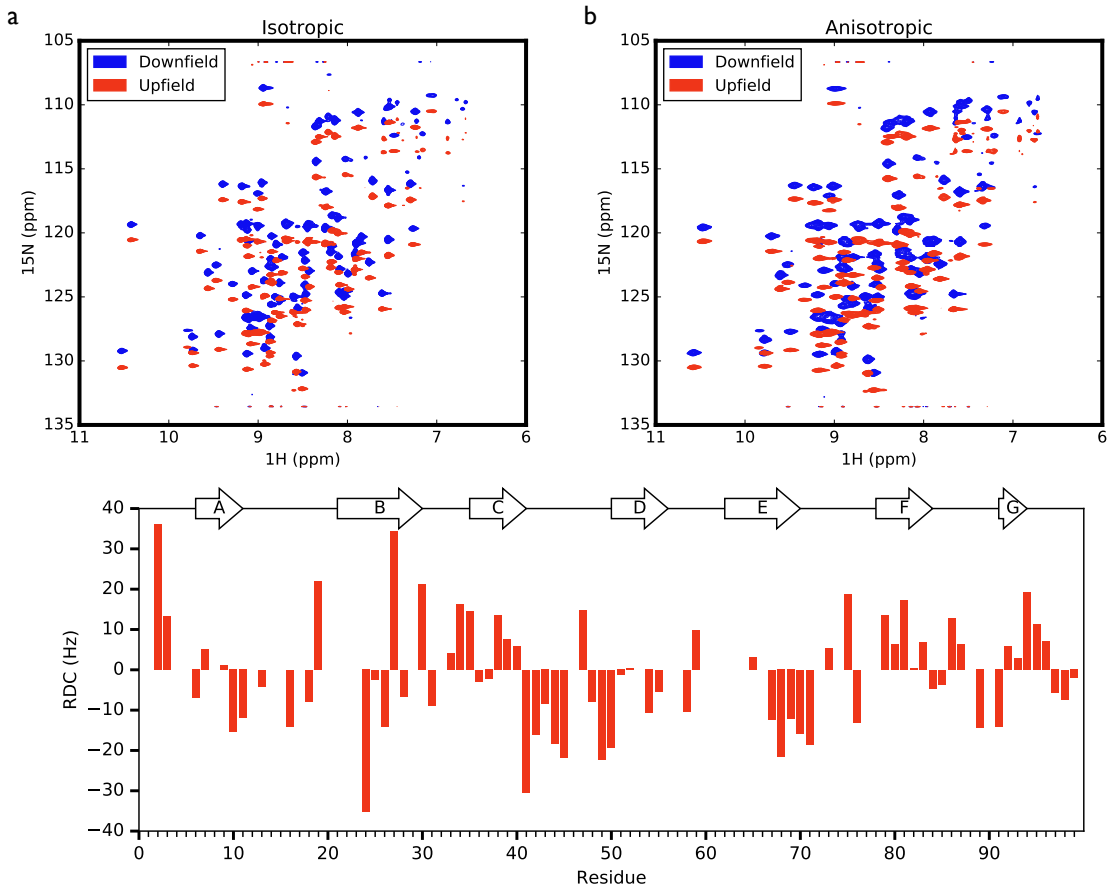


Fig. 5.18: Measurement of D76N RDCs at pH 7.4.
Sensitivity-enhanced in-phase and antiphase (IPAP) J-coupled ^{1}H-^{15}N-HSQC spectra of 300 µM D76N, 90.3 mM NaCl, 5 % D2O, 0.02% *(w/v)*NaN_3. The unaligned (isotropic) sample (a) contained no alignment media and the aligned (anisotropic) sample (b) was aligned with 15mg/ml Pf1 phage. Acquired with a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and Bruker Avance III HD console. Per-residue RDC values for D76N, pH 7.4 are shown in (c).

Table 5.1: Per-residue RDC values for D76N. Acquired using sensitivity-enhanced in-phase and antiphase (IPAP) J-coupled ^{1}H-^{15}N-HSQC spectra on a 300 μM D76N, 90.3 mM NaCl, 5 % D2O, 0.02% *(w/v)*NaN_3 sample in 25 mM sodium phosphate buffer at pH 7.4. Aligned with 15mg/ml Pf1 phage. Acquired with a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and Bruker Avance III HD console.

Residue	RDC (Hz)	Residue	RDC (Hz)	Residue	RDC (Hz)
2	35.9671	38	13.6213	71	-18.5331
3	13.2423	39	7.53452	73	5.41212
6	-7.0115	40	5.7229	75	18.6468
7	4.99522	41	-30.4337	76	-13.0679
9	1.05362	42	-16.0848	79	13.6288
10	-15.395	43	-8.44412	80	6.23076
11	-11.9764	44	-18.2526	81	17.3506
13	-4.21448	45	-21.7622	82	0.32594
16	-14.1064	47	14.8795	83	6.7462
18	-7.8453	48	-7.8453	84	-4.57074
19	22.0123	49	-22.2246	85	-3.6763
24	-35.3076	50	-19.2305	86	12.6889
25	-2.44834	51	-1.28102	87	6.28382
26	-13.9851	52	0.28046	89	-14.3868
27	34.2919	54	-10.6954	91	-14.1443
28	-6.57186	55	-5.39696	92	5.84418
30	21.2846	58	-10.3391	93	2.94862
31	-8.82312	59	9.75546	94	19.1243
33	4.10078	65	3.13812	95	11.3548
34	16.2439	67	-12.3251	96	7.14036
35	14.4247	68	-21.6257	97	-5.7608
36	-2.93346	69	-12.2038	98	-7.3526
37	-2.30432	70	-15.8043	99	-2.0087

Design of CVs for use in MM-MD simulations

In addition to the data used in the MM-MD simulations a range of collective variables were selected to aid exploration of the D76N conformational landscape. The collective variables are detailed in Table 5.2 and shown on the D76N crystal structure in Figure 5.19.

Table 5.2: Collective variables used in D76N MM-MD simulations

Collective Variable	**Description**
N-term	Dihedral correlation of phi-psi and psi-phi angles in the N-terminal 6 residues. This collective variable refers to movement in the N-term.
Pro32	Calculates torsion angle of the omega angle in Pro32 (amide bond between His31 and Pro32) which reports on *cis-trans* isomerisation.
AB-loop	Dihedral correlation of phi-psi and psi-phi angles in the AB loop.
Distance76-42	Distance between N76:1HD2 and N42:OD1.
Distance76-73	Distance between N76:1HD2 and T73:OG1.
D-strand-APBR	Anti parallel beta sheet content in the D and E strands, residues 50 to 70.
Pro32:N-term	Calculates the number of contacts between two groups of atoms. In this case the contacts between. Pro32 and the N-term
Pro32:DE-loop	Calculates the number of contacts between two groups of atoms. In this case the contacts between. Pro32 and the DE-loop

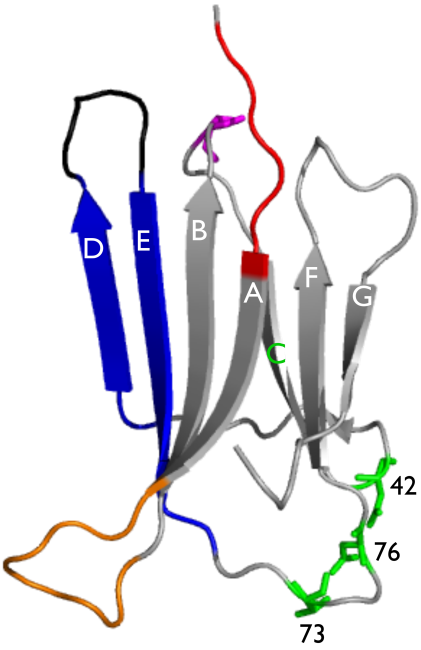


Fig. 5.19: Collective variables used in MM-MD simulations.
CVs on D76N (PDB:4FXL [183]) crystal structure. N-term (red), Pro32 (cyan), AB-loop (orange), Distance76-42 and Distance76-73 (green), D-strand-APBR (blue).

Metadynamics was performed with the collective variables (Table 5.2) in parallel over all 24 replicas in the simulation, where a history-dependent bias was applied in the form of gaussian functions. The gaussian bias potentials were converted into free energy landscapes as a function of each CV. The free energy landscapes for each CV are presented in Section 5.2.2

Monodimensional free energy landscapes from MM-MD simulations

The D76N AB-loop is dynamic both on the timescales observable in the R_2 relaxation data in Section 5.2.1 and in the CPMG-RD experiments in Section 5.2.1. The AB-loop in D76N β_2m takes up different arrangements in various structures. In the crystal structure of monomeric D76N, the AB-loop occupies an arrangement where the loop points out into solution [183], however in the MHC-bound structure of D76N, the AB-loop takes an alternative arrangement [221], where the loop points towards the EF-loop (Figure 5.20). These AB-loop arrangements are also observed for WT β_2m [221]. The 'AB-loop' collective variable was therefore designed to ensure the entire ensemble of possible structures in the AB-loop were explored.



Fig. 5.20: Conformational differences in D76N crystal structures.
Monomeric D76N (PDB:4FXL [183]) crystal structure (grey) and D76N structure from crystal structure of a D76N-MHC complex (PDB:4L3C [221]) (cyan).

In Figure 5.21 the reconstructed free energy landscape for the D76N AB-loop collective variable shows that the AB-loop occupies a wide, single energy minimum. This result

indicates that the AB-loop is capable of taking up many structural arrangements, equally energetically favourable, without large energy barriers between them.

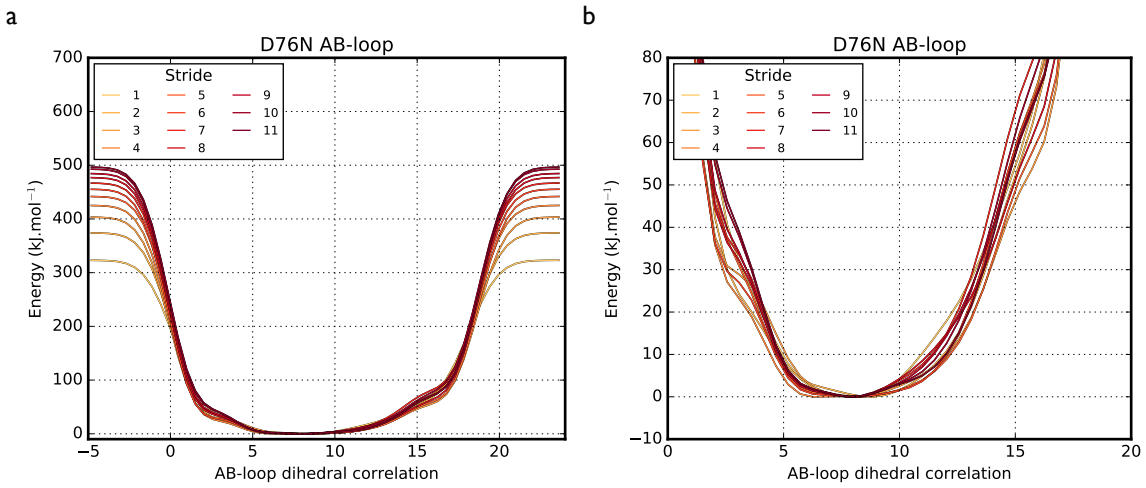


Fig. 5.21: Monodimensional free energy landscape as a function of the AB-loop CV. The full reconstructed energy landscape for the collective variable 'AB-loop' is shown in (a) and a focused view of the energy minimum is shown in (b). The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence.

Like the AB-loop, the D-strand, DE-loop and E-strand are region which are dynamic in T_2 relaxation (Section 5.2.1) and in CPMG-RD experiments (Section 5.2.1). The region encompassing residues 50-70 (D-strand, DE-loop, and E-strand) is included in the 'D-strand-APBR' collective variable. This collective variable is designed to explore the possible arrangements of this area, by modulating the anti-parallel β-content of the region. This area also differs between the monomeric and MHC-bound D76N crystal structures presented in Figure 5.20, where the D-strand is a straight β-strand in the monomeric form and occupies a β-bulge in the MHC-bound structure.

The 'D-strand-APBR' energy landscape in Figure 5.22 shows that residues 50-70 occupy an ordered state (RMSD values close to 0 nm) and a disordered state (RMSD values close to 3 nm), and a state in between (RMSD value approx 1.5 nm). This result suggests that this area of the protein exchanges between multiple states, with different levels of anti-parallel β-content, consistent with the observations made about the region in the CPMG-RD experiments in Section 5.2.1. This region however does not appear to have

fully converged yet as the energy minima are still changing relative to one another. In later segments (strides), as the simulation progresses, the state with an RMSD value approx 1.5 nm becomes less populated, in favour of the two other states.

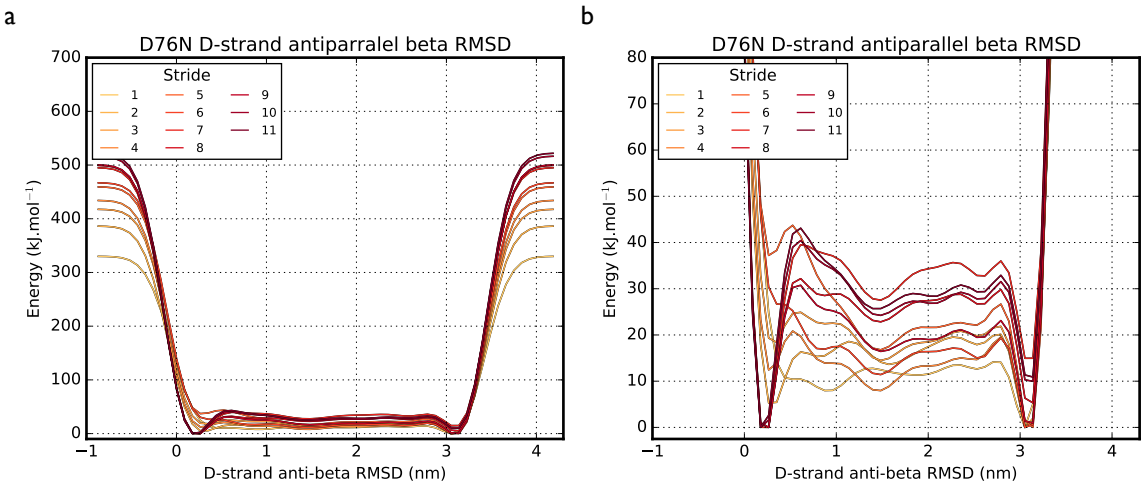


Fig. 5.22: Monodimensional free energy landscape as a function of the D-strand-APBR CV.
The full reconstructed energy landscape for the collective variable 'D-strand-APBR' is shown in (a) and a focused view of the energy minimum is shown in (b). The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence.

The collective variables 'Pro32', 'N-term', 'Pro32:N-term', and 'Pro32:DE-loop' are all designed to encourage exploration of the D76N I_T-state. The 'Pro32' collective variable encourages isomerisation of the Pro32, while the other CVs ('N-term', 'Pro32:N-term', and 'Pro32:DE-loop') provide conformational change in the area to facilitate the Pro32 isomerisation. The N-terminal 6 residues are known to have a role in determining Pro32 isomerisation since deletion of the first six residues prevents the ΔN6 variant form forming a *cis* Pro32 required to fold into the native state and the P5G variant occupies the I_T-state to approximately 60% presumably due to conformational flexibility introduced into the N-terminus [177].

The energy landscape in Figure 5.23 shows that Pro32 can exchange between *cis* (0 radians) and *trans* (π radians) conformations, through the direct modulation of the omega bond angle between His31 and Pro32. The native *cis*-conformation is lower energy

than the *trans* conformation, consistent with observations made in Chapter 4, regarding the equilibrium population and stability of the D76N I_T-state.

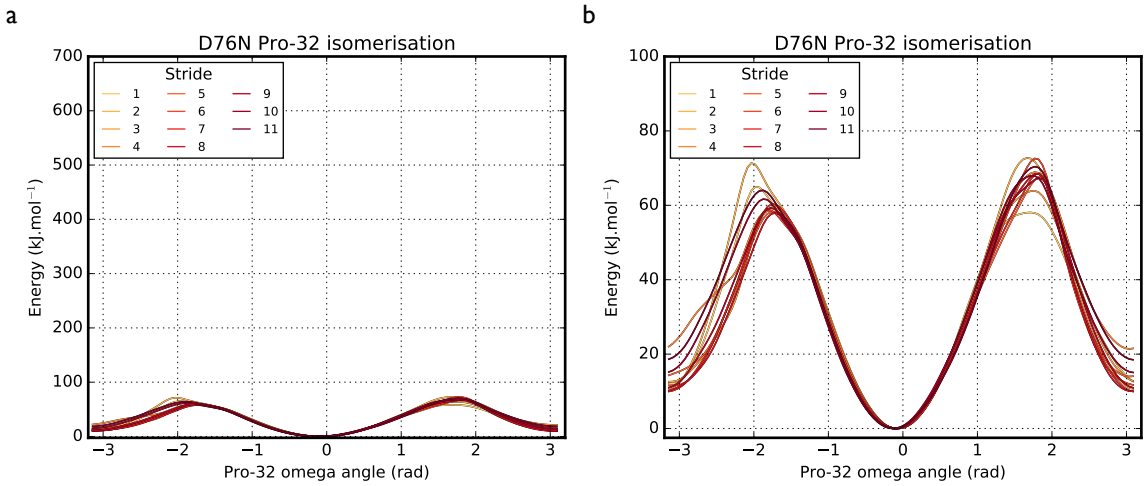


Fig. 5.23: Monodimensional free energy landscape as a function of the Pro32 CV. The full reconstructed energy landscape for the collective variable 'Pro32' is shown in (a) and a focused view of the energy minimum is shown in (b). The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence. The minimum at 0 radians corresponds to the *cis* configuration of Pro32 while the minimum at ±3.14 radians corresponds to the *trans* configuration.

In Figure 5.24 the energy landscape of the N-terminus (Residues 1-6) is explored, and shows that the N-terminus occupies a low energy conformation with an energy barrier of approximately 15 kJ.mol^{-1} to a higher energy state.

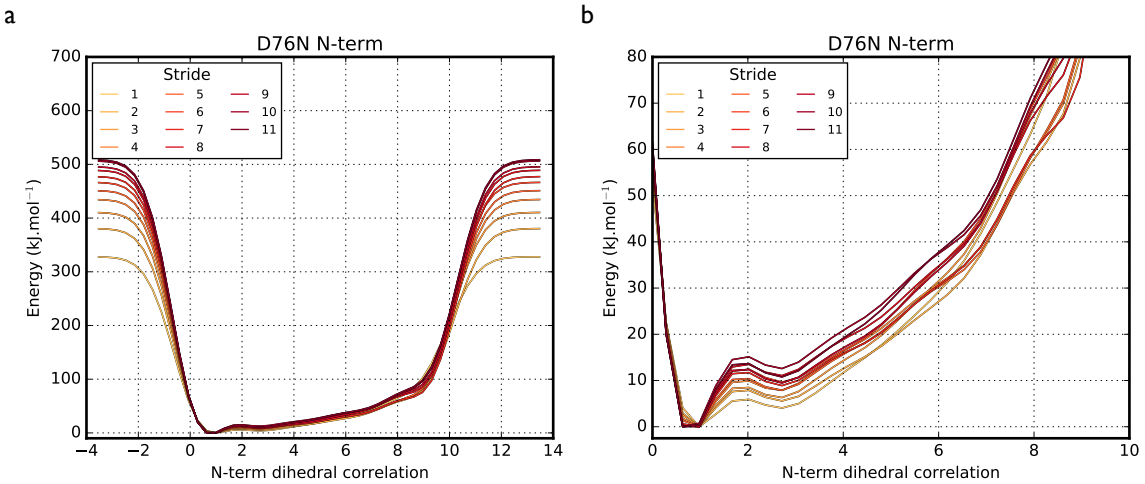


Fig. 5.24: Monodimensional free energy landscape as a function of the N-term CV. The full reconstructed energy landscape for the collective variable 'N-term' is shown in (a) and a focused view of the energy minimum is shown in (b). The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence.

The in 'Pro32:N-term' landscape in Figure 5.25c,d has a complex landscape with two low energy species. The 'Pro32:DE-loop' landscape in Figure 5.25a,b, however has a single low energy species and a many high energy conformations.

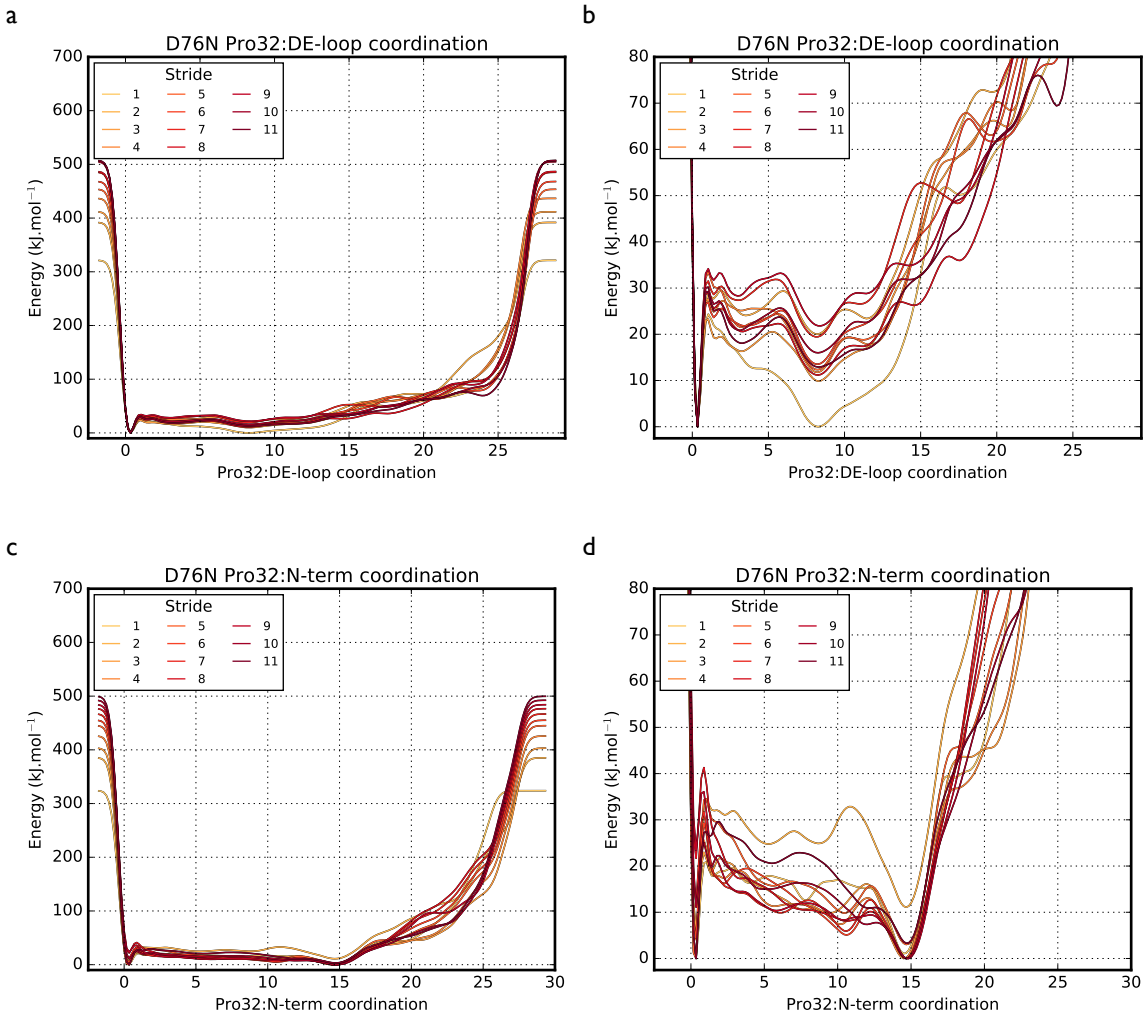


Fig. 5.25: Monodimensional free energy landscapes as a function of the Pro32:N-term and Pro32:DE-loop CVs.
The full reconstructed energy landscape for the collective variable 'Pro32:N-term' is shown in (a) and a focused view of the energy minimum is shown in (b). The 'Pro32:DE-loop' landscapes are shown in c and d, as full and focused views, respectively. The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence.

The final set of collective variables refers to residue 76 and surrounding residues thought to form side chain:side chain hydrogen bonds with residues 76 [184, 185] (N42 and T73). The distance from the residue 76 side chain and the side chains of N42 and T73 was modulated as a measure of hydrogen bonding. The landscapes for 76-42 and 76-73 distances are shown in Figure 5.26. Both of these landscapes have a dominant, low energy

population at approximately 2 Å (0.2 nm in Figure 5.26) consistent with hydrogen bond formation, with a further minimum at approximately 6 Å (0.6 nm, Figure 5.26) suggesting that the bond is now broken. There are also many high energy states with even greater distances, which may be due to greater a conformational change than a change in side chain hydrogen bonding. These collective variables were chosen to identify correlation between the side chain interactions of residue 76 and other conformational changes in the protein.

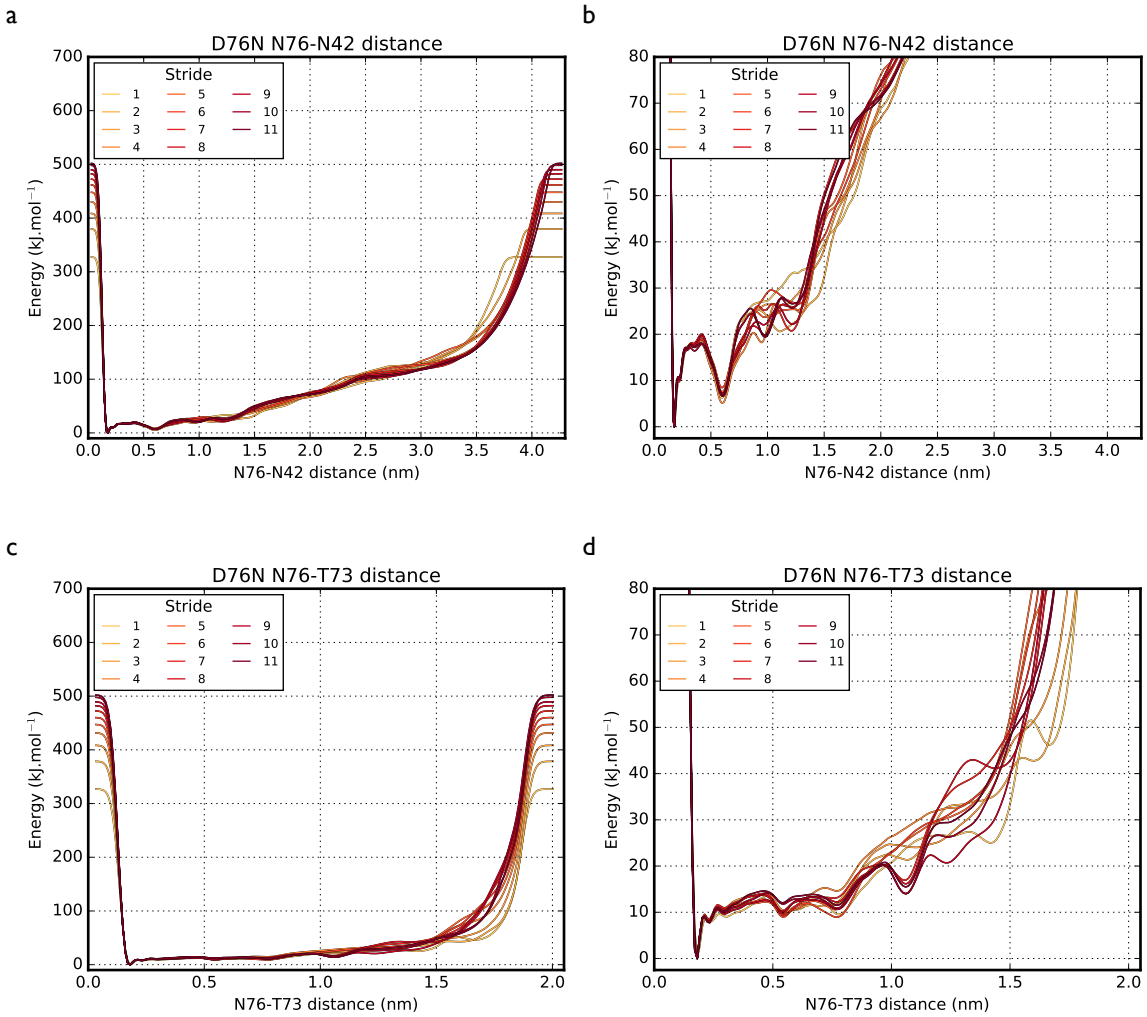


Fig. 5.26: Monodimensional free energy landscapes as a function of the Distance76-42 and Distance76-73 CVs.
The full reconstructed energy landscape for the collective variable 'Distance76-42' is shown in (a) and a focused view of the energy minimum is shown in (b). The 'Distance76-73' landscapes are shown in c and d, as full and focused views, respectively. The darkening red lines correspond to sequential groups of 40000 deposited gaussian bias functions over the full trajectory, which serves as a qualitative measure of convergence.

Multidimensional free energy landscapes from MM-MD simulations

Analysis of multidimensional energy landscapes can reveal relationships between different collective variables; for many of the CVs used in the D76N simulations, a number of correlated behaviours can be identified.

In Figure 5.27 the 'AB-loop' and 'D-strand-APBR' energy landscapes are plotted together as a multidimensional energy landscape. The data in Figure 5.27 show that the species with an ordered D-strand (0-0.25 nm D-strand anti-β RMSD) has a narrower distribution of AB-loop dihedral correlation than the states with anti-β RMSD values above approximately 0.5 nm, this observation suggests that there is a relationship between the AB-loop and the D-strand, DE-loop and E strand area of the protein.

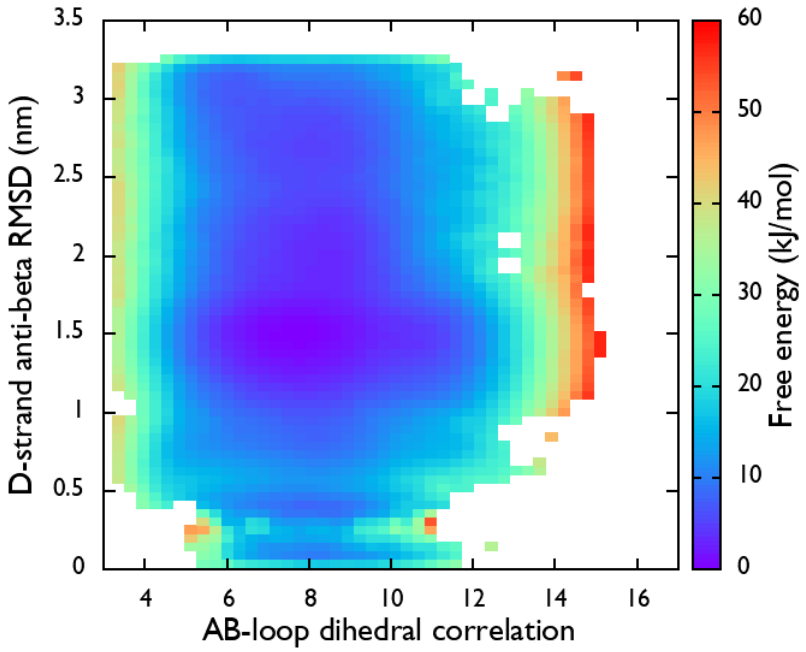


Fig. 5.27: Free energy surface as a function of the 'AB-loop' and 'D-strand-APBR' CVs. The 'AB-loop' and 'D-strand-APBR' CVs occupy the x and y axes, while the free energy is plotted as a heat map in the range 0-60 kJ.mol^{-1}, values outside of this range are not shown.

The 2D landscape in Figure 5.28 shows both of the distance CVs ('Distance76-42' and 'Distance76-73') show that the two H-bonds are mutually exclusive, and that when

one bond is formed (approximately 1-2 Å distance) the other bond distance is most often approximately 6 Å but no closer. This result shows that residue N76 forms side chain hydrogen bonds between either residues N42 or T73 most of the time, but not both together. These H-bonds do not appear to correlate with states found in the other energy landscapes presented in Section 5.2.2.

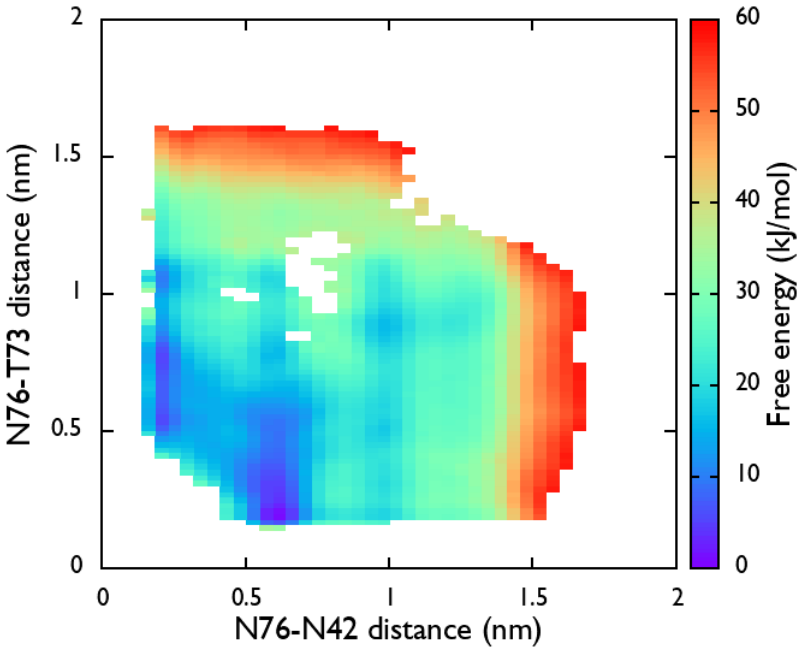


Fig. 5.28: Free energy surface as a function of the 'Distance76-42' and 'Distance76-73' CVs.
The Distance76-42' and 'Distance76-73'CVs occupy the x and y axes, while the free energy is plotted as a heat map in the range 0-60 kJ.mol^{-1}, values outside of this range are not shown.

Figure 5.29 shows that P32 isomerisation is linked to the interactions with the P32 residue and the N-terminal six residues, since there is an energy minimum present in the *cis* configuration, which is not present when P32 is *trans*, this result is promising as it replicates observations made about the P32 isomerisation in β_2m and the role of the N-terminal region in the process [177, 88].

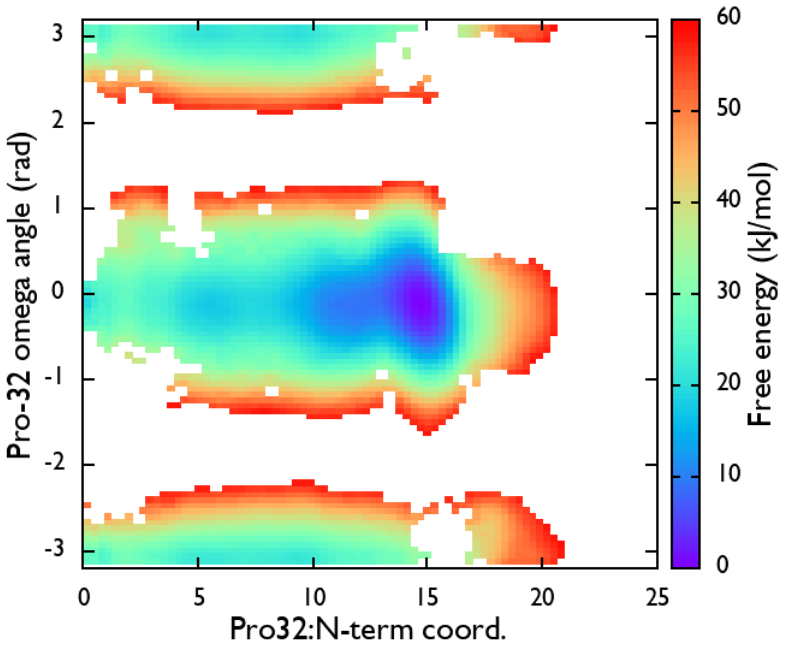


Fig. 5.29: Free energy surface as a function of the 'Pro32' and 'Pro32:N-term' CVs. The 'Pro32' and 'Pro32:N-term' CVs occupy the x and y axes, while the free energy is plotted as a heat map in the range 0-60 kJ.mol^{-1}, values outside of this range are not shown. The minimum at 0 radians corresponds to the *cis* configuration of Pro32 while the minimum at ±3.14 radians corresponds to the *trans* configuration.

Finally, the energy landscape in Figure 5.30 shows that the *trans*-P32 species is not permitted to form an ordered D-strand, DE-loop and E-strand region (RMSD values below to 0.5 nm), yet the *cis*-P32 species can occupy full range of RMSD values in this region.

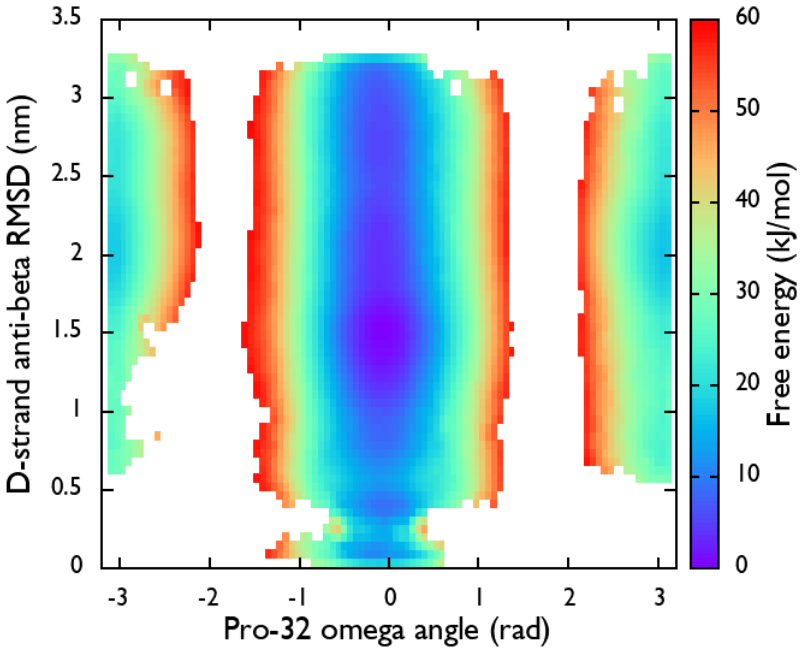


Fig. 5.30: Free energy surface as a function of the 'Pro32' and 'D-strand-APBR' CVs. The 'Pro32' and 'D-strand-APBR' CVs occupy the x and y axes, while the free energy is plotted as a heat map in the range 0-60 kJ.mol^{-1}, values outside of this range are not shown. The minimum at 0 radians corresponds to the *cis* configuration of Pro32 while the minimum at ±3.14 radians corresponds to the *trans* configuration.

Clustering analysis of substates present in the MM-MD structural ensemble

Next, the structures of D76N explored in the MM-MD simulations were analysed through a clustering analysis described in Methods 2.9. Where approximately 25000 frames from the full simulation trajectory (approximately 800000 frames) were grouped based on backbone RMSD. The clustering analysis aims to sort the multitude of structures observed in the simulation into distinct states, which provide insight into the structures of different conformational states present in the structural ensemble.

In the clustering protocol a 2Å cutoff was applied and 136 clusters were generated. The ten most populated clusters are summarised in Table 5.3.

Table 5.3: The number of frames per cluster generated with a 2Å cut-off

Cluster ID	No. of frames
0	16258
1	889
2	854
3	760
4	716
5	572
6	553
7	504
8	397
9	238
10	228

Each frame in the trajectory was then weighted based on the metadynamics energy potential and converted into populations, summarised in Table 5.4 and Figure 5.31

Table 5.4: Equilibrium population of clusters 1-10, weighted based on metadynamics bias potential

Cluster ID	Population (3 sf)
0	0.863
1	0.000
2	0.000
3	0.003
4	0.051
5	0.010
6	0.000
7	0.045
8	0.000
9	0.000
10	0.000

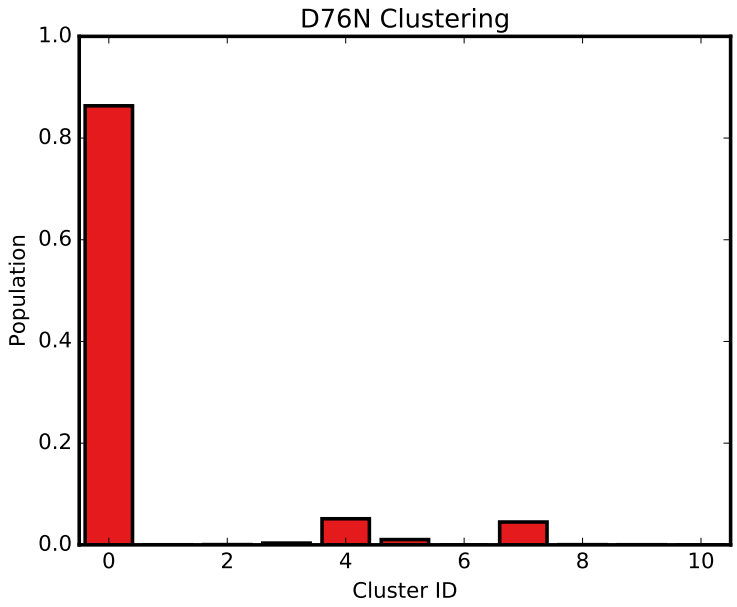


Fig. 5.31: Populations of the predominant D76N species identified in the clustering analysis.

The representative structure of the three most populated clusters (0, 4 and 7) are shown in Figure 5.32. Cluster 0 (Figure 5.32a), populated to 86% has a straight D-strand and therefore has high anti-parallel β-sheet content in the D-strand, DE-loop and E-strand region, consistent with the ordered species observed in the energy landscape for the 'D-

strand-APBR' CV. Additionally, in Cluster 0, the AB-loop occupies a position where the loop is pointing away from the protein into solution. The Cluster 0 structure resembles the crystal structure of monomeric D76N (PDB:4FXL [183]), and the two structures are compared in Figure 5.33a.

Clusters 4 and 7, which are populated to 5.1% and 4.5%, respectively (Figure 5.32b,c) do not have the straight D-strand observed in Cluster 0, and in the monomeric D76N crystal structure [183]. Instead the D-strand in Clusters 4 and 7 occupy a β-bulge structure observed in the crystal structure of D76N in the MHC complex (PDB:4L3C [221]). Clusters 4 and 7 have different AB-loop arrangements, where the Cluster 4 AB-loop points into solution, like in Cluster 0, and the Cluster 7 AB-loop points towards the other loops, resembling the AB-loop arrangement found in the crystal structure of D76N in the MHC complex (Figure 5.33b).

Interestingly, all of the representative structures (chosen to represent the centroid of the cluster) shown in Figure 5.32 are in the *cis*-Pro32 configuration. However, the free energy surface in Figure 5.30 shows that the *trans*-Pro32 species can not form an ordered D-strand, so are presumably not present in Cluster 0, suggesting that the I_T-state population of D76N is indeed low, as expected, based on observations in Chapters 3 and 4.

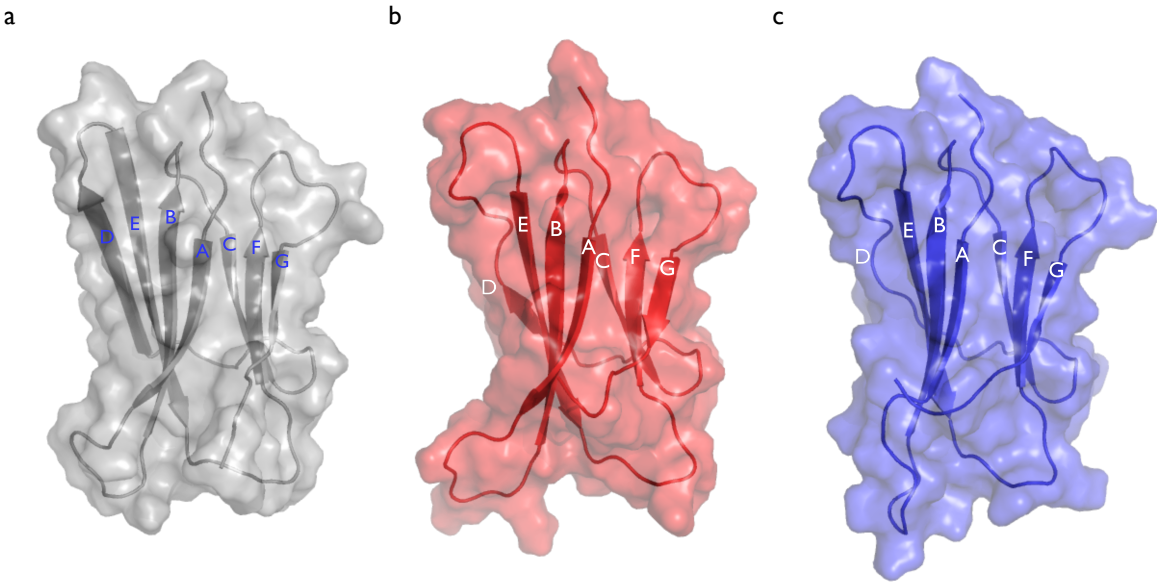


Fig. 5.32: Visualisation of the predominant D76N species identified in the clustering analysis.
Cluster 0 (a, grey) is populated to 86%, Cluster 4 (b, red) to 5.1% and Cluster 7 (c, blue) to 4.5%. The free energy difference between Cluster 0 and Cluster 4 is 7.06 kJ.mol^{-1} and the free energy difference between Cluster 0 and Cluster 7 is 7.37 kJ.mol^{-1} (calculated with Equation 2.12).

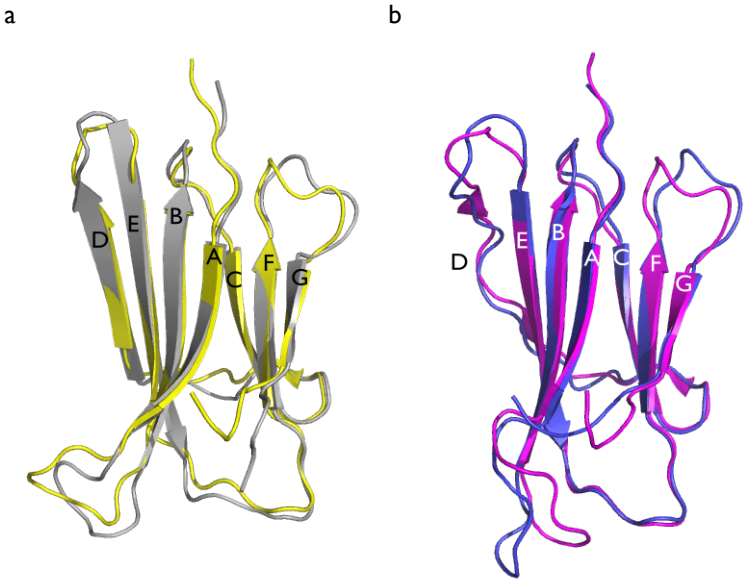


Fig. 5.33: D76N species identified in the clustering analysis replicate structural heterogeneity in D76N crystal structures.
(a) Cluster 0 (grey) is compared to the monomeric D76N crystal structure (yellow) (PDB:4FXL [183]). (b) Cluster 7 (blue) compared to the crystal structure of D76N in the MHC complex (magenta) (PDB:4L3C [221]).

5.3 Discussion

The NMR experiments described in this chapter indicated that both native WT and D76N β_2m have regions which appear to undergo conformational exchange with alternative structures. Areas of the protein which appear particularly dynamic include the D-strand, DE-loop and E-strand area and the AB-loop. This prior knowledge of the protein was used to design collective variables to carry out MM-MD simulations using ensemble averaged data to explore the D76N conformational landscape. The MM-MD simulations revealed that the N76 residue still makes many hydrogen bonding interactions with the surrounding loops in the structure. Interestingly, the D76A variant was shown to have universally faster R_2 relaxation rates than the WT, D76N and D76E proteins, which may reflect an inability of this particular variant to remain in conformations reliant on these interactions, given that the D76A variant is less aggregation-prone than D76N.

The MM-MD simulations on D76N revealed the presence of multiple species in equilibrium with the native state, which have been previously observed in various crystal structures of D76N both as a monomer and as part of the MHC complex, this is a promising result as it verifies that the simulations are indeed accessing known structures. The simulations, however, can utilise the metadynamics bias potentials to calculate equilibrium populations for the structures observed. The clustering analysis in Section 5.2.2 shows that the predominant structure occupied by D76N in solution resembles the monomeric crystal structure with a straight D-strand and an AB-loop which points into solution away from the other loops (Figure 5.33). The minor states have a β-bulge in the D-strand, and the AB-loop can be seen to take up both AB-loop conformations previously observed in the D76N-MHC complex (Figure 5.33).

Importantly, the WT β_2m D-strand is also straight in the crystal structure of the monomeric protein (PDB:1LDS, [151]) and occupies the β-bulge structure also observed for WT β_2m while bound to the MHC (PDB:4L3C [221]). However, the solution NMR structure of WT β_2m (PDB:2XKS, [88]) shows the WT D-strand as a β-bulge and not the ordered configuration observed in the monomeric crystal structure. Also, the WT β_2m AB-loop also occupies the configuration seen in the MHC bound complex (PDB:4L29 [221]) where the loop is tucked in and pointed towards the other loops in the structure rather

than pointing out into solution (Figure 5.34). ^{1}H-^{1}H NOESY experiments carried out on WT β_2m [151] show that even in conditions used for crystallisation of the WT monomer, the WT D-strand does not appear to be ordered in solution. Instead the ^{1}H-^{1}H NOESY data is consistent with a β-bulge structure showing that the WT crystal structure has instead trapped a rare conformer of β_2m with an ordered D-strand. The formation of β-bulges in protein edge strands is thought to be a negative design strategy to avoid edge-to-edge aggregation [242].

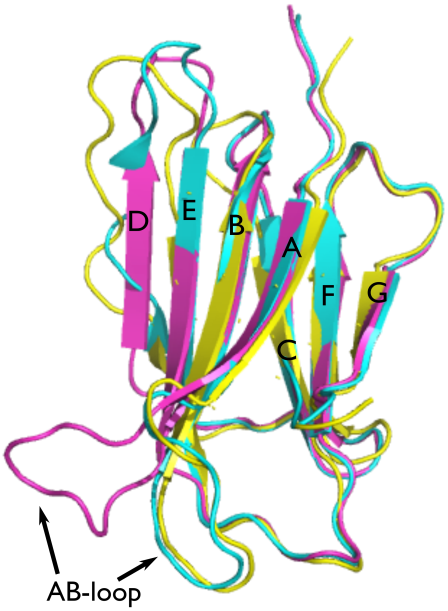


Fig. 5.34: WT β_2m structural heterogeneity.
The monomeric WT β_2m crystal structure (magenta) (PDB:1LDS [151]) compared to the crystal structure of WT β_2m in the MHC complex (cyan) (PDB:4L29 [221]) and the WT β_2m solution NMR structure (yellow) (PDB:2XKS [88]).

The simulations in this chapter suggest that D76N does occupy a major species (86%) with a straight edge strand, and two minor species (5.1% and 4.5%) which have β-bulges in the D-strand. While these results do not directly inform about the role of the D-strand in aggregation, they suggest a testable hypothesis. One such hypothesis is that this species may represent an alternative mechanism for D76N which is independent of the I_T-state, consistent with both the alternative disease progression [183] and the

different fibril structure resulting from D76N aggregation [223]. However, further work is required to test the hypothesis that the ordered edge strand in D76N is a factor in the aggregation pathway, through conducting MM-MD simulations on WT β_2m and on D76N using data obtained at pH 6.2 to compare the populations of the different states observed (these simulations are ongoing). Further verification should include the production of a D76N variant containing a D-strand mutation able to retard aggregation akin to DE-loop variants such as W60G which are less aggregation prone than WT β_2m and D59P which is more aggregation prone, presumably due to altered conformational flexibility in the D-strand/DE-loop region [243, 244, 245].

Chapter 6

Summary and conclusions

6.1 Success and limitations

In this section, the questions raised in Chapter 1 are addressed and the success and limitations of the results in Chapters 3 to 5 are discussed.

6.1.1 What is the role of the I_T-state in D76N aggregation?

The studies in Chapter 3 showed that the native structure of D76N is similar to the WT β_2m structure, as measured using both far-UV CD and ^{1}H-^{15}N HSQC spectra, consistent with the similarity of the WT and D76N β_2m crystal structures [184]. The stability of the D76N variant, however, is impaired. The $\Delta\Delta G^{\circ}_{un}$ of D76N compared with WT, measured by equilibrium urea denaturation is 9.5 kJ.mol^{-1} and the $T_{m,app}$ difference measured by thermal denaturation is 8.9°C (at pH 7.4). These observations are suggestive of a change in the energy landscape of the D76N variant, where the native, major state is destabilised, therefore lowering the energy barrier to other states, some of which may be involved in aggregation.

Since the folding intermediate, known as the I_T-state, present at equilibrium in WT β_2m landscape (to approx 5%) [184, 177] is a known precursor to aggregation; others have studied the role of the I_T-state in D76N aggregation and proposed that the I_T-state is responsible for D76N aggregation. Mangione et al. [184] suggest that D76N aggregation is due to an increased equilibrium population of the I_T-state (25%), measured by GdnHCl denaturation/refolding experiments. Chong et al. [186] suggest the D76N I_T-state is

structurally different and more aggregation prone, based upon molecular dynamics simulations, while Rui et al. [187] suggest that the D76N can take two separate structures, one of which is particularly aggregation-prone, again through molecular dynamics simulations. The work in Chapters 3 and 4, however, show that neither the population, the overall fold or the aggregation propensity of the D76N I_T-state is significantly different from the WT I_T-state.

The far-UV CD experiments in Chapter 3 indicate that the population of I_T-state in both WT and D76N is similar by comparison of their respective far-UV CD spectra in comparison to ΔN6 and a simulated 25% I_T-state spectrum.

The refolding experiments in Chapter 4 show that the folding pathway and refolding rate constants for D76N and other D76 variants remain the same. The ^{1}H-^{15}N HSQC spectra of the I_T-state for WT and D76N, captured during a real-time refolding reaction are similar which suggests that the overall fold of the WT and D76N I_T-states are similar. Peaks corresponding to the I_T-state can not be detected in the equilibrium D76N spectrum, which also suggests that the equilibrium I_T-state population is not increased.

These observations about the D76N I_T-state are reinforced by studies of a ΔN6-D76N variant in Chapter 4. The ΔN6-D76N variant, which is not present in the D76N associated disease [183], is destabilised by 11.8°C compared with D76N (pH 6.2). This stability difference between ΔN6-D76N and D76N is similar to that between ΔN6-WT and ΔN6, which is 10.3°C (pH 6.2). These stability differences suggest that the relative populations of native state and I_T-state is the same in WT and D76N, consistent with the observations in Chapter 3.

The stability difference between the WT and D76N variant and between ΔN6-WT and ΔN6-D76N are 10.9 and 12.4°C, respectively (pH 6.2). These stability differences suggest that the effect of the ΔN6 truncation and the D76N substitution are non-cooperative as the stability differences are additive. This is also consistent with the observations in Chapter 4 which show that the refolding rate is unaffected by D76 substitutions.

Finally, aggregation assays carried out in Chapter 3 and Chapter 4 show that the ΔN6-D76N and ΔN6 variants do not aggregate as rapidly as the full-length D76N variant,

suggesting that there is another conformational state in D76N that is more aggregation prone than the ΔN6/I_T-state species in both WT and D76N.

6.1.2 How important is the identity of residue 76 for aggregation?

The split-β-lactamase assay presented in Chapter 3, facilitated the saturation mutagenesis of position 76 in β_2m without the need to purify all 18 additional variants. The split-β-lactamase assay had previously been used to characterise the stability of various Im7 variants [218] and the split-β-lactamase assay was also later used to test the effect of small molecule inhibitors on amyloidogenic IDPs [246] (islet amyloid polypeptide and amyloid β). The results of the split-β-lactamase assay in Chapter 3 showed that any change in the identity of position 76 impaired the reduced the split-β-lactamase assay scores for β_2m. The D76E and D76A variants performed better than D76N, but not as well as WT, while all other variants were the same or worse than D76N. However, since the split-β-lactamase assay has previously been used to probe protein stability and aggregation in two different studies, the reason for the reduced split-β-lactamase assay score was not clear. The purification of D76E and D76A variants was therefore carried out. However, the thermal stability measurements and the aggregation assays both agree with the scores from the split-β-lactamase assay. Thus, while the assay did indeed predict the effect of the D76 amino acid substitutions in D76E and D76A, what the assay is most sensitive to is still not clear. The correlation of the scores with the hydrophobicity and size of the residues in Chapter 3 show that the assay is measuring more than a simple physical property of the substituted residue, however, and is indeed reporting on the viability of the new variant in some sense. Further work is required to understand the assay further, such as testing mutants which are thermally unstable, while not aggregation prone (V37A in WT sequence or mβ_2m), and other variants which are aggregation-prone, yet not significantly destabilised (P5G) [177].

The split-β-lactamase assay was useful in understanding the importance of residue 76 in determining the stability and aggregation of β_2m. Studies by De Rosa et al. [222] showed that D-N mutations at other sites in β_2m did not affect the structure, stability or aggregation of the protein. However, the results in Chapter 3 showed that any change

of residue at position 76 was detrimental to the protein. The purified D76E and D76A variants were both destabilised and aggregation-prone, although not as severely as D76N. These results are interesting, particularly in the case of D76A, since D76A is not capable of any side chain hydrogen bonding interactions yet the variant is less aggregation-prone and more destabilised than D76N. The D76N variant may, therefore, be more aggregation prone than D76A because of a side chain hydrogen bonding interaction which stabilises an aggregation-prone state. The presence of such a state was investigated in Chapter 5.

6.1.3 Is there an I_T-state independent excited state involved in D76N aggregation? What is the nature of this state?

The investigation into the D76N I_T-state had shown that the D76N and WT I_T-states are remarkably similar and that the full-length D76N protein aggregated more rapidly than the ΔN6-D76N and ΔN6 variants. Therefore, the aggregation of D76N must proceed via an I_T-state independent species. Thus, the conformational landscape of D76N was investigated further, using solution NMR experiments and molecular dynamics simulations in Chapter 5. The HDX-NMR experiments carried out Chapter 5 showed that the difference in protection factors between WT and D76N was non-uniform. In the B and F strands in D76N (held together by the disulphide bond between residues 25 and 80) the protection factors were 5 times lower in D76N. However, in other regions such as the BC-loop and the top of the E-strand, near the DE-loop, the protection factors were even lower in comparison to WT with a 5-20 fold difference. These non-uniform protection factor differences are emblematic of an alternative conformational state, which is less protected from the solvent. However, the presence of the I_T-state, complicates the analysis of this data, as the I_T-state is far less protected from exchange than the native state, and the D76N I_T-state is less stable than the WT I_T-state. Further work, to investigate the HDX behaviour of the ΔN6 and ΔN6-D76N variants could be useful to help tease apart the contributions of various states to the HDX behaviour of the proteins.

Further investigation into the conformational dynamics of D76N was carried out using T_2 relaxation experiments, which are sensitive to both fast timescale dynamics (ps-ns) and slower (μs-ms) motions. In these experiments, the most significant differences between

D76N and WT were observed at pH 6.2, conditions which accelerate D76N aggregation, without noticeable effect on the native structure or thermal stability. These conditions (pH 6.2) may alter the D76N energy landscape in favour of the aggregation-prone state. The T_2 experiments at pH 6.2 show differences predominantly in the AB-loop. The D76E and D76A variants were particularly useful in the T_2 experiments as the larger differences could be observed in the AB-loop and also in the DE-loop region. CPMG-RD experiments also suggest that these areas are in conformational exchange on a µs-ms timescale. While most of the differences observed with the T_2 and CPMG-RD experiments were subtle, the results do suggest that there is indeed conformational changes in these regions.

The information gleaned from these NMR experiments was utilised in the design of CVs for MM-MD [33, 32] simulations to provide structural insights into the observed differences from the NMR experiments. These MM-MD experiments incorporated NMR data (RDCs and chemical shifts) but not data from the dynamics experiments explicitly. The RDCs and chemical shift data do, however, contain structural and dynamic information, and the predicted free energy landscapes generated from these simulations were promisingly consistent with the NMR dynamics results. For example, the free energy landscape for the AB-loop showed that this region occupies a range of structures, and the lack of any significant energy barriers are consistent with ps-ns dynamics observed in the T_2 relaxation experiments. The predicted free energy landscape corresponding to the D-strand through the DE-loop and the E-strand showed the presence of multiple states with energy barriers in the range 10-40 $kJ.mol^{-1}$. These more substantial energy barriers may be responsible for the slow time scale dynamics in the D-strand/DE-loop measured by CPMG-RD. However further experiments and analysis are required to calculate relative populations and energy barriers from the CPMG-RD experiments, to make a more detailed comparison to the predicted free energy landscape generated from the MM-MD simulations.

The clustering analysis in Chapter 5 shows that the MM-MD simulations have been successful in generating a number of structures for D76N, based on RDC and chemical shift data, which can be observed in existing crystal structures. The D76N structural ensemble generated from the MM-MD simulations can be clustered into three confor-

mational states, present at 86%, 5.1% and 4.6%. The major state (86%) resembles the crystal structure of D76N with a straight D-strand and AB-loop pointing into solution. One minor state (4.6%) resembles the MHC-bound D76N structure with a characteristic β-bulge in the D-strand and the AB-loop pointing towards the other loops, including the EF-loop where residue 76 is found. The third state (5.1%) appears to sit between the other two structures, with a partially structured D-strand and the AB-loop into solution. Importantly, the populations of the three states are compatible with study through CPMG-RD experiments, since the low-populated states observed in the MM-MD should be present in a population amenable to CPMG-RD. These states, which are different in the D-strand area and AB-loop region are also consistent with the regions of the protein observed to be undergoing exchange in the CPMG-RD experiments.

The structures of the three states present testable hypotheses revolving around the D-strand region and the connection between the AB-loop and EF-loop. As discussed in Chapter 5, solution NMR experiments could not find evidence for the presence of a straight D-strand akin to the structure observed in the crystal structure of monomeric WT β_2m [151]. Therefore it was proposed that the WT crystal structure had trapped a rare conformational state, which was suggested to be aggregation-prone since ordered edge strands are often linked to aggregation [242]. Further work is required, however, to identify whether any of the species present in the D76N ensemble are aggregation-prone, through mutagenesis and by carrying out MM-MD simulations on the WT protein (ongoing). Through comparison of the WT and D76N energy landscapes and structural ensembles, it may be possible to identify a population shift in the D76N landscape, and identify species unique to D76N that promote aggregation.

6.2 Conclusions

The studies in this thesis have shown that in the case of D76N aggregation there is more than one aggregation-prone state. Clearly, the I_T-state is aggregation-prone, however, only as much as the WT I_T-state is. However, the WT I_T-state is not populated enough at equilibrium to cause aggregation *in vitro*, or *in vivo* without extrinsic factors. These factors presumably alter the conformational energy landscape of the WT protein, tipping

the balance towards aggregation. In the case of the D76N variant, therefore, there must be an I_T-state independent species which is capable of promoting aggregation, and in fact, only a D at position 76 prevents this species from occurring. The NMR and MM-MD experiments in Chapter 5 have begun to detail the energy landscape of D76N and have revealed the presence of two low-populated species in atomic resolution. Exciting questions regarding these states remain regarding their importance in aggregation.

The MM-MD technique shows remarkable synergy with NMR techniques regarding timescales and populations. The MM-MD simulations are effective over a range of timescales from ps-ns fast timescales, to slow μs-ms timescales. These simulations can cross high energy barriers, such as those required for proline isomerisation and are sensitive to low populations. The MM-MD technique therefore shows huge promise in terms of detailing energy landscapes in atomic detail, integrating ensemble averaged data from NMR and also other techniques such as FRET [207], SAXS [207] and even cryo-EM [247].

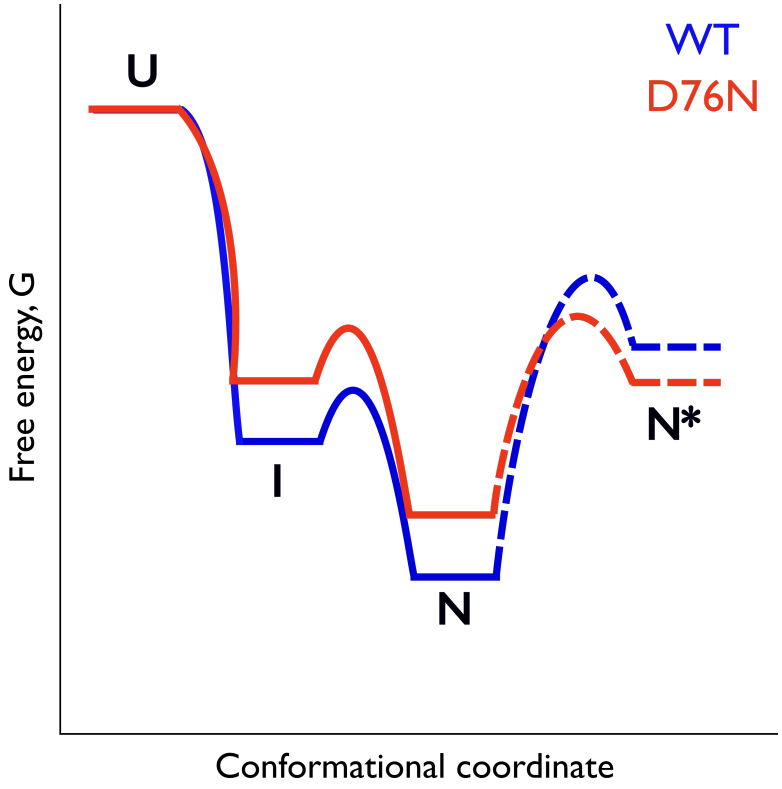


Fig. 6.1: Hypothetical conformational free energy landscape for WT and D76N β_2m In this idealised schematic, simplified free conformational energy landscapes are shown for both WT (blue) and D76N (red). The solid line represents the transition between the N-state and I_T-state and shows that the free energy difference between the N and I_T-state is similar in both WT and D76N which is reflected their similar relative populations. The energy barrier between the N and I_T-state is similar between WT and D76N, which results in similar rates of refolding for the two proteins. A hypothetical N* state is shown (dashed lines) where the relative free energy difference between N and N* are different in WT and D76N, possibly making the D76N N* state more populated at equilibrium. The N* state could be the species which is responsible for aggregation and a greater relative population of the N* state compared to N in D76N may be responsible for the greater aggregation propensity of D76N.

Further exciting questions about D76N aggregation remain, regarding the role of the excited states in the aggregation pathway. Do these states form oligomers important for aggregation? If so, what are the structures of dimers, and higher order species involved in aggregation? Are higher order species involved in D76N aggregation similar to those observed in WT/ΔN6? Or are they new, unique structures, thus explaining the different disease progression for D76N. These details could be investigated through the use of PRE experiments, and further CPMG-RD and CEST experiments at different protein concentrations.

Similar questions apply to the fibril structure. Are the subtle differences in WT and D76N fibril secondary structures (measured by ssNMR [248, 249, 250, 223]) a result of different aggregation pathways and precursor species? Fibril structures of D76N and WT β_2m could be obtained through ssNMR [251, 252, 253, 254, 255, 256] and cryo-EM experiments. Recently, a number of high-resolution fibril structures have been obtained with cryo-EM [257, 258, 259, 260, 261, 262]. The remarkable advances in this field make the possibility of obtaining such structures for β_2m a realistic possibility. The combination of the structural approaches discussed in this thesis, in addition to other techniques, useful in the study of higher order species (such as mass spectrometry [263, 264, 265, 266, 267], cryo-EM and ssNMR) could eventually provide a complete thermodynamic and kinetic description of β_2m aggregation ranging from the structures of excited states to the final fibril species.

References

[1] Beth M Beadle and Brian K Shoichet. "Structural Bases of Stability–function Tradeoffs in Enzymes". In: *Journal of Molecular Biology* 321.2 (2002), pp. 285–296. DOI: 10.1016/S0022-2836(02)00599-5.

[2] Nobuhiko Tokuriki and Dan S. Tawfik. "Protein Dynamism and Evolvability". In: *Science* 324.5924 (2009), pp. 203–207. DOI: 10.1126/science.1169375.

[3] Leo C James and Dan S Tawfik. *Conformational diversity and protein evolution - A 60-year-old hypothesis revisited*. 2003. DOI: 10.1016/S0968-0004(03)00135-X.

[4] Abbas Razvi and J. Martin Scholtz. "Lessons in stability from thermophilic proteins". In: *Protein Science* 15.7 (2006), pp. 1569–1578. DOI: 10.1110/ps.062130306.

[5] Darin M. Taverna and Richard A. Goldstein. "Why are proteins marginally stable?" In: *Proteins: Structure, Function, and Genetics* 46.1 (2002), pp. 105–109. DOI: 10.1002/prot.10016.

[6] P. Zavodszky et al. "Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins". In: *Proceedings of the National Academy of Sciences* 95.13 (1998), pp. 7406–7411. DOI: 10.1073/pnas.95.13.7406.

[7] H Frauenfelder, S. Sligar, and P. Wolynes. "The energy landscapes and motions of proteins". In: *Science* 254.5038 (1991), pp. 1598–1603. DOI: 10.1126/science.1749933.

[8] Katherine Henzler-Wildman and Dorothee Kern. "Dynamic personalities of proteins". In: *Nature* 450.7172 (2007), pp. 964–972. DOI: 10.1038/nature06522.

[9] Andrew J. Baldwin and Lewis E. Kay. "NMR spectroscopy brings invisible protein states into focus". In: *Nature Chemical Biology* 5.11 (2009), pp. 808–814. DOI: 10.1038/nchembio.238.

[10] Lewis E. Kay. "New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy". In: *Journal of Molecular Biology* 428.2 (2016), pp. 323–331. DOI: 10.1016/j.jmb.2015.11.028.

[11] D Flemming Hansen, Pramodh Vallurupalli, and Lewis E Kay. "Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states". In: *Journal of Biomolecular NMR* 41.3 (2008), pp. 113–120. DOI: 10.1007/s10858-008-9251-5.

[12] Nicholas J. Anthis and G. Marius Clore. "Visualizing transient dark states by NMR spectroscopy". eng. In: *Quarterly Reviews of Biophysics* 48.01 (2015), pp. 35–116. DOI: 10.1017/S0033583514000122.

[13] Fabrizio Chiti and Christopher M Dobson. "Amyloid formation by globular proteins under native conditions". In: *Nature Chemical Biology* 5.1 (2009), pp. 15–22. DOI: 10.1038/nchembio.131.

[14] D Beece et al. "Solvent viscosity and protein dynamics". In: *Biochemistry* 19.23 (1980), pp. 5147–5157. DOI: 10.1021/bi00564a001.

[15] R H Austin et al. "Dynamics of ligand binding to myoglobin". In: *Biochemistry* 14.24 (1975), pp. 5355–5373. DOI: 10.1021/bi00695a021.

[16] Katalin Török et al. "Dual Effect of ATP in the Activation Mechanism of Brain Ca 2+ /Calmodulin-Dependent Protein Kinase II by Ca2+ /Calmodulin †". In: *Biochemistry* 40.49 (2001), pp. 14878–14890. DOI: 10.1021/bi010920+.

[17] Chung-Jung Tsai et al. "Folding funnels, binding funnels, and protein function". In: *Protein Science* 8.6 (1999), pp. 1181–1190. DOI: 10.1110/ps.8.6.1181.

[18] Henry Van Den Bedem and James S. Fraser. "Integrative, dynamic structural biology at atomic resolution - It's about time". In: *Nature Methods* 12.4 (2015), pp. 307–318. DOI: 10.1038/nmeth.3324.

[19] Ian R Kleckner and Mark P Foster. "An introduction to NMR-based approaches for measuring protein dynamics". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814.8 (2011), pp. 942–968. DOI: 10.1016/j.bbapap.2010.10.012.

[20] Anthony K Mittermaier and Lewis E Kay. "Observing biological dynamics at atomic resolution using NMR". In: *Trends in Biochemical Sciences* 34.12 (2009), pp. 601–611. DOI: 10.1016/j.tibs.2009.07.004.

[21] Masanori Osawa et al. "Functional dynamics of proteins revealed by solution NMR". In: *Current Opinion in Structural Biology* 22.5 (2012), pp. 660–669. DOI: 10.1016/j.sbi.2012.08.007.

[22] Ashok Sekhar and Lewis E Kay. "NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers". In: *Proceedings of the National Academy of Sciences* 110.32 (2013), pp. 12867–12874. DOI: 10.1073/pnas.1305688110.

[23] Chitra Narayanan et al. "Applications of NMR and computational methodologies to study protein dynamics". In: *Archives of Biochemistry and Biophysics* 628 (2017), pp. 71–80. DOI: 10.1016/j.abb.2017.05.002.

[24] Pramodh Vallurupalli, Guillaume Bouvignies, and Lewis E. Kay. "A Computational Study of the Effects of 13 C- 13 C Scalar Couplings on 13 C CEST NMR Spectra: Towards Studies on a Uniformly 13 C-Labeled Protein". In: *ChemBioChem* 14.14 (2013), pp. 1709–1713. DOI: 10.1002/cbic.201300230.

[25] Pramodh Vallurupalli, Guillaume Bouvignies, and Lewis E. Kay. "Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation". In: *Journal of the American Chemical Society* 134.19 (2012), pp. 8148–8161. DOI: 10.1021/ja3001419.

[26] Sture Forsén and Ragnar A Hoffman. "Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance". In: *The Journal of Chemical Physics* 39.11 (1963), pp. 2892–2901. DOI: 10.1063/1.1734121.

[27] H. Y. Carr and E. M. Purcell. "Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments". In: *Physical Review* 94.3 (1954), pp. 630–638. DOI: 10.1103/PhysRev.94.630.

[28] S. Meiboom and D. Gill. "Modified Spin-Echo Method for Measuring Nuclear Relaxation Times". In: *Review of Scientific Instruments* 29.8 (1958), pp. 688–691. DOI: 10.1063/1.1716296.

[29] Pramodh Vallurupalli, D Flemming Hansen, and Lewis E Kay. "Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy". In: *Proceedings of the National Academy of Sciences* 105.33 (2008), pp. 11766–11771. DOI: 10.1073/pnas.0804221105.

[30] Dmitry M Korzhnev and Lewis E Kay. "Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding". In: *Accounts of Chemical Research* 41.3 (2008), pp. 442–451. DOI: 10.1021/ar700189y.

[31] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. "Metadynamics". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (2011), pp. 826–843. DOI: 10.1002/wcms.31.

[32] Massimiliano Bonomi et al. "Metainference: A Bayesian inference method for heterogeneous systems". In: *Science Advances* 2.1 (2016), e1501177–e1501177. DOI: 10.1126/sciadv.1501177.

[33] Massimiliano Bonomi, Carlo Camilloni, and Michele Vendruscolo. "Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics." In: *Scientific reports* 6.August (2016), p. 31232. DOI: 10.1038/srep31232.

[34] Theodoros K Karamanos et al. "Mechanisms of amyloid formation revealed by solution NMR". In: *Prog Nucl Magn Reson Spectrosc* 88–89 (2015), pp. 86–104. DOI: http://dx.doi.org/10.1016/j.pnmrs.2015.05.002.

[35] J.P Carver and R.E Richards. "A general two-site solution for the chemical exchange produced dependence of T2 upon the Carr-Purcell pulse separation". In: *Journal of Magnetic Resonance (1969)* 6.1 (1972), pp. 89–105. DOI: 10.1016/0022-2364(72)90090-X.

[36] S. Morin et al. "relax: the analysis of biomolecular kinetics and thermodynamics using NMR relaxation dispersion data". In: *Bioinformatics* 30.15 (2014), pp. 2219–2220. DOI: 10.1093/bioinformatics/btu166.

[37] Nikolai R Skrynnikov, Frederick W Dahlquist, and Lewis E Kay. "Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments". In: *Journal of the American Chemical Society* 124.41 (2002), pp. 12352–12360. DOI: 10.1021/ja0207089.

[38] Guillaume Bouvignies et al. "A simple method for measuring signs of 1HN chemical shift differences between ground and excited protein states". In: *Journal of Biomolecular NMR* 47.2 (2010), pp. 135–141. DOI: 10.1007/s10858-010-9418-8.

[39] S W Englander and N R Kallenbach. "Hydrogen exchange and structural dynamics of proteins and nucleic acids." In: *Quarterly reviews of biophysics* 16.4 (1983), pp. 521–655. DOI: 10.1017/S0033583500005217.

[40] Yawen Bai et al. "Protein Folding Intermediates: Native-State Hydrogen Exchange." In: *Science (New York, N.Y.)* 269.5221 (1995), pp. 192–197. DOI: 10.1126/science.7618079.

[41] S Walter Englander et al. "Mechanisms and uses of hydrogen exchange". In: *Current Opinion in Structural Biology* 6.1 (1996), pp. 18–23. DOI: http://dx.doi.org/10.1016/S0959-440X(96)80090-X.

[42] Veniamin Chevelkov et al. "15NH/D-SOLEXSY experiment for accurate measurement of amide solvent exchange rates: application to denatured drkN SH3". In: *Journal of Biomolecular NMR* 46.3 (2010), pp. 227–244. DOI: 10.1007/s10858-010-9398-8.

[43] Nicholas C. Fitzkee, Dennis A. Torchia, and Ad Bax. "Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain". In: *Protein Science* 20.3 (2011), pp. 500–512. DOI: 10.1002/pro.582.

[44] Joshua I. Friedman et al. "Indirect Detection of Labile Solute Proton Spectra via the Water Signal Using Frequency-Labeled Exchange (FLEX) Transfer". In: *Journal of the American Chemical Society* 132.6 (2010), pp. 1813–1815. DOI: 10.1021/ja909001q.

[45] A.L Schwartz and J.D Cutnell. "One- and two-dimensional NMR studies of exchanging amide protons in glutathione". In: *Journal of Magnetic Resonance (1969)* 53.3 (1983), pp. 398–411. DOI: 10.1016/0022-2364(83)90213-5.

[46] Silvia Spera, Mitsuhiko Ikura, and Ad Bax. "Measurement of the exchange rates of rapidly exchanging amide protons: Application to the study of calmodulin and its complex with a myosin light chain kinase fragment". In: *Journal of Biomolecular Nmr* 1.2 (1991), pp. 155–165. DOI: 10.1007/BF01877227.

[47] Heinrich Roder, Gülnur A. Elöve, and S. Walter Englander. "Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR". In: *Nature* 335.6192 (1988), pp. 700–704. DOI: 10.1038/335700a0.

[48] J Lu and F W Dahlquist. "Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR." In: *Biochemistry* 31.20 (1992), pp. 4749–56. DOI: 10.1021/bi00135a002.

[49] M D Jacobs and R O Fox. "Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy." In: *Proceedings of the National Academy of Sciences* 91.2 (1994), pp. 449–453. DOI: 10.1073/pnas.91.2.449.

[50] Y Bai et al. "Protein folding intermediates: native-state hydrogen exchange". In: *Science* 269.5221 (1995), pp. 192–197. DOI: 10.1126/science.7618079.

[51] Clare K. Woodward. "Hydrogen exchange rates and protein folding". In: *Current Opinion in Structural Biology* 4.1 (1994), pp. 112–116. DOI: 10.1016/S0959-440X(94)90068-X.

[52] Bryan E. Jones and C. Robert Matthews. "Early intermediates in the folding of dihydrofolate reductase from escherichia coli detected by hydrogen exchange and NMR". In: *Protein Science* 4.2 (2008), pp. 167–177. DOI: 10.1002/pro.5560040204.

[53] Stanislaw A. Gorski et al. "Equilibrium Hydrogen Exchange Reveals Extensive Hydrogen Bonded Secondary Structure in the On-pathway Intermediate of Im7". In: *Journal of Molecular Biology* 337.1 (2004), pp. 183–193. DOI: 10.1016/j.jmb.2004.01.004.

[54] Ngoc-Diep Vu, Hanqiao Feng, and Yawen Bai. "The Folding Pathway of Barnase: The Rate-Limiting Transition State and a Hidden Intermediate under Native Conditions". In: *Biochemistry* 43.12 (2004), pp. 3346–3356. DOI: 10.1021/bi0362267.

[55] Ruiai Chu et al. "Relationship between the Native-State Hydrogen Exchange and Folding Pathways of a Four-Helix Bundle Protein". In: *Biochemistry* 41.25 (2002), pp. 7998–8003. DOI: 10.1021/bi025872n.

[56] Hanqiao Feng, Ngoc-Diep Vu, and Yawen Bai. "Detection of a Hidden Folding Intermediate of the Third Domain of PDZ". In: *Journal of Molecular Biology* 346.1 (2005), pp. 345–353. DOI: 10.1016/j.jmb.2004.11.040.

[57] Hidenori Kato, Hanqiao Feng, and Yawen Bai. "The Folding Pathway of T4 Lysozyme: The High-resolution Structure and Folding of a Hidden Intermediate". In: *Journal of Molecular Biology* 365.3 (2007), pp. 870–880. DOI: 10.1016/j.jmb.2006.10.047.

[58] Denis Canet et al. "Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme". In: *Nature Structural Biology* 9.4 (2002), pp. 308–315. DOI: 10.1038/nsb768.

[59] Mireille Dumoulin et al. "Reduced Global Cooperativity is a Common Feature Underlying the Amyloidogenicity of Pathogenic Lysozyme Mutations". In: *Journal of Molecular Biology* 346.3 (2005), pp. 773–788. DOI: 10.1016/j.jmb.2004.11.020.

[60] Theodoros K Karamanos et al. "A population shift between sparsely-populated folding intermediates determines amyloidogenicity." In: *Journal of the American Chemical Society* (2016). DOI: 10.1021/jacs.6b02464.

[61] Junji Iwahara and G. Marius Clore. "Detecting transient intermediates in macromolecular binding by paramagnetic NMR". In: *Nature* 440.7088 (2006), pp. 1227–1230. DOI: 10.1038/nature04673.

[62] I. Solomon. "Relaxation Processes in a System of Two Spins". In: *Physical Review* 99.2 (1955), pp. 559–565. DOI: 10.1103/PhysRev.99.559.

[63] N. Bloembergen and L. O. Morgan. "Proton Relaxation Times in Paramagnetic Solutions. Effects of Electron Spin Relaxation". In: *The Journal of Chemical Physics* 34.3 (1961), pp. 842–850. DOI: 10.1063/1.1731684.

[64] Theodoros K. Karamanos et al. "Visualization of Transient Protein-Protein Interactions that Promote or Inhibit Amyloid Assembly". eng. In: *Molecular Cell* 55.2 (2014), pp. 214–226. DOI: 10.1016/j.molcel.2014.05.026.

[65] A. Bax. "Weak alignment offers new NMR opportunities to study protein structure and dynamics". In: *Protein Science* 12.1 (2003), pp. 1–16. DOI: 10.1110/ps.0233303.

[66] N Tjandra and a Bax. "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." eng. In: *Science (New York, N.Y.)* 278.5340 (1997), pp. 1111–1114. DOI: 10.1126/science.278.5340.1111.

[67] Carlo Camilloni and Michele Vendruscolo. "A Tensor-Free Method for the Structural and Dynamical Refinement of Proteins using Residual Dipolar Couplings". In: *The Journal of Physical Chemistry B* 119.3 (2015), pp. 653–661. DOI: 10.1021/jp5021824.

[68] Martin Blackledge. "Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 46.1 (2005), pp. 23–61. DOI: 10.1016/j.pnmrs.2004.11.002.

[69] M Ottiger, F Delaglio, and A Bax. "Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra". In: *Journal of Magnetic Resonance* 131.2 (1998), pp. 373–378. DOI: 10.1006/jmre.1998.1361.

[70] J. R. Tolman et al. "NMR evidence for slow collective motions in cyanometmyoglobin". In: *Nature Structural Biology* 4.4 (1997), pp. 292–297. DOI: 10.1038/nsb0497-292.

[71] Frank Delaglio, Georg Kontaxis, and Ad Bax. "Protein structure determination using molecular fragment replacement and NMR dipolar couplings". In: *Journal of the American Chemical Society* 122.9 (2000), pp. 2142–2143. DOI: 10.1021/ja993603n.

[72] G Marius Clore, Angela M Gronenborn, and Nico Tjandra. "Direct Structure Refinement against Residual Dipolar Couplings in the Presence of Rhombicity of Unknown Magnitude". In: *Journal of Magnetic Resonance* 131.1 (1998), pp. 159–162. DOI: 10.1006/jmre.1997.1345.

[73] J. H. Prestegard, C. M. Bougault, and A. I. Kishore. "Residual dipolar couplings in structure determination of biomolecules". In: *Chemical Reviews* 104.8 (2004), pp. 3519–3540. DOI: 10.1021/cr030419i.

[74] Jean-Christophe Hus, Dominique Marion, and Martin Blackledge. "Determination of Protein Backbone Structure Using Only Residual Dipolar Couplings". In: *Journal of the American Chemical Society* 123.7 (2001), pp. 1541–1542. DOI: 10.1021/ja005590f.

[75] G. Marius Clore and Charles D. Schwieters. "Amplitudes of protein backbone dynamics and correlated motions in a small α/β protein: Correspondence of dipolar coupling and heteronuclear relaxation measurements". In: *Biochemistry* 43.33 (2004), pp. 10678–10691. DOI: 10.1021/bi049357w.

[76] Scott A. Showalter and Rafael Brüschweiler. "Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints". In: *Journal of the American Chemical Society* 129.14 (2007), pp. 4158–4159. DOI: 10.1021/ja070658d.

[77] Paul Guerry et al. "Mapping the population of protein conformational energy substates from NMR dipolar couplings". In: *Angewandte Chemie - International Edition* 52.11 (2013), pp. 3181–3185. DOI: 10.1002/anie.201209669.

[78] Alfonso De Simone et al. "Structures of the Excited States of Phospholamban and Shifts in Their Populations upon Phosphorylation". In: *Biochemistry* 52.38 (2013), pp. 6684–6694. DOI: 10.1021/bi400517b.

[79] Alfonso De Simone et al. "Characterization of the Interdomain Motions in Hen Lysozyme Using Residual Dipolar Couplings as Replica-Averaged Structural Restraints in Molecular Dynamics Simulations". In: *Biochemistry* 52.37 (2013), pp. 6480–6486. DOI: 10.1021/bi4007513.

[80] Markus Zweckstetter and Ad Bax. "Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR". In: *Journal of the American Chemical Society* 122.15 (2000), pp. 3791–3792. DOI: 10.1021/ja0000908.

[81] Miguel X. Fernandes et al. "An Analytical Solution to the Problem of the Orientation of Rigid Particles by Planar Obstacles. Application to Membrane Systems and to the Calculation of Dipolar Couplings in Protein NMR Spectroscopy". In: *Journal of the American Chemical Society* 123.48 (2001), pp. 12037–12047. DOI: 10.1021/ja011361x.

[82] Andrew Almond and Jacob B Axelsen. "Physical Interpretation of Residual Dipolar Couplings in Neutral Aligned Media". In: *Journal of the American Chemical Society* 124.34 (2002), pp. 9986–9987. DOI: 10.1021/ja026876i.

[83] Judit A Losonczi et al. "Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition". In: *Journal of Magnetic Resonance* 138.2 (1999), pp. 334–342. DOI: 10.1006/jmre.1999.1754.

[84] Lewis E Kay, Dennis a Torchia, and Ad Bax. "Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease". eng. In: *Biochemistry* 28.23 (1989), pp. 8972–8979. DOI: 10.1021/bi00449a003.

[85] Gerhard Wagner. "NMR relaxation and protein mobility". In: *Current Opinion in Structural Biology* 3.5 (1993), pp. 748–754. DOI: 10.1016/0959-440X(93)90059-T.

[86] Arthur G Palmer. "Dynamic properties of proteins from NMR spectroscopy." In: *Current opinion in biotechnology* 4.4 (1993), pp. 385–91. DOI: 10.1016/0958-1669(93)90002-E.

[87] Arthur M. Mandel, Mikael Akke, and Arthur G. Palmer. "Backbone dynamics of Escherichia coli ribonuclease HI: Correlations with structure and function in an active enzyme". In: *Journal of Molecular Biology* 246.6 (1995), pp. 144–163. DOI: 10.1006/jmbi.1994.0073.

[88] Timo Eichner et al. "Conformational Conversion during Amyloid Formation at Atomic Resolution". eng. In: *Molecular Cell* 41.2 (2011), pp. 161–172. DOI: 10.1016/j.molcel.2010.11.028.

[89] Wim Robberecht and Thomas Philips. "The changing scene of amyotrophic lateral sclerosis". In: *Nature Reviews Neuroscience* 14.4 (2013), pp. 248–264. DOI: 10.1038/nrn3430.

[90] Ashok Sekhar et al. "Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways". In: *eLife* 4 (2015), pp. 1–33. DOI: 10.7554/eLife.07296.

[91] Caterina Bendotti et al. "Dysfunction of constitutive and inducible ubiquitin-proteasome system in amyotrophic lateral sclerosis: Implication for protein aggregation and immune response". In: *Progress in Neurobiology* 97.2 (2012), pp. 101–126. DOI: 10.1016/j.pneurobio.2011.10.001.

[92] Sheng Chen et al. "Autophagy Dysregulation in Amyotrophic Lateral Sclerosis". In: *Brain Pathology* 22.1 (2012), pp. 110–116. DOI: 10.1111/j.1750-3639.2011.00546.x.

[93] Samer Abou Ezzi, Makoto Urushitani, and Jean-Pierre Julien. "Wild-type superoxide dismutase acquires binding and toxic properties of ALS-linked mutant forms through oxidation". In: *Journal of Neurochemistry* 102.1 (2007), pp. 170–178. DOI: 10.1111/j.1471-4159.2007.04531.x.

[94] Rishi Rakhit et al. "Oxidation-induced Misfolding and Aggregation of Superoxide Dismutase and Its Implications for Amyotrophic Lateral Sclerosis". In: *Journal of Biological Chemistry* 277.49 (2002), pp. 47551–47556. DOI: 10.1074/jbc.M207356200.

[95] Melissa S. Rotunno and Daryl A. Bosco. "An emerging role for misfolded wild-type SOD1 in sporadic ALS pathogenesis". In: *Frontiers in Cellular Neuroscience* 7 (2013), p. 253. DOI: 10.3389/fncel.2013.00253.

[96] S Matsumoto et al. "Sporadic amyotrophic lateral sclerosis with dementia and Cu/Zn superoxide dismutase-positive Lewy body-like inclusions". In: *Clin Neuropathol* 15.1 (1996), pp. 41–46.

[97] Rishi Rakhit and Avijit Chakrabartty. "Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1762.11-12 (2006), pp. 1025–1037. DOI: 10.1016/j.bbadis.2006.05.004.

[98] M. J. Lindberg et al. "Folding of human superoxide dismutase: Disulfide reduction prevents dimerization and produces marginally stable monomers". In: *Proceedings of the National Academy of Sciences* 101.45 (2004), pp. 15893–15898. DOI: 10.1073/pnas.0403979101.

[99] James R. Lepock et al. "Structural analyses of various Cu2+, Zn2+-superoxide dismutases by differential scanning calorimetry and Raman spectroscopy". In: *Archives of Biochemistry and Biophysics* 241.1 (1985), pp. 243–251. DOI: 10.1016/0003-9861(85)90380-7.

[100] Henry J Forman and Irwin Fridovich. "On the stability of bovine superoxide dismutase. The effects of metals." In: *The Journal of biological chemistry* 248.8 (1973), pp. 2645–9.

[101] Richard W. Strange et al. "The Structure of Holo and Metal-deficient Wild-type Human Cu, Zn Superoxide Dismutase and its Relevance to Familial Amyotrophic Lateral Sclerosis". In: *Journal of Molecular Biology* 328.4 (2003), pp. 877–891. DOI: 10.1016/S0022-2836(03)00355-3.

[102] Yuewei Sheng et al. "SOD1 Aggregation and ALS: Role of Metallation States and Disulfide Status". In: *Current Topics in Medicinal Chemistry* 12.22 (2013), pp. 2560–2572. DOI: 10.2174/15680266112122220010.

[103] Helen R Broom, Jessica A.O. Rumfeldt, and Elizabeth M Meiering. "Many roads lead to Rome? Multiple modes of Cu,Zn superoxide dismutase destabilization, misfolding and aggregation in amyotrophic lateral sclerosis". In: *Essays In Biochemistry* 56 (2014), pp. 149–165. DOI: 10.1042/bse0560149.

[104] Yoshiaki Furukawa and Thomas V O'Halloran. "Amyotrophic Lateral Sclerosis Mutations Have the Greatest Destabilizing Effect on the Apo- and Reduced Form of SOD1, Leading to Unfolding and Oxidative Aggregation". In: *Journal of Biological Chemistry* 280.17 (2005), pp. 17266–17274. DOI: 10.1074/jbc.M500482200.

[105] Dorothee Kern, Elan Z. Eisenmesser, and Magnus Wolf-Watz. "Enzyme Dynamics During Catalysis Measured by NMR Spectroscopy". In: *Methods in Enzymology*. Vol. 394. 2005, pp. 507–524. DOI: 10.1016/S0076-6879(05)94021-4.

[106] David D Boehr et al. "The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis". In: *Science* 313.5793 (2006), pp. 1638–1642. DOI: 10.1126/science.1130258.

[107] Magnus Wolf-Watz et al. "Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair". In: *Nature Structural & Molecular Biology* 11.10 (2004), pp. 945–949. DOI: 10.1038/nsmb821.

[108] Rieko Ishima et al. "Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function". In: *Structure* 7.9 (1999), 1047–S12. DOI: 10.1016/S0969-2126(99)80172-5.

[109] Elan Zohar Eisenmesser. "Enzyme Dynamics During Catalysis". In: *Science* 295.5559 (2002), pp. 1520–1523. DOI: 10.1126/science.1066176.

[110] Roger Cole and J. Patrick Loria. "Evidence for Flexibility in the Function of Ribonuclease A †". In: *Biochemistry* 41.19 (2002), pp. 6072–6081. DOI: 10.1021/bi025655m.

[111] Eric D Watt et al. "The mechanism of rate-limiting motions in enzyme function". In: *Proceedings of the National Academy of Sciences* 104.29 (2007), pp. 11981–11986. DOI: 10.1073/pnas.0702551104.

[112] James G. Kempf et al. "Dynamic Requirements for a Functional Protein Hinge". In: *Journal of Molecular Biology* 368.1 (2007), pp. 131–149. DOI: 10.1016/j.jmb.2007.01.074.

[113] Pramodh Vallurupalli and Lewis E Kay. "Complementarity of ensemble and single-molecule measures of protein motion: A relaxation dispersion NMR study of an enzyme complex". In: *Proceedings of the National Academy of Sciences* 103.32 (2006), pp. 11910–11915. DOI: 10.1073/pnas.0602310103.

[114] M. B. Pepys et al. "Human lysozyme gene mutations cause hereditary systemic amyloidosis". In: *Nature* 362.6420 (1993), pp. 553–557. DOI: 10.1038/362553a0.

[115] Sophie Valleix et al. "Hereditary renal amyloidosis caused by a new variant lysozyme W64R in a French family". In: *Kidney International* 61.3 (2002), pp. 907–912. DOI: 10.1046/j.1523-1755.2002.00205.x.

[116] Masahide Yazaki, Sandra A Farrell, and Merrill D Benson. "A novel lysozyme mutation Phe57Ile associated with hereditary renal amyloidosis". In: *Kidney International* 63.5 (2003), pp. 1652–1657. DOI: 10.1046/j.1523-1755.2003.00904.x.

[117] Christoph Röcken et al. "ALys amyloidosis caused by compound heterozygosity in Exon 2 (Thr70Asn) and Exon 4 (Trp112Arg) of the lysozyme gene". In: *Human Mutation* 27.1 (2006), pp. 119–120. DOI: 10.1002/humu.9393.

[118] Michiro Muraki et al. "Origin of Carbohydrate Recognition Specificity of Human Lysozyme Revealed by Affinity Labeling † , ‡". In: *Biochemistry* 35.42 (1996), pp. 13562–13567. DOI: 10.1021/bi9613180.

[119] *The PyMOL Molecular Graphics System*.

[120] Mireille Dumoulin, Janet R Kumita, and Christopher M Dobson. "Normal and Aberrant Biological Self-Assembly: Insights from Studies of Human Lysozyme and Its Amyloidogenic Variants". In: *Accounts of Chemical Research* 39.9 (2006), pp. 603–610. DOI: 10.1021/ar050070g.

[121] Denis Canet et al. "Mechanistic Studies of the Folding of Human Lysozyme and the Origin of Amyloidogenic Behavior in Its Disease-Related Variants †". In: *Biochemistry* 38.20 (1999), pp. 6419–6427. DOI: 10.1021/bi983037t.

[122] Erica Frare et al. "Identification of the Core Structure of Lysozyme Amyloid Fibrils by Proteolysis". In: *Journal of Molecular Biology* 361.3 (2006), pp. 551–561. DOI: 10.1016/j.jmb.2006.06.055.

[123] Russell J.K. Johnson et al. "Rationalising Lysozyme Amyloidosis: Insights from the Structure and Solution Dynamics of T70N Lysozyme". In: *Journal of Molecular Biology* 352.4 (2005), pp. 823–836. DOI: 10.1016/j.jmb.2005.07.040.

[124] Kwang Hun Lim et al. "Localized Structural Fluctuations Promote Amyloidogenic Conformations in Transthyretin". In: *Journal of Molecular Biology* 425.6 (2013), pp. 977–988. DOI: 10.1016/j.jmb.2013.01.008.

[125] Merrill D Benson and Tomoyuki Uemichi. "Transthyretin amyloidosis". In: *Amyloid* 3.1 (1996), pp. 44–56. DOI: 10.3109/13506129609014354.

[126] Merrill D. Benson and John C. Kincaid. "The molecular biology and clinical features of amyloid neuropathy". In: *Muscle & Nerve* 36.4 (2007), pp. 411–423. DOI: 10.1002/mus.20821.

[127] Wilfredo Colon and Jeffery W Kelly. "Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro". In: *Biochemistry* 31.36 (1992), pp. 8654–8660. DOI: 10.1021/bi00151a036.

[128] Zhihong Lai, Wilfredo Colón, and Jeffery W Kelly. "The Acid-Mediated Denaturation Pathway of Transthyretin Yields a Conformational Intermediate That Can Self-Assemble into Amyloid †". In: *Biochemistry* 35.20 (1996), pp. 6470–6482. DOI: 10.1021/bi952501g.

[129] Alexandre Quintas et al. "Tetramer Dissociation and Monomer Partial Unfolding Precedes Protofibril Formation in Amyloidogenic Transthyretin Variants". In: *Journal of Biological Chemistry* 276.29 (2001), pp. 27207–27213. DOI: 10.1074/jbc.M101024200.

[130] P. Hammarstrom. "Trans-Suppression of Misfolding in an Amyloid Disease". In: *Science* 293.5539 (2001), pp. 2459–2462. DOI: 10.1126/science.1062245.

[131] P. Hammarstrom. "Prevention of Transthyretin Amyloid Disease by Changing Protein Misfolding Energetics". In: *Science* 299.5607 (2003), pp. 713–716. DOI: 10.1126/science.1079589.

[132] Andreas Hörnberg et al. "A comparative analysis of 23 structures of the amyloidogenic protein transthyretin". In: *Journal of Molecular Biology* 302.3 (2000), pp. 649–669. DOI: 10.1006/jmbi.2000.4078.

[133] Fumitake Gejyo et al. "Beta 2-microglobulin: a new form of amyloid protein associated with chronic hemodialysis". In: *Kidney Int* 30.3 (1986), pp. 385–390. DOI: 10.1038/ki.1986.196.

[134] Timo Eichner and Sheena E. Radford. "A Diversity of Assembly Mechanisms of a Generic Amyloid Fold". eng. In: *Molecular Cell* 43.1 (2011), pp. 8–18. DOI: 10.1016/j.molcel.2011.05.012.

[135] Fabrizio Chiti and Christopher M. Dobson. "Protein Misfolding, Functional Amyloid, and Human Disease". In: *Annual Review of Biochemistry* 75.1 (2006), pp. 333–366. DOI: 10.1146/annurev.biochem.75.101304.123901.

[136] Jean D. Sipe et al. "Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines". In: *Amyloid* 23.4 (2016), pp. 209–213. DOI: 10.1080/13506129.2016.1257986.

[137] M Sunde and C Blake. "The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction". In: *Advances in Protein Chemistry Volume 50*. Vol. 50. 1997, pp. 123–159. DOI: 10.1016/S0065-3233(08)60320-4.

[138] Sumner Makin. "The structure of Amyloid". In: *Fibre Diffraction Review* 12.12 (2004), p. 29. DOI: 10.1382/s20041229.

[139] Per Westermark et al. "A primer of amyloid nomenclature". In: *Amyloid* 14.3 (2007), pp. 179–183. DOI: 10.1080/13506120701460923.

[140] Ritu Khurana et al. "Is Congo Red an Amyloid-specific Dye?" In: *Journal of Biological Chemistry* 276.25 (2001), pp. 22715–22721. DOI: 10.1074/jbc.M011499200.

[141] P S Vassar and C F Culling. "Fluorescent stains, with special reference to amyloid and connective tissues." In: *Archives of pathology* 68 (1959), pp. 487–498. DOI: 10.1002/ijc.28334.

[142] Tadato Ban et al. "Direct observation of amyloid fibril growth monitored by thioflavin T fluorescence". In: *Journal of Biological Chemistry* 278.19 (2003), pp. 16462–16465. DOI: 10.1074/jbc.C300049200.

[143] Matthew Biancalana and Shohei Koide. "Molecular mechanism of Thioflavin-T binding to amyloid fibrils". eng. In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804.7 (2010), pp. 1405–1412. DOI: 10.1016/j.bbapap.2010.04.001.

[144] Harry LeVine. "Thioflavine T interaction with amyloid β-sheet structures". In: *Amyloid* 2.1 (1995), pp. 1–6. DOI: 10.3109/13506129509031881.

[145] Thomas CT Michaels et al. "Chemical Kinetics for Bridging Molecular Mechanisms and Macroscopic Measurements of Amyloid Fibril Formation". In: *Annual Review of Physical Chemistry* (2018). DOI: 10.1146/annurev-physchem.

[146] Peter Cresswell et al. "Mechanisms of MHC class I-restricted antigen processing and cross-presentation". In: *Immunological Reviews* 207.1 (2005), pp. 145–157. DOI: 10.1111/j.0105-2896.2005.00316.x.

[147] R Germain. "The ins and outs of antigen processing and presentation". In: *Nature* 322.6081 (1986), pp. 687–689. DOI: 10.1038/322687a0.

[148] Elaine C. Campbell, Antony N. Antoniou, and Simon J. Powis. "The multi-faceted nature of HLA class I dimer molecules". In: *Immunology* 136.4 (2012), pp. 380–384. DOI: 10.1111/j.1365-2567.2012.03593.x.

[149] Timo Eichner and Sheena E. Radford. "Understanding the complex mechanisms of β2-microglobulin amyloid assembly". eng. In: *FEBS Journal* 278.20 (2011), pp. 3868–3883. DOI: 10.1111/j.1742-4658.2011.08186.x.

[150] Oleg Y. Borbulevych, Priscilla Do, and Brian M. Baker. "Structures of native and affinity-enhanced WT1 epitopes bound to HLA-A0201: Implications for WT1-based cancer therapeutics". In: *Molecular Immunology* 47.15 (2010), pp. 2519–2524. DOI: 10.1016/j.molimm.2010.06.005.

[151] C. H. Trinh et al. "Crystal structure of monomeric human -2-microglobulin reveals clues to its amyloidogenic properties". In: *Proceedings of the National Academy of Sciences* 99.15 (2002), pp. 9771–9776. DOI: 10.1073/pnas.152337399.

[152] David P. Smith and Sheena E. Radford. "Role of the single disulphide bond of β 2 -microglobulin in amyloidosis in vitro". In: *Protein Science* 10.9 (2001), pp. 1775–1784. DOI: 10.1110/ps.4901.

[153] D E Isenman, R H Painter, and K J Dorrington. "The structure and function of immunoglobulin domains: studies with beta-2-microglobulin on the role of the intrachain disulfide bond." In: *Proceedings of the National Academy of Sciences* 72.2 (1975), pp. 548–552. DOI: 10.1073/pnas.72.2.548.

[154] J Floege et al. "Clearance and synthesis rates of beta 2-microglobulin in patients undergoing hemodialysis and in normal subjects". eng. In: *Journal of Laboratory and Clinical Medicine* 118.2 (1991), pp. 153–165.

[155] Reinhold P. Linke et al. "Production of recombinant human β2-microglobulin for scintigraphic diagnosis of amyloidosis in uremia and hemodialysis". In: *European Journal of Biochemistry* 267.3 (2000), pp. 627–633. DOI: 10.1046/j.1432-1327.2000.01071.x.

[156] Jürgen Floege and Götz Ehlerding. "Beta-2-Microglobulin-Associated Amyloidosis". In: *Nephron* 72.1 (1996), pp. 9–26. DOI: 10.1159/000188801.

[157] Ph. Morinière et al. "Destructive Spondyloarthropathy with β2-Microglobulin Amyloid Deposits in a Uremic Patient before Chronic Hemodialysis". In: *Nephron* 59.4 (1991), pp. 654–657. DOI: 10.1159/000186661.

[158] M E Ullian et al. "Beta-2-microglobulin-associated amyloidosis in chronic hemodialysis patients with carpal tunnel syndrome." In: *Medicine* 68.2 (1989), pp. 107–15.

[159] J Kay. "Beta2-microglobulin amyloidosis in renal failure: Understanding this recently recognized condition". In: *Cleveland Clinic Journal of Medicine* 66.3 (1999), pp. 145–147.

[160] S. Otsubo et al. "Characteristics of dialysis-related amyloidosis in patients on haemodialysis therapy for more than 30 years". In: *Nephrology Dialysis Transplantation* 24.5 (2009), pp. 1593–1598. DOI: 10.1093/ndt/gfn706.

[161] Sarah L. Myers et al. "A Systematic Study of the Effect of Physiological Factors on β 2 -Microglobulin Amyloid Formation at Neutral pH †". In: *Biochemistry* 45.7 (2006), pp. 2311–2321. DOI: 10.1021/bi052434i.

[162] Suguru Yamamoto et al. "Low Concentrations of Sodium Dodecyl Sulfate Induce the Extension of β 2 -Microglobulin-Related Amyloid Fibrils at a Neutral pH †". In: *Biochemistry* 43.34 (2004), pp. 11075–11082. DOI: 10.1021/bi049262u.

[163] N. A. Athanasou, B. Puddle, and B. Sallie. "Highly sulphated glycosaminoglycans in articular cartilage and other tissues containing β2microglobulin dialysis amyloid deposits". In: *Nephrology Dialysis Transplantation* 10.9 (1995), pp. 1672–1678. DOI: 10.1093/oxfordjournals.ndt.a091405.

[164] Tadakazu Ookoshi et al. "Lysophospholipids induce the nucleation and extension of β2-microglobulin-related amyloid fibrils at a neutral pH". In: *Nephrology Dialysis Transplantation* 23.10 (2008), pp. 3247–3255. DOI: 10.1093/ndt/gfn231.

[165] Kazuhiro Hasegawa et al. "Growth of β 2 -microglobulin-related amyloid fibrils by non-esterified fatty acids at a neutral pH". In: *Biochemical Journal* 416.2 (2008), pp. 307–315. DOI: 10.1042/BJ20080543.

[166] Charles J Morgan et al. "Kidney dialysis-associated amyloidosis: a molecular role for copper in fiber formation". In: *Journal of Molecular Biology* 309.2 (2001), pp. 339–345. DOI: 10.1006/jmbi.2001.4661.

[167] James Villanueva et al. "Increase in the conformational flexibility of beta 2-microglobulin upon copper binding: a possible role for copper in dialysis-related amyloidosis." In: *Protein science : a publication of the Protein Society* 13.3 (2004), pp. 797–809. DOI: 10.1110/ps.03445704.

[168] A Relini et al. "Collagen plays an active role in the aggregation of beta(2)-microglobulin under physiopathological conditions of dialysis-related amyloidosis". In: *Journal of Biological Chemistry* 281.24 (2006), pp. 16521–16529. DOI: 10.1074/jbc.M51387200.

[169] Annalisa Relini et al. "Heparin strongly enhances the formation of ??2-microglobulin amyloid fibrils in the presence of type I collagen". In: *Journal of Biological Chemistry* 283.8 (2008), pp. 4912–4920. DOI: 10.1074/jbc.M702712200.

[170] Vittorio Bellotti et al. "Beta 2-Microglobulin Can Be Refolded Into a Native State From Ex Vivo Amyloid Fibrils". eng. In: *European Journal of Biochemistry* 258.1 (1998), pp. 61–67. DOI: 10.1046/j.1432-1327.1998.2580061.x.

[171] G. Esposito et al. "Removal of the N-terminal hexapeptide from human β2-microglobulin facilitates protein aggregation and fibril formation". eng. In: *Protein Science* 9.5 (2000), pp. 831–845. DOI: 10.1110/ps.9.5.831.

[172] Fabrizio Chiti et al. "Detection of two partially structured species in the folding process of the amyloidogenic protein β2-microglobulin". In: *Journal of Molecular Biology* 307.1 (2001), pp. 379–391. DOI: 10.1006/jmbi.2000.4478.

[173] Michiko Sakata et al. "Kinetic Coupling of Folding and Prolyl Isomerization of β2-Microglobulin Studied by Mutational Analysis". eng. In: *Journal of Molecular Biology* 382.5 (2008), pp. 1242–1255. DOI: 10.1016/j.jmb.2008.08.003.

[174] Atsushi Kameda et al. "Nuclear Magnetic Resonance Characterization of the Refolding Intermediate of β2-Microglobulin Trapped by Non-native Prolyl Peptide Bond". eng. In: *Journal of Molecular Biology* 348.2 (2005), pp. 383–397. DOI: 10.1016/j.jmb.2005.02.050.

[175] Fabrizio Chiti et al. "A Partially Structured Species of β2-Microglobulin Is Significantly Populated under Physiological Conditions and Involved in Fibrillogenesis". In: *Journal of Biological Chemistry* 276.50 (2001), pp. 46714–46721. DOI: 10.1074/jbc.M107040200.

[176] Thomas R Jahn et al. "Amyloid formation under physiological conditions proceeds via a native-like folding intermediate". eng. In: *Nature Structural & Molecular Biology* 13.3 (2006), pp. 195–201. DOI: 10.1038/nsmb1058.

[177] Timo Eichner and Sheena E. Radford. "A Generic Mechanism of β2-Microglobulin Amyloid Assembly at Neutral pH Involving a Specific Proline Switch". In: *Journal of Molecular Biology* 386.5 (2009), pp. 1312–1326. DOI: 10.1016/j.jmb.2009.01.013.

[178] Catherine M. Eakin, Andrea J. Berman, and Andrew D. Miranker. "A native to amyloidogenic transition regulated by a backbone trigger". In: *Nature Structural and Molecular Biology* 13.3 (2006), pp. 202–208. DOI: 10.1038/nsmb1068.

[179] Matthew F. Calabrese et al. "A regulatable switch mediates self-association in an immunoglobulin fold". In: *Nature Structural and Molecular Biology* 15.9 (2008), pp. 965–971. DOI: 10.1038/nsmb.1483.

[180] Nicholas B. Borotto et al. "Increased β-Sheet Dynamics and D–E Loop Repositioning Are Necessary for Cu(II)-Induced Amyloid Formation by β-2-Microglobulin". In: *Biochemistry* 56.8 (2017), pp. 1095–1104. DOI: 10.1021/acs.biochem.6b01198.

[181] Y Miura et al. "Radiolucent bone cysts and the type of dialysis membrane used in patients undergoing long-term hemodialysis." In: *Nephron* 60.3 (1992), pp. 268–73. DOI: 10.1159/000186764.

[182] C van Ypersele de Strihou et al. "Effect of dialysis membrane and patient's age on signs of dialysis-related amyloidosis. The Working Party on Dialysis Amyloidosis." In: *Kidney international* 39.5 (1991), pp. 1012–9.

[183] Sophie Valleix et al. "Hereditary Systemic Amyloidosis Due to Asp76Asn Variant β 2 -Microglobulin". eng. In: *New England Journal of Medicine* 366.24 (2012), pp. 2276–2283. DOI: 10.1056/NEJMoa1201356.

[184] P. P. Mangione et al. "Structure, Folding Dynamics, and Amyloidogenesis of D76N 2-Microglobulin: ROLES OF SHEAR FLOW, HYDROPHOBIC SURFACES, AND -CRYSTALLIN". In: *Journal of Biological Chemistry* 288.43 (2013), pp. 30917–30930. DOI: 10.1074/jbc.M113.498857.

[185] Chandrasekaran P and Rajasekaran R. "A systematic molecular dynamics approach to the structural characterization of amyloid aggregation propensity of β2-microglobulin mutant D76N". In: *Mol. BioSyst.* (2016). DOI: 10.1039/C5MB00759C.

[186] Song-Ho Chong et al. "Structural and Thermodynamic Characteristics of Amyloidogenic Intermediates of β-2-Microglobulin". In: *Scientific Reports* 5.1 (2015), p. 13631. DOI: 10.1038/srep13631.

[187] Rui J. S. Loureiro et al. "A tale of two tails: The importance of unstructured termini in the aggregation pathway of β2-microglobulin". In: *Proteins: Structure, Function, and Bioinformatics* 85.11 (2017), pp. 2045–2057. DOI: 10.1002/prot.25358.

[188] Levon Halabelian et al. "A covalent homodimer probing early oligomers along amyloid aggregation". In: *Scientific Reports* 5 (2015), p. 14651. DOI: 10.1038/srep14651.

[189] Neil M Kad et al. "β2-microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies in vitro". eng. In: *Journal of Molecular Biology* 313.3 (2001), pp. 559–571. DOI: 10.1006/jmbi.2001.5071.

[190] Elisabeth Gasteiger et al. "Protein Identification and Analysis Tools on the ExPASy Server". In: *The Proteomics Protocols Handbook*. Totowa, NJ, 2005, pp. 571–607. DOI: 10.1385/1-59259-890-0:571.

[191] Sharon M. Kelly, Thomas J. Jess, and Nicholas C. Price. "How to study proteins by circular dichroism". eng. In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1751.2 (2005), pp. 119–139. DOI: 10.1016/j.bbapap.2005.06.005.

[192] Norma J Greenfield. "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions". In: *Nature Protocols* 1.6 (2007), pp. 2527–2535. DOI: 10.1038/nprot.2006.204.

[193] Markus Niklasson et al. "Robust and convenient analysis of protein thermal and chemical stability". In: *Protein Science* 24.12 (2015), pp. 2055–2062. DOI: 10.1002/pro.2809.

[194] Samuel I.A. Cohen et al. "From macroscopic measurements to microscopic mechanisms of protein aggregation". In: *Journal of Molecular Biology* 421.2-3 (2012), pp. 160–171. DOI: 10.1016/j.jmb.2012.02.031.

[195] Paolo Arosio, Tuomas P.J. Knowles, and Sara Linse. "On the lag phase in amyloid fibril formation". In: *Physical Chemistry Chemical Physics* 17.12 (2015), pp. 7606–7618. DOI: 10.1039/c4cp05563b.

[196] Frank Delaglio et al. "NMRPipe: A multidimensional spectral processing system based on UNIX pipes". eng. In: *Journal of biomolecular NMR* 6.3 (1995), pp. 277–293. DOI: 10.1007/BF00197809.

[197] Wim F. Vranken et al. "The CCPN data model for NMR spectroscopy: Development of a software pipeline". eng. In: *Proteins: Structure, Function, and Bioinformatics* 59.4 (2005), pp. 687–696. DOI: 10.1002/prot.20449.

[198] Alexandra Ahlner et al. "PINT: a software for integration of peak volumes and extraction of relaxation rates". In: *Journal of Biomolecular NMR* 56.3 (2013), pp. 191–202. DOI: 10.1007/s10858-013-9737-7.

[199] Jonathan J Helmus and Christopher P Jaroniec. "Nmrglue: an open source Python package for the analysis of multidimensional NMR data." eng. In: *Journal of biomolecular NMR* 55.4 (2013), pp. 355–367. DOI: 10.1007/s10858-013-9718-x.

[200] Ewen Lescop, Paul Schanda, and Bernhard Brutscher. "A set of BEST triple-resonance experiments for time-optimized protein resonance assignment". In: *Journal of Magnetic Resonance* 187.1 (2007), pp. 163–169. DOI: 10.1016/j.jmr.2007.04.002.

[201] S Grzesiek and A Bax. "Amino-acid type determination in the sequential assignment procedures of uniformly C-13/N-15-enriched proteins". In: *Journal of Biomolecular NMR* 3.2 (1993), pp. 185–204.

[202] Yu-Zhu Zhang. *Protein and peptide structure and interactions studied by hydrogen exchange and NMR*. 1995.

[203] Marius Clore and Jennifer Potts, eds. *Recent Developments in Biomolecular NMR*. RSC Biomolecular Sciences. Cambridge, 2012, FP001–347. DOI: 10.1039/9781849735391.

[204] Keyang Ding and Angela M. Gronenborn. "Sensitivity-enhanced 2D IPAP, TROSY-anti-TROSY, and E.COSY experiments: Alternatives for measuring dipolar 15N-1HN couplings". In: *Journal of Magnetic Resonance* 163.2 (2003), pp. 208–214. DOI: 10.1016/S1090-7807(03)00081-8.

[205] Janet C Saunders et al. "An in vivo platform for identifying inhibitors of protein aggregation". In: *Nature Chemical Biology* 12.2 (2016), pp. 94–101. DOI: 10.1038/nchembio.1988.

[206] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (2015), pp. 19–25. DOI: 10.1016/j.softx.2015.06.001.

[207] Gareth A. Tribello et al. "PLUMED 2: New feathers for an old bird". In: *Computer Physics Communications* 185.2 (2014), pp. 604–613. DOI: 10.1016/j.cpc.2013.09.018.

[208] Massimiliano Bonomi et al. "PLUMED: A portable plugin for free-energy calculations with molecular dynamics". In: *Computer Physics Communications* 180.10 (2009), pp. 1961–1972. DOI: 10.1016/j.cpc.2009.05.011.

[209] Abil E. Aliev et al. "Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics". In: *Proteins: Structure, Function and Bioinformatics* 82.2 (2014), pp. 195–215. DOI: 10.1002/prot.24350.

[210] William L. Jorgensen. "Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water". In: *Journal of the American Chemical Society* 103.2 (1981), pp. 335–340. DOI: 10.1021/ja00392a016.

[211] Berk Hess et al. "LINCS: A Linear Constraint Solver for molecular simulations". In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[212] Kai J Kohlhoff et al. "Suppplementary Information: Fast and accurate predictions of protein NMR chemical shifts from interatomic distances." In: *Journal of the American Chemical Society* 131.39 (2009), pp. 13894–5. DOI: 10.1021/ja903772t.

[213] Jim Pfaendtner and Massimiliano Bonomi. "Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics". In: *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5062–5067. DOI: 10.1021/acs.jctc.5b00846.

[214] Paolo Raiteri et al. "Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics". In: *Journal of Physical Chemistry B* 110.8 (2006), pp. 3533–3539. DOI: 10.1021/jp054359r.

[215] Naveen Michaud-Agrawal et al. "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations". In: *Journal of Computational Chemistry* 32.10 (2011), pp. 2319–2327. DOI: 10.1002/jcc.21787.

[216] Richard J Gowers et al. "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations". In: *Proceedings of the 15th Python in Science Conference*. Ed. by Sebastian Benthall and Scott Rostrup. 2016, pp. 98–105.

[217] David P Smith et al. "A Systematic Investigation into the Effect of Protein Destabilisation on Beta 2-Microglobulin Amyloid Formation". In: *Journal of Molecular Biology* 330.5 (2003), pp. 943–954. DOI: 10.1016/S0022-2836(03)00687-9.

[218] Linda Foit et al. "Optimizing Protein Stability In Vivo". In: *Molecular Cell* 36.5 (2009), pp. 861–871. DOI: 10.1016/j.molcel.2009.11.022.

[219] Wouter G. Touw et al. "A series of PDB-related databanks for everyday needs". In: *Nucleic Acids Research* 43.D1 (2015), pp. D364–D368. DOI: 10.1093/nar/gku1028.

[220] Wolfgang Kabsch and Christian Sander. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.12 (1983), pp. 2577–2637. DOI: 10.1002/bip.360221211.

[221] Levon Halabelian et al. "Class I Major Histocompatibility Complex, the Trojan Horse for Secretion of Amyloidogenic β 2 -Microglobulin". In: *Journal of Biological Chemistry* 289.6 (2014), pp. 3318–3327. DOI: 10.1074/jbc.M113.524157.

[222] Matteo de Rosa et al. "Decoding the Structural Bases of D76N ß2-Microglobulin High Amyloidogenicity through Crystallography and Asn-Scan Mutagenesis". In: *Plos One* 10.12 (2015), e0144061. DOI: 10.1371/journal.pone.0144061.

[223] Tanguy Le Marchand et al. "Conformational dynamics in crystals reveal the molecular bases for D76N beta-2 microglobulin aggregation propensity". In: *Nature Communications* (2018), pp. 1–11. DOI: 10.1038/s41467-018-04078-y.

[224] Victoria J. McParland et al. "Partially Unfolded States of β 2 -Microglobulin and Amyloid Formation in Vitro †". In: *Biochemistry* 39.30 (2000), pp. 8735–8746. DOI: 10.1021/bi000276j.

[225] Wei-feng Xue, Steve W Homans, and Sheena E Radford. "Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.26 (2008), pp. 8926–31. DOI: 10.1073/pnas.0711664105.

[226] Suguru Yamamoto et al. "Glycosaminoglycans Enhance the Trifluoroethanol-Induced Extension of β2-Microglobulin-Related Amyloid Fibrils at a Neutral pH". In: *Journal of the American Society of Nephrology* 15.1 (2004), pp. 126–133. DOI: 10.1097/01.ASN.0000103228.81623.C7.

[227] Carlo Camilloni et al. "Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability". In: *Scientific Reports* 6.1 (2016), p. 25559. DOI: 10.1038/srep25559.

[228] Christian Jelsch et al. "Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 A resolution." In: *Proteins* 16.4 (1993), pp. 364–83. DOI: 10.1002/prot.340160406.

[229] Oscar D. Monera et al. "Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix". In: *Journal of Peptide Science* 1.5 (1995), pp. 319–329. DOI: 10.1002/psc.310010507.

[230] N J Darby and T E Creighton. *Protein structure*. In focus. 1993.

[231] Kentaro Iwata et al. "High-resolution crystal structure of β2-microglobulin formed at pH 7.0". In: *Journal of Biochemistry* 142.3 (2007), pp. 413–419. DOI: 10.1093/jb/mvm148.

[232] Monica Stoppini et al. "Proteomics of β2-microglobulin amyloid fibrils". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1753.1 (2005), pp. 23–33. DOI: 10.1016/j.bbapap.2005.07.019.

[233] E. Rennella et al. "Oligomeric states along the folding pathways of β2-microglobulin: Kinetics, thermodynamics, and structure". In: *Journal of Molecular Biology* 425.15 (2013), pp. 2722–2736. DOI: 10.1016/j.jmb.2013.04.028.

[234] Alessandra Corazza et al. "Native-unlike Long-lived Intermediates along the Folding Pathway of the Amyloidogenic Protein β 2 -Microglobulin Revealed by Real-time Two-dimensional NMR". In: *Journal of Biological Chemistry* 285.8 (2010), pp. 5827–5835. DOI: 10.1074/jbc.M109.061168.

[235] Enrico Rennella et al. "Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate". In: *Journal of the American Chemical Society* 134.19 (2012), pp. 8066–8069. DOI: 10.1021/ja302598j.

[236] Rémi Franco et al. "Probing Conformational Exchange Dynamics in a Short-Lived Protein Folding Intermediate by Real-Time Relaxation–Dispersion NMR". In: *Journal of the American Chemical Society* 139.3 (2017), pp. 1065–1068. DOI: 10.1021/jacs.6b12089.

[237] Rémi Franco et al. "Optimized fast mixing device for real-time NMR applications". In: *Journal of Magnetic Resonance* 281 (2017), pp. 125–129. DOI: 10.1016/j.jmr.2017.05.016.

[238] S Walter Englander and Leland Mayne. "Using Hydrogen-Exchange Labeling and Two Dimensional Nmr". In: *Annu. Rev. Biophys. Biomol. Struct.* 21 (1992), pp. 243–265.

[239] Paul Schanda, Ēriks Kupče, and Bernhard Brutscher. "SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds". eng. In: *Journal of Biomolecular NMR* 33.4 (2005), pp. 199–211. DOI: 10.1007/s10858-005-4425-x.

[240] Enrico Rennella et al. "Equilibrium Unfolding Thermodynamics of β 2 -Microglobulin Analyzed through Native-State H/D Exchange". In: *Biophysical Journal* 96.1 (2009), pp. 169–179. DOI: 10.1529/biophysj.108.142448.

[241] Jim Pfaendtner and Massimiliano Bonomi. "Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics". In: *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5062–5067. DOI: 10.1021/acs.jctc.5b00846.

[242] Jane S Richardson and David C Richardson. "Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.5 (2002), pp. 2754–9. DOI: 10.1073/pnas.052706099.

[243] Carlo Santambrogio et al. "DE-loop mutations affect β2 microglobulin stability, oligomerization, and the low-pH unfolded form". In: *Protein Science* 19.7 (2010), pp. 1386–1394. DOI: 10.1002/pro.419.

[244] Stefano Ricagno et al. "DE loop mutations affect β2-microglobulin stability and amyloid aggregation". In: *Biochemical and Biophysical Research Communications* 377.1 (2008), pp. 146–150. DOI: 10.1016/j.bbrc.2008.09.108.

[245] Stefano Ricagno et al. "DE loop mutations affect β2-microglobulin stability and amyloid aggregation". In: *Biochemical and Biophysical Research Communications* 377.1 (2008), pp. 146–150. DOI: 10.1016/j.bbrc.2008.09.108.

[246] Janet C Saunders et al. "An in vivo platform for identifying inhibitors of protein aggregation". In: *Nature Chemical Biology* 12.2 (2016), pp. 94–101. DOI: 10.1038/nchembio.1988.

[247] Massimiliano Bonomi, Riccardo Pellarin, and Michele Vendruscolo. "Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy". In: *Biophysical Journal* 114.7 (2018), pp. 1604–1613. DOI: 10.1016/j.bpj.2018.02.028.

[248] Emeline Barbet-Massin et al. "Fibrillar vs Crystalline Full-Length β-2-Microglobulin Studied by High-Resolution Solid-State NMR Spectroscopy". In: *Journal of the American Chemical Society* 132.16 (2010), pp. 5556–5557. DOI: 10.1021/ja1002839.

[249] Galia T. Debelouchina et al. "Magic Angle Spinning NMR Analysis of β 2 -Microglobulin Amyloid Fibrils in Two Distinct Morphologies". In: *Journal of the American Chemical Society* 132.30 (2010), pp. 10414–10423. DOI: 10.1021/ja102775u.

[250] Yongchao Su et al. "Secondary Structure in the Core of Amyloid Fibrils Formed from Human β 2 m and its Truncated Variant ΔN6". In: *Journal of the American Chemical Society* 136.17 (2014), pp. 6313–6325. DOI: 10.1021/ja4126092.

[251] Robert Tycko. "Solid-State NMR Studies of Amyloid Fibril Structure". In: *Annual Review of Physical Chemistry* 62.1 (2011), pp. 279–299. DOI: 10.1146/annurev-physchem-032210-103539.

[252] Beat H Meier, Roland Riek, and Anja Böckmann. "Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR". In: *Trends in Biochemical Sciences* 42.10 (2017), pp. 777–787. DOI: 10.1016/j.tibs.2017.08.001.

[253] Antoine Loquet et al. "3D structure determination of amyloid fibrils using solid-state NMR spectroscopy". In: *Methods* 138-139 (2018), pp. 26–38. DOI: 10.1016/j.ymeth.2018.03.014.

[254] Akira Naito and Izuru Kawamura. "Solid-state NMR as a method to reveal structure and membrane-interaction of amyloidogenic proteins and peptides". In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1768.8 (2007), pp. 1900–1912. DOI: 10.1016/j.bbamem.2007.03.025.

[255] Ming Tang, Gemma Comellas, and Chad M. Rienstra. "Advanced Solid-State NMR Approaches for Structure Determination of Membrane Proteins and Amyloid Fibrils". In: *Accounts of Chemical Research* 46.9 (2013), pp. 2080–2088. DOI: 10.1021/ar4000168.

[256] Robert Tycko. "Solid-State NMR Studies of Amyloid Fibril Structure". In: *Annual Review of Physical Chemistry* 62.1 (2011), pp. 279–299. DOI: 10.1146/annurev-physchem-032210-103539.

[257] Lothar Gremer et al. "Fibril structure of amyloid-β(1–42) by cryo–electron microscopy". In: *Science* 358.6359 (2017), pp. 116–119. DOI: 10.1126/science.aao2825.

[258] Anthony W P Fitzpatrick et al. "Cryo-EM structures of tau filaments from Alzheimer's disease". In: *Nature* 547.7662 (2017), pp. 185–190. DOI: 10.1038/nature23002.

[259] Marielle Aulikki Wälti et al. "Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril". In: *Proceedings of the National Academy of Sciences* 113.34 (2016), E4976–E4984. DOI: 10.1073/pnas.1600749113.

[260] Michael T. Colvin et al. "Atomic Resolution Structure of Monomorphic Aβ 42 Amyloid Fibrils". In: *Journal of the American Chemical Society* 138.30 (2016), pp. 9663–9674. DOI: 10.1021/jacs.6b05129.

[261] Benjamin Falcon et al. "Structures of filaments from Pick's disease reveal a novel tau protein fold". In: *Nature* 561.7721 (2018), pp. 137–140. DOI: 10.1038/s41586-018-0454-y.

[262] Ricardo Guerrero-Ferreira et al. "Cryo-EM structure of alpha-synuclein fibrils". In: *eLife* 7 (2018). DOI: 10.7554/eLife.36402.

[263] L.A. Woods, S.E. Radford, and A.E. Ashcroft. "Advances in ion mobility spectrometry–mass spectrometry reveal key insights into amyloid assembly". In: *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1834.6 (2013), pp. 1257–1268. DOI: 10.1016/j.bbapap.2012.10.002.

[264] Marius Ionuţ Iuraşcu et al. "Structural characterization of ß-amyloid oligomer-aggregates by ion mobility mass spectrometry and electron spin resonance spectroscopy". In: *Analytical and Bioanalytical Chemistry* 395.8 (2009), pp. 2509–2519. DOI: 10.1007/s00216-009-3164-3.

[265] Alison E. Ashcroft. "Mass spectrometry and the amyloid problem—How far can we go in the gas phase?" In: *Journal of the American Society for Mass Spectrometry* 21.7 (2010), pp. 1087–1096. DOI: 10.1016/j.jasms.2010.02.026.

[266] David P Smith, Sheena E Radford, and Alison E Ashcroft. "Elongated oligomers in 2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry". In: *Proceedings of the National Academy of Sciences* 107.15 (2010), pp. 6794–6798. DOI: 10.1073/pnas.0913046107.

[267] Charlotte A. Scarff, Alison E. Ashcroft, and Sheena E. Radford. "Characterization of Amyloid Oligomers by Electrospray Ionization-Ion Mobility Spectrometry-Mass Spectrometry (ESI-IMS-MS)". In: *Methods in molecular biology (Clifton, N.J.)* Vol. 1345. 2016, pp. 115–132. DOI: 10.1007/978-1-4939-2978-8_8.

Appendix A

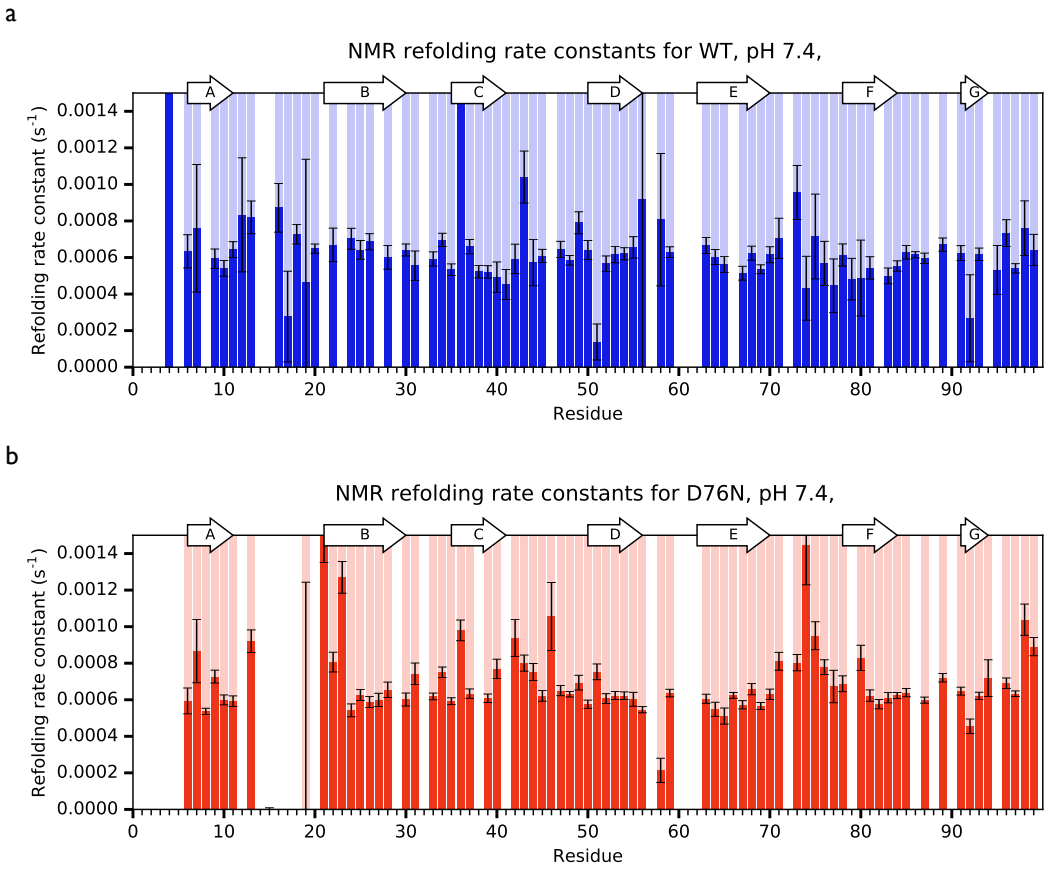


Fig. A.1: Single residue refolding rates for N-state peaks, without filtering fits based on errors.
The rate constants for all residues are shown in (a) and (b) for WT and D76N, respectively.

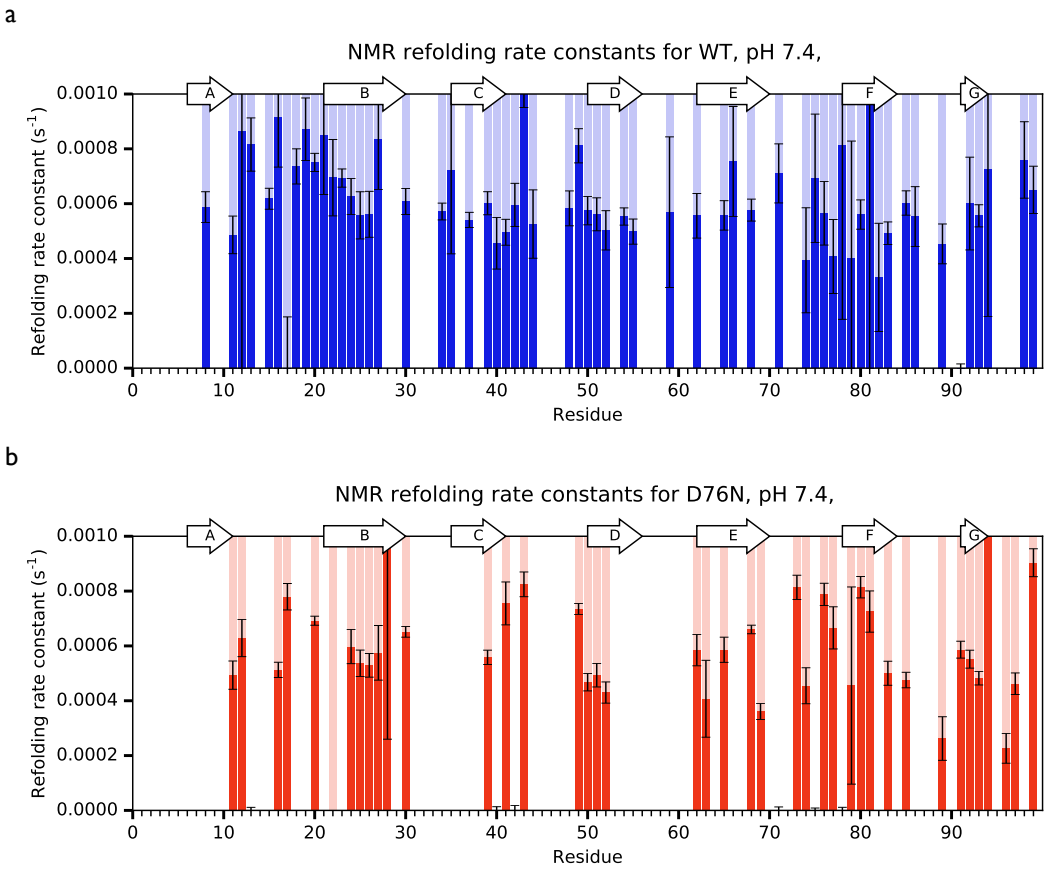


Fig. A.2: Single residue refolding rates for I_T-state peaks, without filtering fits based on errors.
The rate constants for all residues are shown in (a) and (b) for WT and D76N, respectively.

Appendix B

B.1 CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

The residues which were deemed to be undergoing exchange are listed in Table B.1. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test.

Table B.1: Exchanging residues in D76N at pH 7.4, measured at 600 MHz. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test. Data were acquired with 500 μM samples at pH 7.4, recorded with a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

Residue	Residue	Residue
7	34	56
9	35	59
10	36	63
11	49	64
12	50	65
13	51	67
26	53	87
27	54	97
30	55	98
31		

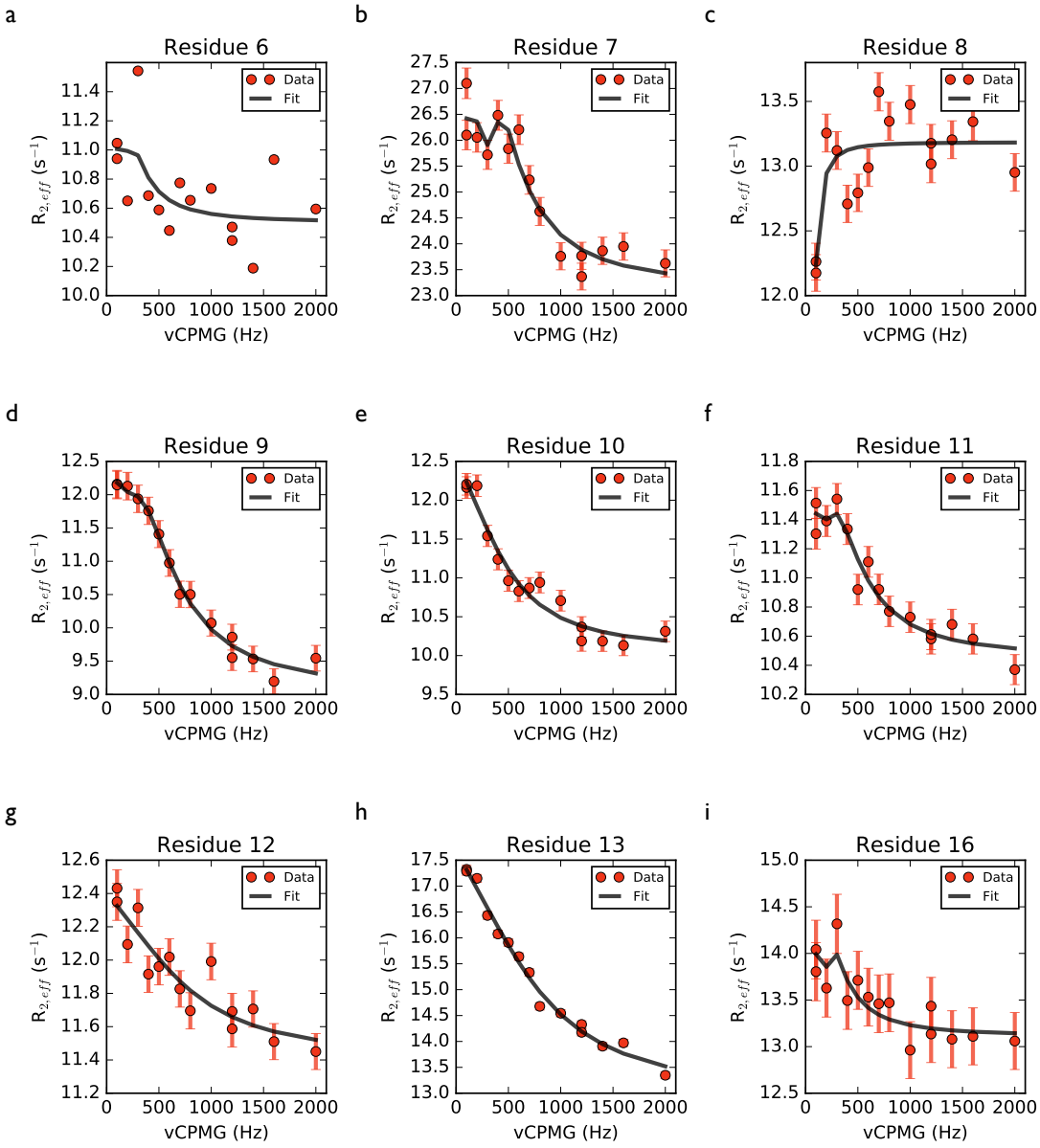


Fig. B.1: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

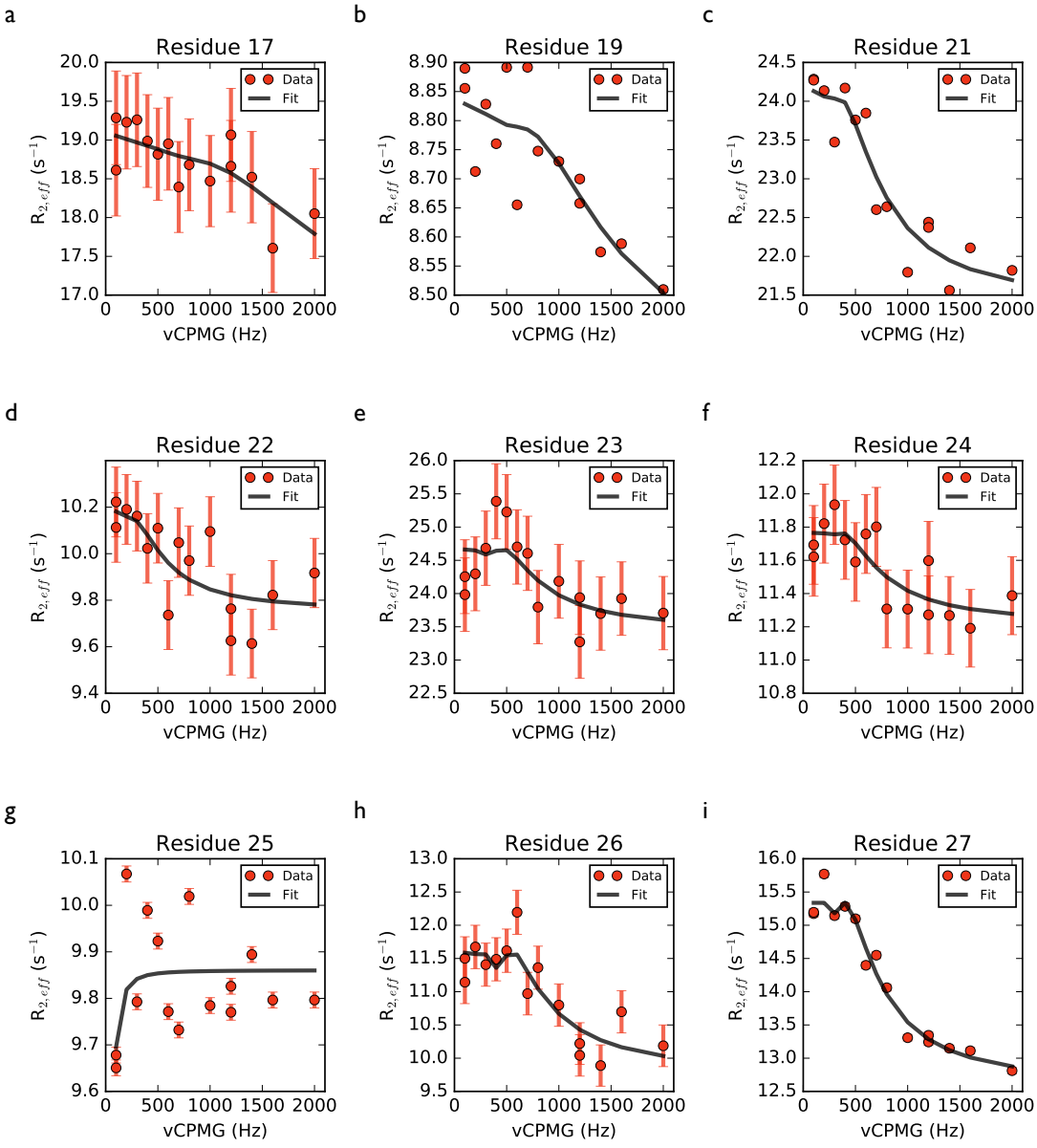


Fig. B.2: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

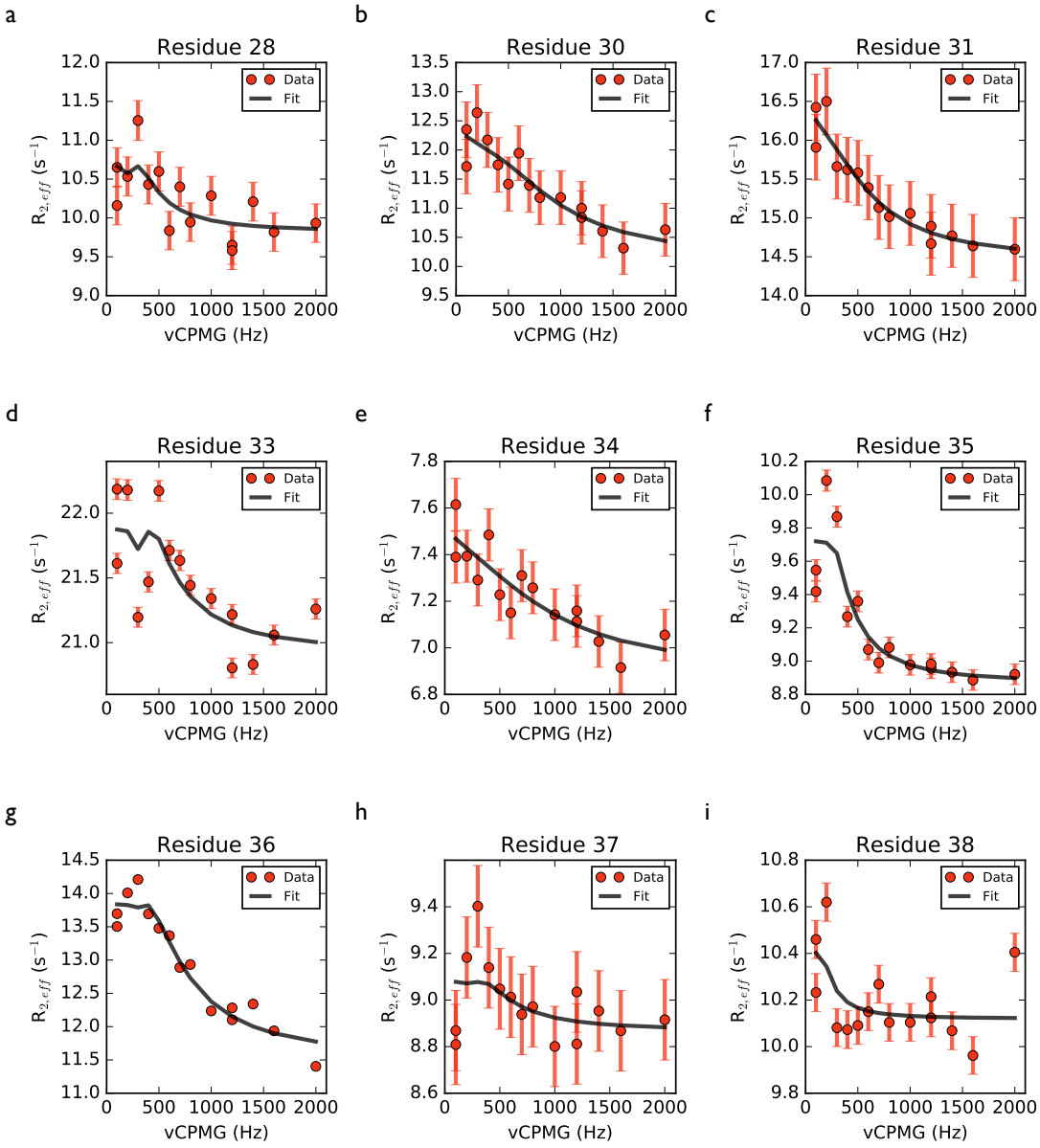


Fig. B.3: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

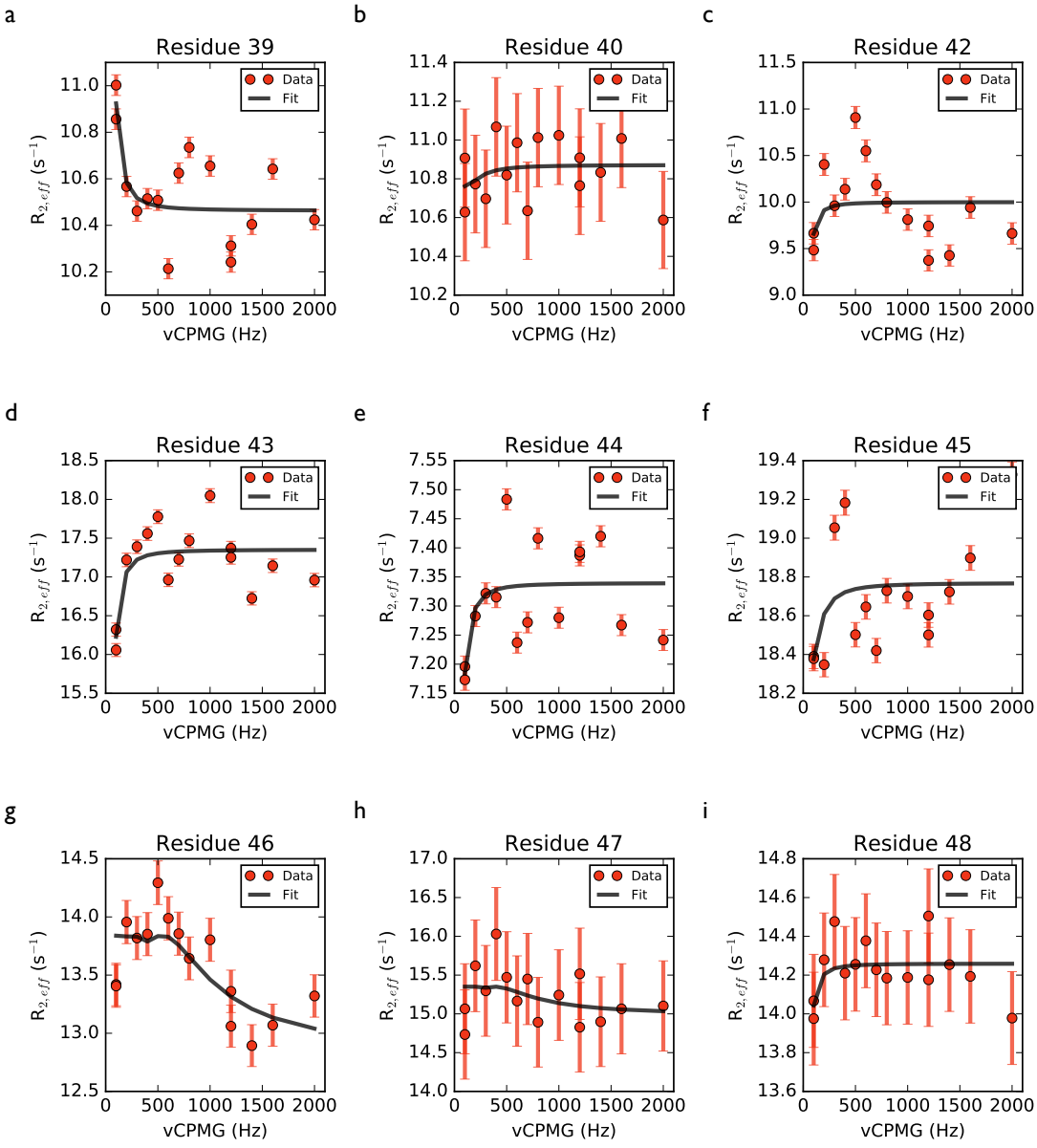


Fig. B.4: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

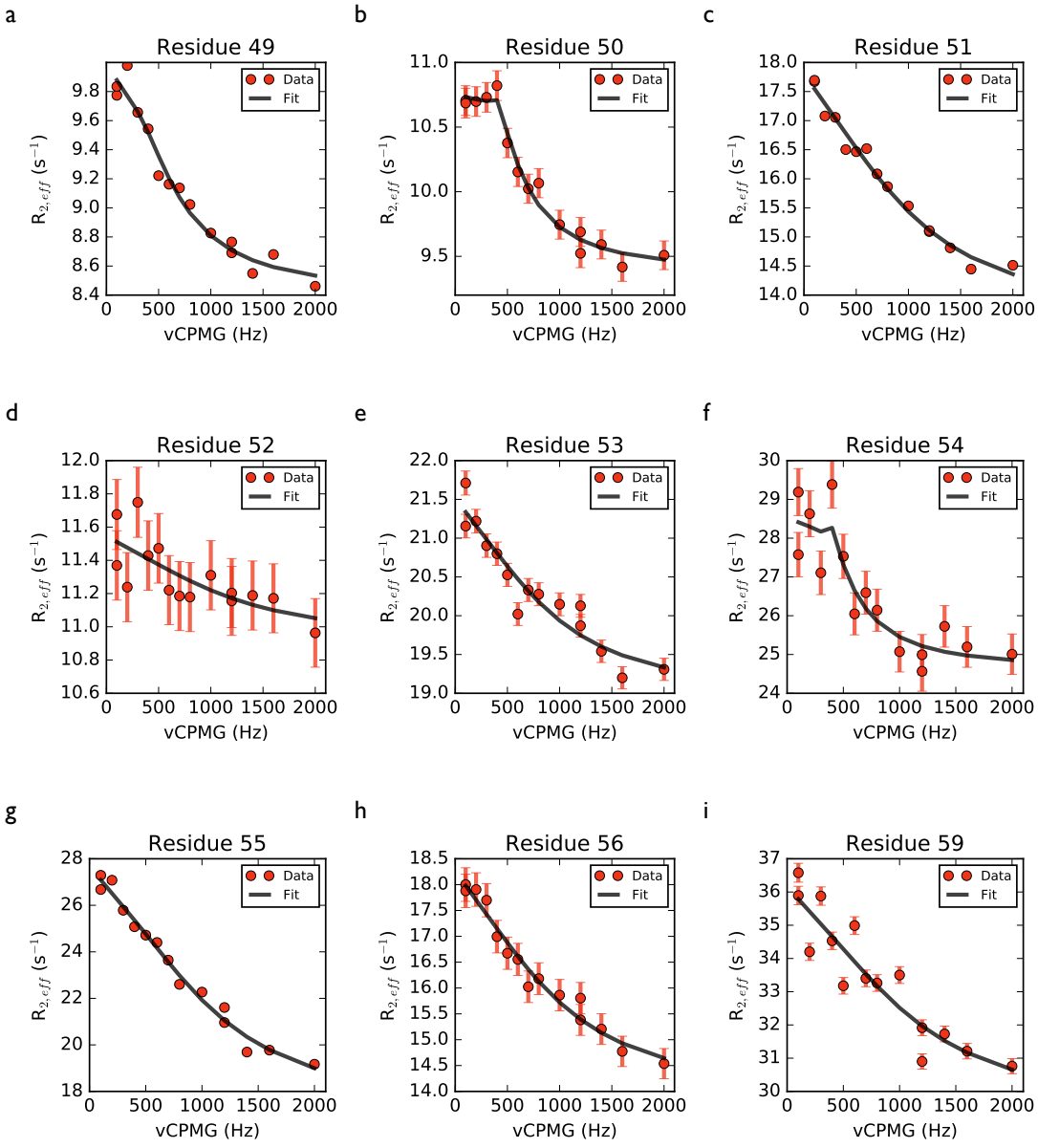


Fig. B.5: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

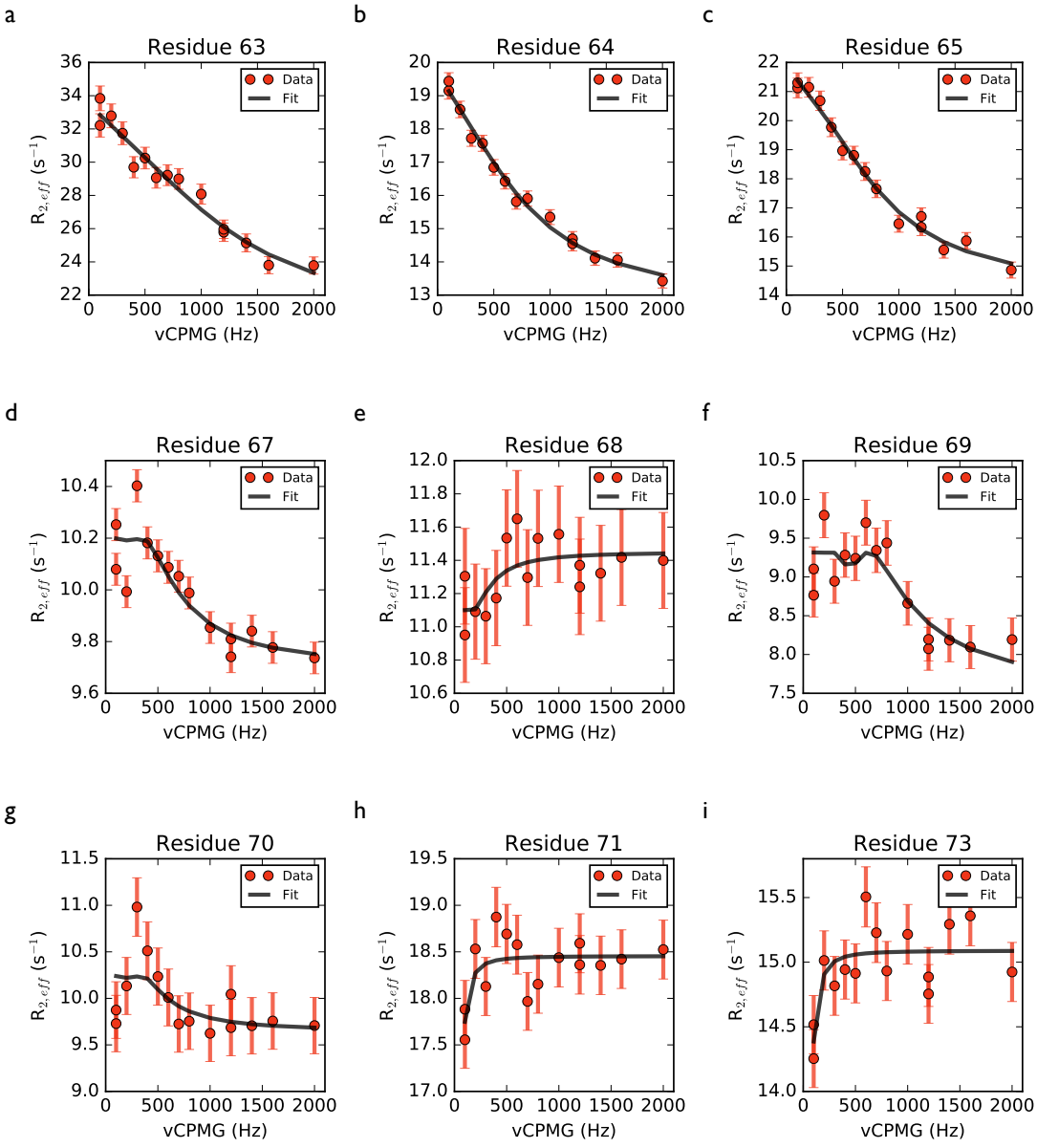


Fig. B.6: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

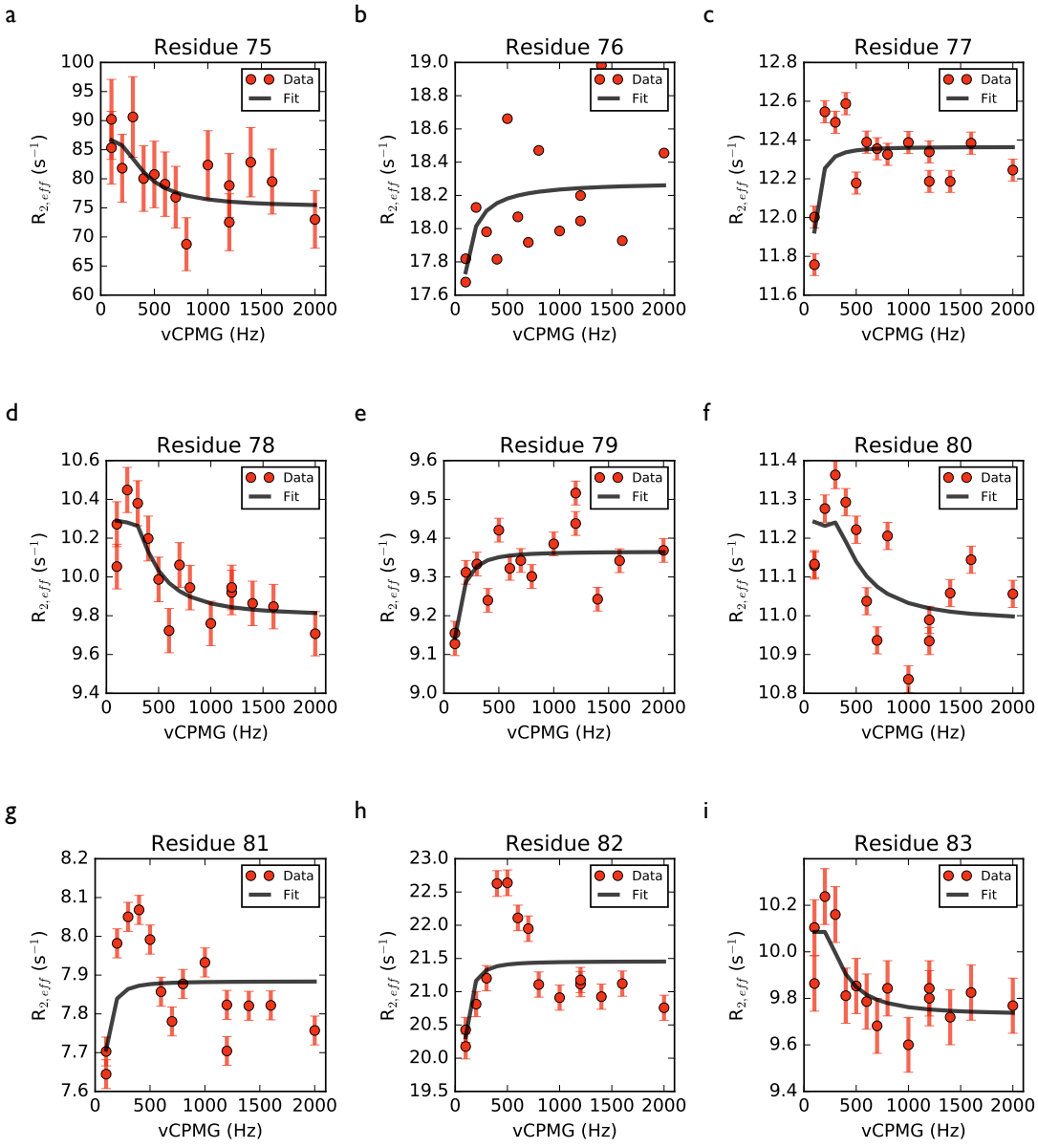


Fig. B.7: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

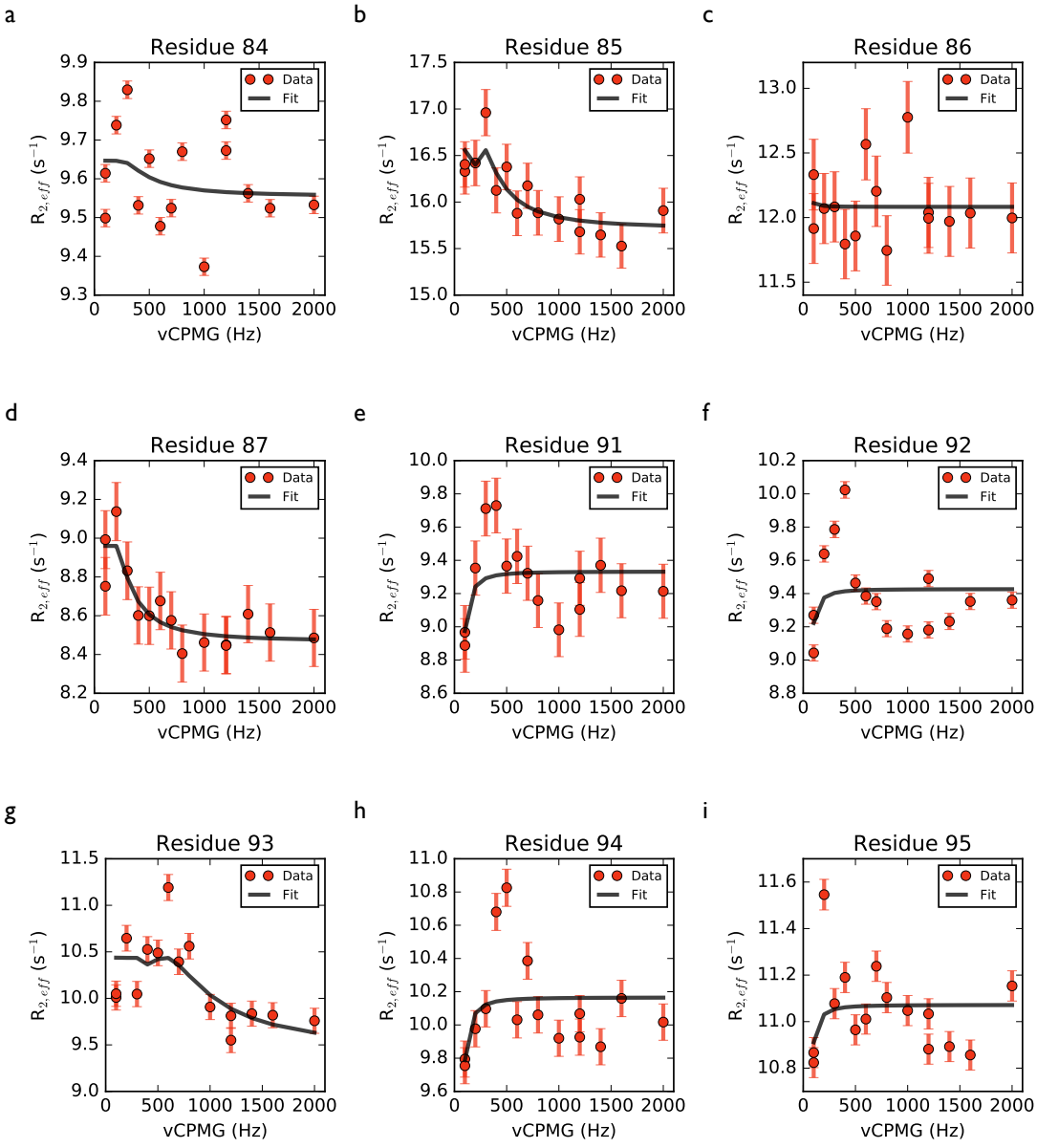


Fig. B.8: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

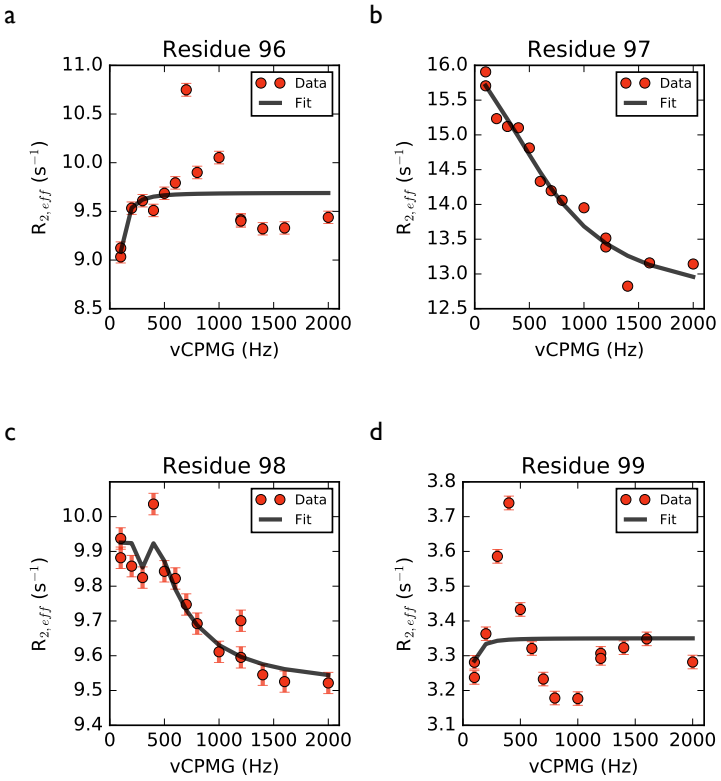


Fig. B.9: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 600MHz

B.2 CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

The residues which were deemed to be undergoing exchange are listed in Table B.2. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test.

Table B.2: Exchanging residues in WT at pH 7.4, measured at 600 MHz. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test. Data were acquired with 500 μM samples at pH 7.4, recorded with a 600 MHz QCI-P-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

Residue	Residue
4	36
8	49
9	5
10	55
11	56
13	63
20	64
23	65
27	97
31	

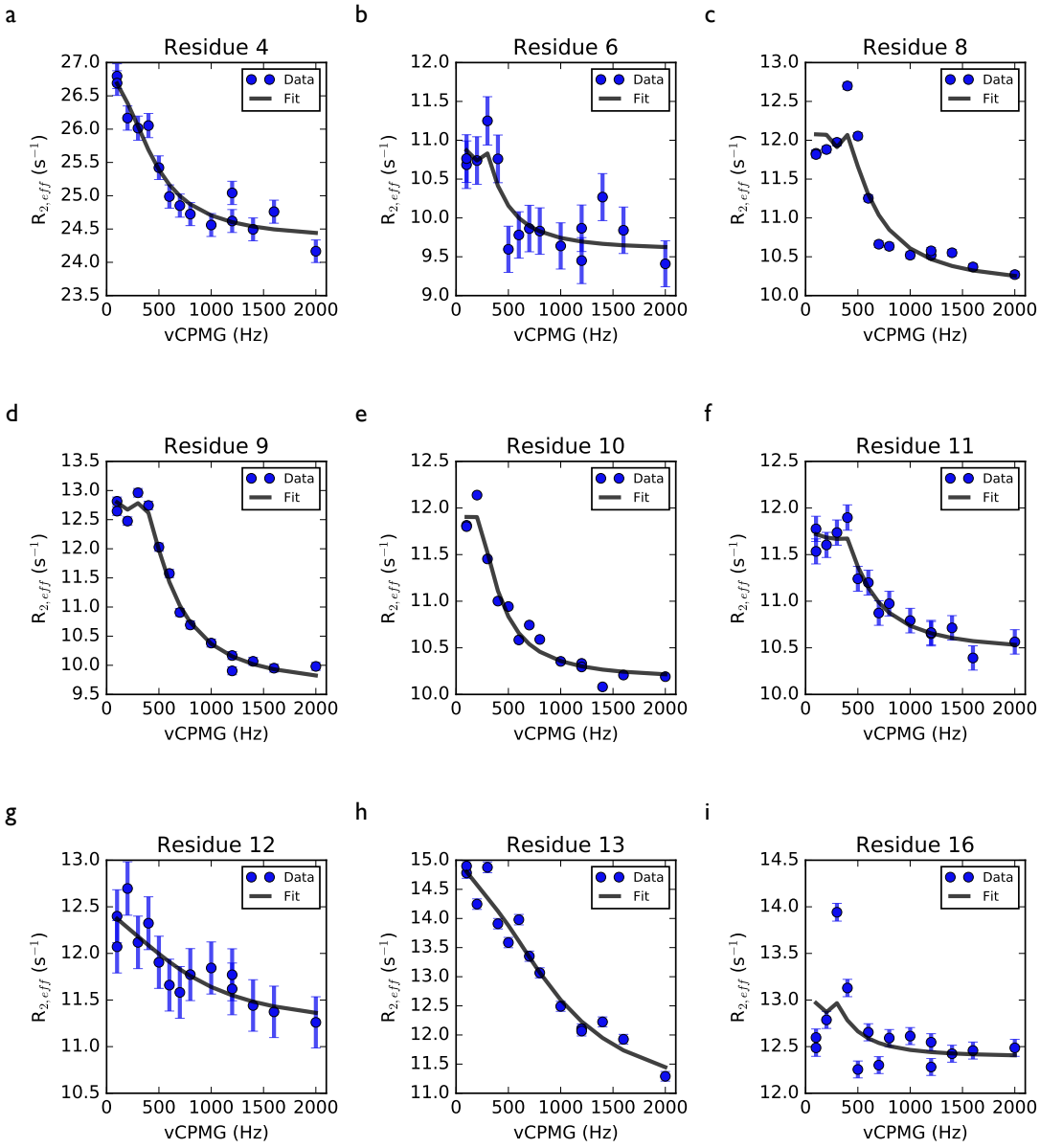


Fig. B.10: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

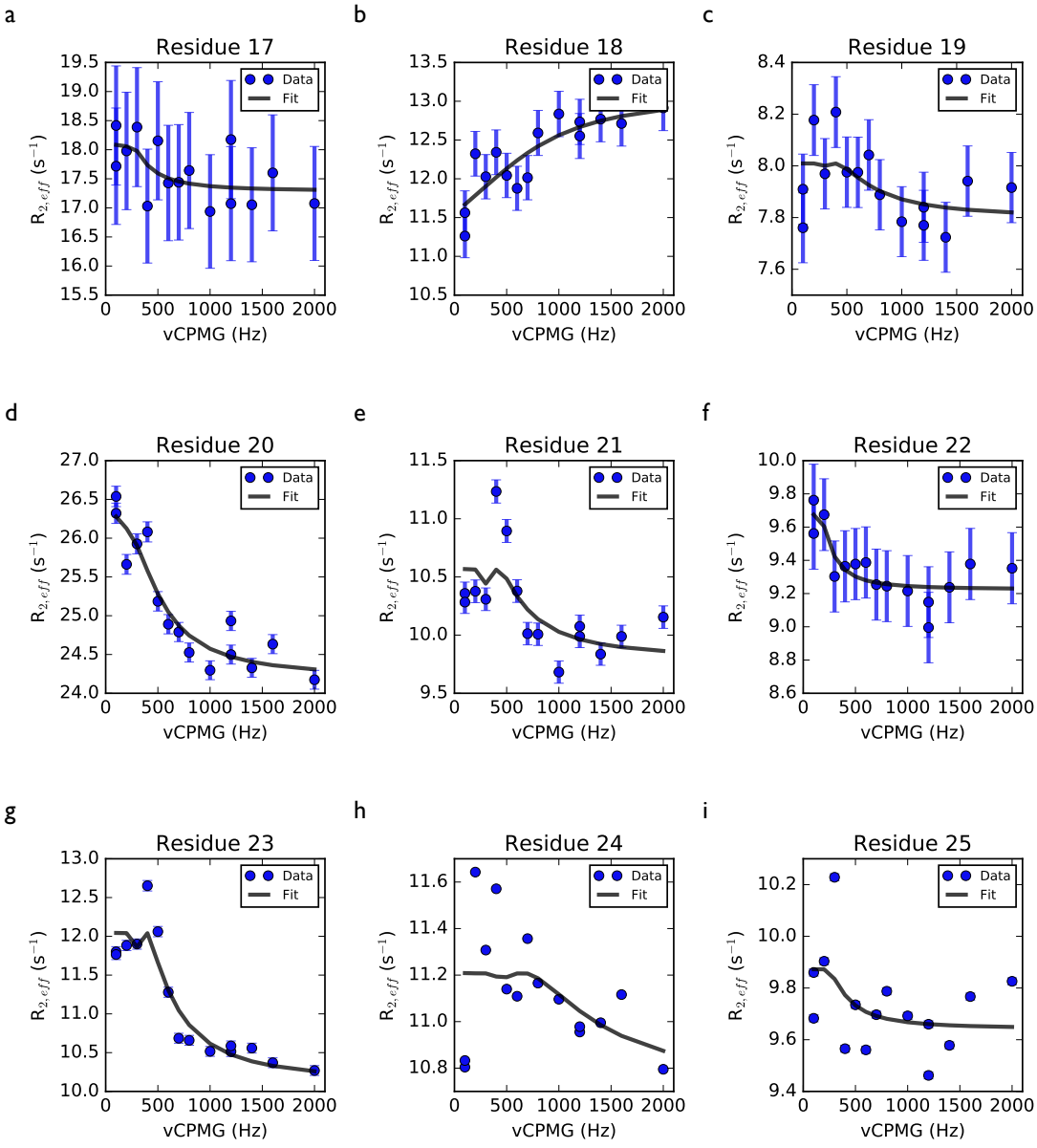


Fig. B.11: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

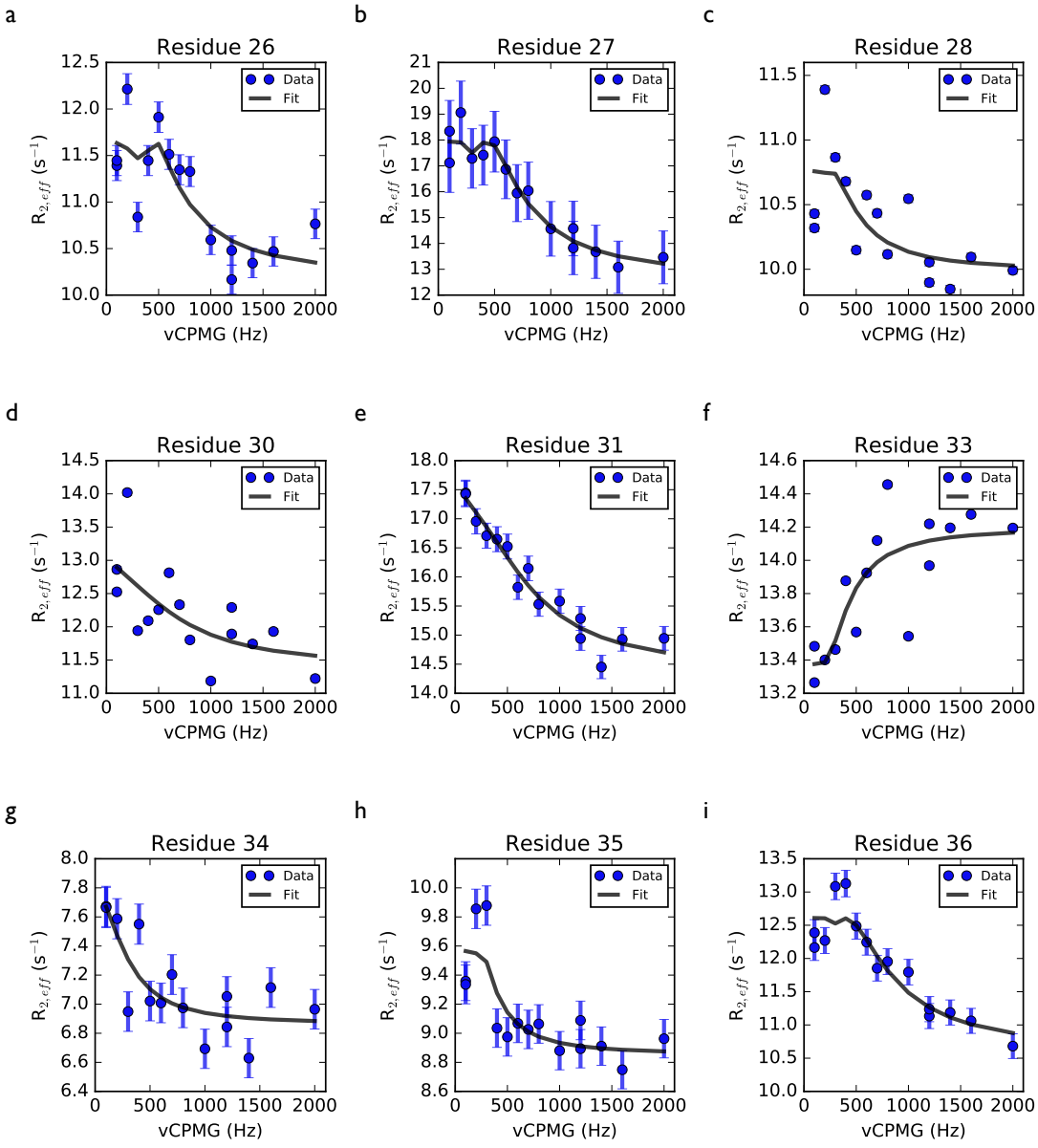


Fig. B.12: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

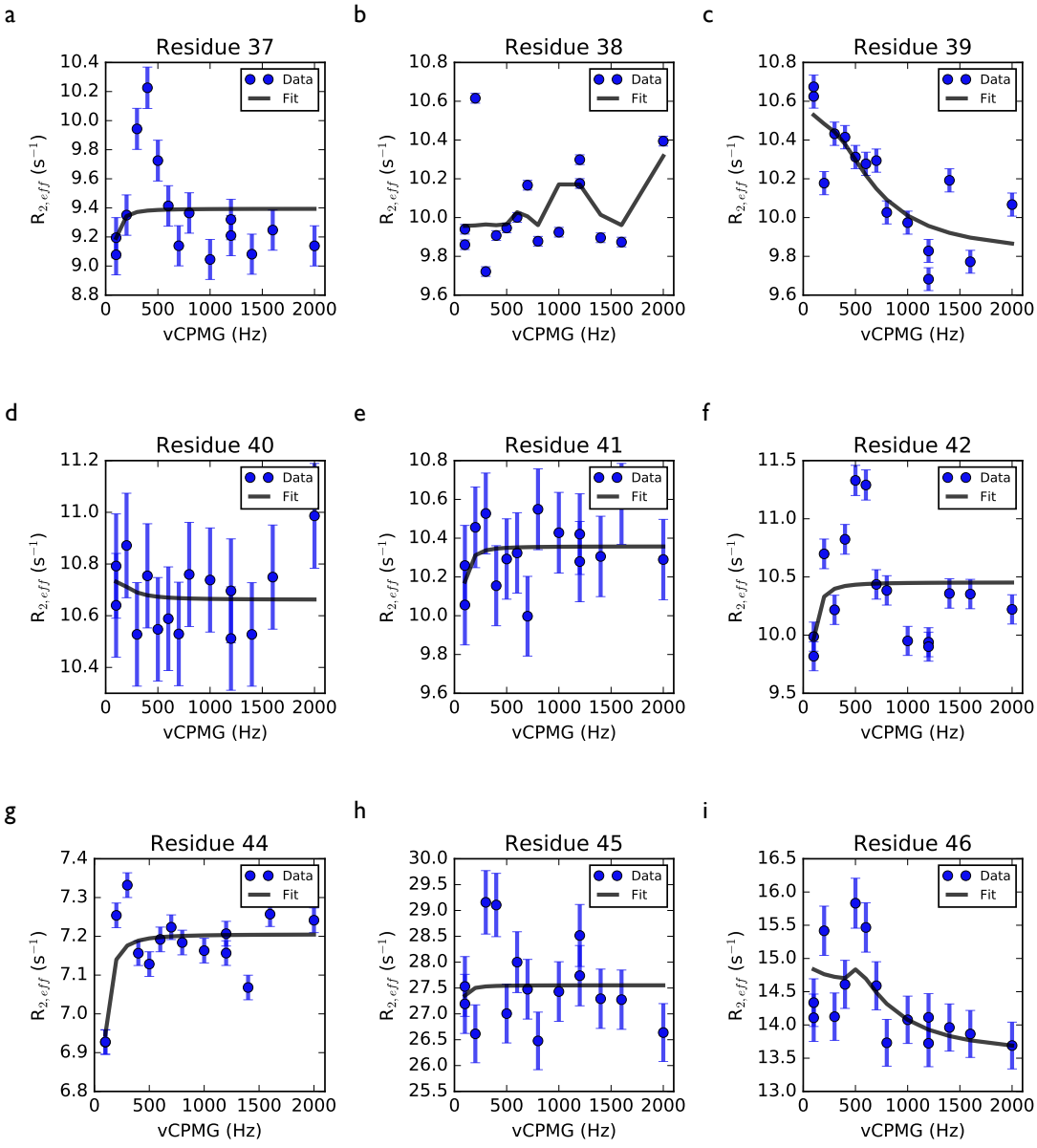


Fig. B.13: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

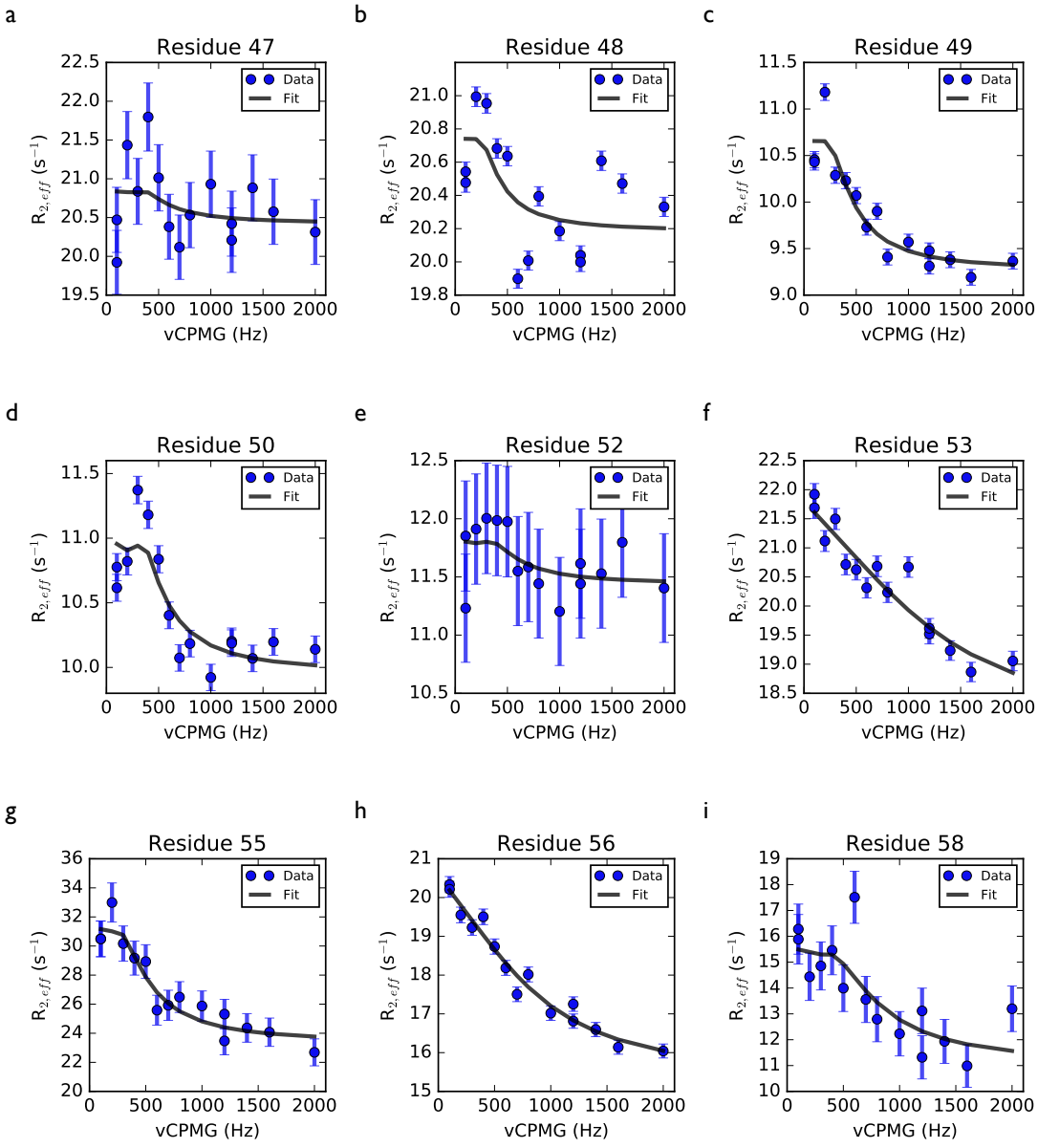


Fig. B.14: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

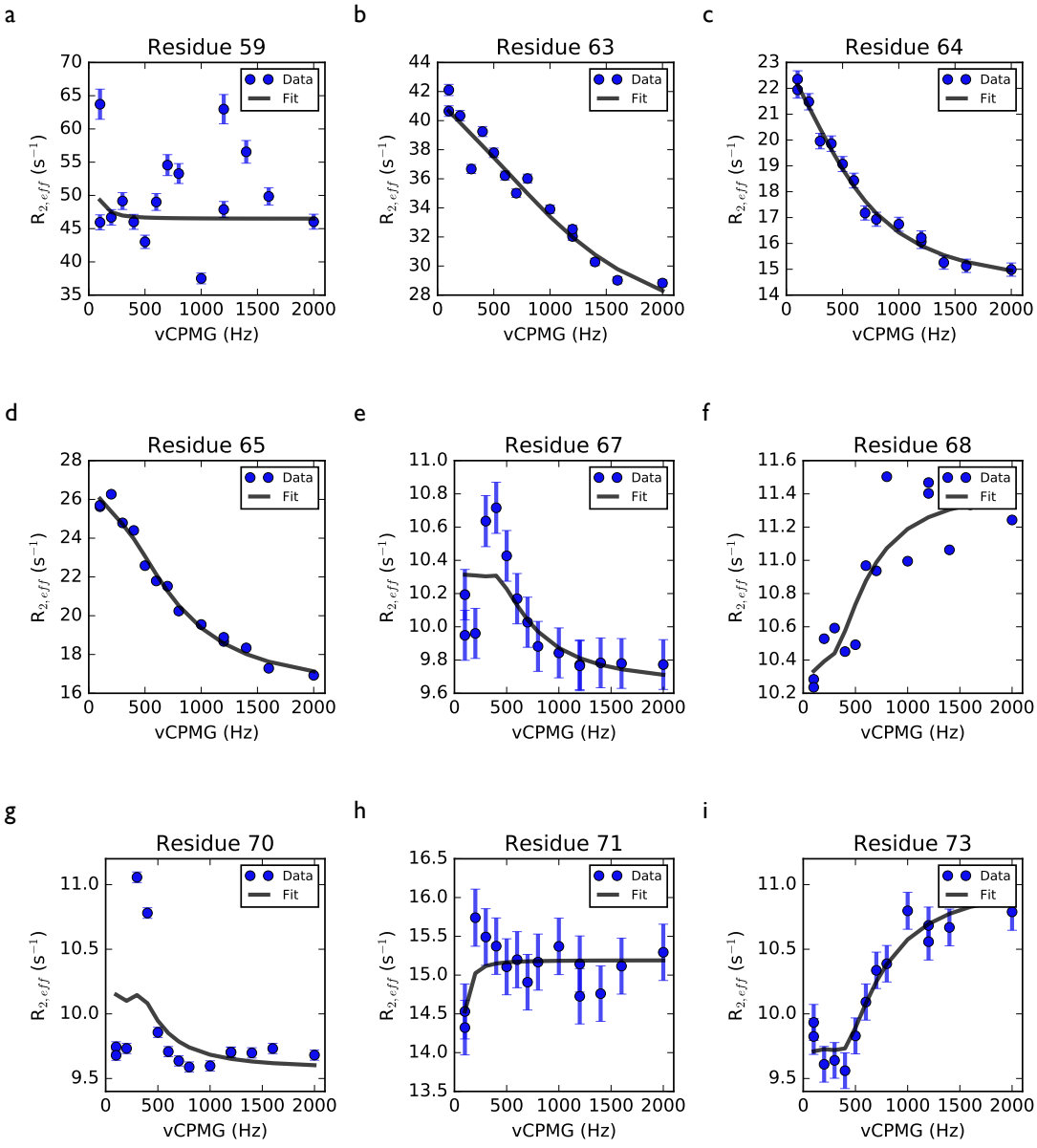


Fig. B.15: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

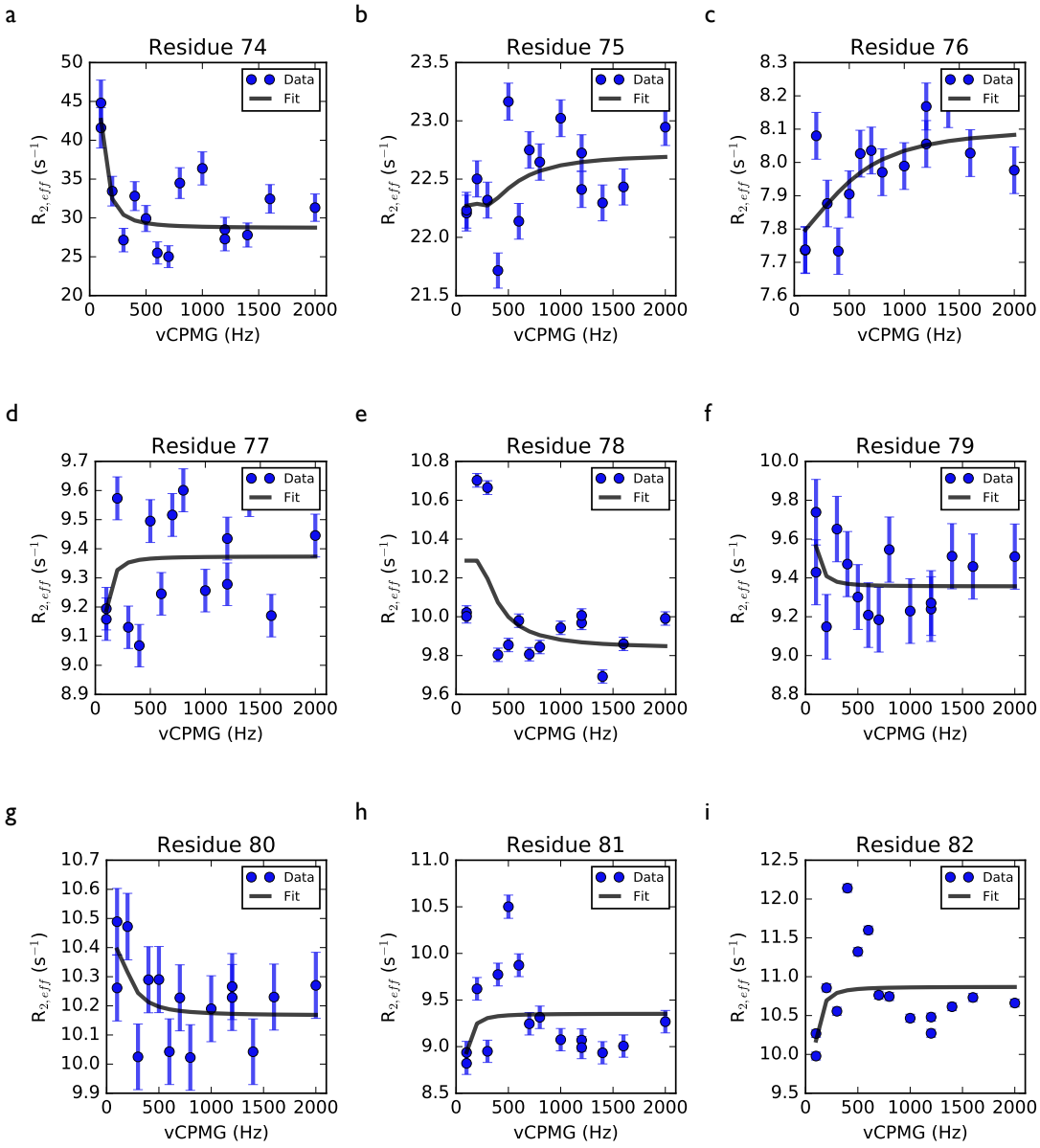


Fig. B.16: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

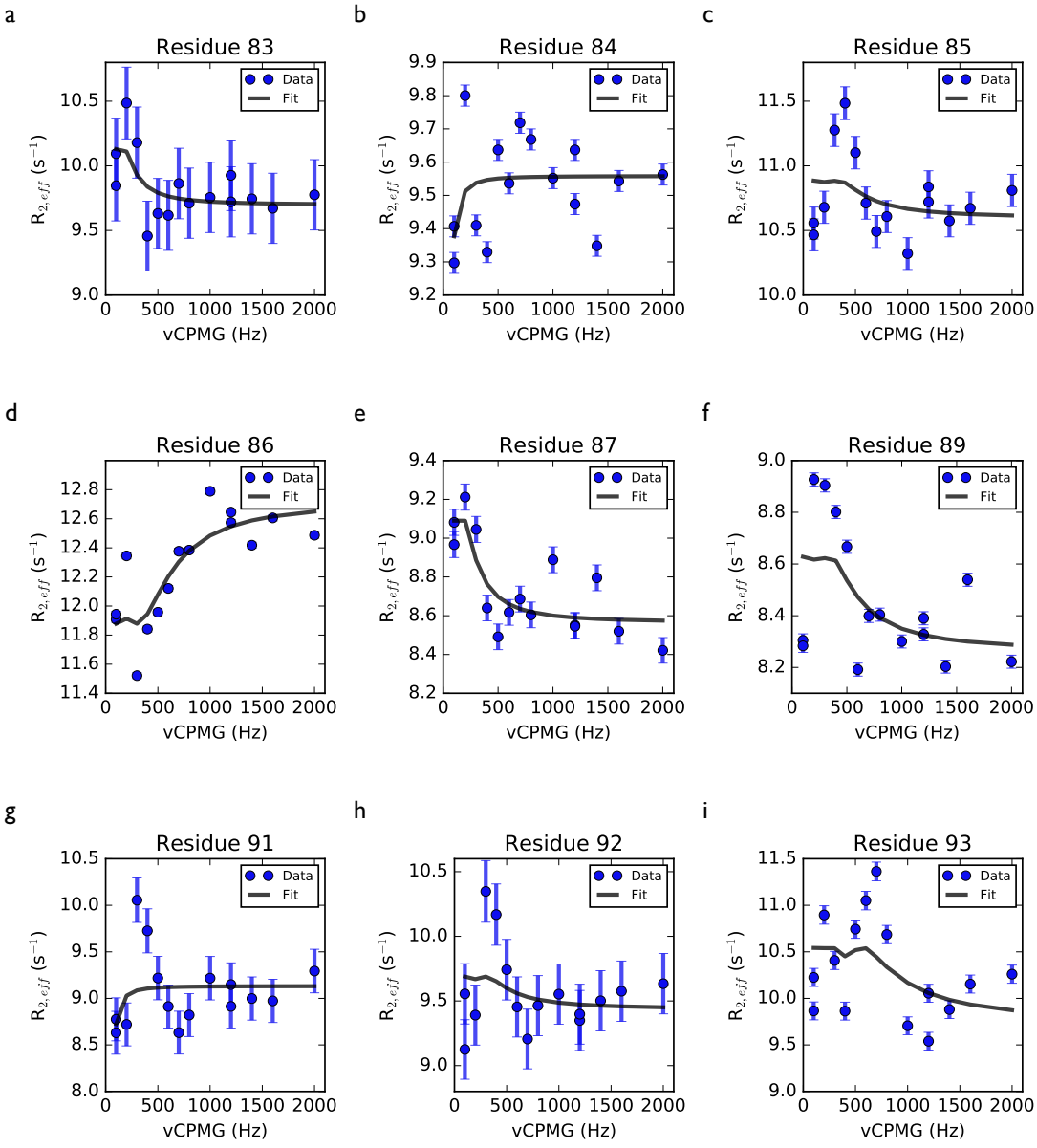


Fig. B.17: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

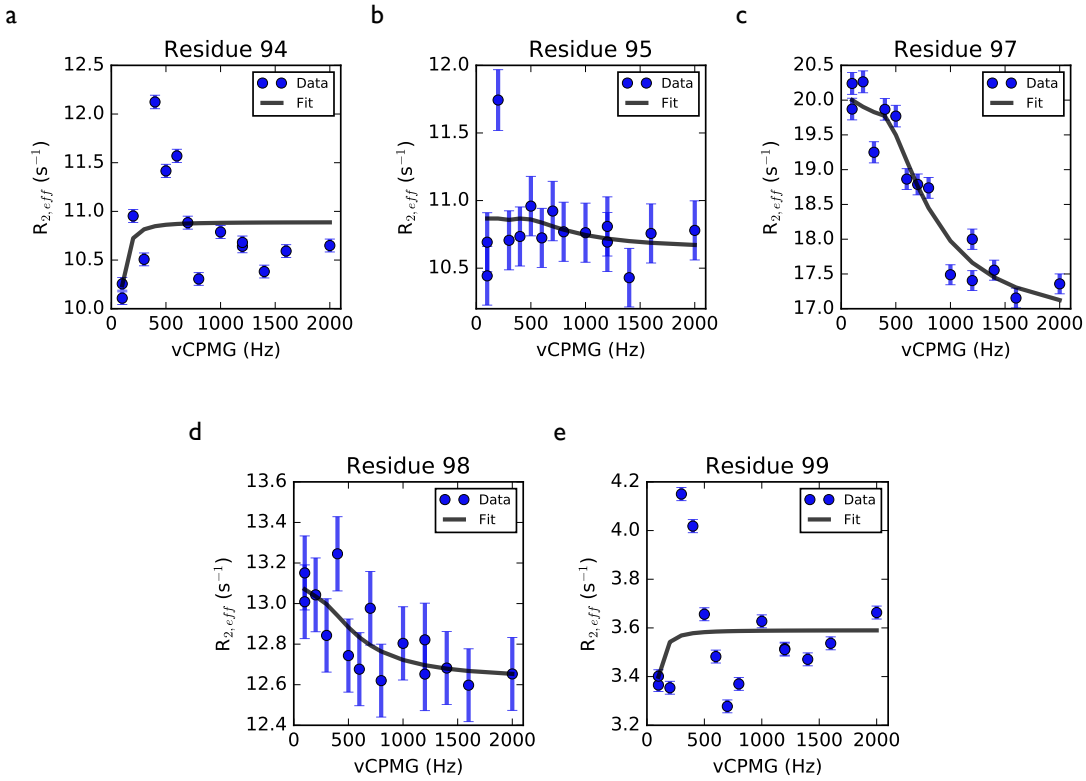


Fig. B.18: CPMG-RD dispersion curves for WT at pH 7.4 measured at 600MHz

B.3 CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

The residues which were deemed to be undergoing exchange are listed in Table B.3. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test.

Table B.3: Exchanging residues in D76N at pH 7.4, measured at 750 MHz. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test. Data were acquired with 500 μM samples at pH 7.4, recorded with a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

Residue	Residue
2	33
3	36
6	39
7	49
9	50
10	51
11	53
12	54
13	55
16	56
19	59
24	63
26	64
27	65
30	97
31	98

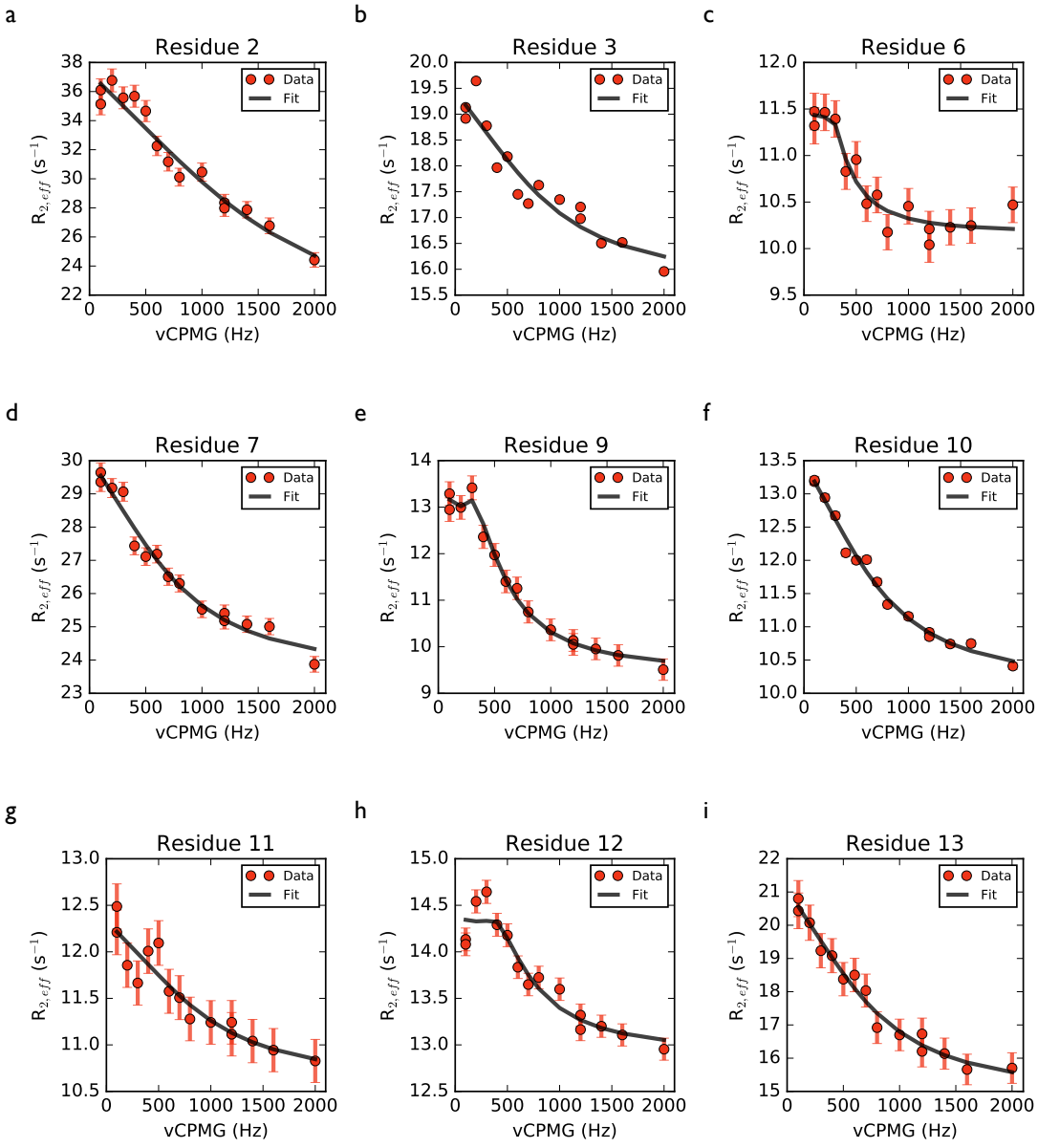


Fig. B.19: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

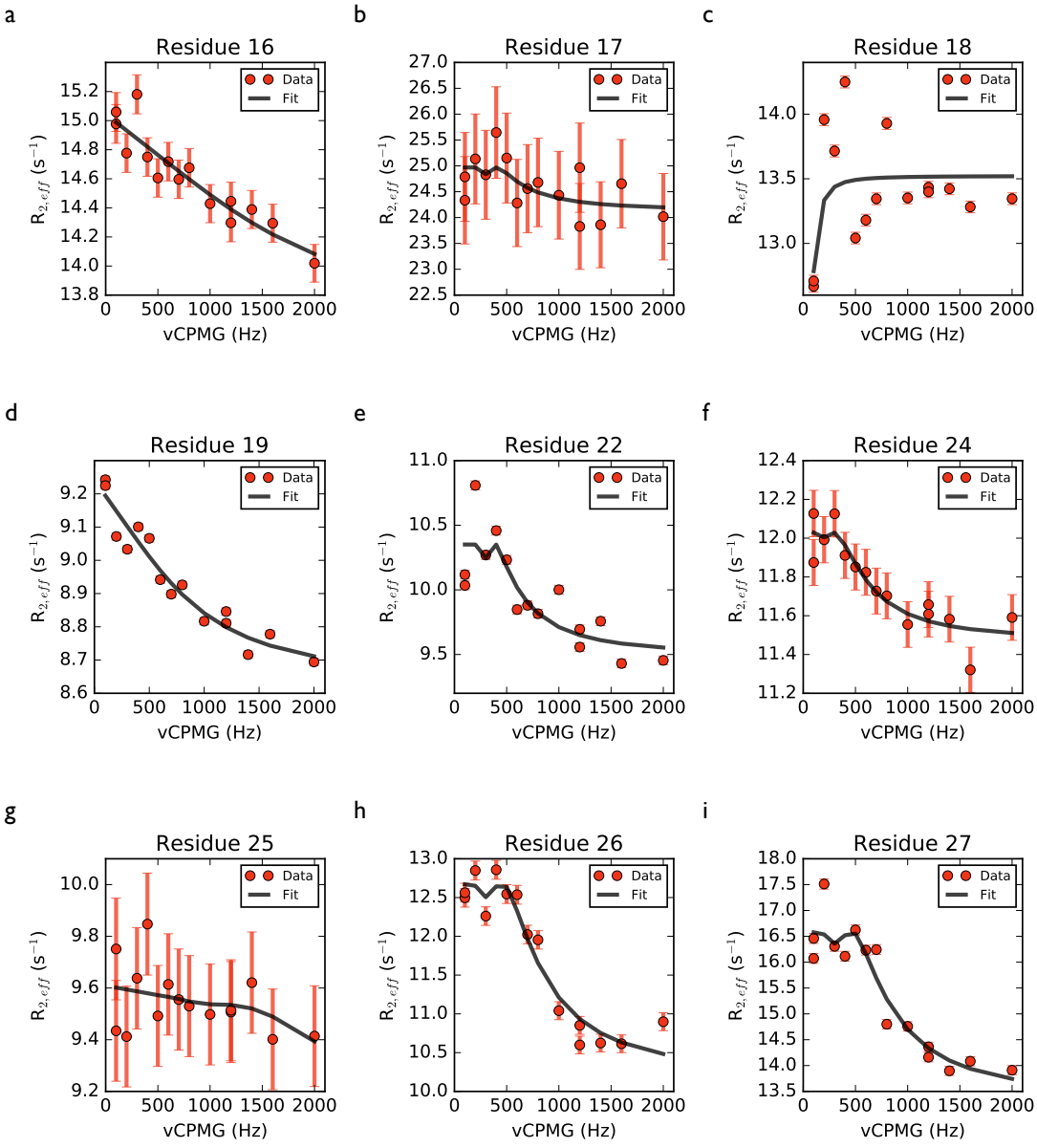


Fig. B.20: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

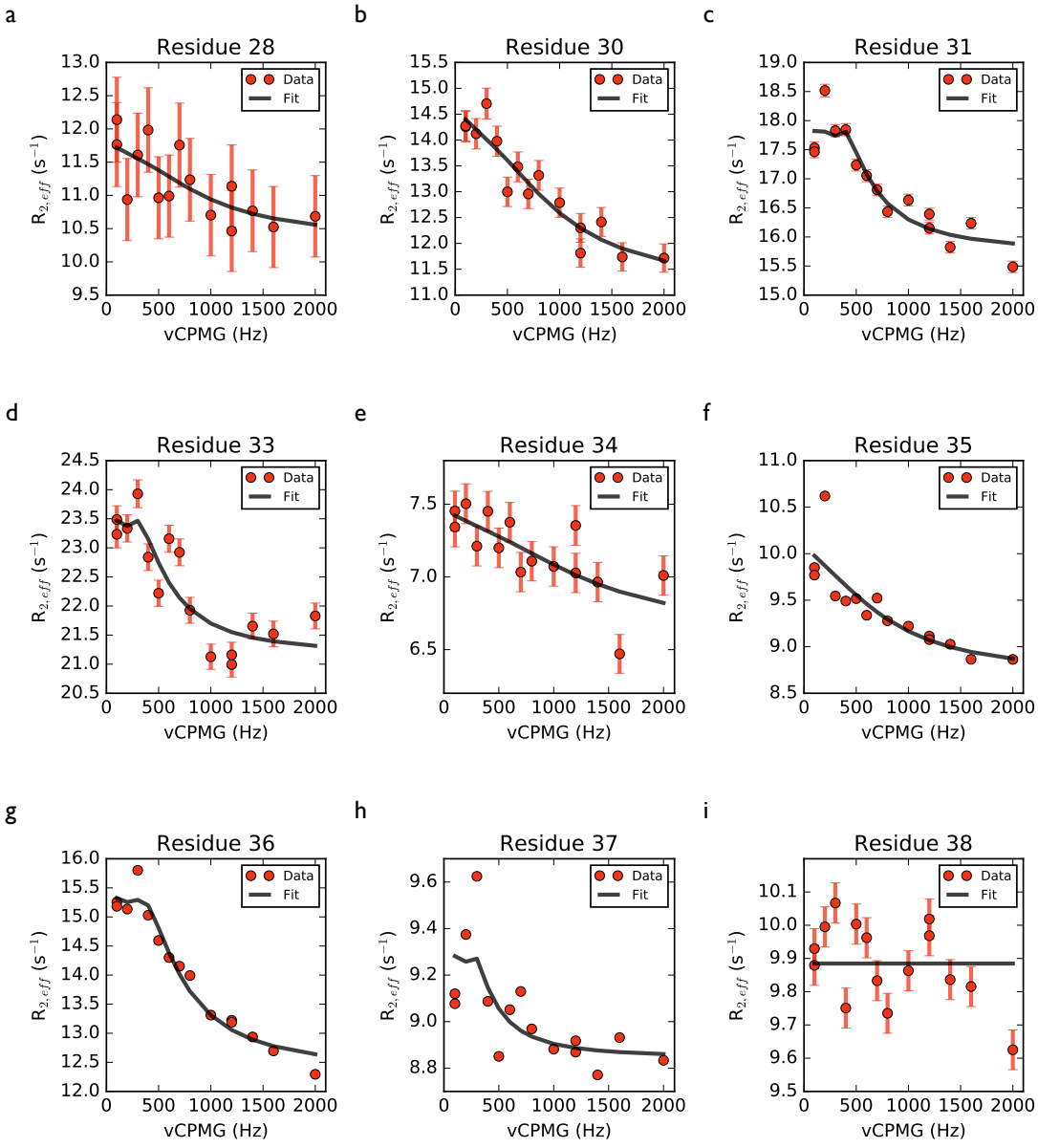


Fig. B.21: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

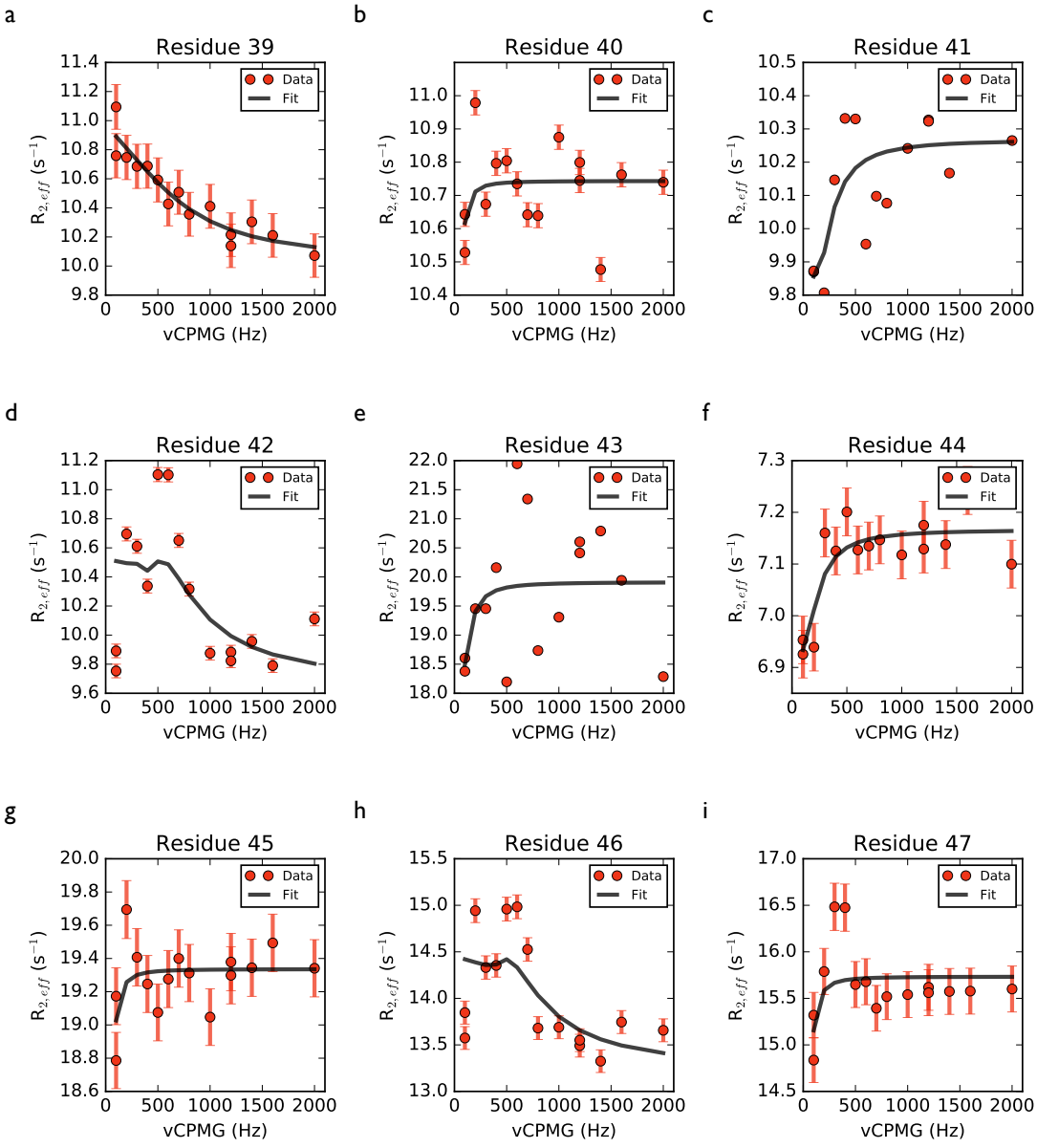


Fig. B.22: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

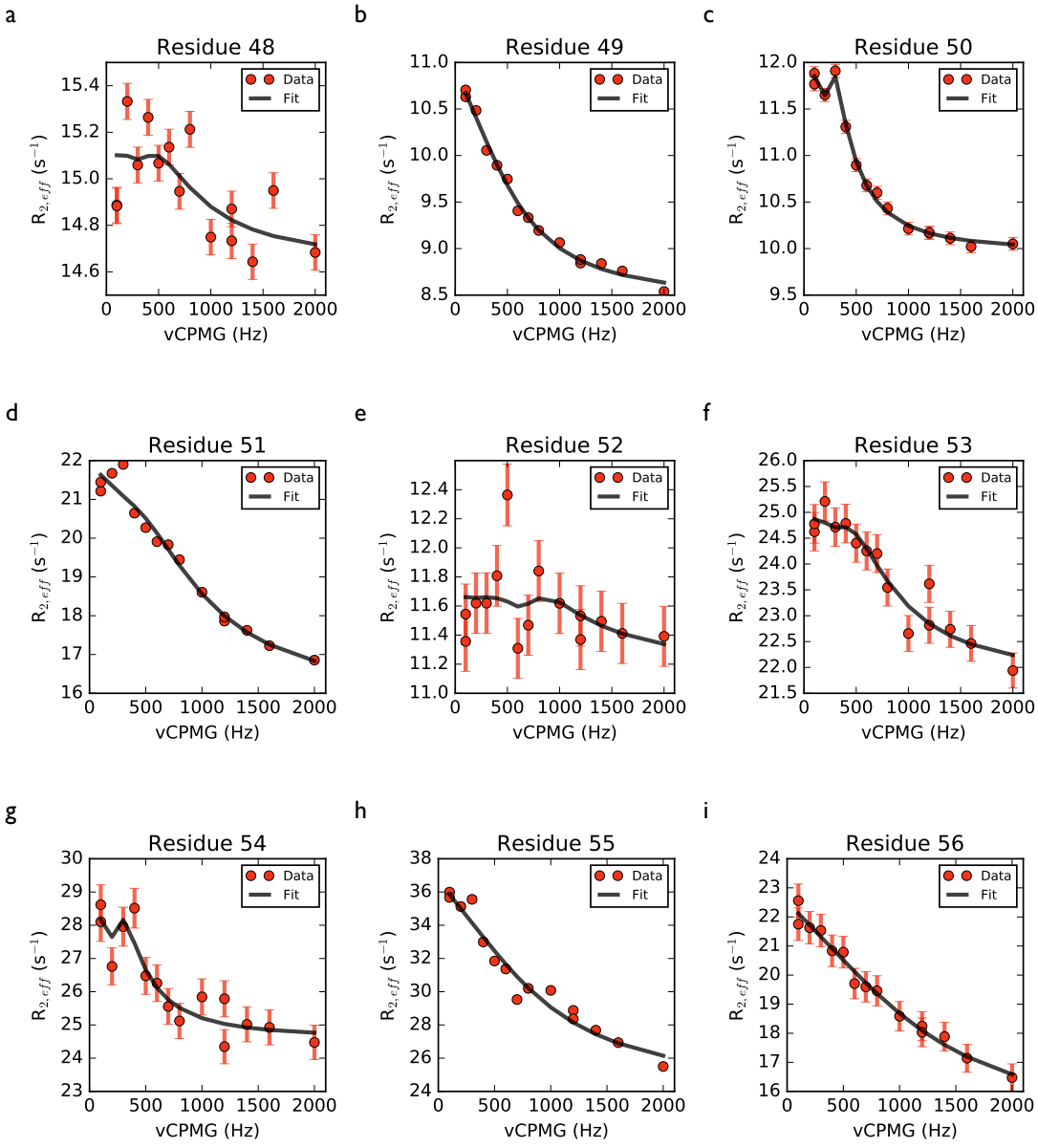


Fig. B.23: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

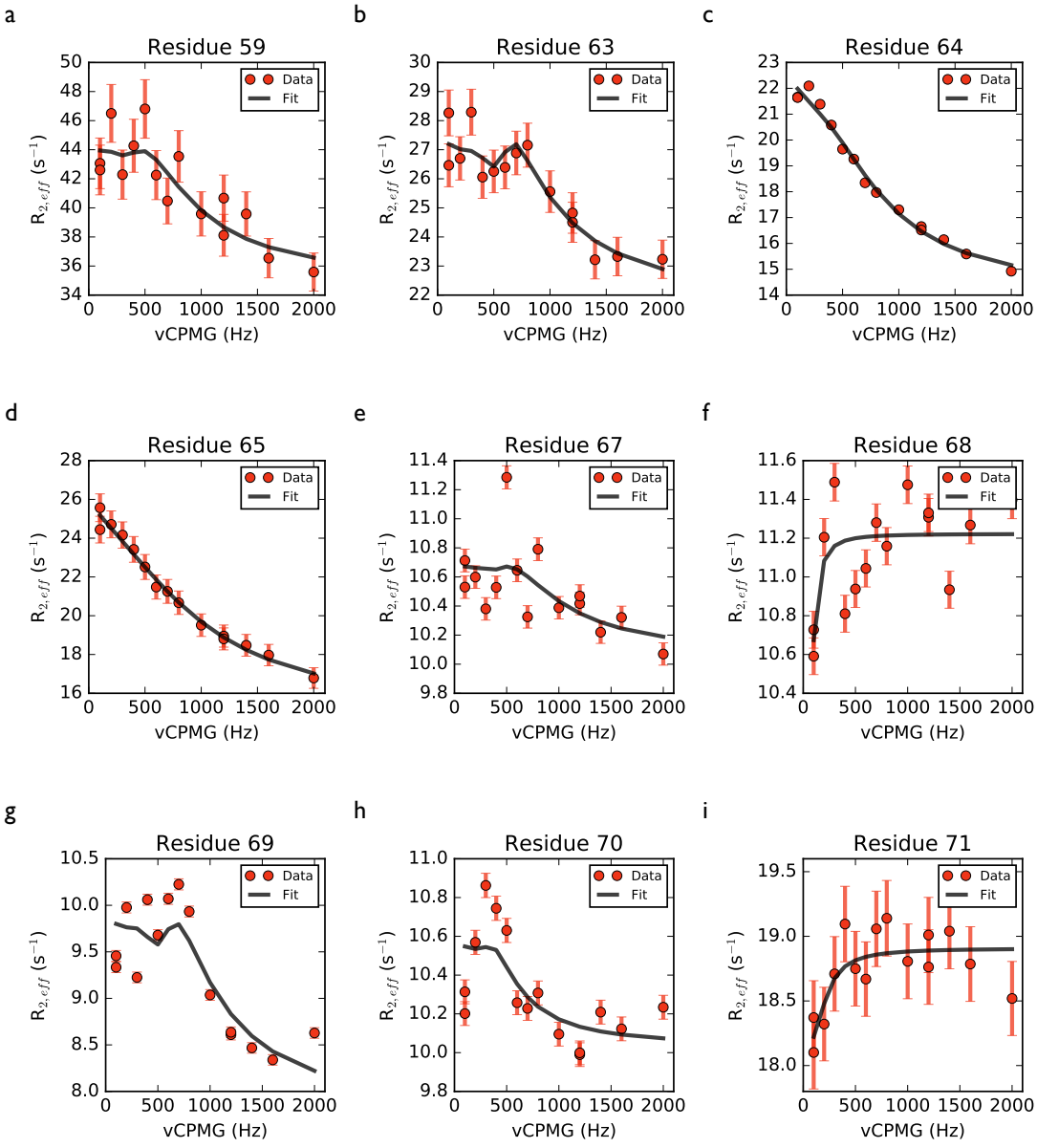


Fig. B.24: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

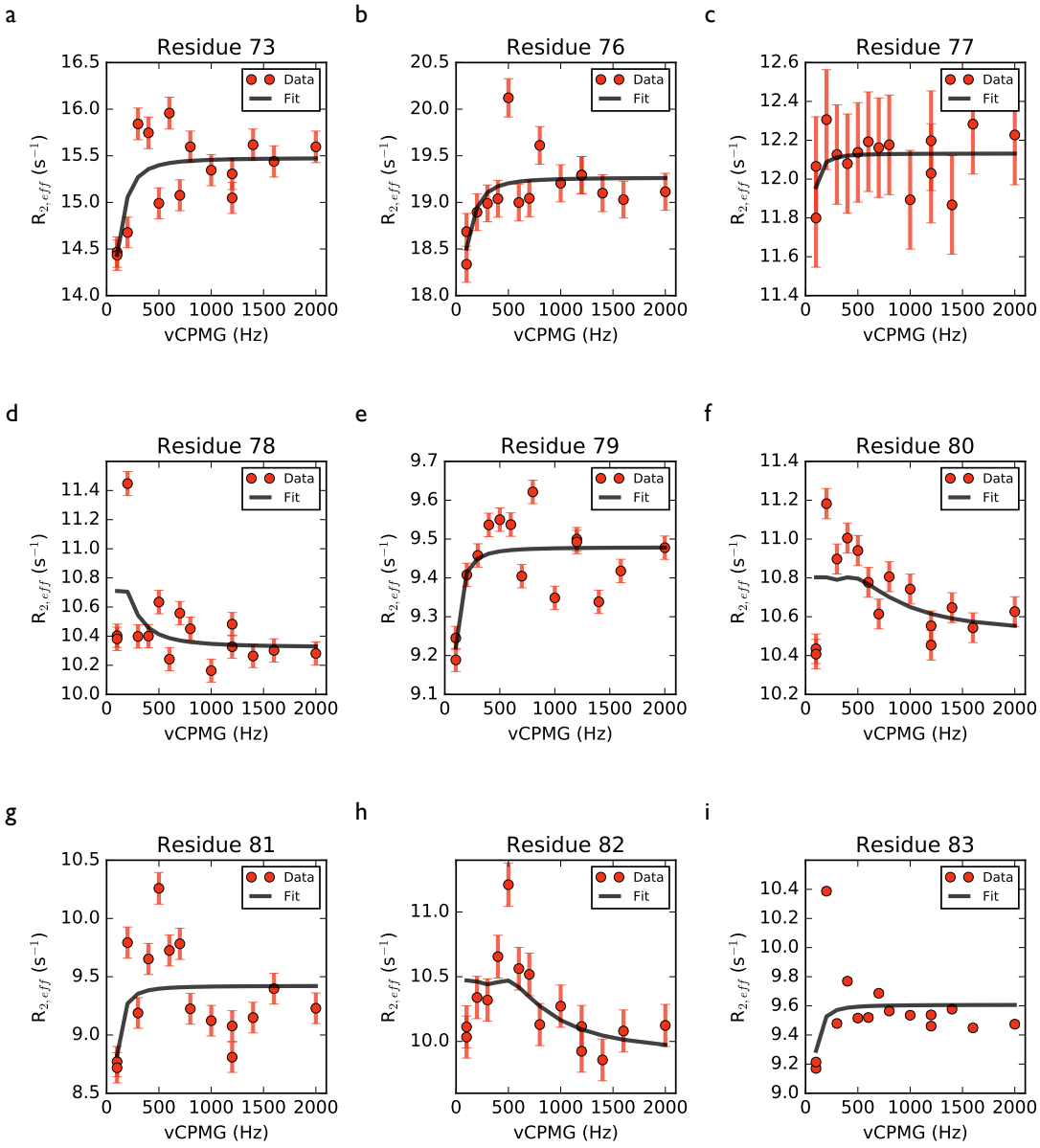


Fig. B.25: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

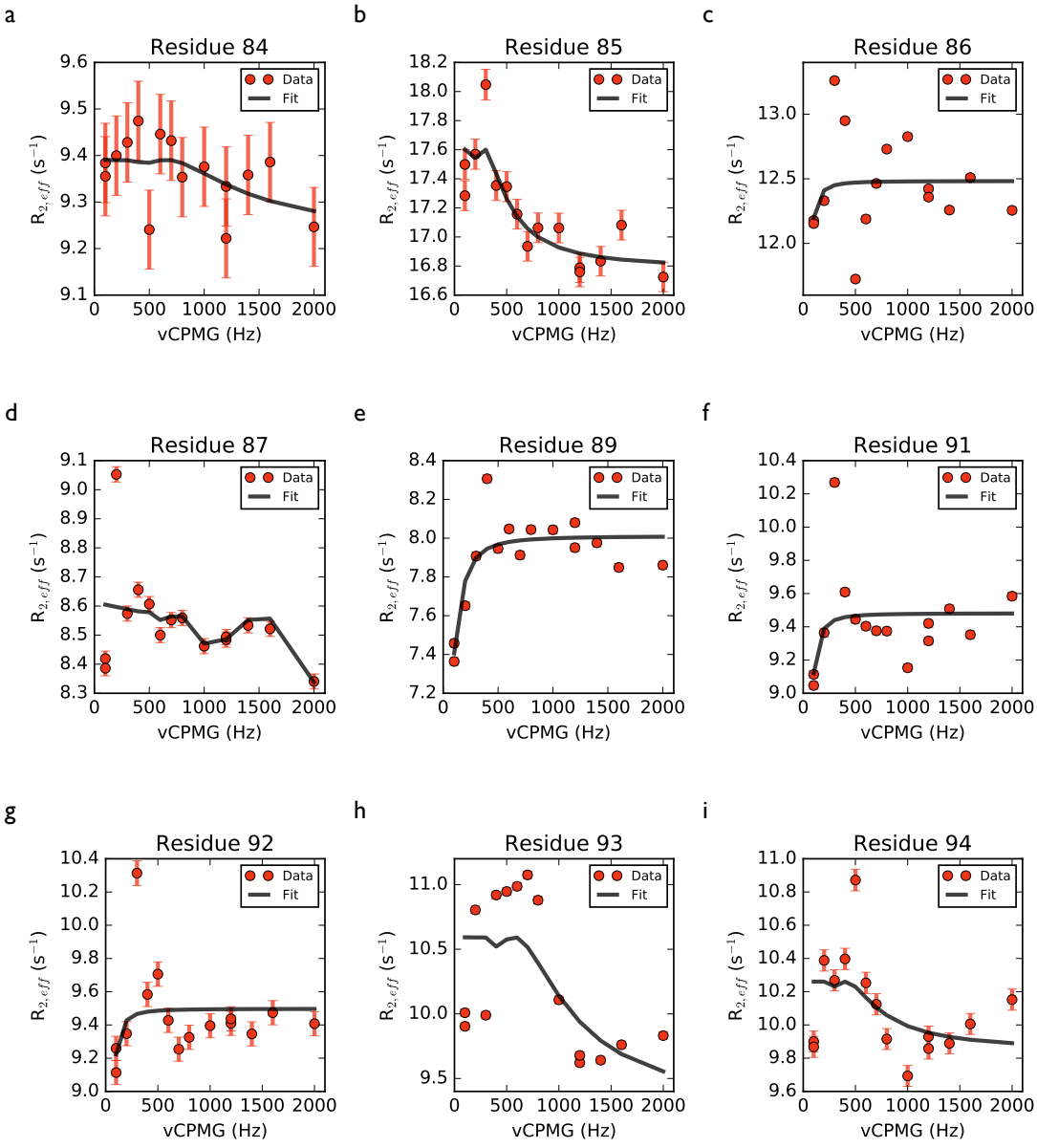


Fig. B.26: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

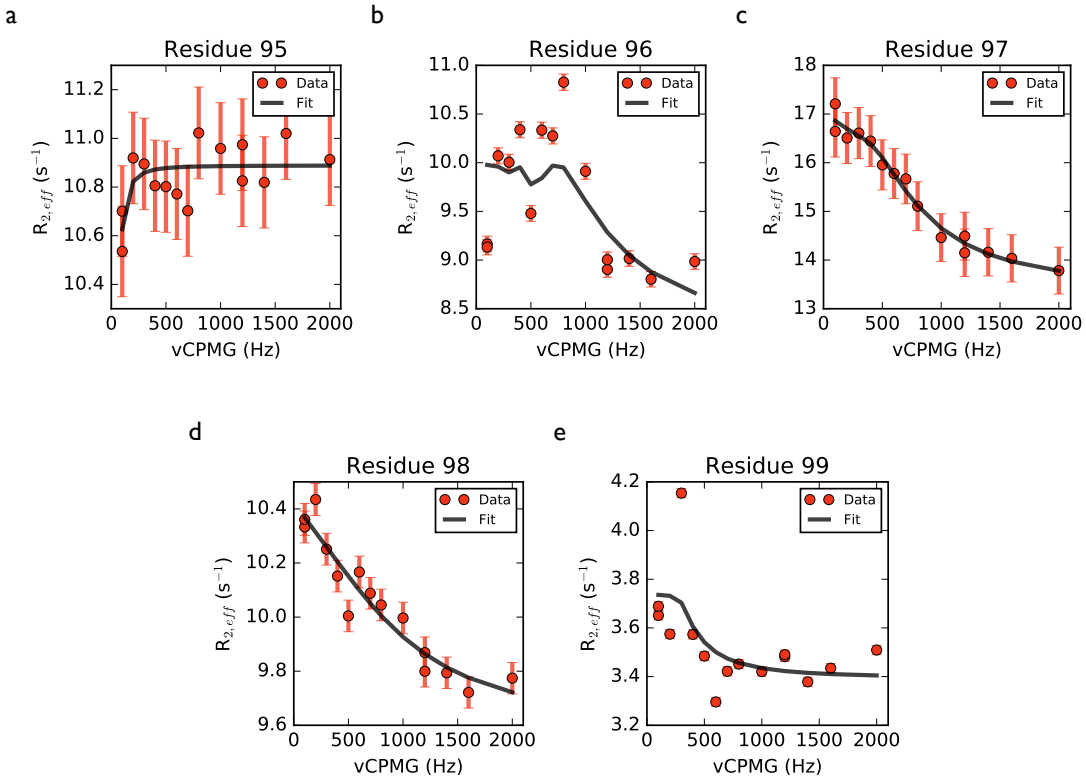


Fig. B.27: CPMG-RD dispersion curves for D76N at pH 7.4 measured at 750MHz

B.4 CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

The residues which were deemed to be undergoing exchange are listed in Table B.4. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test.

Table B.4: Exchanging residues in D76N at pH 6.2, measured at 750 MHz. Residues were deemed to be undergoing exchange when the fit to the Bloch-McConnell equations for two-site exchange was preferred over the fit to a constant model (no exchange) after the performance of an F-test. Data were acquired with 500 μM samples at pH 6.2, recorded with a 750 MHz TCI-cryoprobe; proton-optimised quadruple resonance NMR 'inverse' probe and a Bruker Avance III HD console.

Residue	Residue	Residue
6	22	56
7	26	59
8	30	63
9	33	64
10	35	65
11	36	67
12	46	69
13	48	75
16	49	84
17	50	97
19	52	98
21	55	

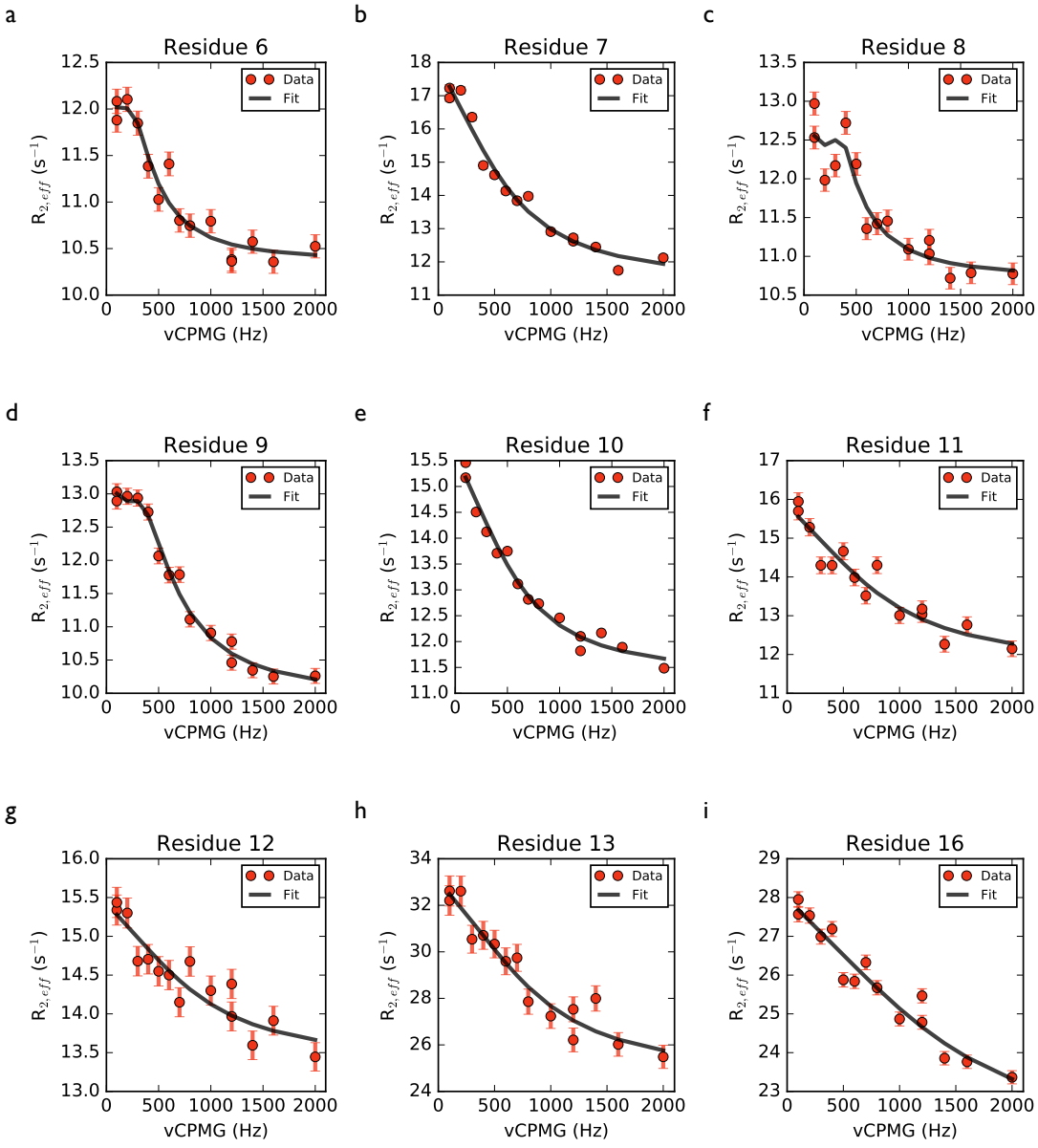


Fig. B.28: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

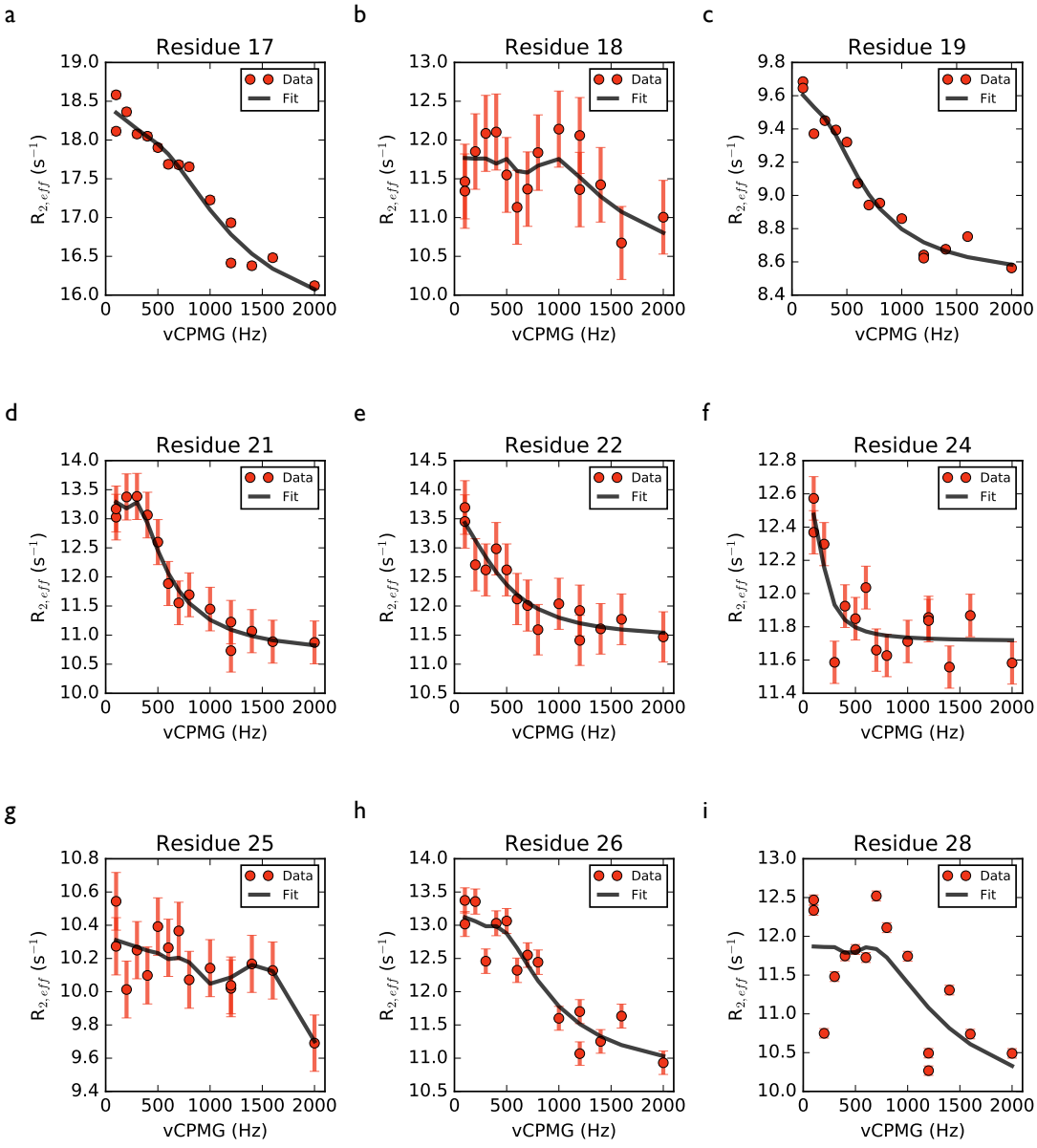


Fig. B.29: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

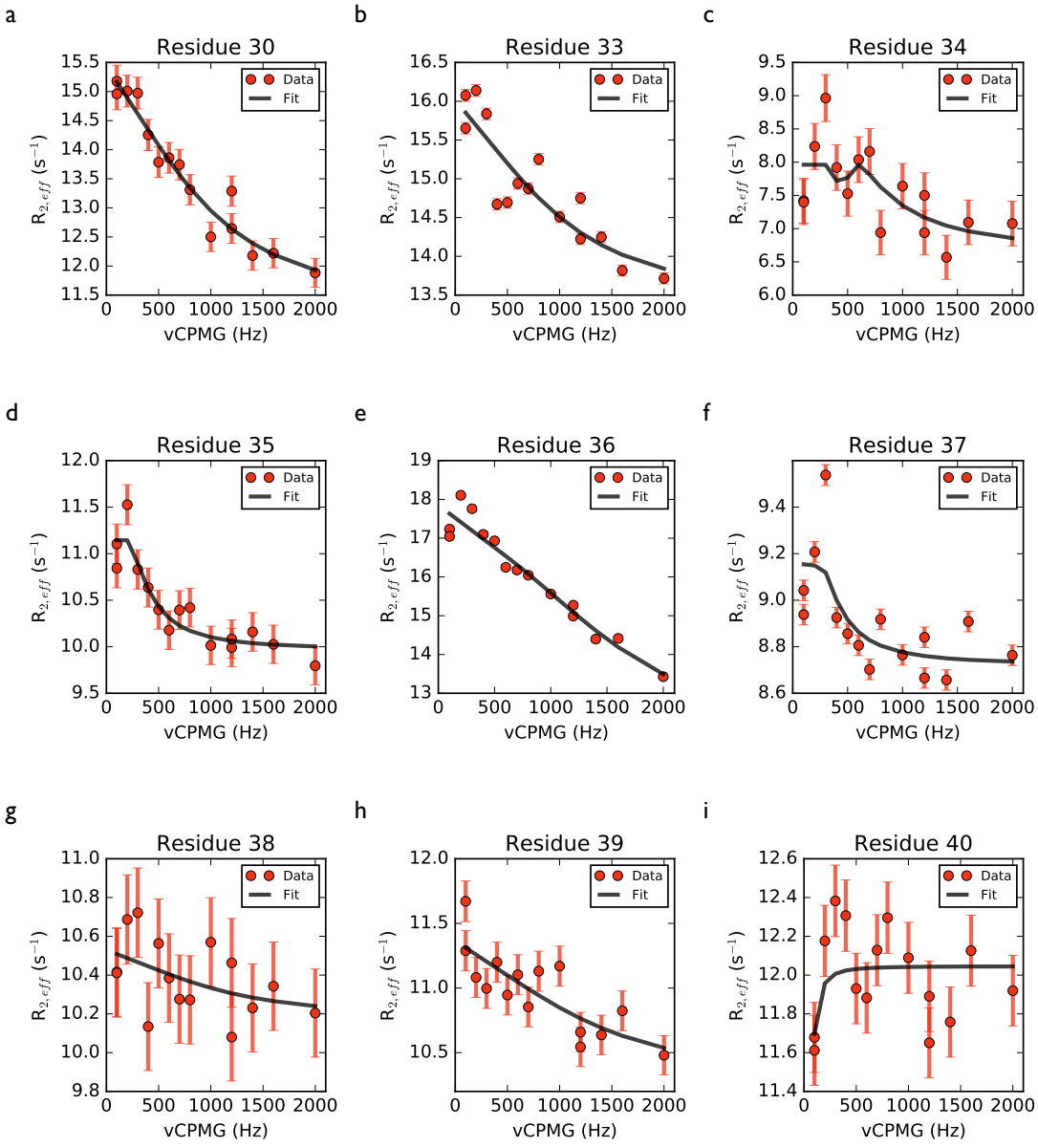


Fig. B.30: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

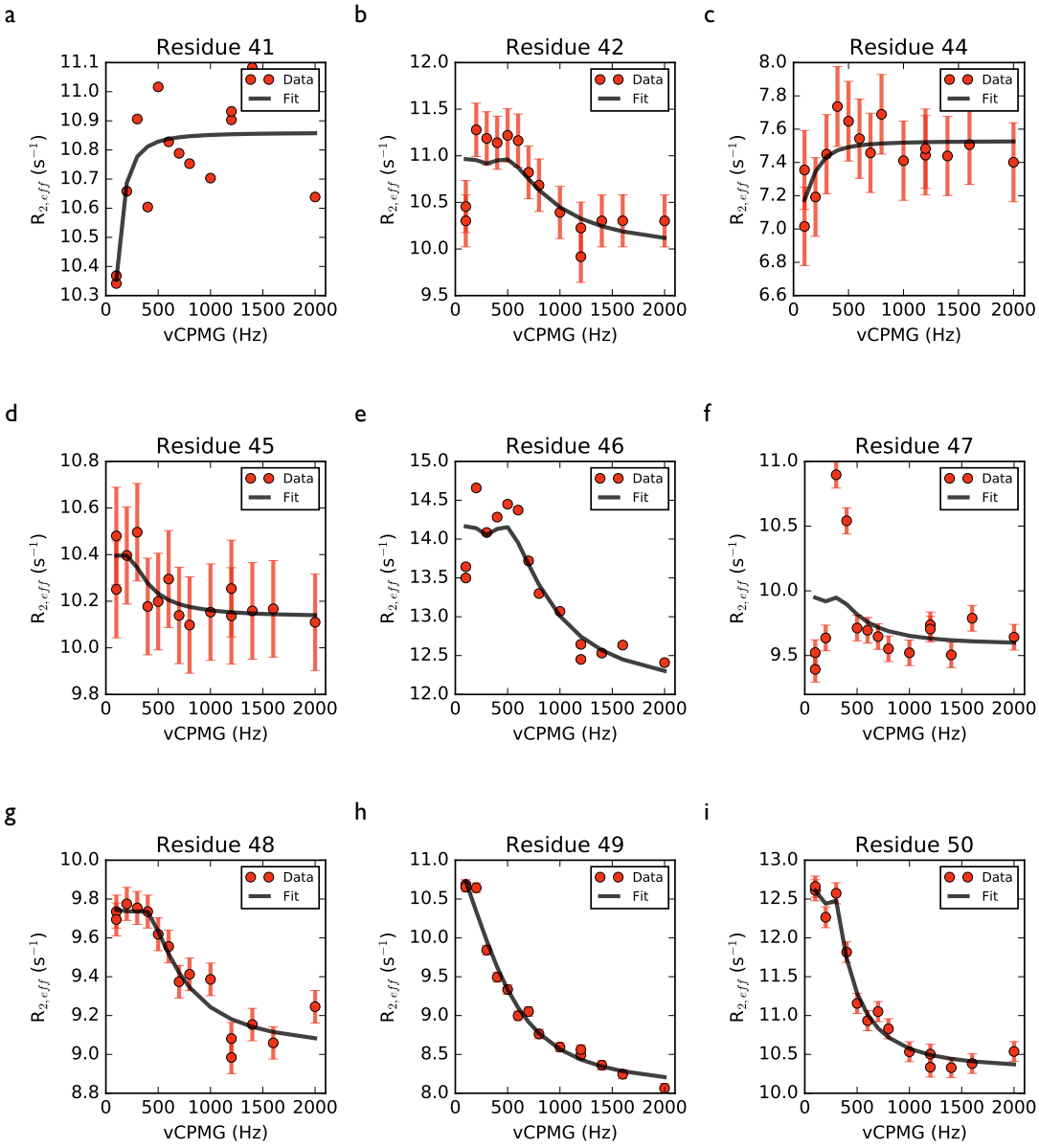


Fig. B.31: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

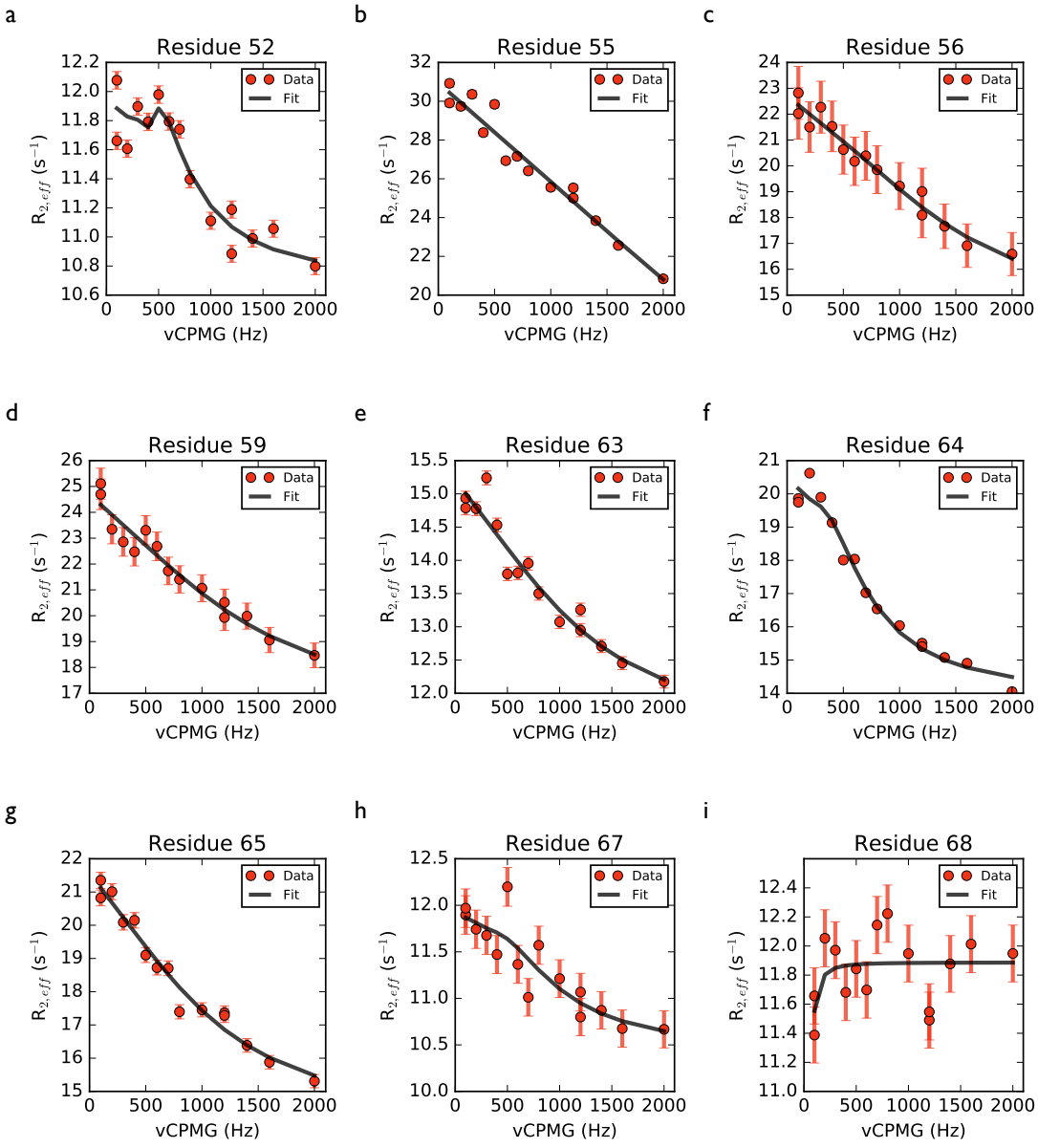


Fig. B.32: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

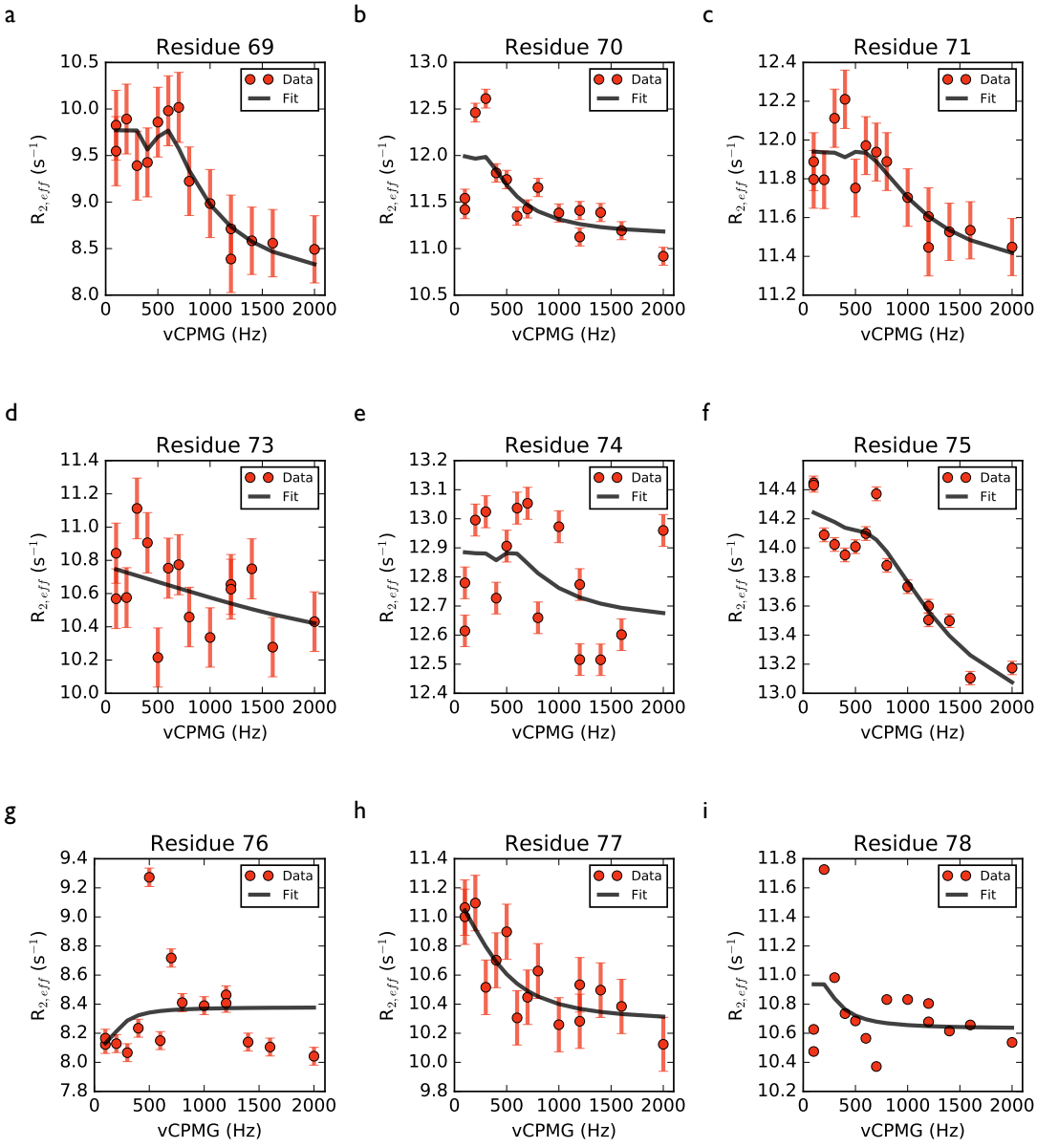


Fig. B.33: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

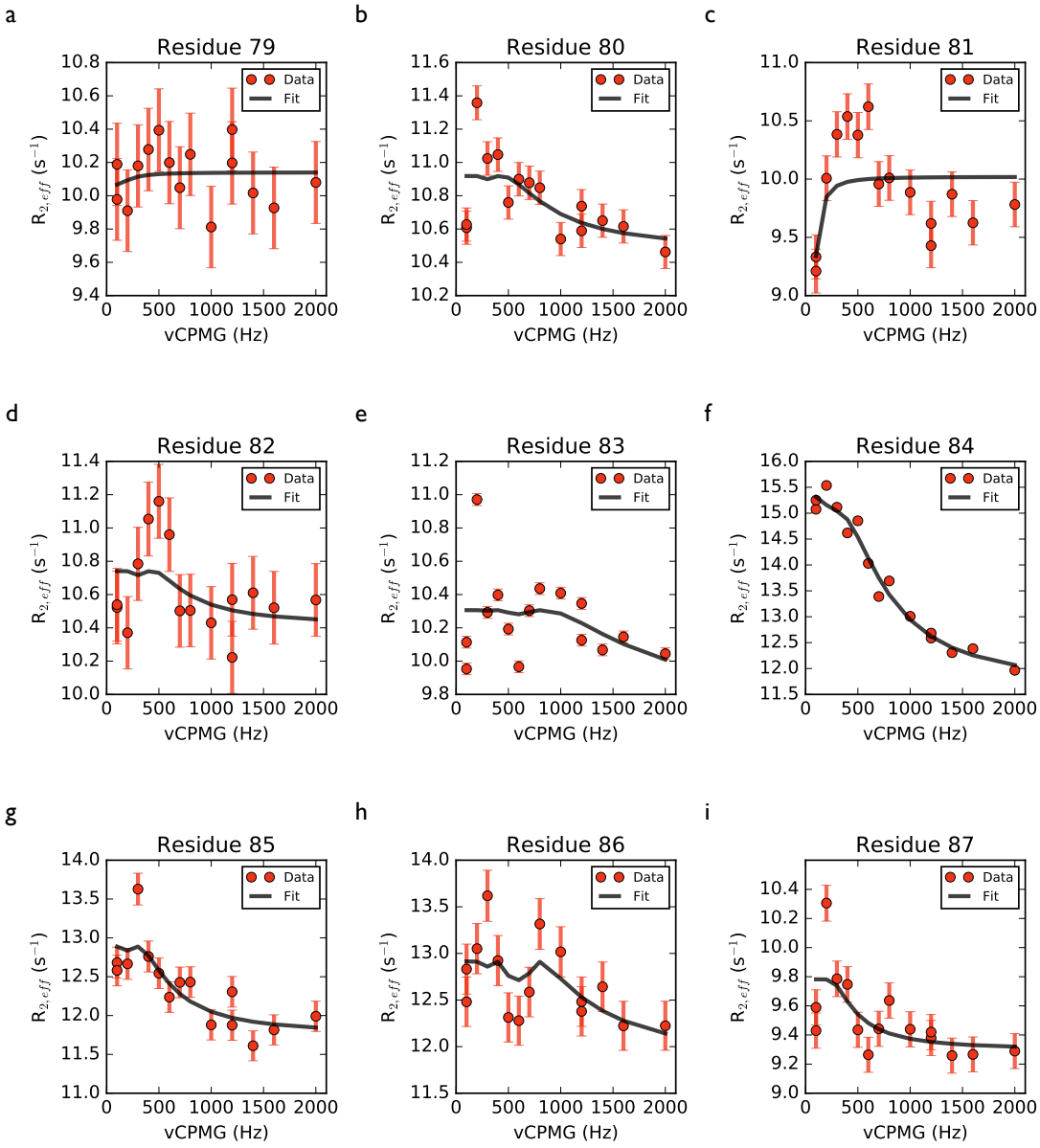


Fig. B.34: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

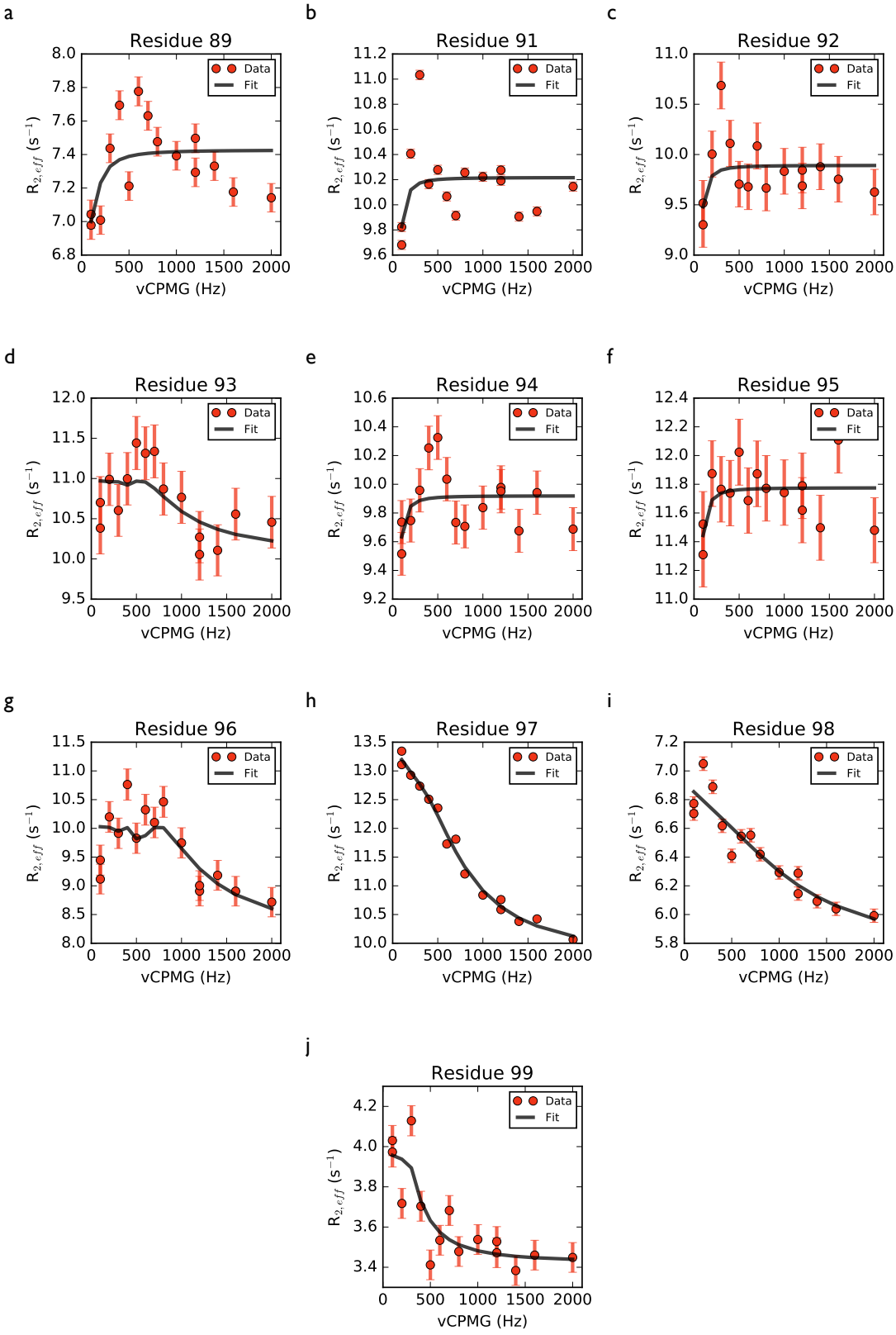


Fig. B.35: CPMG-RD dispersion curves for D76N at pH 6.2 measured at 750MHz

Appendix C

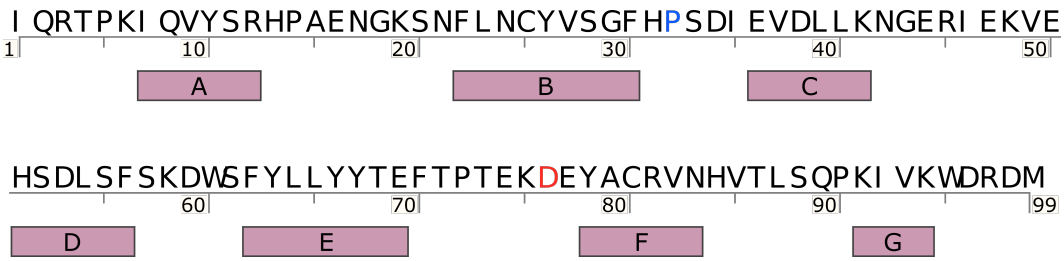


Fig. C.1: Protein sequence of WT β_2m.
The protein sequence is labelled with strands (purple boxes). Pro32 is coloured blue and position 76 is coloured red.